

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

**Studien zu Ursprung, Funktion und Sequenzierung
von RNA- und DNA-Nukleosiden**

Martin Rossa

aus

Berlin, Deutschland

2021

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.

Berlin, 06.06.2021

.....
Martin Rossa

Dissertation eingereicht am: 16.06.2021

1. Gutachter: Prof. Dr. Thomas Carell

2. Gutachterin: Prof. Dr. Stefanie Kellner

Mündliche Prüfung am: 13.07.2021

meiner Familie

Publikationen

Teile dieser Arbeit wurden bereits publiziert:

- „Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides”
S. Becker, J. Feldmann, S. Wiedemann, H. Okamura, C. Schneider, K. Iwan, A. Crisp, M. Rossa, T. Amatov, T. Carell, *Science*. **2019**, 366, 76–82.
- „Non-canonical Bases in the Genome: The Regulatory Information Layer in DNA”
T. Carell, M. Q. Kurz, M. Müller, M. Rossa, F. Spada, *Angew. Chemie - Int. Ed.* **2018**, 57, 4296–4312.
- „N6-methyladenosine (m⁶A) recruits and repels proteins to regulate mRNA homeostasis”
R. R. Edupuganti, S. Geiger, R. G. H. Lindeboom, H. Shi, P. J. Hsu, Z. Lu, S. Y. Wang, M. P. A. Baltissen, P. W. T. C. Jansen, M. Rossa, M. Müller, H. G. Stunnenberg, C. He, T. Carell, M. Vermeulen, *Nat. Struct. Mol. Biol.* **2017**, 24, 870–878.
- „Functional impacts of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine at a single hemi-modified CpG dinucleotide in a gene promoter”
N. Kitsera, J. Allgayer, E. Parsa, N. Geier, M. Rossa, T. Carell, A. Khobta, *Nucleic Acids Res.* **2017**, 45, 11033–11042.

Konferenzbeiträge

- Posterpräsentation „*Towards a deeper understanding of the functional role of different mRNA modifications*“, Synthetic Biology III, Landshut, Deutschland, Jul. 2018
- Vortrag „*Towards a deeper understanding of the functional role of different modifications: m⁶A & mC, hmC, fC, caC*“, SPP1784 PhD Meeting, Dortmund, Deutschland, Jun. 2018
- Posterpräsentation „*Biological Significance and Occurrence of Modified Bases in RNA and DNA*“ SPP1784 PhD Meeting, Dortmund, Deutschland, Jun. 2018
- Posterpräsentation „*A hypermodification in mRNA – The quest for function and sequence context of ms²i⁶A*“, 1st Symposium on Nucleic Acid Modifications, Mainz, Deutschland, Sept. 2017
- Posterpräsentation „*A hypermodification in mRNA – The quest for function and sequence context of ms²i⁶A*“, 1st International Pre-conference on Molecular Systems Engineering (ICMSE), Basel, Schweiz, Aug. 2017
 - **Best Poster Award**
- Vortrag & Posterpräsentation „*Dehydroalanine – A key intermediate of ms²i⁶A biosynthesis?*“, SPP1784 PhD Meeting, Dortmund, Deutschland, Mai 2016
- Posterpräsentation „*ms²i⁶A – More than a tRNA modification?!*“, VII. Nucleinsäurechemie-Treffen, Berlin, Deutschland, Sept. 2015

Inhaltsverzeichnis

Zusammenfassung	V
Summary	VII
A Einleitung	1
1 Der chemische Ursprung der RNA-Nukleoside	1
1.1 RNA-Welt-Hypothese.....	2
1.2 Präbiotische Chemie im Hadaikum	4
1.2.1 Die Bedingungen präbiotischer Reaktionen.....	4
1.2.2 Die chemische Triebkraft präbiotischer Reaktionen.....	8
1.3 Bildung kanonischer RNA-Nukleoside in präbiotischen Modellreaktionen	10
1.4 Die Bedeutung nicht-kanonischer Nukleoside auf der präbiotischen Erde	12
1.4.1 Thio-RNA-Nukleoside	13
1.4.2 Methylierte RNA-Nukleoside	16
2 Die mRNA-Modifikation m⁶A	18
3 Epigenetik und die Sequenzierung von Cytosin-Modifikationen der DNA	20
3.1 Epigenetik.....	20
3.2 Methoden zur Sequenzierung modifizierter Cytosine	25
B Ergebnisse	33
4 Untersuchungen zum chemischen Ursprung kanonischer und nicht-kanonischer RNA-Nukleoside	33
4.1 Vereinigte präbiotisch plausible Synthese von Pyrimidin- und Purin-Nukleotiden der RNA	33
4.2 Studien zur Bildung nicht-kanonischer RNA-Nukleoside unter präbiotisch plausiblen Bedingungen.....	42
4.2.1 Studien zur Bildung von m ⁵ C und m ⁵ U unter präbiotisch plausiblen Bedingungen .	42
4.2.1.1 Synthesestrategie	42
4.2.1.2 präbiotische Modellreaktionen	43
4.2.1.2.1 Methacrylnitril (32).....	43
4.2.1.2.2 Epoxidierung zu 2-Methyloxiran-2-carbonitril (33).....	43

4.2.1.2.3 Isoxazol-Synthese zu 3-Amino-4-methylisoxazol (34)	47
4.2.1.2.3.1 Untersuchungen zur Stabilität von 2-Methyloxiran- 2-carbonitril (33).....	48
4.2.1.2.3.2 präbiotische Modellreaktionen	49
4.2.1.2.4 Carbamoylierung zu 4-Methylisoxazol-3-ylharnstoff (36).....	54
4.2.1.2.5 Voruntersuchungen zu Reduktion und Ringschluss	56
4.2.1.3 Synthese der Referenzverbindungen.....	57
4.2.1.3.1 Methacrylnitril (32).....	57
4.2.1.3.2 2-Methyloxiran-2-carbonitril (33)	58
4.2.1.3.3 3-Amino-4-methylisoxazol (34)	59
4.2.1.3.4 4-Methylisoxazol-3-ylharnstoff (36)	59
4.2.1.4 Zusammenfassung der Studien zur Bildung von m ⁵ C (2) und m ⁵ U (3) und Ausblick	60
4.2.2 Studien zur Bildung von s ² U unter präbiotisch plausiblen Bedingungen	62
4.2.2.1 Synthesestrategie	62
4.2.2.2 präbiotische Modellreaktionen	64
4.2.2.2.1 Thiocarbamoylierung zu <i>N</i> -(Isoxazol-5-yl)thioharnstoff (57)	64
4.2.2.2.1.1 Umsetzungen mit Thioharnstoff, CS ₂ oder Thiocyanat. 65	
4.2.2.2.1.2 Umsetzungen mit P ₄ S ₁₀	65
4.2.2.2.1.3 Umsetzungen mit <i>N</i> ¹ -Methyl- <i>N</i> ¹ -(nitroso)thioharnstoff (66).....	67
4.2.2.2.1.4 Zusammenfassung der Thiocarbamoylierung.....	73
4.2.2.2.2 Ribosylierung zu <i>N</i> -Ribosyl- <i>N</i> -(isoxazol-5-yl)thioharnstoff (58). 74	
4.2.2.2.3 Reduktion und Ringschluss	78
4.2.2.3 Synthese der Referenzverbindungen.....	81
4.2.2.3.1 <i>N</i> -(Isoxazol-5-yl)thioharnstoff (57).....	81
4.2.2.3.2 <i>N</i> -Ribosyl- <i>N</i> -(isoxazol-5-yl)thioharnstoff (58)	81
4.2.2.4 Zusammenfassung und Ausblick der Studien zur Bildung von s ² U (1).....	85
5 Protein-Interaktoren der mRNA-Modifikation m⁶A.....	88
6 Cytosin-Modifikationen der DNA: Funktionsstudien und Verbesserung der Sequenzierung	101
6.1 <i>in vitro</i> -Funktionsänderung eines Promotors, durch hmdC, fdC oder cadC im hemi-CpG-Kontext.....	101
6.2 Methoden zur Verbesserung der Sequenzierung von mdC, hmdC, fdC und cadC	113
6.2.1 Haarnadel-Oligonukleotide: Referenzen zur Sequenzierung komplementärer DNA-Einzelstränge	113
6.2.1.1 Konzept	114

6.2.1.2 Einordnung, Zusammenfassung und Ausblick	116
6.2.1.3 Experimentelle Details	117
6.2.1.3.1 Oligonukleotid-Synthese.....	117
6.2.1.3.2 Entschützung.....	118
6.2.1.3.3 Aufreinigung	119
6.2.2 <i>Spike-ins</i> : Referenz-Oligonukleotide für die Sequenzierung von mdC, hmdC, fdC und cadC	121
6.2.2.1 Konzept	121
6.2.2.2 Sequenzierung	123
6.2.2.3 Zusammenfassung, Einordnung und Ausblick	125
6.2.2.4 Experimentelle Details	125
6.2.2.4.1 Sequenz-Kontexte der <i>spike-ins</i>	125
6.2.2.4.2 Oligonukleotid-Synthese.....	128
6.2.2.4.3 Aufreinigung	130
6.2.2.4.4 <i>spike-in</i> -Auswertung mittels SNUPE	135
6.2.3 Eine überlegene Methode zur fdC-Sequenzierung: fdC-HA-Seq	137
6.2.3.1 Hydroxylamin-Reagenz.....	137
6.2.3.2 Konzept	138
6.2.3.3 Validierung des Hydroxylamin-Reagenzes 81 für die fdC-HA-Seq	140
6.2.3.4 Anwendung.....	143
6.2.3.5 Zusammenfassung, Einordnung und Ausblick	144
6.2.3.6 Experimentelle Details	146
C allgemeine Methoden	148
7 Methoden der chemischen Synthese.....	148
8 DNA-Oligonukleotid-Herstellung.....	148
8.1 Synthese	149
8.2 Entschützung und Abspaltung.....	151
8.3 Aufreinigung und Charakterisierung	153
9 Massenspektrometrie	153
9.1 HR-HESI-LC-MS	153
9.2 HR-ESI-MS	154
9.3 HR-EI-MS.....	154
9.4 LR-(LC-)ESI-MS.....	155

9.5 MALDI-TOF-MS.....	155
10 Spektroskopie	156
10.1 NMR-Spektroskopie	156
10.2 Infrarot-Spektroskopie	157
11 Chromatographie.....	157
11.1 Dünnschichtchromatographie	157
11.2 Säulenchromatographie	157
11.3 HPLC	157
12 Nomenklatur.....	158
13 Synthesevorschriften.....	159
13.1 Synthesen bezüglich m ⁵ C (2) und m ⁵ U (3).....	159
13.1.1 Methacrylnitril (32).....	159
13.1.2 2-Methyloxiran-2-carbonitril (33)	160
13.1.3 3-Amino-4-methylisoxazol (34)	162
13.1.4 <i>N</i> -(4-Methylisoxazol-3-yl)harnstoff (36).....	163
13.1.5 5-Methylcytosin (52).....	164
13.2 Synthesen bezüglich s ² U (1).....	165
13.2.1 5-Aminoisoxazol (55).....	165
13.2.2 <i>N</i> ¹ -Benzoyl- <i>N</i> ³ -(isoxazol-5-yl)thioharnstoff (60).....	166
13.2.3 <i>N</i> -(Isoxazol-5-yl)thioharnstoff (57)	167
13.2.4 <i>N</i> ¹ -Methyl- <i>N</i> ¹ -(nitroso)thioharnstoff (66).....	169
13.2.5 <i>S</i> -Methyl- <i>N</i> -(isoxazol-5-yl)thioharnstoff (67)	170
13.2.6 <i>N</i> ¹ -(β-D-Ribosyl)- <i>N</i> ³ -(isoxazol-5-yl)thioharnstoff (β-58a).....	171
13.2.7 2-Thiouridin (β-1a).....	173
D Anhang.....	174
14 Abkürzungsverzeichnis.....	174
15 Literaturverzeichnis.....	181

Zusammenfassung

Im Zentrum dieser Arbeit stehen nicht-kanonische Nukleoside sowie Modifikationen der RNA und DNA. Die Studie dieser zentralen Funktionsträger erfolgte auf drei sehr unterschiedlichen Gebieten: a) Auf dem Gebiet der präbiotischen Chemie nicht-kanonischer und kanonischer RNA-Nukleoside, b) der Funktionsstudien rezenter RNA- und DNA-Modifikationen sowie c) der Sequenzierung von Cytosin-Modifikationen der DNA.

Die Studien zum präbiotischen Ursprung ermöglichten Einblicke in potenzielle Reaktionswege zur Bildung von RNA-Nukleosiden auf der präbiotischen Erde. So konnten erstmals kompatible Synthesen kanonischer Purin- und Pyrimidin-Nukleoside beschrieben werden. Für die nicht-kanonischen Nukleoside m⁵C, m⁵U sowie s²U wurden die Grundlagen für deren Synthese unter präbiotisch plausiblen Bedingungen geschaffen.

Im Fokus der Funktionsstudien standen die zentrale mRNA-Modifikation m⁶A sowie die DNA-Modifikationen mdC, hmdC, fdC und cadC im hemi-CpG-Kontext. Bezüglich m⁶A gelang die erste systematische Analyse der mit m⁶A interagierenden Proteine. Neben der Identifikation einer Vielzahl bisher unbekannter m⁶A-bindender Proteine, wurden erstmalig auch von m⁶A abgestoßene Proteine identifiziert. Zudem konnte so die erste bekannte Verbindung zwischen einer mRNA-Modifikation und einer Krankheit des Autismus-Spektrums identifiziert werden. Im zweiten Teil der Funktionsstudien wurde der Einfluss von mdC, hmdC, fdC oder cadC im hemi-CpG-Kontext auf die Genexpression untersucht. Dabei zeigten mdC und hmdC einerseits, sowie fdC und cadC andererseits, sehr unterschiedliche Einflüsse. Im Vergleich zu dC reduzierten mdC und hmdC die Genexpression nur leicht. Ein Effekt, der sich auf eine gewisse Abstoßung des Transkriptionsfaktors zurückführen ließ. Im Gegensatz dazu reduzierten fdC und cadC die Genexpression zunehmend, bis zum nahezu vollständigen Stillstand. Interessanterweise trat dieser Effekt ausschließlich in Zusammenhang mit TDG-vermittelter Basenexzisionsreparatur (BER) von fdC und cadC auf.

Die Sequenzierung der DNA-Modifikationen mdC, hmdC, fdC und cadC kann durch die in dieser Arbeit entwickelten Methoden nun deutlich zuverlässiger erfolgen. Der Fokus lag dabei insbesondere auf der anspruchsvollen Sequenzierung von fdC sowie von hemi- und hybrid-modifizierten Positionen der DNA. Dazu wurden zunächst Haarnadel-Oligonukleotide entwickelt, die eine Sequenzierung von mdC, hmdC, fdC sowie cadC in hemi- und hybrid-modifizierten Positionen ermöglichen. Anschließend wurden diese Konstrukte, gemeinsam mit einem im AK Carell entwickelten Hydroxylamin-Reagenz, zur Bisulfit-Sequenzierung eingesetzt. So gelang die Entwicklung einer Methode zur Sequenzierung von fdC mit höchster Genauigkeit.

Erste Studien demonstrierten die klare Überlegenheit dieser Methode gegenüber fCAB, der aktuell gebräuchlichsten Methode zur fdC-Sequenzierung. Darüber hinaus wurden Referenz-Oligonukleotide entwickelt, sogenannte *spike-ins*, die vielfältig zur Verbesserung von Sequenzierungen der dC-Modifikationen einsetzbar sind.

Zusammengenommen konnte mit dieser Arbeit das Verständnis des Ursprungs nicht-kanonischer und kanonischer RNA-Nukleoside erweitert werden. Darüber hinaus wurden bisher unbekannte Funktionen zentraler RNA- und DNA-Modifikationen identifiziert. Außerdem wurden Methoden und Techniken zur Sequenzierung von dC-Modifikationen entwickelt, die eine Grundlage bieten für weitere Aufklärungen der Funktionen dieser zentralen Funktionsträger.

Summary

This work focuses on non-canonical nucleosides and modifications of the RNA and DNA. The study of these central functional entities was conducted in three very different areas: **a)** prebiotic chemistry of non-canonical and canonical RNA nucleosides, **b)** functional studies of extant RNA and DNA modifications, and **c)** sequencing of DNA cytosine modifications.

The studies on prebiotic origins provided insights into possible reaction pathways for the formation of RNA nucleosides on the prebiotic earth. For the first time, compatible syntheses of canonical purine and pyrimidine nucleosides were demonstrated. Moreover, for the non-canonical nucleosides m⁵C, m⁵U as well as s²U, foundations for their syntheses under prebiotically plausible conditions were established.

The functional studies focused on the central mRNA modification m⁶A and the DNA modifications mdC, hmdC, fdC and cadC in the hemi-CpG-context. Regarding m⁶A, the first systematic analysis of proteins interacting with m⁶A was demonstrated. In addition to the identification of many previously unknown protein interactors, also proteins repelled by m⁶A were identified for the first time. Furthermore, the first known link between an mRNA modification and a disease of the autism spectrum was identified. In the second part of the functional studies, the influence of mdC, hmdC, fdC or cadC on gene expression was investigated, when these modifications were present in the hemi-CpG-context. The influence of mdC and hmdC on the one hand, and fdC and cadC on the other hand, was very different: Compared to dC, mdC and hmdC only slightly reduced gene expression. An effect that could be attributed to reduced transcription factor affinity. In contrast, fdC and cadC reduced gene expression increasingly until it came to an almost complete hold. Interestingly, this effect occurred exclusively in combination with TDG-mediated base excision repair (BER) of fdC and cadC.

Sequencing of the DNA modifications mdC, hmdC, fdC and cadC can now be performed much more reliably, based on the methods developed in this thesis. This part of my work focussed especially on the challenging sequencing of fdC positions as well as of hemi- and hybrid-modified loci in DNA. For this purpose, hairpin constructs were developed to allow strand-specific sequencing of mdC, hmdC, fdC and cadC positions. These constructs were then used in bisulfite sequencing together with a hydroxylamine reagent developed at AK Carell. Thereby, enabling us to develop a method for sequencing fdC with highest accuracy. Initial studies demonstrated the clear superiority of this method over fCAB, currently the most common method for fdC sequencing. Furthermore, reference oligonucleotides, so-called *spike-ins*, were developed which can be used in many ways to improve sequencing of dC modifications.

In sum, this work has broadened our understanding of the origin of non-canonical and canonical RNA nucleosides. Furthermore, previously unknown functions of central RNA and DNA modifications were identified. Moreover, methods and techniques for sequencing dC modifications were developed, which provide a basis for further elucidation of the functions of these central functional entities.

A Einleitung

„Wo kommt das eigentlich her und was macht das hier? Das sehe ich mir genauer an!“ So würden Kinder wohl die drei Kern-Themen dieser Dissertation zusammenfassen: *Ursprung, Funktion* und *Sequenzierung* von RNA- und DNA-Modifikationen. Wie die gesamte Dissertation gliedert sich diese Einleitung erdgeschichtlich chronologisch in drei Teile: Kapitel 1 behandelt die Entstehung von RNA-Nukleosiden auf der präbiotischen Erde sowie die Rolle der RNA-Nukleoside bei der Entstehung des Lebens. In Kapitel 2 werden die Funktionen der wichtigen mRNA-Modifikation m⁶A behandelt. Abschließend gibt Kapitel 3 einen Überblick zur Epigenetik sowie zur Sequenzierung von DNA-Modifikationen des Cytosins.

Die Bezeichnungen „Modifikation“ und „nicht-kanonisch“ beziehen sich auf verschiedene Ursprünge eines Nukleosids. In dieser Arbeit bezeichnet der Begriff Nukleosid-Modifikation ein chemisch verändertes Nukleosid, dessen Struktur von den rezenten kanonischen Nukleosiden abweicht. Durch biotische Prozesse gebildete Nukleosid-Modifikationen werden in dieser Arbeit als „Modifikationen“ bezeichnet. Hingegen bezeichnet der in der präbiotischen Chemie übliche Ausdruck „nicht-kanonisch“ in dieser Arbeit Nukleoside abiotischen Ursprungs.

1 Der chemische Ursprung der RNA-Nukleoside

Der Ursprung des Lebens auf der Erde ist eine der großen ungelösten Fragen der Menschheit, nicht zuletzt, weil Evidenz nur schwer zu bilden ist. Das erste Leben der Erde entstand vermutlich in einer präbiotischen Welt organischer Moleküle. Die geringen Stabilitäten organischer Moleküle gegenüber Hydrolyse lassen keine molekularen Fossilien aus dieser Zeit erwarten.^[1] Deshalb basiert Evidenz über früheste organische Reaktionen hauptsächlich auf Isotopenstudien, sofern aussagekräftige Isotopenmuster erhalten blieben. Weitere Anhaltspunkte zu den Bedingungen der präbiotischen Erde stammen aus Studien geologischer Prozesse und Beobachtungen zur Planetenbildung. Auf Basis dieses Fundaments stellen Chemiker in Modellreaktionen die Bedingungen auf der frühen Erde nach. Durch Einsatz von Reagenzien, die präbiotisch plausibel sind, lassen sich so Prozesse finden, die zur Entstehung des Lebens auf der Erde geführt haben könnten. Die Bedeutung der so gewonnenen Erkenntnisse scheint zunächst im Verständnis unseres eigenen Ursprungs zu liegen. Aber ihre Bedeutung geht weit darüber hinaus: Zu verstehen, wo im Universum und unter welchen Bedingungen Leben entstehen kann, stellt womöglich eine Grundlage für das Fortbestehen der Menschheit über deren Existenz auf der Erde hinaus dar.

Nukleoside bildeten vermutlich die Basis des ersten Lebens dieser Erde. Kapitel 1 stellt den Zusammenhang zwischen RNA-Nukleosiden und allen rezenten Lebewesen der Erde dar. Es gibt einen Überblick zu Gegebenheiten unserer präbiotischen Erde und es fasst präbiotische Modellreaktionen zur Bildung von RNA-Nukleosiden zusammen.

1.1 RNA-Welt-Hypothese

Die *RNA-Welt-Hypothese* beschreibt plausibel den Ursprung allen rezenten Lebens auf der Erde. Bereits 1962 schlug ALEXANDER RICH vor, Leben auf der frühen Erde könnte ausschließlich auf RNA basiert haben.^[2] Anfang der 1980er Jahre entdeckten die Gruppen um THOMAS CECH^[3] und SIDNEY ALTMAN^[4] katalytisch aktive RNA-Spezies in rezenten Organismen, sogenannte Ribozyme. Dies galt als letzter noch fehlender Beleg für die Theorie, dass RNA allein die beiden Kernfunktionen der heutigen RNA-DNA-Protein-Welt übernehmen kann: Die Speicherung genetischer Informationen sowie katalytische Aktivität. Diese Erkenntnisse fasste WALTER GILBERT 1986 als *RNA-Welt-Hypothese* zusammen. Nach dieser stand ein ausschließlich RNA-basierter Organismus, am Anfang der Evolution aller rezenten Lebewesen der Erde.^[5] Der Einbau von Fehlern in das Erbgut dieses Organismus und die Replikation dieser Fehler ermöglichten so Darwin'sche Evolution auf der molekularen Ebene.^[6] Gleichzeitig scheint es plausibel, dass RNA-basierte Organismen nicht die ersten und einzigen Lebensformen auf der frühen Erde waren.^[7,8] Mögliche RNA-Vorläufer und -Alternativen, werden unter dem Begriff proto-RNA zusammengefasst.^[9-11] Abb. 1 ordnet die „RNA-Welt“ chronologisch ein.

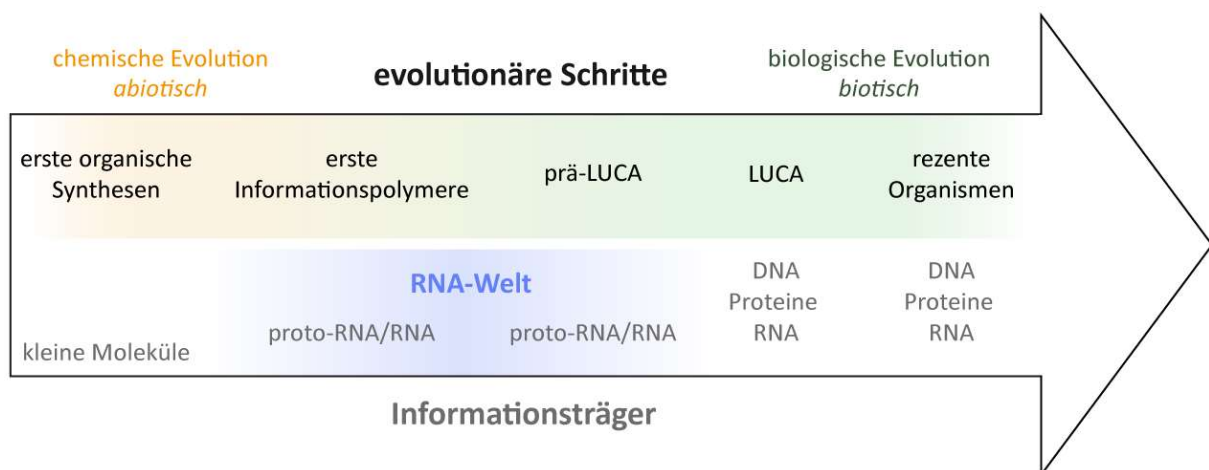


Abb. 1 Zeitliche Abfolge wichtiger Schritte sowie der Informationsträger bei der Entwicklung des Lebens auf der Erde. Die Abbildung basiert auf dem in dieser Arbeit dargelegten Kenntnisstand.

Der Ursprung aller rezenten Organismen wird auf einen gemeinsamen Urvorfahr zurückgeführt: *LUCA*. Das erste Leben auf der Erde basierte gemäß der RNA-Welt-Hypothese ausschließlich auf (proto-)RNA. Diese ersten Organismen könnten als prä-LUCA-Organismen zusammengefasst

werden, siehe Abb. 1. Im Anschluss an diese Phase entwickelten sich Organismen die sowohl RNA als auch DNA und Proteine verwendeten. Darunter auch die Population des letzten gemeinsamen Urvorfahren aller rezenten Organismen, LUCA (*the last universal common ancestor*). Eine Reihe rezenter RNA-Modifikationen scheinen seit LUCA konserviert zu sein, siehe Abb. 2.^[12] Dabei handelt es sich insbesondere um Methyl- und Thio-Modifikationen.^[12] Beide Modifikations-Klassen legen nahe, dass Methan bzw. Sulfide in LUCAs Lebensraum präsent waren.^[12] Dies weist darauf hin, dass ozeanische Hydrothermalfelder LUCAs Lebensraum darstellten.^[12] Welche Zeiträume zwischen der ersten (proto-)RNA, dem ersten Leben auf der Erde und LUCA liegen ist bisher gänzlich unklar. Die zur Zeit ältesten Hinweise auf Leben, sogenannte Biosignaturen, werden auf 3,8^[13] und 4,1 Mrd. Jahre^[14] vor unserer Zeit datiert. Die ältesten unumstrittenen Belege für marines Leben auf 3,42 Mrd. Jahre vor unserer Zeit.^[15] Damit liegt der Zeitraum präbiotischer Reaktionen, der dem ersten Leben vorausging, wahrscheinlich größtenteils im Erdzeitalter des Hadaikums.

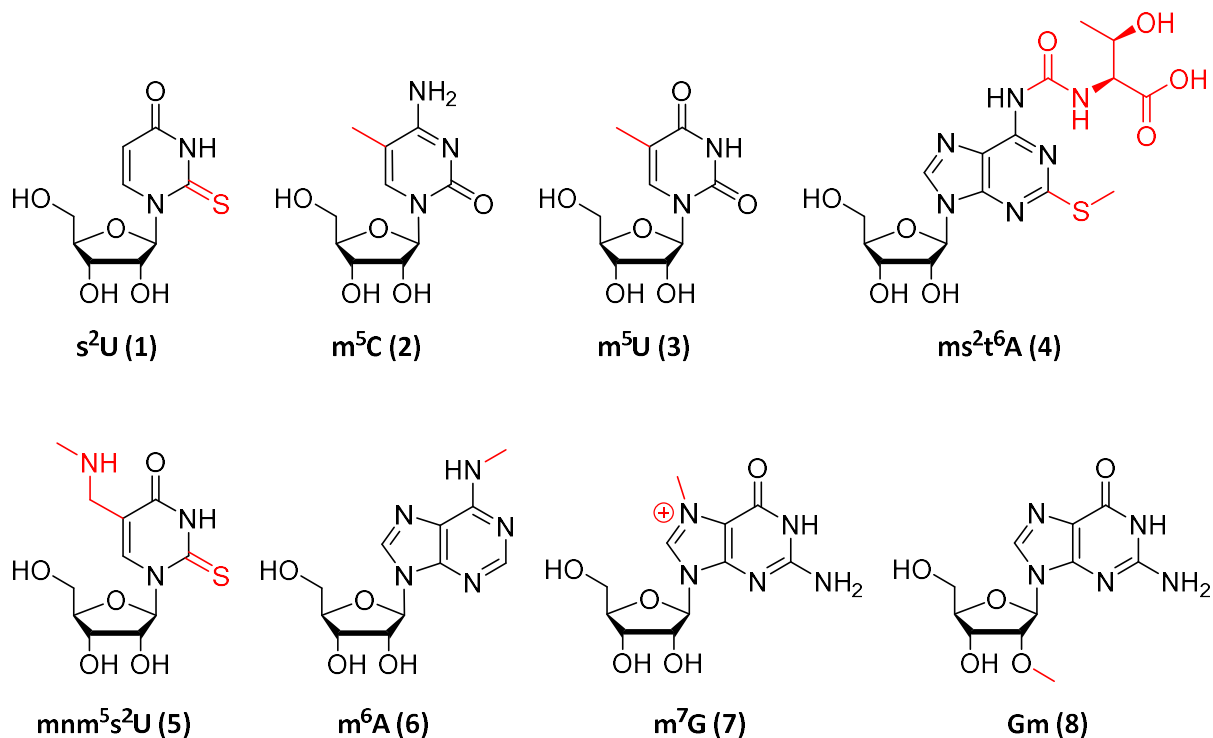


Abb. 2 Auswahl an RNA-Modifikationen die mit hoher Wahrscheinlichkeit seit LUCA in allen Domänen des Lebens konserviert sind.^[12] **rot:** Von den kanonischen RNA-Nukleosiden abweichende Strukturen. s²U: 2-Thiouridin (1), m⁵C: 5-Methylcytidin (2), m⁵U: 5-Methyluridin (3), ms²t⁶A: 2-Methylthio-N⁶-threonyl-carbamoyladenine (4), mnm⁵s²U: 5-Methylaminomethyl-2-thiouridin (5), m⁶A: N⁶-Methyladenine (6), m⁷G: N⁷-Methylguanin (7), Gm: 2'-O-Methylguanin (8).

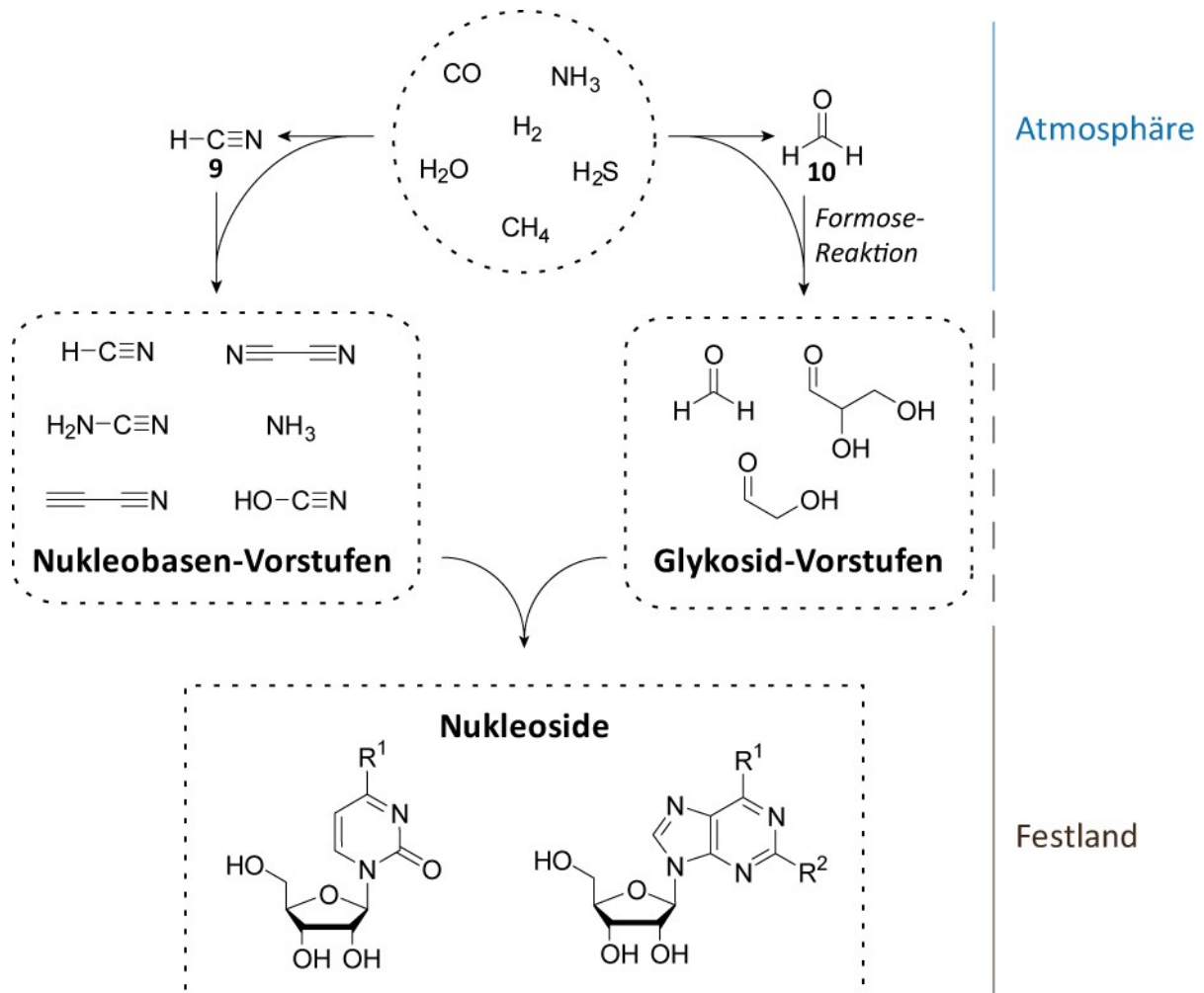
1.2 Präbiotische Chemie im Hadaikum

Plausibilität ist das zentrale Maß zur Beurteilung präbiotischer Reaktionen. Die präbiotische Plausibilität misst sich daran, ob sowohl Reaktionsbedingungen als auch Edukte mit anerkannten Gegebenheiten der präbiotischen Erde konsistent sind. Auch extraterrestrische Quellen können dabei in Erwägung gezogen werden.^[16,17] Die Bedingungen im Hadaikum sind in vielen Belangen noch immer unklar. Ein einflussreicher Faktor ist der potenziell lange Zeitraum präbiotischer Reaktionen über einige 100 Mio. Jahre: So könnten auch Reaktionen die nach Labor-Standards als sehr langsam gelten, präbiotisch plausibel sein. Zudem erlaubt ein Zeitraum dieser Größenordnung auch die Kombination von Szenarien geringer Wahrscheinlichkeit. Neben der zum Teil fehlenden Evidenz führt dies zusätzlich zur Steigerung der Anzahl präbiotisch plausibler Szenarien. Um Experimente durchführen zu können, sind Chemiker demnach oft gezwungen aus mehreren plausiblen Szenarien zu wählen. Bei der Abwägung sollten Szenarien die als häufiger gelten können und solche die auf robusteren Modellen basieren anderen vorgezogen werden.^[6,18] Zudem können nur Prozesse als präbiotisch plausibel gelten, die kontinuierlich, d. h. ohne Eingriffe, ablaufen bzw. auf der präbiotischen Zeitskala kontinuierlich abgelaufen sein könnten.^[6] Zusammengefasst ist präbiotische Plausibilität das Maß an Übereinstimmung einer Modellreaktion mit den Bedingungen der präbiotischen Erde.

1.2.1 Die Bedingungen präbiotischer Reaktionen

Die Bedingungen des Hadaikums wurden insbesondere durch die Entstehung der Erde sowie durch unser junges Sonnensystem geprägt. Das Hadaikum erstreckte sich von der Entstehung der Erde vor 4,53 Mrd. Jahren, bis zum Beginn des Archaikums vor 4,0 Mrd. Jahren, vgl. Abb. 1. Allgemeine Aussagen über das Hadaikum sind aufgrund der wechselvollen Gegebenheiten kaum möglich. Klar scheint, dass direkt nach der Entstehung der Erde Einschläge anderer Protoplaneten die Bedingungen dominierten.^[19] Je nach Größe des Impaktors führte dies zum vollständigen oder anteiligen Wiederaufschmelzen des Erdmantels und zur Zerstörung möglicher organischer Verbindungen. Der Verlauf der Temperatur der Erdoberfläche im Hadaikum lässt sich bis heute nur schwer abschätzen. Neue Modellierungen weisen auf eine gefrorene Erdoberfläche vor 4,3 Mrd. Jahren hin.^[20] Andere Studien halten moderate Temperaturen bis zu 100 °C für möglich.^[21] Aus der Beobachtung anderer Sterne scheint es wahrscheinlich, dass die UV-Intensität der jungen Sonne um etwa 30 % geringer war als heute.^[22] Zusammengefasst dominierten direkt nach der Erd-Entstehung Einschläge die Bedingungen auf der Erde, nach dieser Phase waren Temperaturen und UV-Strahlung vermutlich moderat.

Präbiotische Synthesen liefen in der Atmosphäre und an Land ab, Wasser war dabei das zentrale Lösungsmittel. Die Atmosphäre des Hadaikums gilt als einer der zentralen Orte präbiotischer Synthesen.^[23] So wird die Bildung der Nukleosid-Bausteine HCN und Formaldehyd in der Atmosphäre verortet. Über Zwischenstufen könnten beide Bausteine die Basis der präbiotischen Nukleosid-Synthese dargestellt haben,^[24] siehe Schema 1. In präbiotischen Modellreaktionen lassen sich auf Basis von HCN Nukleobasen bilden, Formaldehyd stellt hingegen die Grundlage zur Synthese der Kohlenhydrate dar. So gebildete komplexere Moleküle sinken aus der Atmosphäre ab.^[23] Deshalb gilt eine feste Erdoberfläche als Voraussetzung für deren Anreicherung und weitere Syntheseschritte. Auch Wasser gilt als Grundlage präbiotischer Synthesen. So konnte es als Reaktand und Lösungsmittel fungieren. Dessen pH-Wert wurde insbesondere durch die Interaktion des Wassers mit Gasen und Mineralien bestimmt. So bildeten SO₂ und CO₂ saure Lösungen, NH₃ hingegen basische.^[25-27] Auch Mineralien konnten mit Wasser, je nach Art, sauer bis basisch reagieren.^[20,28,29] Neben Wasser könnten auch Formamid^[30] und Harnstoff^[31] als Lösungsmittel fungiert haben. Zusammengenommen liefen präbiotische Reaktionen vermutlich sowohl in der Atmosphäre als auch an Land ab, insbesondere in Wasser. In welchen Umgebungen präbiotische Reaktionen außerhalb der Atmosphäre abgelaufen sein könnten, wird hingegen kontrovers diskutiert.^[23]



Schema 1 Potenziell präbiotische Nucleosid-Synthesen: Die Bildung der zentralen Nucleosid-Bausteine HCN (**9**) und Formaldehyd (**10**) verlief wohl in der Atmosphäre. Ausgehend von diesen zentralen Bausteinen wurden vermutlich weitere Nucleosid-Vorstufen gebildet, die sich auf dem Festland anreichern konnten. Die präbiotische Bildung der Nucleoside war vermutlich auf dem Festland verortet. R_1 : OH oder NH_2 , R_2 : OH, R_3 : NH_2 oder R_2 : NH_2 , R_3 : H.

Extraterrestrische Quellen ermöglichten eventuell präbiotische Synthesen, ozeanische Hydrothermalfelder wahrscheinlich nicht. Neben Atmosphäre und Landflächen werden auch extraterrestrische Quellen und Hydrothermalfelder häufig als Orte präbiotischer Synthesen genannt. Um extraterrestrisches Material anreichern zu können, ist eine ausreichend große Landfläche nötig.^[23] Modellierungen deuten aber darauf hin, dass im Hadaikum maximal ein sehr kleiner Teil der Kontinentalkruste oberhalb des Meeresspiegels lag, z. B. in der Größe des heutigen Australiens.^[23,32] Dies lässt eine nützliche Anreicherung extraterrestrischer Moleküle fraglich erscheinen, wenn auch nicht unmöglich.^[6] Des Weiteren wurden ozeanische Hydrothermalfelder als Orte präbiotischer Reaktionen untersucht.^[33,34] Chemische Reaktionen im Ozean haben den Nachteil starker Verdünnung. Notwendige Mechanismen zur Konzentration organischer Verbindungen in Hydrothermalfeldern konnten bisher nicht gefunden werden.^[23] Auch die

wichtige Reaktionsklasse der Kondensationen ist in wässriger Lösung stark gehemmt, so dass ozeanische Hydrothermalfelder nicht mehr als wahrscheinliche Orte präbiotischer Synthesen gelten.

Heute gelten Mulden als wahrscheinlichste Orte präbiotischer Chemie außerhalb der Atmosphäre. Gesteinsmulden könnten im Hadaikum an Flanken von Vulkaninseln oder auf dem Festland verortet gewesen sein,^[32] siehe Abb. 3. Im Gegensatz zu Hydrothermalfeldern im Ozean, könnten hohe Konzentrationen der Bestandteile durch Verdunstung oder Einfrieren erreicht worden sein. Zudem könnten Mulden in räumlicher Nähe zueinander unterschiedliche Bedingungen aufgewiesen haben, z. B. bezüglich pH-Wert, Verbindungen, Metallionen, Mineralien, Temperatur, Volumen etc. Solche Mulden hätten zeitweise miteinander sowie mit dem Ozean in Verbindung stehen können. So könnte der zeitweise Austausch von Reagenzien ermöglicht worden sein. Zusätzlich könnten nahegelegene Vulkane hohe lokale Konzentrationen reduzierter Spezies geliefert haben, die zur Synthese komplexerer Moleküle notwendig waren. Neben Vulkanausbrüchen könnten diese durch Blitzentladungen sowie stille Entladungen gebildet worden sein.^[32,35] Lokale Temperaturerhöhungen durch vulkanische Aktivität hätten darüber hinaus auch bei einer überfrorenen Erde präbiotische Reaktionen erlaubt.

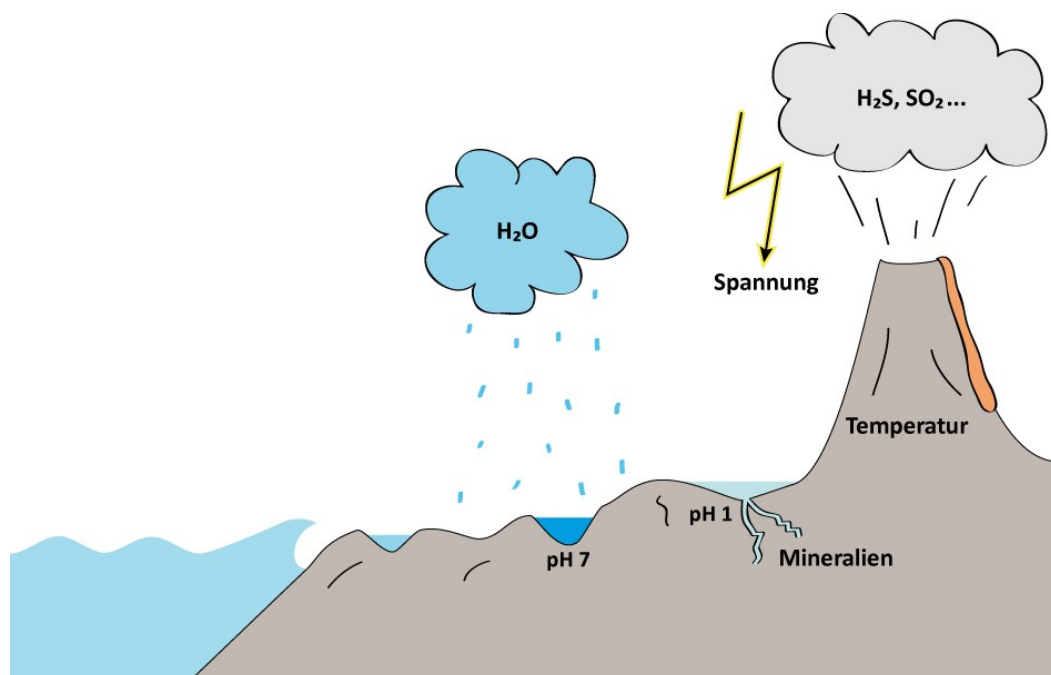


Abb. 3 Mulden an Vulkanflanke gelten heute als wahrscheinlichste Orte präbiotischer Chemie außerhalb der Atmosphäre. Eintrocknen oder Gefrieren hätten hohe Konzentrationen ermöglicht. Vulkanische Aktivität hätte neben erhöhten Temperaturen auch reduzierte Spezies, Blitzentladungen und Regen verursachen können. Das Wasser verschiedener Mulden hätte sich unterscheiden können bzgl. Mineraliengehalt, pH-Wert, Metallionen, Volumina etc.

Mineralien waren von zentraler Bedeutung für präbiotische Synthesen. Mineralien konnten präbiotische Reaktionen auf einer Vielzahl von Wegen beeinflussen: Durch Modulation des pH-Wertes, als Edukt, durch Katalyse, Stabilisierung, Adsorption usw.; dementsprechend war eine lokale mineralische Diversität für präbiotische Prozesse von zentraler Bedeutung.^[6] Einige Mineralien und Metalle können, z. B. durch Datierung, als klar präbiotisch plausibel eingeordnet werden. Bei vielen anderen ist dieser Status nicht klar.^[6] Eine Übersicht potentieller Mineralien des Hadaikums wurde von HAZEN und Kollegen zusammengestellt.^[36,37] Biologisch wichtige Elemente wie B, Mo und P traten als Hauptbestandteile nur in präbiotisch seltenen Mineralien auf. Interessanterweise scheinen diese seltenen Mineralien für die Verfügbarkeit von B, Mo und P nicht zwingend notwendig gewesen zu sein. Stattdessen standen diese Elemente als Nebenbestandteile präbiotisch sehr häufiger Mineralien zur Verfügung.^[37] Auch extraterrestrische Reaktionen laufen größtenteils auf Mineralien ab.^[6]

1.2.2 Die chemische Triebkraft präbiotischer Reaktionen

Die Atmosphäre des Hadaikums war oxidativ neutral, die Vorläufer der präbiotischen Nukleosid-Synthesen hingegen reduziert. Die Atmosphäre des Hadaikums gilt als zentraler Ort zur präbiotischen Bildung der Nukleosid-Vorläufer,^[23] siehe Abb. 4. Die Zusammensetzung dieser Atmosphäre war durch die oxidierten Spezies CO_2 , H_2O , SO_2 und N_2 bestimmt. Reduziertere Spezies wie CO , H_2 , H_2S , und NH_3 waren nur in geringen Mengen präsent.^[23] Damit war die Atmosphäre des Hadaikums oxidativ neutral, ebenso wie der Erdmantel^[38]. Nahezu alle präbiotischen Nukleosid-Vorläufer waren hingegen reduziert,^[24] vgl. Schema 1. Reduziert bezieht sich dabei auf das Verhältnis von H-Atomen zu Hetero-Atomen am Kohlenstoff-Gerüst. Je größer der Anteil an H-Atomen, desto reduzierter ist die Verbindung. Sowohl die Atmosphäre als auch der Erdmantel waren demnach stärker oxidiert als die zentralen Vorläufer der präbiotischen Nukleosid-Synthese.^[23] Woher stammte also die chemische Triebkraft?

Die chemische Triebkraft präbiotischer Reaktionen lässt sich nicht allein durch UV-Strahlung, Entladungen und Serpentinisierung erklären. Quellen reduktiver Verbindungen wie UV-Strahlung, Blitzentladungen und stille Entladungen bilden lediglich Redox-Paare. Sie ermöglichten aber keine anhaltende Verschiebung des Redox-Gleichgewichts.^[23] Eine andere häufig genannte Quelle von Reduktionsäquivalenten ist die Serpentinisierung. Dieser Prozess beschreibt die Reduktion von Wasser durch heißes Gestein, z. B. in Hydrothermalfeldern.^[29,39,40] Modellierungen zeigen aber, dass die so gebildeten Reduktionsäquivalente kaum ausreichend gewesen wären, um eine atmosphärische HCN-Synthese in nützlichem Umfang zu ermöglichen.^[40-43] HCN gilt als Edukt nahezu aller präbiotischen Vorläufer der Nukleobasen, siehe Schema 1. Zusammengenommen

wäre demnach die Bildung komplexer Nukleosid-Vorläufer in großem Umfang wohl nicht auf Basis von Serpentinisierung, UV-Strahlung und Entladungen möglich gewesen.^[23]

Ein extraterrestrisches Objekt könnte ausreichend chemische Triebkraft für präbiotische Nukleosid-Synthesen geliefert haben. Der Einschlag eines einzigen Impaktors von der Größe des heutigen Erdmondes hätte genügend Reduktionskraft für alle präbiotischen Nukleosid-Synthesen liefern können. Der Einschlag wäre erfolgt, nachdem die Erde vor 4,50 Mrd. Jahren erstmals abkühlte und dadurch von flüssigem Wasser bedeckt gewesen sein könnte. Der zerschellte Fe⁰-Kern des Impaktors hätte Wasser zu H₂ reduziert und so große Mengen an Reduktionsäquivalent erzeugt. Für die präbiotischen Umsetzungen in Folge des Einschlags wird ein Zeitraum vor 4,36 Mrd. Jahren postuliert, siehe Abb. 4. Dieser Zeitraum beginnt nach der Abkühlung der Erde in Folge des Einschlags. Beendet worden wäre dieser Zeitraum durch das stetige Entweichen des H₂ aus dem Gravitationsfeld der Erde. So hätte sich ein stärker reduziertes Redox-Gleichgewicht eingestellt. Interessanterweise hätte der Einschlag auch das im Hadaikum seltene P⁰ verfügbar gemacht. Zusammengenommen könnte der Einschlag eines großen Impaktors die Bildung von Nukleosiden und (proto-)RNA-Oligomeren durch Verschiebung des Redox-Gleichgewichts plausibel erklären.^[23,44–46]

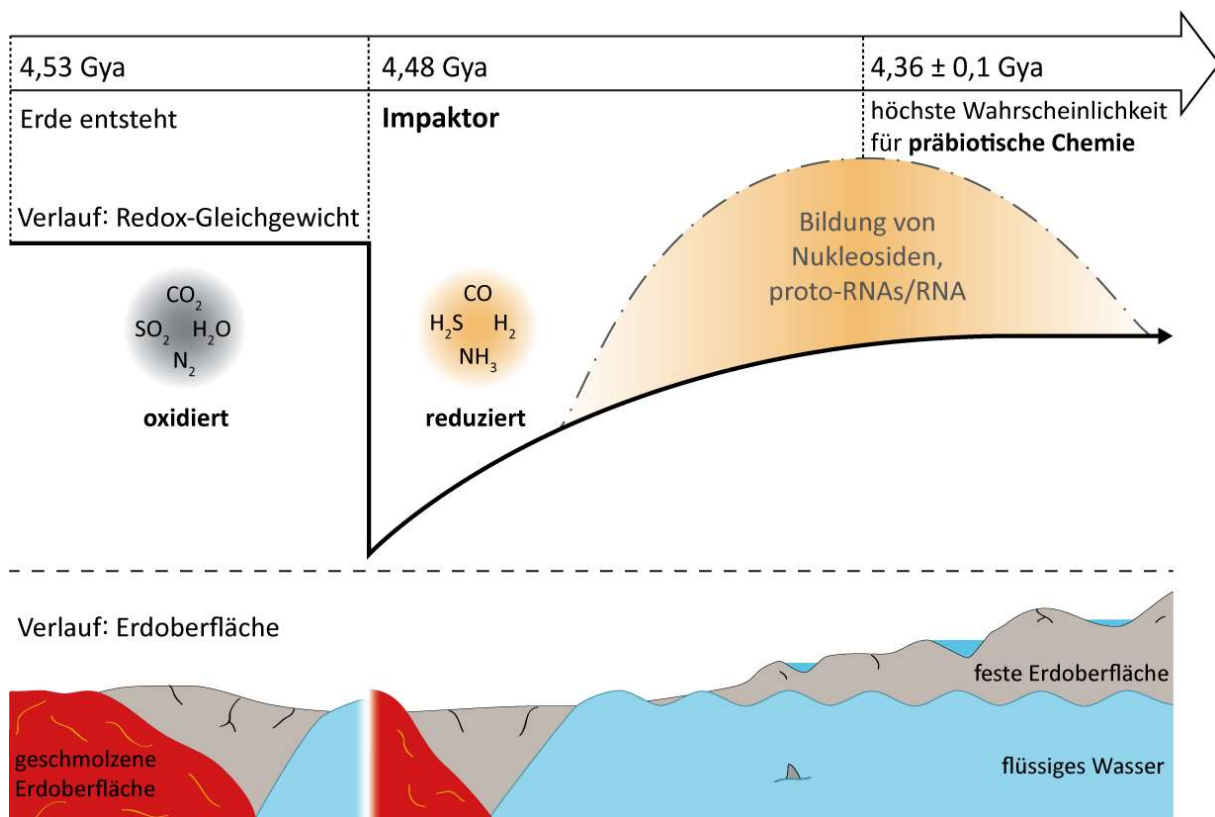


Abb. 4 Verlauf des Redox-Gleichgewichts (oben) auf der frühen Erde sowie Veränderung der Erdoberfläche (unten). Ein extraterrestrischer Impaktor vor 4,48 Mrd. Jahren hätte genügend Reduktionsäquivalente erzeugen können, um präbiotische Nukleosid-Chemie zu ermöglichen. **Punkt-Strich:** Potentieller zeitlicher Verlauf der Wahrscheinlichkeit präbiotischer Nukleosid-Chemie in Folge des Impaktors. **Gya:** Milliarden Jahre vor unserer Zeit.

1.3 Bildung kanonischer RNA-Nukleoside in präbiotischen Modellreaktionen

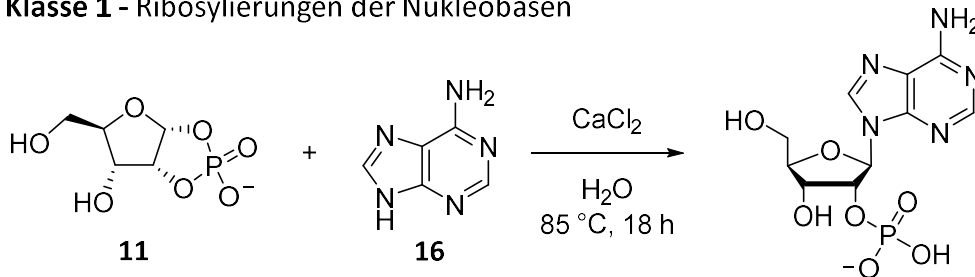
Die Ribosylierung der Nucleobasen ist ein zentrales Problem der präbiotischen Chemie. Die präbiotische Bildung der RNA-Nucleoside wurde zuerst von der Gruppe um ORGEL in den 1970er Jahren untersucht. Durch Umsetzung der Nucleobasen mit Ribose (Ribosylierung) gelang die Bildung von Adenosin (Ausbeute 4 %) und Guanosin (8 %).^[47] Pyrimidin-Nucleoside konnten hingegen nicht gebildet werden.^[48] Die geringe Reaktivität zeigte bereits, dass die Ribosylierung von Nucleobasen in der Regel anspruchsvoll ist. Das lässt sich zum einen auf die meist geringe Nucleophilie der N^1 -Position der Nucleobase zurückführen. Zum anderen gehören Ribosylierungen zu den Kondensationsreaktionen, die in wässriger Lösung aufgrund des Prinzips von LE CHATELIER stark gehemmt sind. In den folgenden 50 Jahren wurden daraufhin eine Vielzahl alternativer Synthesewege untersucht.

Die derzeit bekannten präbiotischen Modellreaktionen zur Synthese kanonischer RNA-Nucleoside lassen sich nach Art der Ribosylierung in drei Klassen einteilen: **1** Ribosylierungen der Nucleobasen, **2** Ribosylierung eines Nucleobasen-Vorläufers und **3** Synthesewege ohne Ribosylierung. Die jeweiligen Ausbeuten sind in Klammern angegeben, falls bekannt. Es sollte beachtet werden, dass Ausbeuten verschiedener Klassen nur begrenzt miteinander vergleichbar sind.

- **Klasse 1** umfasst *Ribosylierungen der Nucleobasen*, siehe Schema 2. In dieser Klasse wurden bisher drei verschiedene Ansätze publiziert: **a)** Über Phosphat-aktivierte Ribose **11** werden A (15 %), C (7 %), G (< 1 %) und U (< 1 %) erhalten.^[49,50] **b)** Die Ribosylierung in Mikro-Tröpfchen liefert A (2,5 %), C (0,7 %) und U (2,5 %), jedoch kein G.^[51] **c)** Unter entwässernden Bedingungen werden A, C und G erhalten. Nur A bildete sich dabei in hoher Ausbeute.^[52]
- **Klasse 2** umfasst die *Ribosylierung eines Nucleobasen-Vorläufers*, siehe Schema 3. Die Ribosylierung von Formamidopyrimidin-Vorläufern **12** (FaPys) ermöglicht die Synthese der kanonischen Purin-Nucleoside A (17 %) und G (4 %).^[53-55] Im Gegensatz zur Ribosylierung der Nucleobasen verläuft die FaPy-Ribosylierung mit quantitativer Regioselektivität.
- **Klasse 3** umfasst *Synthesewege ohne Ribosylierung*, siehe Schema 4. In dieser Klasse wurden bisher zwei Ansätze publiziert: **a)** Die Umsetzung von Formamid mit Meteoritenpulver und Protonenstrahlung liefert A, C und U in Spuren.^[56] Der Mechanismus dieser Reaktionen ist bisher nicht bekannt. **b)** Ausgehend vom Nucleosid-Vorläufer 2-Aminooxazol (**13**) können die kanonischen Pyrimidin-Nucleoside C (26 %) und U (1,9 %) erhalten werden.^[55,57] 2-Aminooxazol (**13**) bildet sich aus Cyanamid (**14**) und Glycolaldehyd (**15**) unter präbiotisch plausiblen Bedingungen.^[55,58] Da 2-Aminooxazol (**13**) bereits die spätere glykosidische Bindung enthält, verläuft die Synthese ohne Einsatz von Ribose sowie Nucleobase.

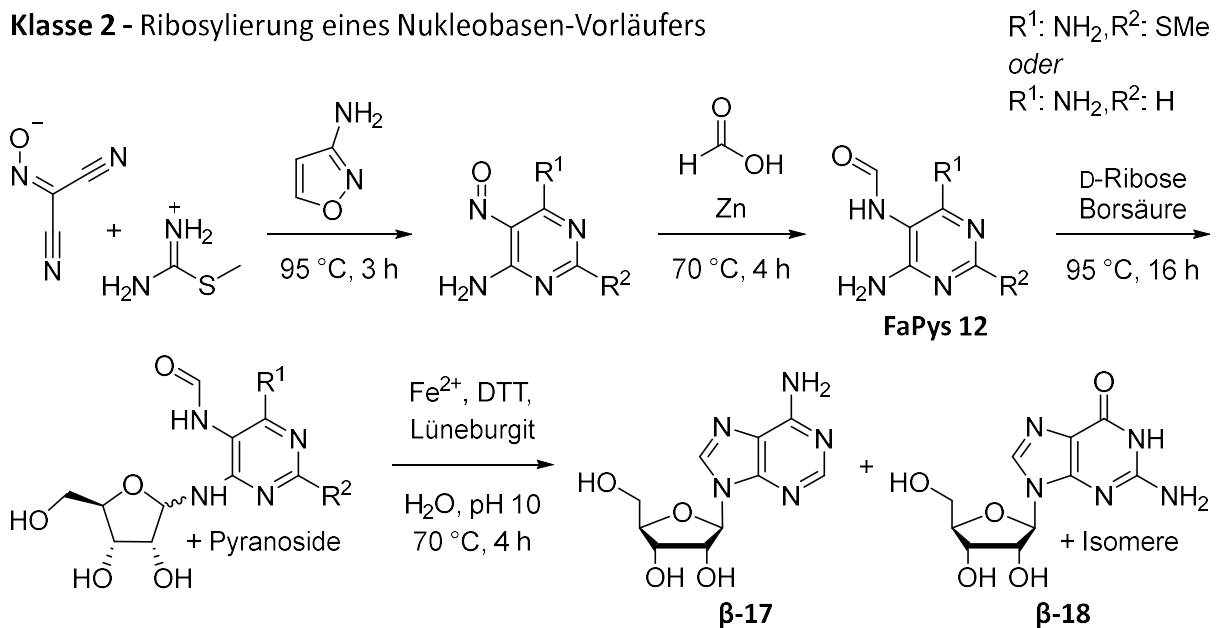
Zusammengenommen wurde die Bildung einzelner kanonischer Nucleoside in verschiedenen präbiotischen Modellreaktionen erreicht. Eine vereinigte Synthese aller kanonischen RNA-Nucleoside – präbiotisch plausibel und in nützlichen Ausbeuten – war hingegen zu Beginn dieser Arbeit nicht bekannt.

Klasse 1 - Ribosylierungen der Nucleobasen

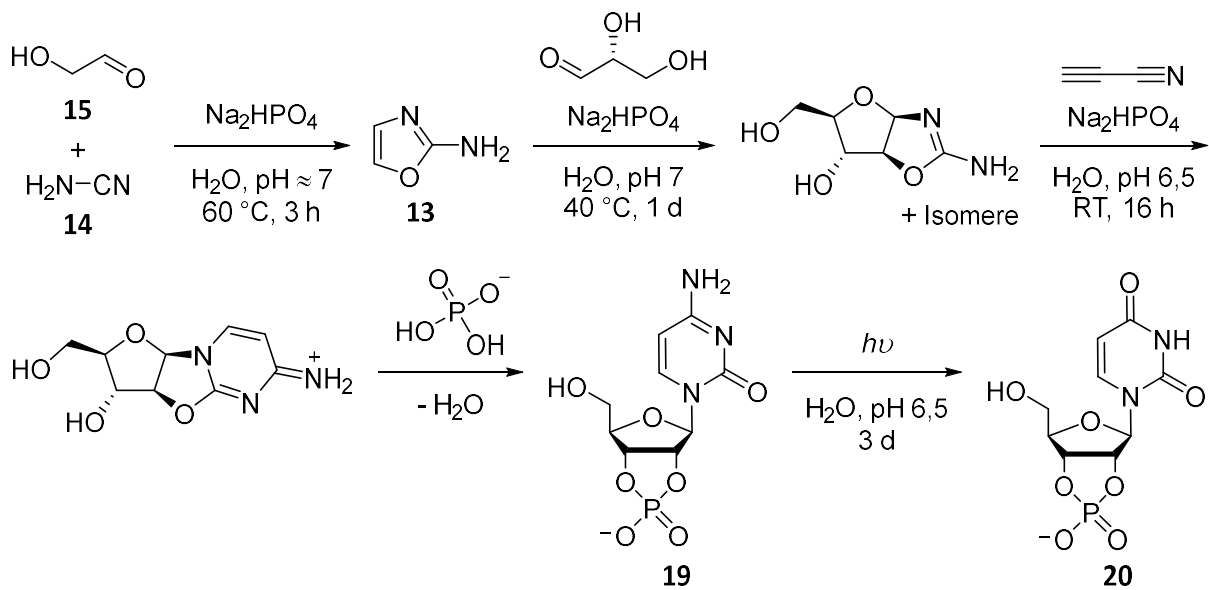


Schema 2 Klasse 1 - Ribosylierung der Nucleobasen: Ribosylierung von Adenin (**16**) mittels 1',2'-Phosphat-aktivierter D-Ribose **11**, als beispielhafte Reaktion dieser Klasse.^[49,50,55]

Klasse 2 - Ribosylierung eines Nucleobasen-Vorläufers



Schema 3 Klasse 2 - Ribosylierung eines Nucleobasen-Vorläufers: Ribosylierung von Formamido-pyrimidin-Vorläufern **12 (FaPys)** ermöglicht die Synthese von β -Adenosin (**β -17**), β -Guanosin (**β -18**) und weiteren Isomeren.^[53-55]

Klasse 3 - Synthesen ohne Ribosylierung

Schema 4 Klasse 3 – Synthesen ohne Ribosylierung: Die Synthese von Cytidin- (**19**) sowie Uridin-2',3'-cyclophosphat (**20**) via 2-Aminooxazol (**13**), als beispielhafte Reaktion dieser Klasse.^[55,57,58]

1.4 Die Bedeutung nicht-kanonischer Nukleoside auf der präbiotischen Erde

Nicht-kanonische Nukleoside nahmen beim Ursprung des Lebens auf der Erde vermutlich eine zentrale Rolle ein. Viele nicht-kanonische Nukleoside bilden sich selbst in wässriger Lösung spontan in hohen Ausbeuten.^[59] Deren hohe Reaktivität, auch gegenüber Ribose, lässt vermuten, dass sie auf der präbiotischen Erde in deutlich größerem Umfang entstanden sein könnten als kanonische Nukleoside.^[59] Neben einer robusten Synthese weisen viele nicht-kanonische Nukleoside weitere Eigenschaften auf, die deren zentrale Rolle beim Ursprung des Lebens auf der Erde nahelegen: Dazu gehört eine stabilere Basenpaarung, sowohl auf Ebene einzelner Nucleobasen als auch auf Oligonukleotid-Ebene.^[59] Auch andere Kernprozesse der Entstehung des Lebens, wie die nicht-enzymatische RNA-Replikation und der Translationsprozess, waren vermutlich in hohem Maße von nicht-kanonischen Nukleosiden bzw. RNA-Modifikationen abhängig.

Ein Zusammenhang zwischen LUCAs RNA-Modifikationen und nicht-kanonischen Nukleosiden der präbiotischen Erde wurde bisher nicht untersucht. Der letzte gemeinsame Urvorfahr aller rezenten Lebewesen (LUCA) nutzte wahrscheinlich diverse RNA-Modifikationen, von denen mindestens einige bis heute konserviert sind.^[12] Unter den seit LUCA konservierten RNA-Modifikationen sind Methyl-Modifikationen besonders häufig, aber auch Thio-Modifikationen sind in großem Umfang konserviert,^[12] vgl. Abb. 2. Beide Klassen stellen vermutlich eine Verbindung zu chemischen Prozessen in ozeanischen Hydrothermalfeldern her, LUCAs Lebensraum.^[12] Ob ein

Zusammenhang zwischen den RNA-Modifikationen von LUCA und nicht-kanonischen Nukleosiden der präbiotischen Erde besteht, ist bisher nicht geklärt. Erste (proto-)RNAs bildeten sich auf Basis der präbiotisch verfügbaren Nukleoside. Inwieweit die ersten (proto-)RNA-basierten Organismen aber weiterhin von abiotisch gebildeten Nukleosiden abhängig waren ist unklar. Ebenso unklar ist die Beziehung zwischen diesen Organismen und LUCA. Zu verstehen, ob LUCAs RNA-Modifikationen präbiotisch verfügbar gewesen sein könnten, würde eine erste Grundlage zur Beurteilung dieses Zusammenhangs liefern. Aus diesem Grund fassen die folgenden beiden Abschnitte den aktuellen Wissensstand zu Thio- und Methyl-Modifikationen bzw. -Nukleosiden in Bezug auf die Entstehung des Lebens zusammen.

1.4.1 Thio-RNA-Nukleoside

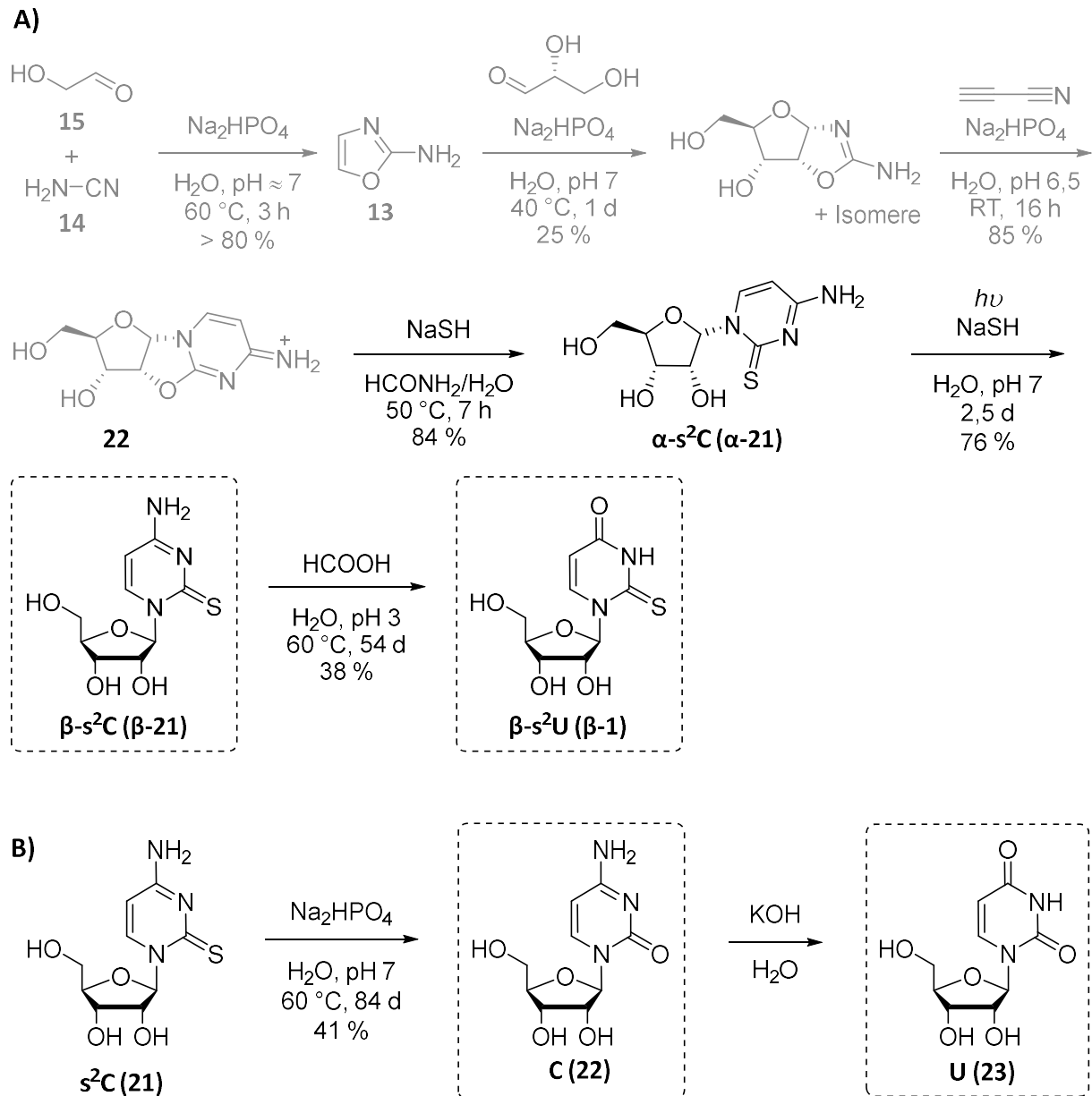
Thio-Nukleoside stellen womöglich eine zentrale Klasse präbiotischer RNA-Nukleoside dar. Diese Hypothese basiert zum einen auf der vermutlich großen Häufigkeit von Schwefel auf der frühen Erde, sowohl in verschiedenen Oxidationsstufen als auch Aggregatzuständen.^[60] Zum anderen weisen Thio-Modifikationen der RNA eine Vielzahl von Eigenschaften auf, die bei der Entstehung des Lebens Selektionsvorteile geboten haben könnten.

Thio-Modifikationen erhöhen die RNA-Stabilität. So weist das RNA-Basenpaar s^2U-A eine deutlich höhere Stabilität auf als das Basenpaar $U-A$.^[61] Gleichzeitig zeigt das s^2U-G -wobble-Basenpaar eine geringere Stabilität, so dass die Genauigkeit der Basenpaarung steigt.^[62] Beide Eigenschaften, Duplexstabilität und Genauigkeit der Basenpaarung, waren vermutlich zentral für Bildung und Replikation der ersten RNA-Oligonukleotide. Darüber hinaus führt die große Elektronenzahl des Schwefels zu einer höheren Polarisierbarkeit sowie stärkeren Dispersions-Wechselwirkungen. Beides könnte zu verstärktem *stacking* beitragen,^[62-64] ein weiterer zentraler Bestandteil der RNA-Duplexbildung.^[65] Das methylierte s^2U -Analog s^2T stabilisiert vermutlich durch eine C3'-endo-Konformation der Ribose die tRNA hyperthermophiler Organismen.^[66,67] Auch die Ribose von freiem s^2U (**1**) liegt in der C3'-endo-Konformation vor.^[68,69] Thio-Modifikationen könnten so die RNA-Stabilität in heißen Umgebungen der präbiotischen Erde erhöht haben. Zusammengenommen steigern Thio-Modifikationen die RNA-Stabilität. Sie könnten so zur RNA-Duplexbildung und fehlerfreien Replikation bei der Entstehung des Lebens beigetragen haben.

*s^2U (**1**) und s^2C könnten für die Entstehung des Lebens von zentraler Bedeutung gewesen sein.* Der Austausch von U durch s^2U (**1**) in RNA führt zu einer wesentlich genaueren und schnelleren nicht-enzymatischen RNA-Replikation.^[70] Ein Prozess, der als Grundlage der Evolution auf der frühen Erde gilt. Eine Reihe an Studien zeigt zudem, dass verschiedene Thio-Modifikationen eine

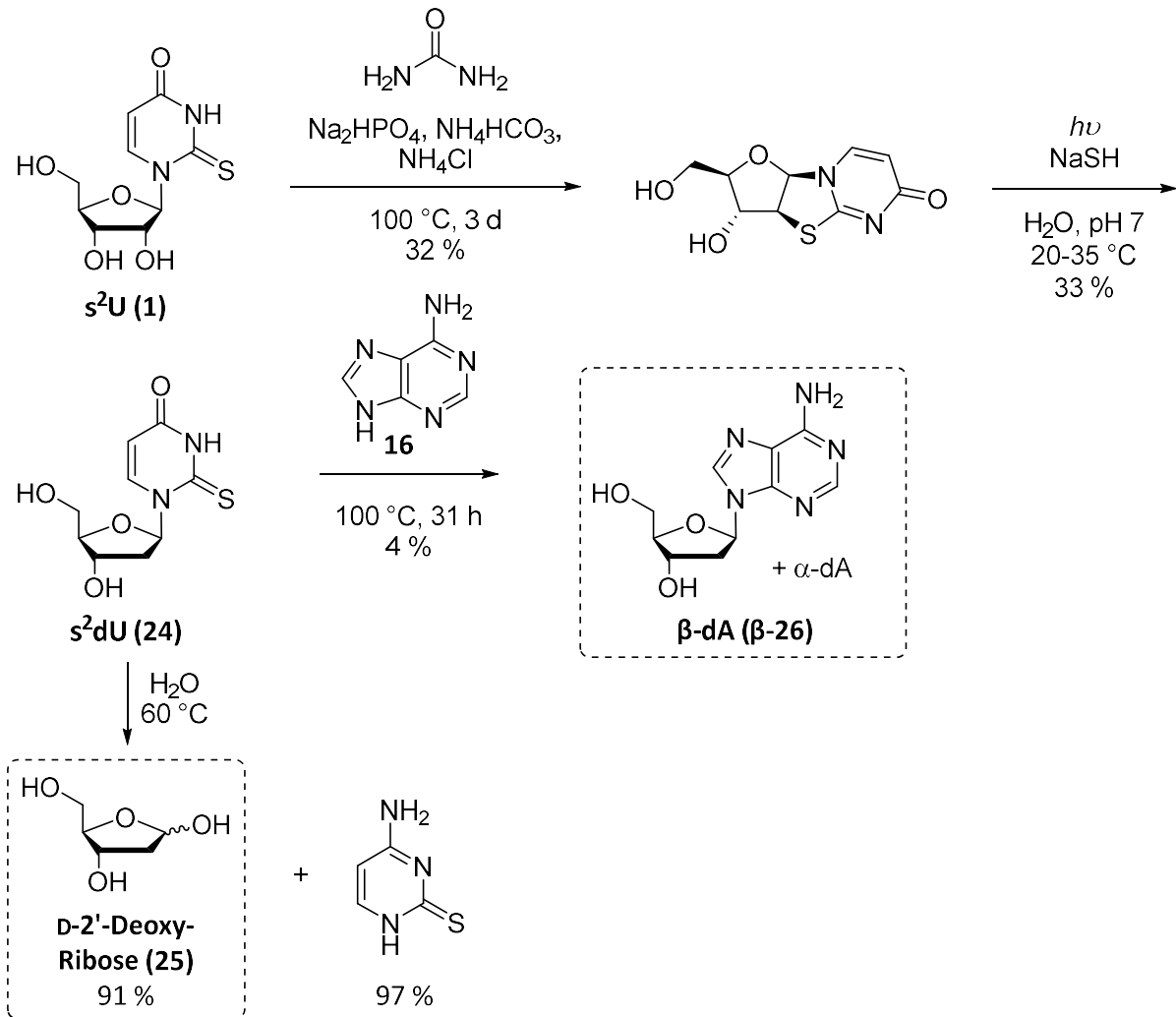
größere Genauigkeit der Proteinbiosynthese ermöglichen.^[64,71-73] So erhöhen s^2U (**1**) und s^2C in rezenten tRNA beispielsweise die Spezifität der mRNA-tRNA-Interaktion im Translationsprozess.

*Die Synthese von s^2U (**1**) und s^2C in präbiotischen Modellreaktionen ist mittels Photoanomerisierung möglich.* 2017 etablierte die Gruppe um SUTHERLAND die Synthese von s^2U (**1**) und s^2C (**21**) ohne Ribosylierung, also ohne Einsatz freier Ribose oder Nukleobase, siehe Schema 5 A).^[57] Analog zur Synthese der kanonischen Pyrimidine, vgl. Schema 47, beginnen die präbiotischen Modellreaktionen mit der Bildung von 2-Amino-Oxazol (**13**) aus simplen Vorläufern. So wird zunächst die spätere glykosidische Bindung erzeugt. In weiteren Schritten werden Ribofuranose und Nukleobase aufgebaut. Durch Thiolyse wird der Nukleosid-Vorläufer **22** in α - s^2C (α -**21**) überführt, welches mittels UV-Bestrahlung zu β - s^2C (β -**21**) photoanomerisiert. β - s^2C (β -**21**) kann anschließend zu β - s^2U (β -**1**) hydrolysieren. Zudem führt die Hydrolyse von s^2C (**21**) in Anwesenheit von Na_2HPO_4 zur Bildung des kanonischen Pyrimidin-Nukleosids C (**22**),^[57] welches weiter zu U (**23**)^[74] hydrolysieren kann, siehe Schema 5 B). Dieser Ansatz ist demnach mit der Bildung der kanonischen Pyrimidin-Nukleoside kompatibel. Bedingungen, die darüber hinaus auch kompatibel mit der Bildung kanonischer Purin-Nukleoside sind, wurden für s^2U (**1**) und s^2C (**21**) bisher nicht gefunden.



Schema 5 A) Präbiotische Modellreaktionen zur Bildung von s²C (**21**) und s²U (**1**).^[55,57]
grau: Reaktionsschritte identisch zur Bildung der kanonischen Nukleoside, dargestellt in Schema 4.
B) Umsetzung von s²C (**21**) in die kanonischen Nukleoside C (**22**)^[57] und U (**23**)^[57,74] mittels Hydrolyse.^[55]
 Zentrale Produkte sind umrahmt.

s²U (**1**) stellt ein Bindeglied zwischen RNA und DNA dar. So lässt sich s²U (**1**) mittels Photolyse in hoher Ausbeute in das entsprechende DNA-Nukleosid s²dU (**24**) überführen, siehe Schema 36.^[75] Zudem liefert dessen Hydrolyse 2'-Deoxy-D-ribose (**25**) in hoher Ausbeute und somit den zentralen Baustein zur Glykosylierung anderer DNA-Basen. s²U (**1**) könnte also eine frühe Quelle des DNA-Zuckers **25** dargestellt haben. Darüber hinaus ermöglicht s²U (**1**) die Synthese des kanonischen DNA-Nukleosids dA (**26**) durch Transribosylierung mit Adenin (**16**).^[75] Zusammengefasst scheint damit eine zentrale präbiotische Bedeutung von Thio-Nukleosiden bzw. -Modifikationen naheliegend.



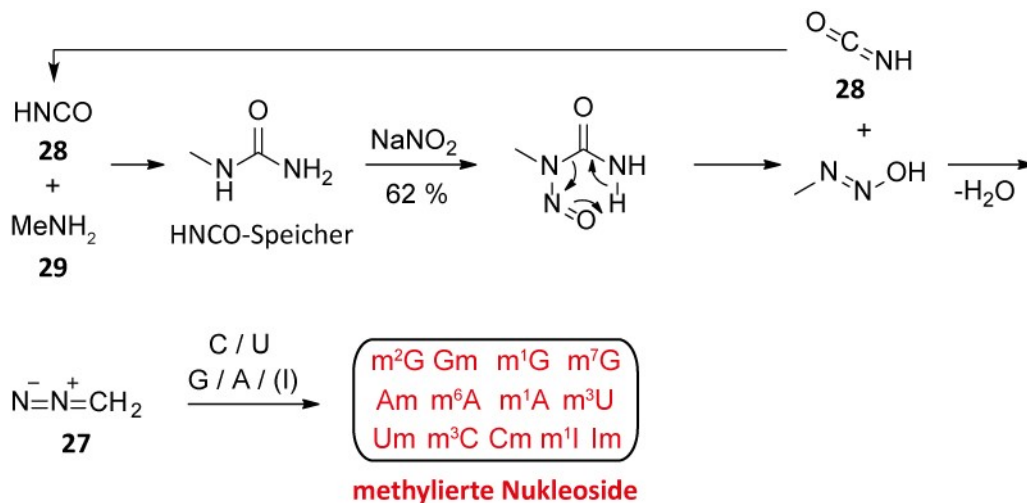
Schema 6 Umsetzung von s^2U (1) in das kanonische DNA-Nukleosid β -dA (β -26), mittels Transribosylierung.^[75] Zusätzlich ergibt die Hydrolyse von s^2dU (24) 2'-Deoxy-D-ribose (25), den zentralen Baustein zur Glykosylierung der DNA-Nukleobasen.^[75] Abbildung basiert auf ^[55]. Zentrale Produkte sind umrahmt.

1.4.2 Methylierte RNA-Nukleoside

Methyl-Modifikationen könnten beim Ursprung des Lebens auf der Erde wichtige Funktionen ausgeübt haben. Ebenso wie Thio-Modifikationen übernehmen auch Methyl-Modifikationen heute wichtige Funktionen im Translationsprozess: So ist m^1G eine in allen drei Domänen des Lebens streng konservierte tRNA-Modifikation, die zentral für die Leseraster-Erhaltung der Translation ist.^[76] Eine Funktion, die auch bei den ersten Translations-Vorgängen unerlässlich gewesen sein musste. Im Gegensatz zu Fehleinbauten führen Verschiebungen des Leserasters mit hoher Wahrscheinlichkeit zur Biosynthese stark abweichender Peptide, so dass die eigentliche Funktion des Proteins in der Regel nicht erhalten bleibt. Auch andere methylierte Nukleoside übernehmen heute wichtige Funktionen: So kann m^5U (3) als RNA-Analog und möglicher Vorläufer des kanonischen DNA-Nukleosids dT betrachtet werden. m^5C (2) ist das RNA-Analog und ein

möglicher Vorläufer der DNA-Modifikation m⁵dC, die als zentraler epigenetische Schalter der DNA in Eukaryoten fungiert, vgl. Kapitel 3.

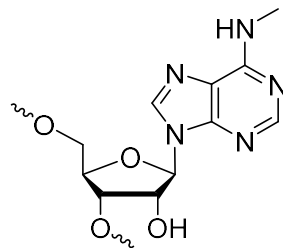
Die Synthese methylierter RNA-Nukleoside wird erst seit kurzem in präbiotischen Modellreaktionen untersucht. So wurde 2018 die Bildung methylierter Purin-Nukleoside in präbiotischen Modellreaktionen von der Gruppe um CARELL untersucht.^[54] Dabei wurden neben m¹G, m²G, m²G, m²A und ms²A auch die kanonischen Purin-Nukleoside erhalten. Ebenfalls 2018 wurde die Synthese des Methylierungsreagenzes Diazomethan (**27**) in einer präbiotischen Modellreaktion von der Gruppe um CARELL etabliert, siehe Schema 7.^[77] In Gegenwart der kanonischen Nukleoside führt diese Synthese zur Bildung der methylierten Purin-Nukleoside m¹G, m²G, m⁷G (**7**), Gm (**8**) sowie m¹A, m⁶A (**6**) und Am. In Anwesenheit von m²G, m⁶A (**6**) und Inosin (I) wird zusätzlich m²G, m⁶A, m¹I und Im gebildet. Auch die methylierten Pyrimidin-Nukleoside m³U, Um, m³C und Cm konnten so gebildet werden. Die Synthese des potenziell wichtigen dT-Analogs m⁵U (**3**) gelang in präbiotischen Modellreaktionen bisher nur in Spuren.^[52,56] Auch eine mögliche präbiotische Synthese von m⁵C (**2**) wurde bisher nicht gefunden. Zusammengenommen wurden die Synthesen einer Vielzahl methylierter Purin-Nukleoside bereits in präbiotischen Modellreaktionen demonstriert. Synthesen methylierter Pyrimidin-Nukleoside wurden bisher hingegen kaum untersucht.



Schema 7 Diazomethan (**27**) kann in einer präbiotischen Modellreaktion aus Isocyanensäure (**28**) und Methylamin (**29**) gebildet werden. In Anwesenheit von RNA-Nukleosiden ermöglicht dies die Bildung methylierter RNA-Nukleoside. Abbildung basiert auf ^[77].

2 Die mRNA-Modifikation m⁶A

m⁶A beeinflusst eine Vielzahl zellulärer Prozesse und Krankheiten. Seit LUCA, dem letzten gemeinsamen Urvorfahr, sind Modifikationen der RNA in allen Domänen des Lebens konserviert.^[78] Heute sind weit über 100 RNA-Modifikationen in diversen RNA-Klassen bekannt.^[79,80] Nach der Entdeckung der ersten RNA-Modifikation 1951,^[81] gab es in den 1970er Jahren erste Berichte über m⁶A (N⁶-Methyladenosin) in polyadenylierter RNA (**30**, Abb. 5).^[82-85] Der anfangs umstrittenen^[86] Rolle dieser mRNA-Modifikation steht heute ein wachsendes Verständnis um dessen zentrale Bedeutung entgegen: So beeinflusst m⁶A zentrale zelluläre Prozesse wie die Translation^[87-89], das Spleißen^[90,91] und den circadianen Rhythmus^[92]. Auch die Zusammenhänge zwischen m⁶A und einer Vielzahl an Krankheiten werden klarer, darunter neurodegenerative Krankheiten,^[93,94] Diabetes mellitus Typ 2^[95,96] und diverse Krebsformen^[97].



30

Abb. 5 Ausschnitt einer m⁶A-modifizierten mRNA (**30**).

Die m⁶A-Synthese ist reversibel, wird reguliert und stellt die häufigste Modifikation in mRNA dar. Trotz der frühen Entdeckung von m⁶A in mRNA (**30**) in den 1970er Jahren, führten erst Erkenntnisse in den 2000er Jahren zu einem breiten Interesse der Forschungsgemeinschaft: Zu dieser Zeit entdeckten verschiedene Gruppen, dass m⁶A (**30**) sowohl eine reversible^[98,99] als auch regulierte^[100,101] mRNA-Modifikation darstellt. Eigenschaften die bis dahin ausschließlich für epigenetische DNA-Modifikationen bekannt waren. Diese Entdeckungen führten zur Prägung des Begriffs *Epitranskriptom*.^[102] 2012 folgten die ersten Transkriptom-weiten Sequenzierungen von m⁶A.^[103,104] Diese zeigten, dass m⁶A (**30**) in tausenden unterschiedlichen mRNAs auftritt und vor allem im DRACH-Sequenzmotiv vorliegt.^[103-106] Der größte Anteil der Modifikation findet sich in der Nähe des Stopcodons.^[103,104] Darüber hinaus ist m⁶A auch in der 5'-UTR, der kodierenden Sequenz sowie der 3'-UTR abundant.^[107] Heute ist bekannt, dass m⁶A die häufigste Modifikation eukaryotischer mRNA darstellt.^[107,108]

Die Funktionen von m⁶A werden über Proteine vermittelt. Seit Reversibilität und Regulation von m⁶A (**30**) in den 2000er Jahren entdeckt wurden, hat auch das Verständnis um die mit m⁶A (**30**)

wechselwirkenden Proteine große Fortschritte gemacht. Bis 2011 war lediglich ein einziger mit m⁶A wechselwirkender Protein-Komplex bekannt (METTL3-METTL14).^[109] Heute sind sowohl m⁶A-installierende (METTL3-METTL14)^[110] als auch -entfernende Proteine (ALKBH5,^[99] FTO^[98]) identifiziert, siehe Abb. 6 A). Zudem konnte gezeigt werden, dass eine dritte Protein-Klasse existiert: m⁶A-Effektoren.^[108] Diese Proteine vermitteln durch Interaktion mit m⁶A (30) dessen Funktionen. Als m⁶A-Effektoren wurden YTH-Domänen-Proteine^[90,103,111-113] sowie einige Proteine der hnRNP-Familie^[114-116] identifiziert, siehe Abb. 6 B). Darüber hinaus waren zu Beginn dieser Arbeit keine m⁶A-bindenden Proteine bekannt.^[117] Während für verschiedene Cytosin-DNA-Modifikationen neben bindenden auch spezifisch abgestoßene Proteine entdeckt wurden,^[118] war zu Beginn dieser Arbeit unbekannt ob auch m⁶A (30) Proteine abstößt.

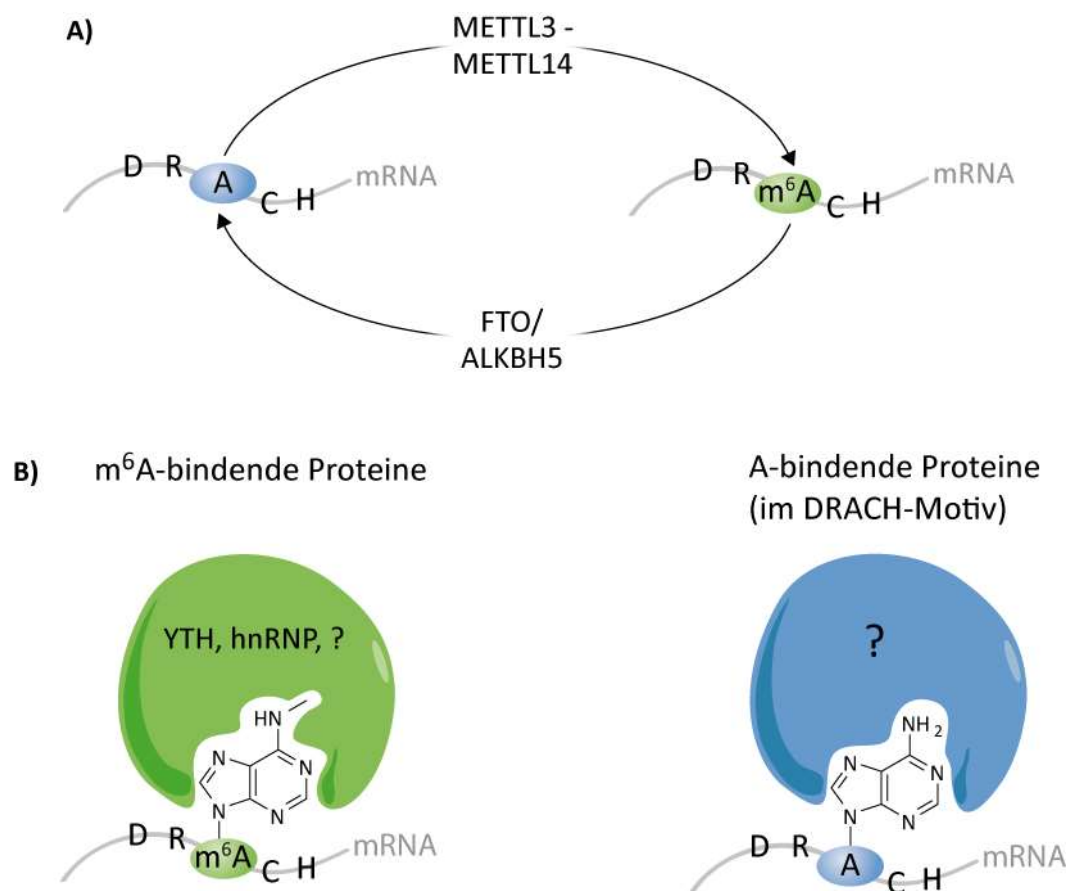


Abb. 6 A) Die Methylierung von A zu m⁶A in mRNA erfolgt durch den Proteinkomplex METTL3-METTL14. Die Demethylierung von m⁶A kann durch ALKBH5 oder FTO erfolgen. **B)** Mit m⁶A bzw. A im DRACH-Sequenzmotiv interagierende Proteine. Beide Abbildungen repräsentieren den Erkenntnisstand zu Beginn dieser Arbeit.

3 Epigenetik und die Sequenzierung von Cytosin-Modifikationen der DNA

3.1 Epigenetik

Epigenetik ist für Eukaryoten von grundlegender Bedeutung. Die Vorsilbe *epi-* stammt aus der altgriechischen Sprache und bedeutet „dazu“ oder „außerdem“.^[119] Diese Bedeutung beschreibt passend eine Kerneigenschaft der Epigenetik: Sie ist neben der Genetik (DNA-Sequenz) die zweite Informationsebene der DNA. Über Informationsträger beeinflusst Epigenetik die Aktivität von Genen und damit die Entwicklung der Zelle. Die drei etablierten Klassen epigenetischer Informationsträger sind **a)** Modifikationen der DNA-Base Cytosin, **b)** Modifikationen der Histone sowie **c)** nicht-kodierende RNAs. Wie die DNA-Sequenz selbst, können auch die epigenetischen Informationsträger vererbt werden. Damit stellen Genetik und Epigenetik die beiden Formen vererbbarer Informationen dar.^[120]

Epigenetische Informationsträger reagieren dynamisch auf Umwelteinwirkungen und spielen eine zentrale Rolle bei Krankheiten. Anpassungen des Organismus auf Ebene der DNA-Sequenz erfolgen über sehr lange Zeiträume.^[121,122] Die epigenetischen Informationsträger hingegen können auf Umwelteinwirkungen vergleichsweise dynamisch reagieren.^[122,123] Damit spielen sie eine zentrale Rolle bei der Anpassung des Organismus an mittelfristige Veränderungen der Umwelt. So können sowohl Ernährung^[124] als auch Stress^[125] und Schadstoffe^[126] unsere epigenetischen Informationsträger verändern. Die Fehlregulation epigenetischer Informationsträger bildet die Grundlage einer Vielzahl an Erkrankungen, darunter diverse Formen von Krebs^[127] sowie psychische Erkrankungen wie Depression,^[127] Angst-Störungen,^[128] Suizidneigung^[125] und Abhängigkeit^[129,130].

Methyliertes Cytosin ist der zentrale epigenetische Schalter der DNA und tritt hauptsächlich in CpGs auf. mC (5-Methylcytosin) ist in Eukaryoten der zentrale epigenetische Schalter der DNA.^[131] In Säugetieren findet sich mC hauptsächlich in CpGs (CpG-Dinukleotiden)^[132] und wird durch DNMT-Methyltransferasen gebildet^[133]. Etwa 75 % aller Methylierungen im Menschen sind CpG-Methylierungen.^[134,135] Seltener tritt mC auch im CpH-^[135-137] und GpC-Kontext^[138,139] auf. CpGs treten naturgemäß in Dyaden auf, also in miteinander gepaarten CpGs, siehe Abb. 7 A). Sequenzen mit hoher CpG-Dichte werden CpG-Inseln genannt. Betrachtet man die Gesamtheit aller CpGs, also CpG-Inseln und einzelne CpGs, so ist der Großteil der somatischen CpGs des Menschen methyliert (60-80 %).^[140] CpG-Inseln hingegen treten meist unmethyliert auf und finden sich vor allem in Promotor-Regionen. So enthalten etwa 70 % der menschlichen Promotoren CpG-Inseln, siehe Abb. 7 B).^[141,142]

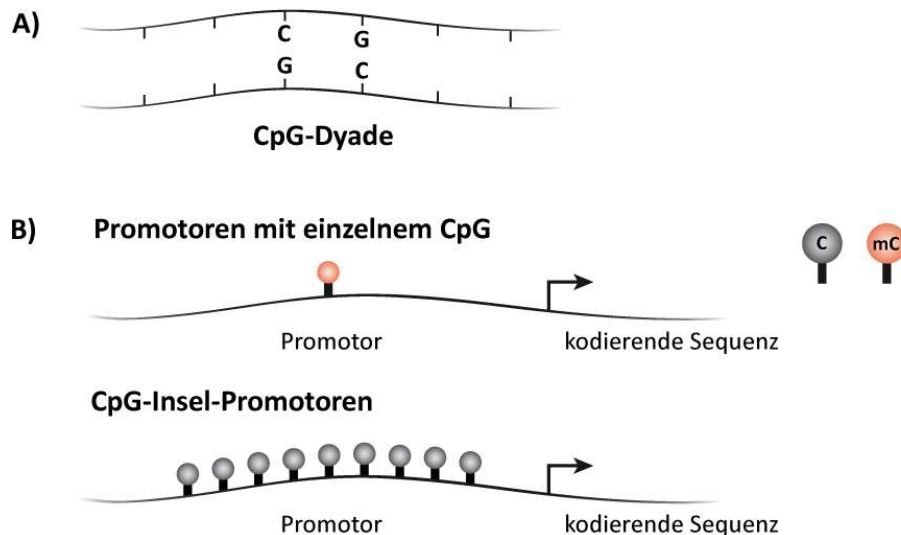


Abb. 7 A) CpG-Dinukleotide treten naturgemäß in Dyaden auf. **B)** Die CpG-vermittelte Regulation der Genexpression lässt sich nach der Art des Promotors klassifizieren: Promotoren mit einem individuellem CpG und CpG-Insel-Promotoren. CpG-Inseln treten größtenteils unmethyliert auf, einzelne CpGs im Vergleich eher methyliert.

CpG-Methylierungen regulieren die Gen-Aktivität je nach Kontext unterschiedlich. Häufig führen CpG-Methylierungen zu einer verringerten Gen-Aktivität.^[143] Relevant scheint dafür insbesondere deren Lage zu sein: So treten CpG-Methylierungen sowohl im Promotor als auch im Genkörper auf. Genkörper-Methylierungen scheinen im Vergleich eine höhere Aussagekraft in Bezug auf die Gen-Aktivität aufzuweisen.^[144] Die Regulation der Gen-Aktivität durch CpG-Methylierung im Promotor, unterscheidet sich zwischen *CpG-Insel-Promotoren* und jenen mit *einzelnen CpGs*: So beeinflussen *einzelne CpGs* hauptsächlich die Bindung des jeweiligen Transkriptionsfaktors (TFs).^[145-148] *CpG-Inseln* können zusätzlich zur Beeinflussung der TF-Bindung^[149-151] auch eine langfristige Gen-Stilllegung bewirken.^[152] Dies wird durch eine Veränderung der Chromatin-Struktur erreicht.^[153] Viele Facetten der Gen-Regulation durch CpG-Methylierungen bleiben aber weiter unklar. Zusammengefasst können CpG-Methylierungen die Gen-Aktivität mindestens über zwei Wege beeinflussen: durch TF-Abstoßung sowie über Veränderung der Chromatin-Struktur.

TET-Proteine oxidieren mdC schrittweise zu hmdC (Hydroxy-), fdC (Formyl-) und cadC (Carboxycytosin), siehe Abb. 8. Die Rolle von mdC als zentraler epigenetischer Schalter der DNA ist fest etabliert. Welchen Einfluss aber die oxidierten mdC-Derivate (ox-mdCs) auf die Gen-Aktivität haben, ist bisher größtenteils unklar. Jüngere Ergebnisse deuten darauf hin, dass auch ox-mdCs eigene epigenetischen Funktionen besitzen.^[154] Die Anreicherung von hmdC in aktiven *Enhancern* und Gen-Körpern deutet auf dessen Assoziation mit erhöhter Gen-Aktivität hin.^[155] hmdC tritt dabei besonders häufig in Neuronen sowie embryonalen Stammzellen auf,^[156] siehe Abb. 9. Auch

fdC wird als stabile^[157] sowie semipermanente^[158] Modifikation angeführt, die Bedeutung des Erhalts von fdC ist jedoch noch nicht geklärt. Interessanterweise zeigen hmdC und cadC eigene Funktionen bei der Feinjustierung der DNA-Replikation. Zudem beeinflusst deren Anwesenheit die Stabilität des Genoms.^[159,160] Wie hmdC, scheint auch cadC mit einer gesteigerter Gen-Aktivität assoziiert zu sein.^[161-163] So finden sich in Brust- und Hirn-Tumoren erhöhte cadC-Level.^[76] Die teilweise sehr geringe Zahl der ox-mdCs, siehe Abb. 9, könnte auf deren hoch-dynamische Installation und Entfernung hinweisen.^[164] Zusammengefasst ist im Vergleich zu mdC bisher nur wenig über die Funktionen der ox-mdCs bekannt. Deren Assoziation mit erhöhter Gen-Aktivität und weiteren epigenetischen Funktionen zeichnet sich aber deutlich ab.

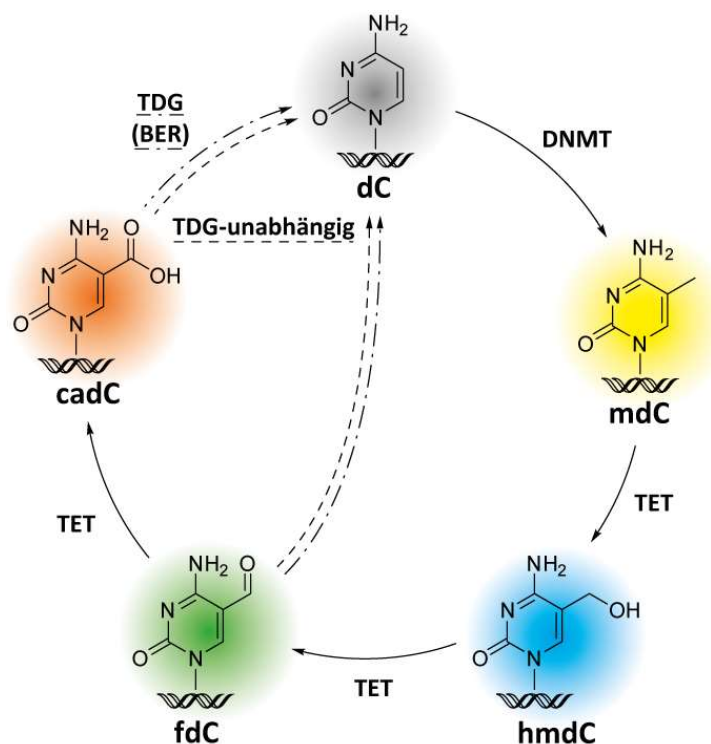


Abb. 8 Methylierung, Oxidation und aktive Demethylierung der DNA: In DNA wird dC durch DNMT-Proteine zu mdC methyliert. Dieses kann durch TET-Proteine schrittweise zu hmdC, fdC und cadC oxidiert werden. Die aktive Demethylierung kann TDG-vermittelt (Punkt-Strich) sowie TDG-unabhängig (gestrichelt) erfolgen. Die TDG-unabhängige Demethylierung erfolgt womöglich durch direkte Deformylierung von fdC oder Decarboxylierung von cadC. Diese Abbildungen basiert auf ^[156]. mdC: 5-Methylcytosin, hmdC: 5-Hydroxycytosin, fdC: 5-Formylcytosin, cadC: 5-Carboxycytosin.

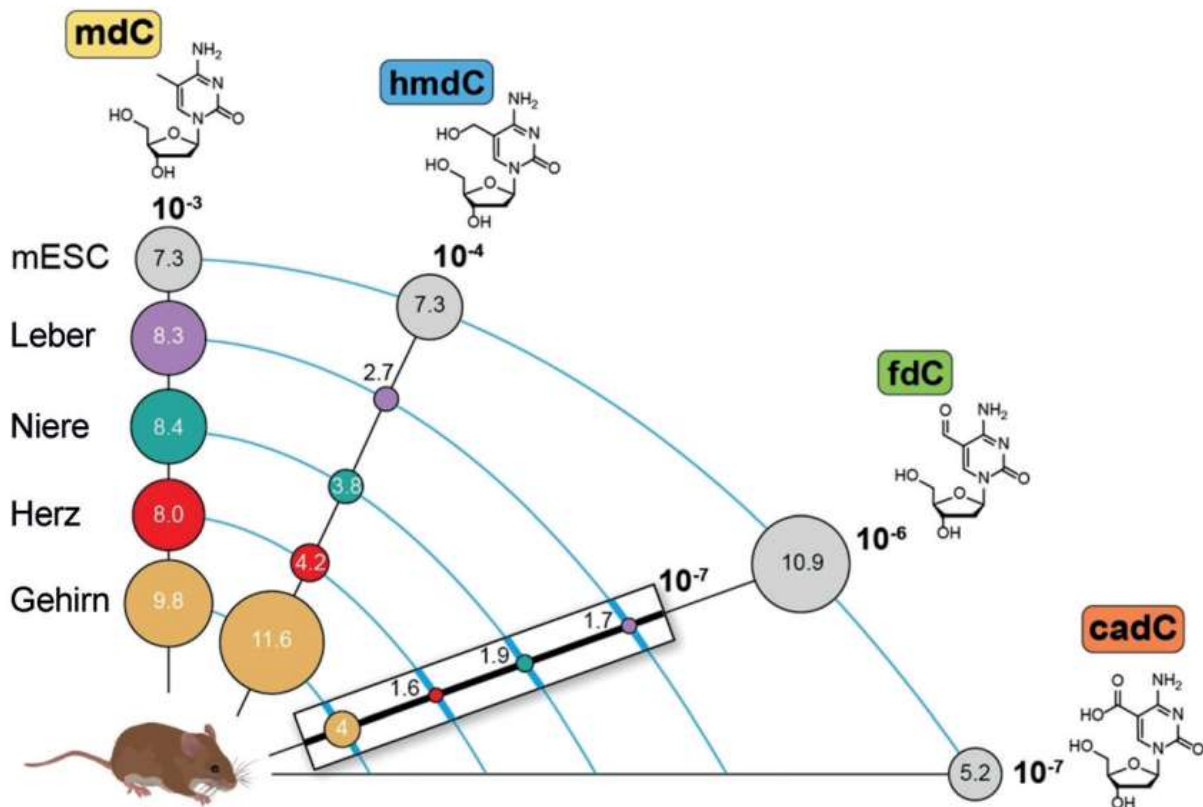


Abb. 9 Anteil der dC-Modifikationen in DNA verschiedener Gewebe sowie embryonaler Stammzellen der Maus. Die Werte geben den Anteil des modifizierten Nucleosids bezogen auf den Anteil an dG an. Werte für das Gehirn stellen die arithmetischen Mittelwerte der Anteile im Hippocampus, der Großhirnrinde und im Kleinhirn dar. Die Darstellung der fdC-Anteile in Geweben (umrahmt) sind um den Faktor 10 vergrößert. Diese Abbildungen basieren auf [156]. mDC: 5-Methylcytosin, hmdC: 5-Hydroxycytosin, fdC: 5-Formylcytosin, cadC: 5-Carboxycytosin, mESC: embryonale Stammzellen der Maus.

DNA-Demethylierung ist die Grundlage dynamischer Genregulation. Die Entfernung von mDC aus dem Genom kann sowohl durch *aktive* als auch *passive* Demethylierung erfolgen: Die *aktive* Demethylierung erlaubt die Regulation der Genaktivität unabhängig von der Zellteilung. Dazu wird zunächst mDC durch TET oxidiert. fdC oder cadC können anschließend durch die Glykosylase TDG entfernt werden. Diese sogenannte BER (Basen-Exzisions-Reparatur) ersetzt die ausgeschnittene Base durch unmodifiziertes Cytosin.^[165] Dabei wird fdC durch TDG deutlich effizienter entfernt als cadC.^[166] Die *aktive* Demethylierung ist auch unabhängig von TDG möglich,^[167] eventuell durch direkte Deformylierung von fdC oder Decarboxylierung von cadC, siehe Abb. 8.^[168-170] Die *Passive* Demethylierung beschreibt das Ausbleiben der Methylierung des neu synthetisierten Stranges bei der DNA-Replikation. Somit entstehen CpGs die nur auf einem DNA-Strang methyliert sind. Solche CpGs werden als hemi-methyliert bezeichnet, siehe Abb. 10 C).^[165]

Hemi-Methylierungen besitzen womöglich eigene epigenetische Funktionen. Nur wenige Studien haben bisher den Einfluss hemi-methylierter CpGs untersucht. Mittlerweile ist aber bekannt, dass während der frühen Embryonalentwicklung 4-18 % aller Methylierungen als hemi-CpGs vorliegen.^[171] In dieser Phase der frühen Embryonalentwicklung ändern sich hemi-Methylierungsmuster mit großer Dynamik.^[171] Interessanterweise wurden veränderte hemi-Methylierungen auch bei verschiedenen Krebsarten entdeckt, darunter in Tumoren der Nieren,^[172] Eierstöcke^[173] sowie womöglich in Brustkrebs-Zellen^[174]. Zudem konnte gezeigt werden, dass hemi-Methylierungen stabil vererbbar sind,^[171] ein Markenzeichen epigenetischer Informationsträger. Hemi-Methylierungen können außerdem zur Aktivierung der Histon-gesteuerten Gen-Stillegung führen.^[175] Sie können also die Kontrolle der Gen-Aktivität übertragen, von mdC auf Chromatin. Damit stellen hemi-Methylierungen ein Bindeglied zwischen beiden epigenetischen Informationsträgern dar.^[175] Daran anknüpfend wurde entdeckt, dass hemi-Methylierungen ebenso Chromatin-Interaktionen regulieren können.^[171] Auch ein Effekt auf die Gen-Aktivität konnte bereits beobachtet werden: So scheinen hemi-Methylierungen im Genkörper mit erhöhter genetischer Aktivität zu korrelieren, im Vergleich zu Genkörpern mit symmetrisch-methylierten Positionen.^[171,175] Hemi-Methylierungen im Promotor hingegen korrelieren mit verringerter genetischer Aktivität.^[171] Zusammengenommen könnten auch hemi-Methylierungen eigene epigenetischen Funktionen aufweisen.^[176]

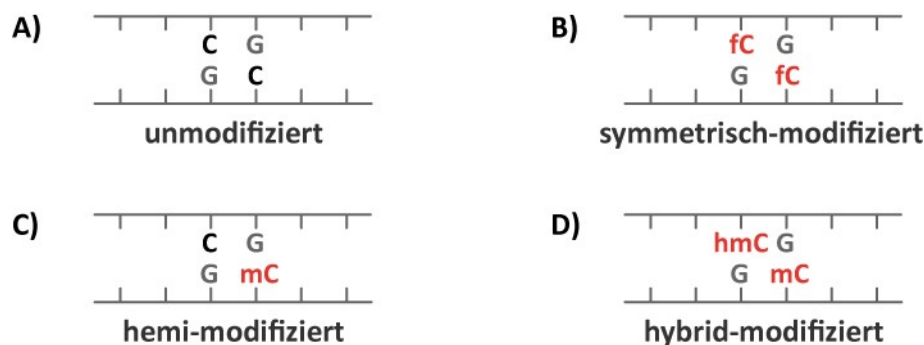


Abb. 10 Die vier Formen der CpG-Modifizierung: **A)** unmodifiziert **B)** symmetrisch-modifiziert **C)** hemi-modifiziert **D)** hybrid-modifiziert. Beispielhafte Auswahl verschiedener dC-Modifikationen. Die dargestellten Modifikationen wurden zufällig gewählt. C: dC, mC: mdC, hmC: hmdC, fc: fdC, G: dG.

Auch hemi-hmdC, -fdC und -cadC besitzen bisher unbekannt Funktionen. Der größte Teil aller hemi-CpGs sind methyliert, sie enthalten also mdC.^[171] Aber auch hmdC findet sich an hemi-modifizierten Positionen: Interessanterweise existieren nahezu alle hmdCs während der frühen Embryonalentwicklung der Maus als hemi-hmdC.^[171] Eine weitere Studie zeigt, dass hmdC in Geweben und embryonalen Stammzellen der Maus hauptsächlich gegenüber mdC aufzutreten scheint, also in hmdC-mdC-Hybriden, siehe Abb. 10 D).^[177] Solche gemischt modifizierten hybrid-CpGs sind auch mit fdC und cadC wahrscheinlich.^[164] Die Häufigkeit dieser

und weiterer hybrid-CpGs ist bisher aber nicht bekannt. Interessanterweise konnte, wie für hemi-mdC, auch für hemi-hmdC eine Vererbbarkeit gezeigt werden.^[178] Zusammengenommen wurden die Funktionen von hemi-hmdC, -fdC und -cadC bisher erst in wenigen Studien untersucht. Deren Ergebnisse deuten aber darauf hin, dass auch diese hemi-Modifikationen eigene Funktionen besitzen, hemi-hmdC sogar epigenetische.

Die nächsten Jahre lassen einen drastischen Wissenszuwachs in Bezug auf hemi-Modifikationen erwarten: Insbesondere durch den Einsatz neuer Technologien wie der Hairpin-Sequenzierung,^[179,180] FRET^[177] sowie Sequenziermethoden der 3. Generation^[181] zusammen mit neuen bioinformatischen Werkzeugen^[182]. Im folgenden Abschnitt werden Methoden zur Sequenzierung verschiedener Cytosin-Modifikationen erläutert.

3.2 Methoden zur Sequenzierung modifizierter Cytosine

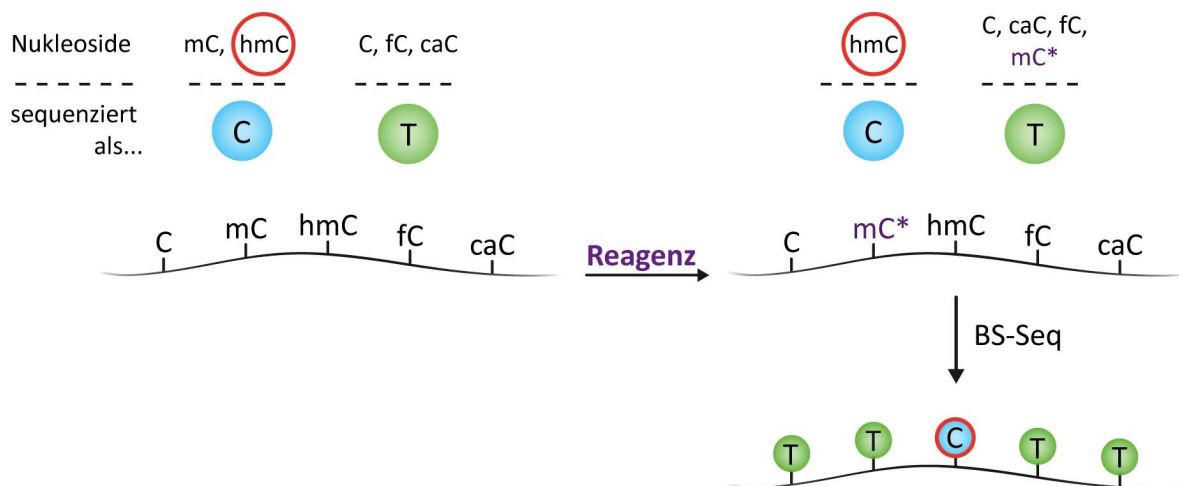
Die konventionelle DNA-Sequenzierung mittels next generation sequencing analysiert ausschließlich die genetische Informationsebene. Die epigenetische Informationsebene kann demnach mit konventionellen Methoden nicht ausgelesen werden. So ist eine Unterscheidung zwischen dC und dessen Modifikationen mdC, hmdC, fdC und cadC nicht möglich. Ein detailliertes Verständnis der Funktionen dieser Modifikationen kann aber nur erreicht werden, wenn deren Position im Genom basengenau bestimmt werden kann.

BS-Seq-Methoden ermöglichen die Sequenzierung von mdC sowie hmdC, fdC und cadC. FROMMER *et al.* entwickelten bereits 1992 eine Methode zur PCR-basierten Sequenzierung von mdC.^[183] Weiterentwicklungen dieser Methode werden unter dem Begriff Bisulfit-Sequenzierung (BS-Seq) zusammengefasst. BS-Seq ermöglicht die Sequenzierung von mdC durch Umwandlung der *epigenetischen* in eine *genetische* Information. Diese Umwandlung basiert auf der BS-vermittelten Desaminierung von dC zu dU. Nach der Behandlung der DNA mit Bisulfit (BS) wird dC somit als T gelesen, mdC aber weiterhin als C.

Heute existieren Methoden zur gezielten Sequenzierung aller ox-mdCs (hmdC, fdC und cadC). Abb. 11 A) zeigt, dass sich die BS-Reaktivität von dC und dessen Modifikationen in zwei Klassen aufteilt: 1.) Basen, die *nicht* durch BS *desaminiert* werden (mdC, hmdC). Diese werden als *C* sequenziert. Sowie 2.) Basen, die durch BS *desaminiert* werden (dC, fdC, cadC). Diese werden als *T* sequenziert. Dieser Unterschied in der BS-Reaktivität stellt die Basis der BS-Seq im Allgemeinen dar, erlaubt in dieser Form aber keine selektive Sequenzierung einzelner ox-mdCs. Zur selektiven Sequenzierung muss die BS-Reaktivität einer oder mehrerer Basen gezielt verändert werden. Dazu erfolgt die Behandlung mit einem Reagenz, welches die BS-Reaktivität eines oder mehrerer ox-mdCs gezielt beeinflusst.

Die BS-Seq einer dC-Modifikation kann entweder durch vergleichende oder direkte Sequenzierung erreicht werden. Zur vergleichenden Sequenzierung wird die BS-Reaktivität der Zielmodifikation mit einem Reagenz geändert, so dass sie in eine andere Klasse fällt, siehe Abb. 11 B). Im Anschluss werden sowohl die behandelte als auch unbehandelte DNA mittels BS-Seq analysiert. Ein Vergleich der Sequenzen von unbehauelter und behandelter DNA zeigt dann die Positionen der Zielmodifikation. Die direkte Sequenzierung einer Modifikation ist hingegen möglich, wenn mit einem Reagenz erreicht werden kann, dass die Zielmodifikation allein in einer Klasse vorliegt. Nach Behandlung mit dem Reagenz, kann die Zielmodifikation dann direkt mittels BS-Seq detektiert werden, siehe Abb. 11 A).

A) direktes Auslesen (Bsp. hmdC)



B) vergleichende Sequenzierung (Bsp. fdC)

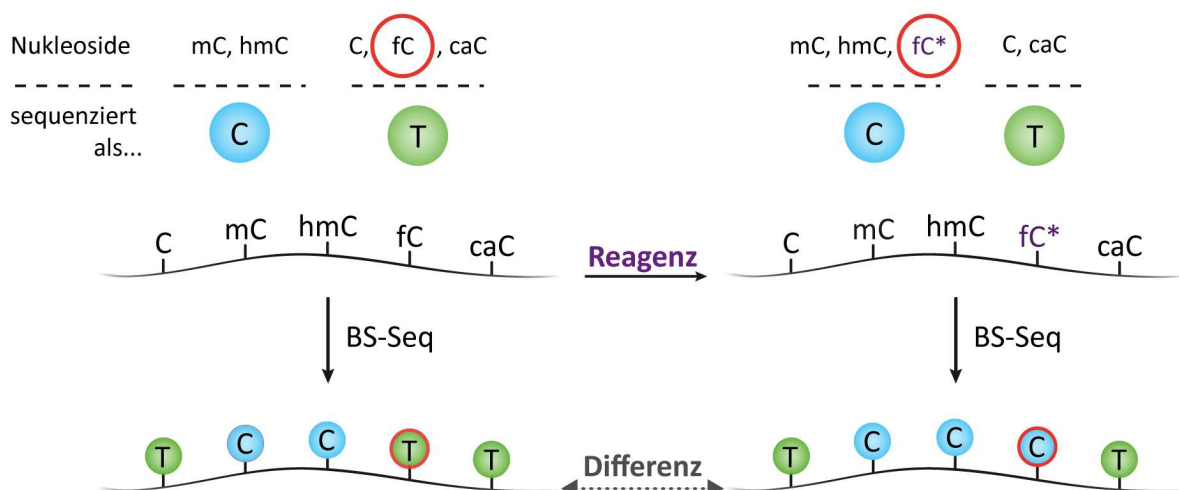


Abb. 11 Zwei Arten der BS-Seq: **A)** direktes Auslesen und **B)** vergleichende Sequenzierung. **A)** direktes Auslesen von hmdC am Beispiel der TAB-Sequenzierung (TET-assistierte BS-Seq): Zunächst wird hmdC mit Hilfe einer Glycosyltransferase vor dem nachfolgenden Reagenz (TET) geschützt. Anschließend wird mdC mittels TET in cadC überführt (mC*), welches BS-vermittelt desaminiert. So verbleibt ausschließlich hmdC als nicht via BS desaminierbares Nucleosid und kann direkt ausgelesen

werden. **B)** Vergleichende Sequenzierung von fdC am Beispiel von fCAB (fdC chemisch assistierte BS-Seq): Das Reagenz überführt fdC in ein Derivat, welches nicht BS-vermittelt desaminiert. Da die Zielmodifikation fdC in dieser Klasse (als C sequenziert) nicht isoliert vorliegt, ist eine Referenz-Sequenzierung derselben unbehandelten DNA nötig. Durch bioinformatischen Abgleich („Differenz“) können so die Zielmodifikation bestimmt werden. * derivatisiertes Nukleosid; C: dC, mc: mC, hmC: hmdC, fc: fdC, caC: cadC; BS-Seq: Bisulfit-Sequenzierung. **Reagenz:** Derivatisierung eines oder mehrerer Nucleoside mit dem Ziel der BS-Seq der Zielmodifikation. **rot umkreist:** Zielmodifikation

Das BS-Reagenz führt zu DNA-Degradation, unvollständiger Umwandlung und Einzelstrang-abhängigkeit. Die Reaktion mit BS (SO_3^{2-}) erfolgt bei pH 5 zwischen 50 °C und 99 °C.^[184] Diese harschen Reaktionsbedingungen verursachen eine sequenzspezifische Degradation der DNA.^[185] Gleichzeitig ist eine möglichst vollständige Umwandlung der Basen, insbesondere von fdC, nur unter harschen Reaktionsbedingungen möglich.^[184] So stehen sich geringe DNA-Degradation und vollständige Desaminierung unvereinbar gegenüber, siehe Abb. 12 A). Gleichzeitig ist die BS-Umwandlung einzelstrangabhängig.^[186-188] Das heißt, Basen in doppelsträngiger DNA reagieren mit BS erst nach DNA-Denaturierung, siehe Abb. 12 B). Da bei konventionellen BS-Seq-Methoden keine Zuordnung der komplementären Einzelstränge nach der Denaturierung möglich ist, können hemi- sowie hybrid-modifizierte CpG-Dyaden (vgl. Abb. 10) nicht als solche detektiert werden. Darüber hinaus führt die BS-vermittelte Desaminierung zu einer Verringerung der als C sequenzierten Positionen. Diese verringerte Sequenz-Komplexität erschwert zusätzlich die Zuordnung des sequenzierten Fragments zu einem Abschnitt des Genoms.

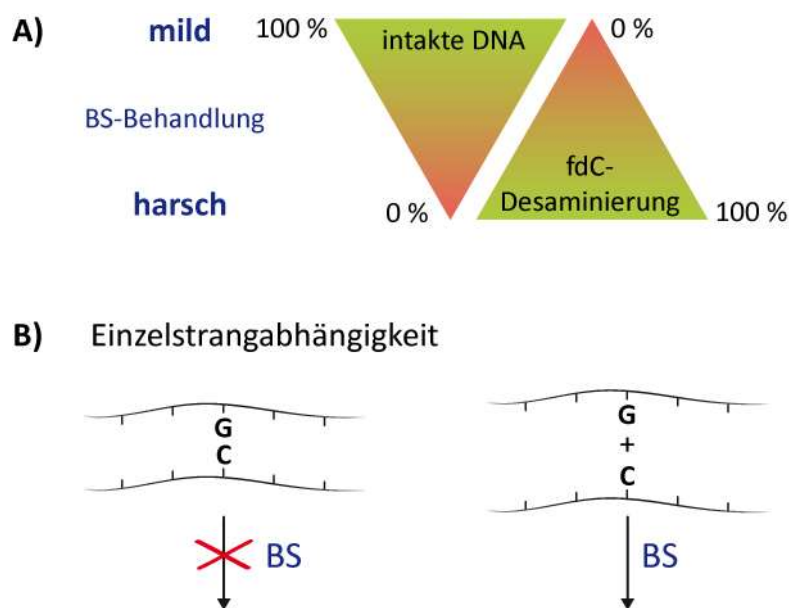


Abb. 12 A) Einfluss milder/harscher BS-Behandlung auf die Integrität der DNA-Probe sowie den Anteil der fdC-Desaminierung. **B)** Die BS-vermittelte Desaminierung verläuft nur an DNA-Einzelsträngen (rechts) mit nützlichen Ausbeuten.^[186-188]

Auch ununterscheidbare reads führen zu Problemen bei der BS-Seq. Konventionelle BS-Seq-Methoden sind auf eine Amplifikation mittels PCR angewiesen. Ein unabhängiger *read* entsteht durch Sequenzierung eines individuellen DNA-Moleküls. Die Sequenzierung mehrerer PCR-Amplicons des gleichen DNA-Moleküls steigert hingegen die Tiefe der Sequenzierung nicht. Eine Amplifikation ohne Markierung der *reads* führt demnach im Ergebnis dazu, dass keine Quantifizierung der Basen mehr möglich ist. Dies beruht darauf, dass PCR-Amplicons desselben DNA-Fragments nicht voneinander unterscheidbar sind. Auch *reads* von unabhängigen DNA-Fragmenten mit gleicher Sequenz sind bei konventioneller BS-Seq nicht voneinander unterscheidbar. Somit bleibt die reale Tiefe der Sequenzierung und damit die Verlässlichkeit der Ergebnisse der BS-Seq unklar. Zusammengenommen können konventionelle BS-Seq-Methoden, ohne Kontroll- und Korrektur-Mechanismen, den tatsächlichen Modifikationsgrad nicht verlässlich bestimmen.

Die BS-Seq lässt sich entsprechend der Sequenziertiefe in drei BS-Seq-Techniken unterteilen: *whole genome* (WGBS), *reduced representation* (RRBS) und lokale Tiefensequenzierung, siehe Abb. 13. Die Tiefe einer Sequenzierung bezieht sich auf die Anzahl unabhängiger *reads* pro Base. Durch die harschen Bedingungen der BS-Seq sowie die verringerte Sequenz-Komplexität, erfordert die Zuordnung des sequenzierten DNA-Fragments nach BS-Seq mehr *reads* als dies für BS-freie Sequenzierungen der Fall ist. BS-Seq des gesamten Genoms werden als WGBS bezeichnet. Diese Methode erfasst alle sequenzierbaren Modifikationen im Genom, wird aber üblicherweise mit der geringsten Sequenziertiefe ausgeführt. RRBS ermöglicht eine größere Sequenziertiefe: Durch enzymatische Fragmentierung des Genoms und Anreicherung CpG-reicher Abschnitte können 10-20 % der 28 Mio. CpGs des menschlichen Genoms sequenziert werden.^[189] Tiefe (> 100 *reads*) und ultra-tiefe (> 1000 *reads*) BS-Seq können nur durch lokale Tiefensequenzierung erreicht werden.^[190] Dazu werden gezielt einzelne Genloci durch Restriktionsenzyme isoliert und mittels BS-Seq analysiert. Die hohe Anzahl an *reads* ermöglicht es so den Modifikationsgrad einer einzelnen Position mit hoher Genauigkeit zu bestimmen.

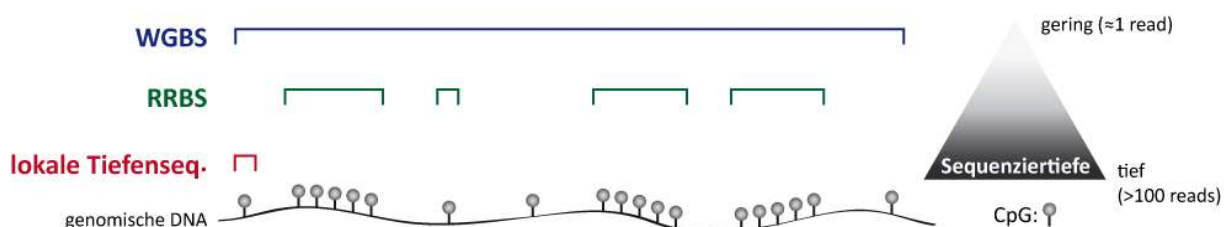


Abb. 13 Abdeckung des Genoms und Sequenziertiefe der wichtigsten BS-Seq-techniken: WGBS (*whole genome*), RRBS (*reduced representation*) und lokale Tiefensequenzierung.

Die Sequenzierung von fdC ist besonders anspruchsvoll. Dies begründet sich insbesondere in der geringen Häufigkeit von fdC im Genom in Kombination mit den generellen Herausforderungen der konventionellen BS-Seq, vgl. Abb. 12. Bisher existieren zwei BS-basierte Methoden: fCAB^[191] und redBS^[192], siehe Abb 14. fCAB basiert auf der Derivatisierung von fdC mit *O*-Ethylhydroxylamin. Das so gebildete Oxim-Derivat desaminiert in der Reaktion mit BS nicht, wird also als C gelesen. redBS reduziert fdC hingegen mit NaBH₄ zu hmdC, welches ebenfalls als C gelesen wird. Daneben existieren zwei Methoden zur basengenauen Bisulfit-freien Sequenzierung von fdC: CLEVER-Seq^[193] sowie eine ddPCR (*droplet digital PCR*)-basierte Methode^[158] (hier: rU*): CLEVER-Seq derivatisiert fdC mit Malononitril über eine FRIEDLÄNDER-Reaktion. Das so gebildete fdC-Derivat wird als T sequenziert. Die ddPCR-Methode basiert auf der Paarung mit einem modifizierten DNA-Strang. Dieser reagiert über einen Hydroxylamin-Linker mit fdC an der Ziel-Position. Keine der vier Methoden konnte sich jedoch bisher zur Sequenzierung von fdC etablieren. Auch kommerzielle Produkte dieser Methoden sind noch immer nicht verfügbar. Das Kernproblem der BS-basierten Methoden ist insbesondere die unvollständige Umsetzung von fdC mit dem jeweiligen Reagenz.^[194,195] Die Bisulfit-freien Methoden sind hingegen anspruchsvoll in der Durchführung:^[194,195] So benötigt der Umgang mit Malononitril idealerweise chemisches Hintergrundwissen. Die ddPCR-Methode verlangt ein teures und wenig verbreitetes ddPCR-Gerät. Zusammengefasst, hat sich bisher keine Methode zur exakten Sequenzierung von fdC etabliert.

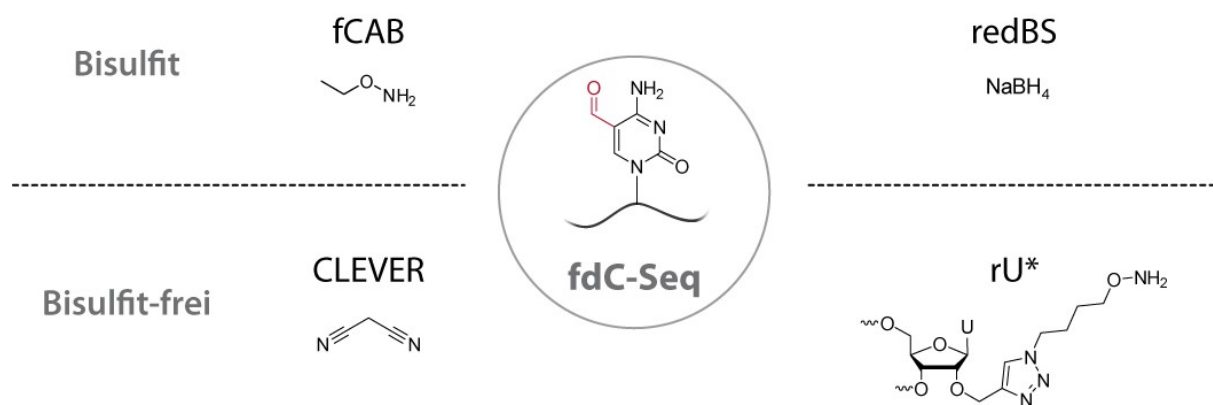


Abb 14 Auswahl an Methoden zur fdC-Sequenzierung. Dargestellt sind jeweils die Reagenzien. fCAB^[191] und redBS^[192] basieren auf Bisulfit-vermittelter Desaminierung; CLEVER-Seq^[193] und die ddPCR (*droplet digital PCR*)-basierte Methode (hier: rU*)^[158] erfolgen ohne Bisulfit. fc: fdC.

Es gibt Möglichkeiten die Probleme der BS-Seq zu adressieren: Referenznucleoside erlauben es die Umsatzraten der Zielmodifikation zu untersuchen. Bei der BS-Seq von ox-mdCs werden häufig synthetische Oligonucleotide als Referenz eingesetzt, sogenannte *spike-ins*. Diese *spike-ins* enthalten die Zielmodifikation als Referenznucleosid, siehe Abb. 15 A). Anhand der Umwandlungsrate der Referenznucleoside erfolgt eine Abschätzung der wahren

Umwandlungsrate in der DNA-Probe. Trotz dieser Vorteile weisen *spike-ins* elementare Probleme auf: So werden *spike-ins* in der Regel mittels PCR hergestellt.^[195] Da PCR keine Unterscheidung zwischen dC und der dC-Zielmodifikation ermöglicht, werden dC-Positionen zufällig durch die dC-Zielmodifikation ersetzt. Dies kann zu einer unnatürlichen Anzahl und Verteilung der Zielmodifikation im *spike-in* führen. Zudem scheint es verbreitet *spike-ins* nicht gemeinsam mit der DNA-Probe der BS-Behandlung zu unterziehen, sondern in einer separaten Reaktion.^[163,193] Teilweise wird das *spike-in* darüber hinaus mit einer anderen Methode sequenziert.^[163,193] Separat behandelte und sequenzierte *spike-ins* stellen folglich durch die gegebenenfalls abweichenden Bedingungen keine verlässliche Referenz dar. Zusammengenommen ermöglichen *spike-ins* nur bei optimalem Einsatz eine Abschätzung der Umwandlungsrate der Zielmodifikation.

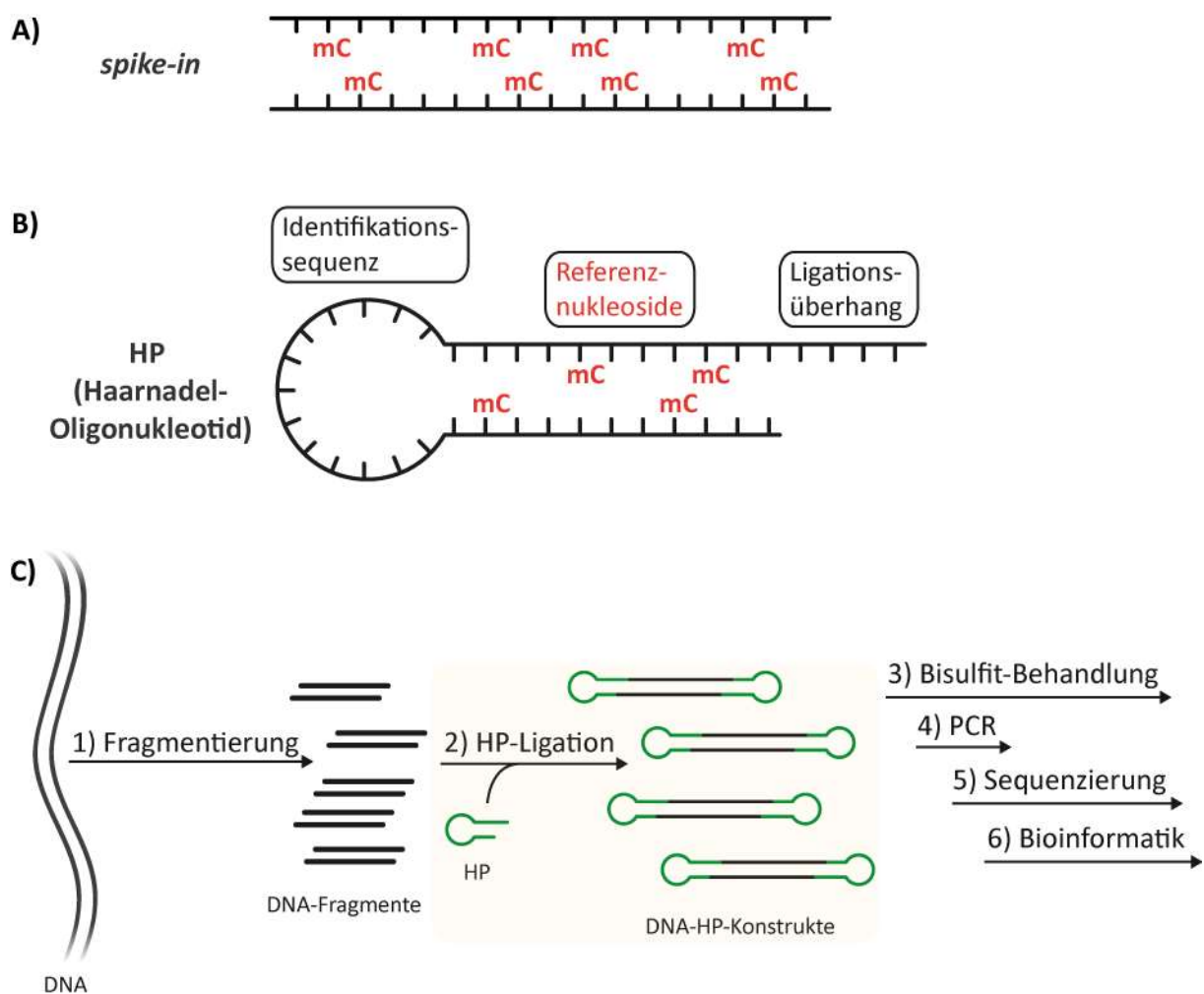


Abb. 15 **A)** Schematische Darstellung eines beispielhaften *spike-ins*, erzeugt mittels PCR. Durch Einsatz eines Triphosphats des Referenznucleosids anstelle des dC-Triphosphats, werden alle dC-Positionen durch das Referenznucleosid ersetzt. **B)** Schematische Darstellung eines beispielhaften Haarnadel-Oligonukleotids (HP) zur Einzelstrang-spezifischen Bisulfit-Sequenzierung (BS-Seq). **C)** Schritte der HPBS (*hairpin BS-Seq*). Die Ligation der HPs (Haarnadel-Oligonukleotide) ist orange hinterlegt. Alle weiteren Schritte entsprechen der konventionellen Bisulfit-Tiefensequenzierung. HP: Haarnadel-Oligonukleotid. mC: mdC.

*Einzelsträngige Haarnadel-Oligonukleotide (HPs) können als Weiterentwicklung der spike-ins betrachtet werden, siehe Abb. 15 B). Sie enthalten sowohl dC als auch die Zielmodifikation als Referenzen im Stamm-Bereich. Darüber hinaus bietet die Schleife der Haarnadel Raum für eine individuelle Identifikationssequenz.^[196] Der 5'-Überhang entspricht einer Endonuklease-Schnittstelle. BS-Seq unter Verwendung von HPs wird als HPBS (*hairpin BS-Seq*) bezeichnet.^[197] Zur HPBS wird die DNA-Probe zunächst mit einer passenden Endonuklease fragmentiert, siehe Abb. 15 C). Anschließend werden die DNA-Abschnitte mittels Ligation kovalent mit den Ligationsüberhängen der HPs verbunden.*

Haarnadel-Oligonukleotide können die Probleme der BS-Seq adäquat adressieren, siehe Abb. 16. Durch Kombination der DNA-Abschnitte mit den individuellen Identifikationssequenzen der HPs, kann jeder DNA-Abschnitt einzeln identifiziert werden. Dies ermöglicht zum einen die exakte Quantifizierung der Zielmodifikation. Zum anderen wird so die Bestimmung der realen Sequenzierertiefe möglich, denn die Identifikationssequenzen der HPs ermöglichen eine Unterscheidung zwischen individuellen DNA-Molekülen und PCR-Amplicons dieser DNA-Moleküle. Darüber hinaus ermöglicht die kovalente Verknüpfung beider DNA-Einzelstränge auch die Analyse von CpG-Dyaden, da die komplementären Stränge einander nun eindeutig zugeordnet werden können. So können sowohl hemi- als auch hybrid-modifizierte CpG-Dyaden detektiert und von voll- oder nicht-modifizierten Dyaden unterschieden werden, vgl. Abb. 10. Die Vorteile der HPBS basieren demnach insbesondere auf der kovalenten Verknüpfung beider DNA-Einzelstränge mittels HPs. Zusammengenommen ermöglicht die HPBS neben einer verlässlichen Beurteilung der Sequenzierungs-Ergebnisse auch die Quantifizierung der Zielmodifikation sowie die Analyse von CpG-Dyaden.

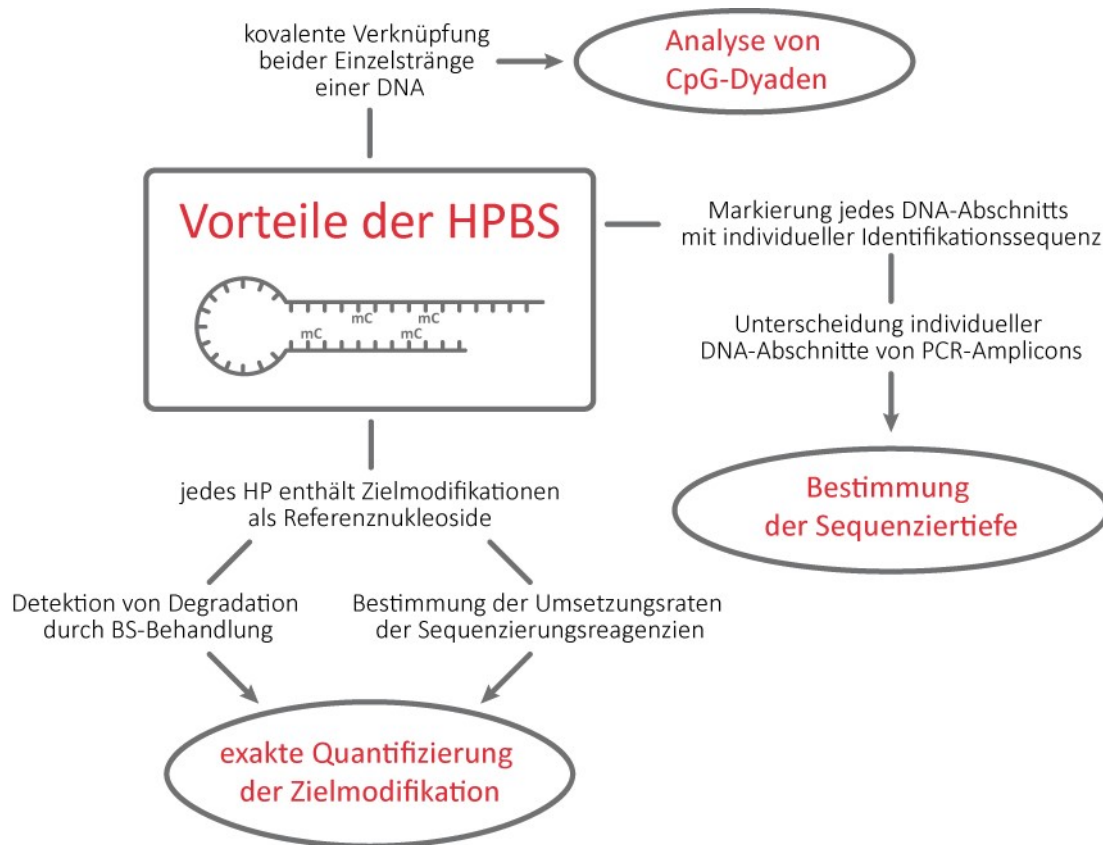


Abb. 16 Vorteile der Haarnadel-Bisulfit-Sequenzierung (HPBS).

Bisher existieren keine HPBS-Methoden zur Sequenzierung von fdC und cadC. LAIRD *et al.* entwickelten 2004 die erste PCR-basierte Analyse von mdC unter Verwendung von HPs.^[179] 2018 folgte die Gruppe um WALTER mit der Entwicklung der HPoxBS, einer BS-Seq von hmdC unter Verwendung von HPs.^[180] Methoden zur BS-Seq von fdC und cadC unter Verwendung von HPs wurden hingegen bisher nicht beschrieben.

B Ergebnisse

4 Untersuchungen zum chemischen Ursprung kanonischer und nicht-kanonischer RNA-Nukleoside

Eine vereinigte Bildung aller kanonischen RNA-Nukleoside konnte trotz vieler Versuche bis heute nicht etabliert werden. Hingegen wurde die Bildung nicht-kanonischer RNA-Nukleoside unter präbiotisch plausiblen Bedingungen bisher nur wenig untersucht. Dieser Teil der Arbeit adressiert mögliche chemische Ursprünge sowohl der kanonischen als auch der nicht-kanonischen RNA-Nukleoside. Es sollten kompatible Synthesen der kanonischen sowie potenziell wichtiger nicht-kanonischer RNA-Nukleoside etabliert werden. Abschnitt 4.1 beschreibt den ersten Zugang zur vereinigten Bildung aller kanonischen RNA-Nukleoside in präbiotischen Modellreaktionen. Abschnitt 4.2 fasst anschließend Ansätze zur Bildung methylierter RNA-Nukleoside sowie Thio-RNA-Nukleoside zusammen.

4.1 Vereinigte präbiotisch plausible Synthese von Pyrimidin- und Purin-Nukleotiden der RNA

„Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides”
S. Becker, J. Feldmann, S. Wiedemann, H. Okamura, C. Schneider, K. Iwan, A. Crisp, M. Rossa, T. Amatov, T. Carell, *Science*. **2019**, *366*, 76–82.

Prolog

Eine vereinigte präbiotisch plausible Bildung kanonischer Pyrimidin- und Purin-Nukleoside der RNA konnten bisher nicht gefunden werden, vgl. Einleitung Abschnitt 1.3. In dieser Arbeit gelang es nun erstmals Bedingungen zu etablieren, die eine vereinigte Bildung von Pyrimidin- und Purin-Nukleosiden in präbiotischen Modellreaktionen ermöglichen. Dazu wurde ein neuartiger Zugang zu den Pyrimidin-Nukleosiden via 3-Aminoisoxazol als Schlüsselmolekül erschlossen. Anschließend gelang die Vereinigung dieses Synthesewegs mit der Formamidopyrimidin-(FaPy-)basierten Bildung der Purin-Nukleoside.^[53] Darüber hinaus konnte die selektive Bildung der Pyrimidin-mono- sowie -dinukleotide in einer Eintopfreaktion erzielt werden.

Autorenbeitrag

Für diese Studie habe ich zum einen Referenzverbindungen mittels herkömmlicher Synthese hergestellt. Dabei handelte es sich um verschiedene Vorläuferverbindungen der Furanosyl- und Pyranosyl-Isomere von *N*¹-Ribosyl-*N*²-(isoxazol-3-yl)harnstoff. Darüber hinaus habe ich

Untersuchungen zu Reduktion und Ringschluss verschiedener Pyrimidin-Nukleosid-Vorläufer unter präbiotisch plausiblen Bedingungen durchgeführt.

Lizenz

Kopie der Publikation mit Erlaubnis des Verlags

PREBIOTIC CHEMISTRY

Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides

Sidney Becker^{1,2*}, Jonas Feldmann^{1*}, Stefan Wiedemann^{1*}, Hidenori Okamura^{1,3}, Christina Schneider¹, Katharina Iwan^{1,4}, Antony Crisp¹, Martin Rossa¹, Tynchtyk Amatov^{1,5}, Thomas Carell^{1†}

Theories about the origin of life require chemical pathways that allow formation of life's key building blocks under prebiotically plausible conditions. Complex molecules like RNA must have originated from small molecules whose reactivity was guided by physico-chemical processes. RNA is constructed from purine and pyrimidine nucleosides, both of which are required for accurate information transfer, and thus Darwinian evolution. Separate pathways to purines and pyrimidines have been reported, but their concurrent syntheses remain a challenge. We report the synthesis of the pyrimidine nucleosides from small molecules and ribose, driven solely by wet-dry cycles. In the presence of phosphate-containing minerals, 5'-mono- and diphosphates also form selectively in one-pot reactions. The pathway is compatible with purine synthesis, allowing the concurrent formation of all Watson-Crick bases.

The discovery of catalytic RNA (1) and the development of replicating RNA systems (2, 3) have lent strong support to the concept of an RNA world (4). The RNA world hypothesis predicts that life started with RNAs that were able to (self-)recognize and replicate. Through a process of chemical evolution, a complex RNA and later RNA-peptide and protein world supposedly evolved, from which life ultimately emerged (4). A prerequisite for the RNA world is the ability to create RNA under prebiotic conditions. This requires as the first elementary step the concurrent formation of pyrimidine and purine nucleosides in the same environment. They must have condensed to form information-carrying polymers able to undergo Darwinian evolution. The question of how the pyrimidine and purine nucleosides could have formed together is an unsolved chemical problem that is under intensive chemical investigation (5–9). Starting from an early atmosphere mainly composed of N₂ and CO₂ (10), the abiotic synthesis of life's building blocks must have occurred on the early Earth in aqueous environments, whose characteristics were determined by the minerals and chemical elements from which the early Earth's crust was made (11, 12). Atmospheric chemistry, impact events, and volcanic activities must have provided the first reactive small molecules. These reacted in surface or deep-sea hydrothermal vents (13–15), on mineral surfaces (16), or in shallow ponds (17).

Within these environments, volcanic activity, and seasonal or day-night cycles caused fluctuations of pH and temperature. Such fluctuations provided wet-dry conditions allowing precipitation or crystallization of chemicals (18). Mixing of microenvironments may have opened up new reaction pathways that led to increasing chemical complexity.

Along these geophysical boundaries, two main reaction pathways have been proposed for the formation of purine and pyrimidine nucleosides. The synthesis of the purines is possible along a continuous pathway based on the reaction of formamidopyrimidine (FaPy) precursors with ribose (6, 18). For the pyrimidines, a reaction sequence involving aminooxazoles has been discovered (5). These pathways provide the corresponding nucleosides under very different and partially incompatible conditions, leaving unanswered the question of how purines and pyrimidines could have formed in the same environment. Here, we report a prebiotically plausible pathway to pyrimidine nucleosides that selectively provides the 5'-mono- and 5'-diphosphorylated nucleosides needed for RNA strand formation. By connecting the pathway with the reported purine route (6, 18), we establish a unifying reaction network that allows for the simultaneous formation of both types of nucleosides in the same environment and that is driven by wet-dry cycles.

Results

Prebiotically plausible synthesis of pyrimidine nucleosides

The chemistry leading to pyrimidines starts from cyanoacetylene **1** as the key building block (Fig. 1A). Compound **1** is observed in interstellar clouds and in the atmosphere of Titan (19). It has been shown to form in large quantities by electric discharge through a CH₄-N₂ atmosphere (20) and is also a product

of the Cu(II)-mediated reaction of HCN and acetylene in water (Fig. 1B) (21). A recent report suggested that molecules such as **1** are plausible prebiotic starting materials which could have formed in surface hydrothermal vents in significant concentrations (13). We found that **1** reacts quickly and cleanly with hydroxylamine **2** or hydroxylurea **3** to give 3-aminoisoxazole **4**. The reaction of **1** with **3** proceeds under slightly basic conditions (pH ~10) with 80 to 90% yield within 2 hours. **3** is formed in almost quantitative yields from the reaction of **2** with cyanate (22). Compound **4** formed robustly even if we varied the temperature (10° to 95°C), the reactant concentrations (10 to 100 mM), or added additional compounds, such as urea **5** and/or different metal ions (see below). Reaction of cyanoacetylene **1** with hydroxylamine **2** produced **4** with 17% yield after 2 hours at pH 10.

While hydroxylamine **2** is an accepted building block for prebiotic amino acid syntheses (23), its potential formation on the early Earth is unclear. We therefore aimed to demonstrate its prebiotic availability. **2** is ultimately produced by reduction from NO, which is formed in large quantities when lightning passes through moist atmospheres containing N₂ and CO₂ (Fig. 1B) (10). NO forms as the main product under these conditions and spontaneously reacts in the presence of water to form nitrite (NO₂⁻) and nitrate (NO₃⁻), and this leads to the assumption that both anions were quite abundant on the early Earth (24–26). With Fe(II) as a plausible prebiotic reductant, NO₂⁻ is converted to NH₃ but not to NH₂OH **2** (26). Formation of the latter requires a partial reduction. We found that this can be achieved with HSO₃⁻, which forms from volcanic SO₂ and water (27). NO₂⁻ and HSO₃⁻ react to **2** with up to 85% yield (Fig. 1B and fig. S1) (28). We confirmed that this reaction gives first hydroxylamine disulfonate **6** (Fig. 1B), which hydrolyzes to hydroxylamine **2** and HSO₄⁻. We found that intermediate **6** reacts with cyanoacetylene **1** as well (88% yield) (Fig. 1B and fig. S2) to give the stable olefin **7**, which upon hydrolysis provides again the key intermediate **4**. The overall yield of **4** via compound **7** is 63% over these two steps. The suggested pyrimidine intermediate **4** is therefore readily available from cyanoacetylene **1** upon reaction with either **2**, **3**, or **6** under prebiotic conditions (Fig. 1B).

When we added urea **5** to a solution of **4**, warming (70° to 95°C) and dry-down resulted in formation of *N*-isoxazolyl-urea **8** (Figs. 1A and 2A) in a spot-to-spot reaction that is catalyzed by Zn²⁺ or Co²⁺. These metal ions were likely present on the early Earth (11, 12). In the presence of Zn²⁺, compound **8** is formed in 88% yield after 2 days at 95°C (at 70°C, the same yield is obtained after ~2 to 3 weeks). With Co²⁺, 68% yield is achieved after 2 days

¹Center for Integrated Protein Science, Department of Chemistry, LMU München, Butenandtstrasse 5-13, 81377 München, Germany.

²Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. ³Institute for Multidisciplinary Research for Advanced Materials, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai, Miyagi 980-8577, Japan. ⁴Centre for Translational Omics, University College London, Great Ormond Street Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK. ⁵Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany.

*These authors contributed equally to this work.

†Corresponding author. Email: thomas.carell@lmu.de

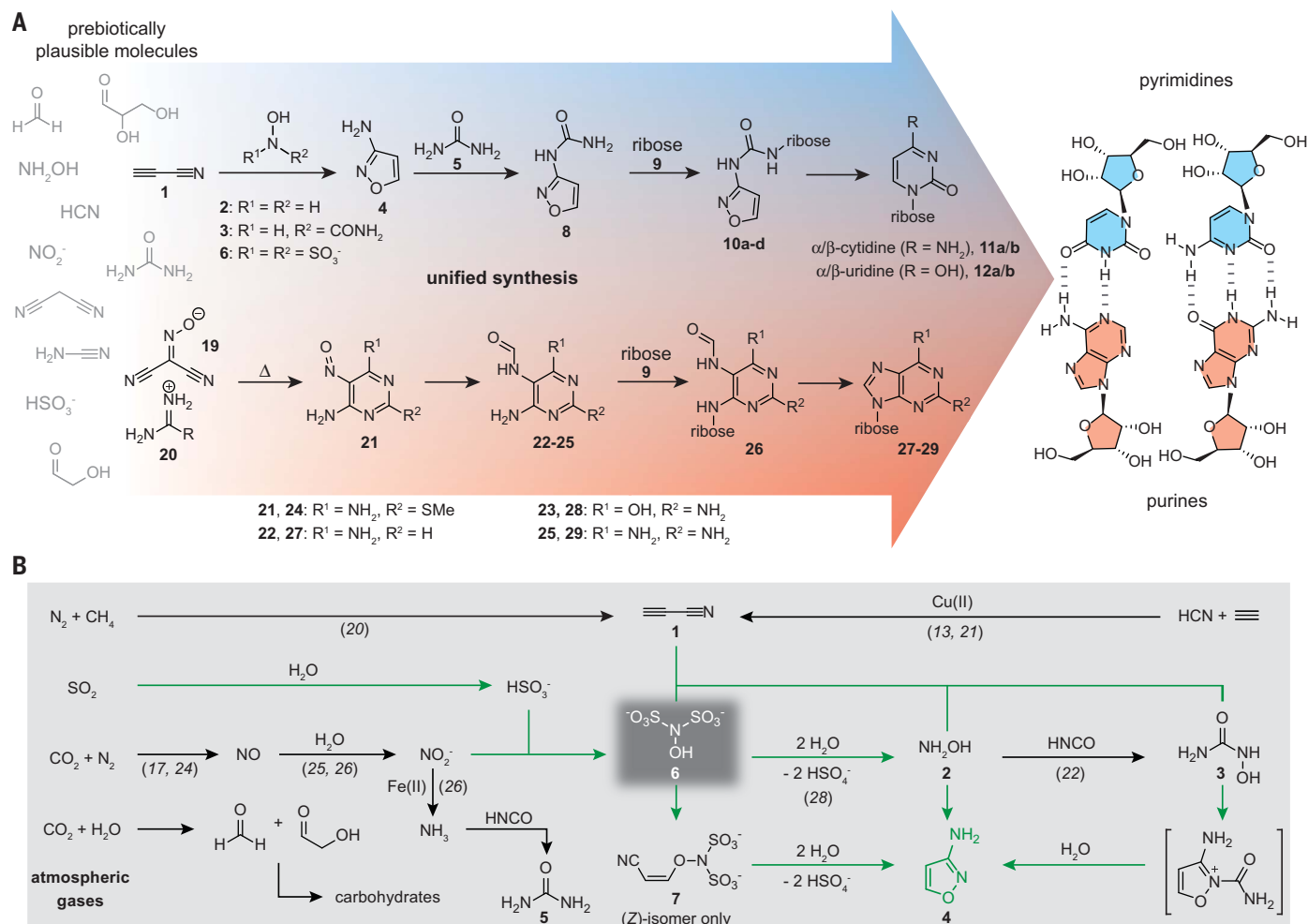


Fig. 1. Unified synthesis of pyrimidine and purine RNA building blocks. (A) Starting from plausible prebiotic molecules, the reaction scheme depicts the route toward the pyrimidines via isoxazolylurea **8** (blue background) and the purines via formamidopyrimidines **22** to **25** (red background) (*18*). (B) Fundamental chemistry

that produces the molecules needed for the pyrimidine pathway. Reactions performed in this work are shown with green arrows, while black arrows represent well-known reactions. Formation of **4** requires reaction of **1** with hydroxylamine **2**, hydroxylurea **3**, or the disulfonate **6** (dark-gray box). **6** is formed from NO₂⁻ and SO₂/HSO₃⁻.

at 95°C. The reaction of **4** to **8** is in all cases a clean process, with the only impurity being unreacted **4** (Fig. 2A). The product **8** can be subsequently physically enriched. Addition of carbonated water to the dried reaction mixture solubilizes **4**, **5**, and **8**, leaving behind the metal ions as hydroxides or carbonates. Subsequent concentration of the supernatant leads to spontaneous crystallization of **8** (55%). This allowed us to obtain a crystal structure of **8** (fig. S3). In order to simulate early Earth chemistry, we performed a one-pot experiment. We mixed **1** with **3**, **5**, and Zn²⁺ or Co²⁺ in a carbonate solution (pH ~10) and obtained compound **4** at 95°C (80 to 90%). Neutralizing the solution to pH ~6 to 7, which may have occurred on the early Earth as a result of acidic rain, followed by dry-down at the same temperature, provided compound **8** with yields between 56% (Zn²⁺) and 40% (Co²⁺). The continuous synthesis of the key building block **8** was consequently achieved in a plau-

sible prebiotic setting that could have existed in hydrothermal vents or near volcanic activity, both of which would be able to provide elevated temperatures (fig. S3). The synthesis is also possible at lower temperatures, but with extended reaction times.

For the final step toward nucleosides, we need to assume that, due to flooding or a mixing of environments, **8** came into contact with ribose **9** (Figs. 1A and 2B) or any other sugar unit, such as threose (for TNA) or glyceraldehyde (for GNA), that was able to form a backbone for a pairing system (*29*, *30*). When we mixed **8** with ribose **9** and warmed the mixture to 95°C in the presence of boric acid, we observed a fast and high-yielding reaction that provided the ribosylated products **10a** to **10d** with 95% yield (fig. S4a). Other borate minerals, such as synthetic lüneburgite {Mg₃[(PO₄)₂][B₂(OH)₆]·6H₂O} (*31*) or borax {Na₂[B₄O₅(OH)₄]·10H₂O} (*32*), were also able to catalyze this reaction with high yields (>70%)

(fig. S5). The major products were initially the α- and β-pyranosides (**10c** and **10d**), which dominate over the α- and β-furanosides (**10a** and **10b**) (fig. S4a). After heating the mixture under slightly basic conditions at 95°C in the presence of borates, the furanosides (54%; **10a** and **10b**) (Fig. 2B) gradually became the dominant products (fig. S4b). Under these conditions, we also observed hydrolysis of **10a** to **10d** to **8** and **9**. The accumulation of the furanosides **10a** and **10b** is best explained by complexation of their *cis*-diols with borate (*32*).

The final step toward pyrimidine nucleosides requires reductive opening of the isoxazole N-O bond, followed by tautomerization, intramolecular cyclization, and water elimination in a cascade-like fashion (Fig. 2, C and D). We found that this reaction occurred rapidly with Fe²⁺ in the presence of thiols (Fig. 2D) (*33*). Liquid chromatography-mass spectrometry (LC-MS) analysis indicated that cytidine nucleosides **11a** to **11d** formed efficiently under

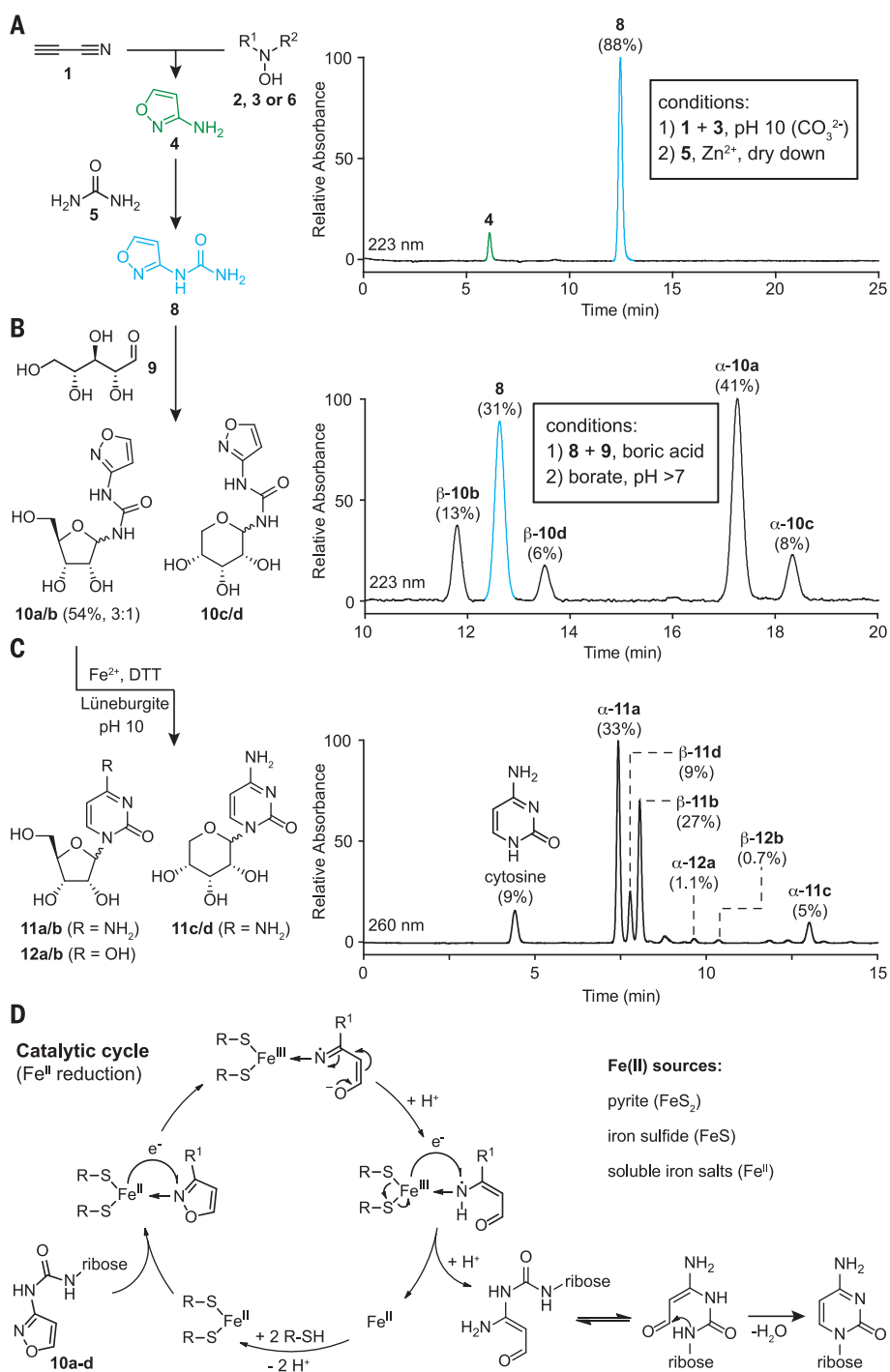


Fig. 2. Formation of pyrimidine nucleosides (11 and 12) from N-isoxazolylurea ribosides 10a and 10b. The different isomers are labeled as follows: **a** = α -furanosyl; **b** = β -furanosyl; **c** = α -pyranosyl; **d** = β -pyranosyl. **(A)** Formation of **4** and its conversion with urea **5** to *N*-isoxazolylurea **8**. **(B)** Ribosylation of **8** with ribose **9** and equilibration of the reaction mixture in the presence of borates gives the furanosidic isomers **10a** and **10b** (54%). **(C)** Pyrimidine nucleoside formation by reductive N-O cleavage from the compound mixture of **10a** and **10b** in the presence of ammonium iron(II) sulfate hexahydrate (0.0005 equiv.). The HPLC results with detection at 260 nm show formation of cytosine (C; **11a** to **11d**) and uridine (U; **12a** and **12b**). **(D)** Proposed catalytic cycle for the Fe²⁺ catalyzed reduction of the N-O bond of the isoxazole moiety.

these conditions, with the furanosidic uridine nucleosides **12a** and **12b** as the corresponding deamination products formed by hydrolysis (Fig. 2C). Reductive pyrimidine formation can be performed with FeS or the mineral pyrite (FeS₂), and both have been discussed in the context of early metabolic pathways (15, 34). Just 0.0001 equiv. of soluble Fe²⁺ in water is sufficient for the reaction. In the absence of Fe²⁺, pyrimidine formation was not observed. The reduction also appears to be independent of the thiol source, as the products **11a** to **11d** and **12a** and **12b** were obtained regardless of whether we used dithiothreitol (DTT), propanedithiol, mercaptoethanol, or cysteine (fig. S6).

Selective one-pot formation of 5'-nucleoside mono- and diphosphates

The addition of naturally occurring minerals such as hydroxyapatite, colemanite, or (synthetic) lüneburgite to the reductive pyrimidine-forming reaction mixture had a strong influence on the distribution of the four cytidine isomers. Synthetic lüneburgite gave a combined high yield of 85% (Fig. 2C). The natural furanosidic β -cytidine (**11b**) and its α -anomer (**11a**) are formed under these conditions with about the same yields, together with small amounts of α - and β -uridine (**12a** and **12b**). We found only small amounts of the α - and β -cytidine pyranosides (**11c** and **11d**), together with the cytosine base. Because synthetic lüneburgite is known to enable nucleotide formation in the presence of urea (Fig. 3A) (31), we simply added urea to the one-pot reaction mixture after pyrimidine formation and allowed the mixture to evaporate to dryness at 85°C over a period of about 20 hours. LC-MS analysis of the reaction mixture showed formation of phosphorylated nucleosides (Fig. 3A) in a substantial 19% yield relative to that for cytosine (Fig. 3B and fig. S7). We assumed that the reaction generated the α - and β -cytidine 5'-monophosphates **13a** and **13b** and the 5'-diphosphorylated cytidines **14a** and **14b**. Owing to hydrolysis, we also expected some α - and β -uridine 5'-monophosphates and 5'-diphosphates **15a** and **15b** and **16a** and **16b**. We isolated the corresponding high-performance LC (HPLC) peaks and removed the phosphate groups enzymatically (Fig. 3, B and C). LC-MS analysis showed the dephosphorylated furanosides **11a** and **11b** and **12a** and **12b** with over 94% in the nucleoside pool, which corresponds to a change of the furanoside/pyranoside ratio from an initial 4:1 to 17:1 (Fig. 3C). The formation of phosphorylated pyranosides **17** are only a minor side reaction. We found no discrimination between α - and β -anomers during the phosphorylation. The furanoside enrichment is best explained by the presence of a primary hydroxyl group in the furanosides which is absent in the pyranosides. The enrichment of 5'-nucleoside monophosphates and diphosphates under these one-pot conditions establishes a further

chemical selection step that favors the furanosides as the components of RNA. We further characterized the structures of the phosphorylated nucleosides and confirmed the formation of the 5'- α - and 5'- β -cytidine mono- and diphosphates (**13a**, **13b**, and **14a**, **14b**; α -/ β -CMP and α -/ β -CDP) (fig. S8). Additional analysis allowed identification of α , β -UDP **16a** and **16b** (fig. S9). 5'-Pyrophosphates are the dominating species within the diphosphorylated nucleoside mixture (fig. S8a).

Compatible formation of pyrimidine and purine RNA nucleosides

We next investigated if the prebiotically plausible pyrimidine and purine nucleoside pathways are compatible with each other so that they can be connected with the goal to form all Watson-Crick building blocks in the same environment, driven solely by wet-dry cycles. The purine synthesis (18) requires as the initial step a reaction of malonitrile **18** with sodium nitrite to give (hydroxyimino)malonitrile **19**. Because malonitrile **18** can be also generated from cyanoacetylene **1**, as shown by Eschenmoser (35), pyrimidines and purines can be traced back to the same chemical root (Fig. 4). Compound **19** forms an organic salt with amidines **20** to give nitroso-pyrimidines **21** and, upon reduction and formylation, FaPys (**22** to **25**). The latter can react with ribose **9** to give ribosylated FaPy **26** and then purine nucleosides **27** to **29** (Fig. 1A) (18). To investigate how the chemical conditions needed for

pyrimidine formation from the urea-isoxazole **8** would affect purine formation, we reacted **8** and the FaPy compounds **22** and **23** with ribose **9** under dry-down conditions. We performed the reaction under identical conditions but in separate reaction vials (Fig. 4). Under these conditions, formation of all four Watson-Crick nucleosides, cytidine **11**, uridine **12**, adenosine **27**, and guanosine **28**, were detected.

We next investigated if pyrimidines and purines can form simultaneously in the same environment (Fig. 5A). For this experiment, we mixed the starting materials cyanoacetylene **1**, hydroxylurea **3**, (hydroxyimino)malonitrile **19**, and amidine **20** under slightly basic conditions (pH ~10). Analysis of the mixture indeed showed formation of **4** with 86% yield, despite the presence of **19** and **20**. It is surprising that the N-OH functionality of compound **19** does not interfere with the formation of **4**. Compound **4** is a liquid that can enrich from a water solution by dry-down, owing to its high boiling point (228°C). Compound **4** can act as a solvent to facilitate the formation of **21** from the reaction of **19** with **20** under milder conditions (50°C to 100°C instead of 126°C), in contrast with results from a previous experiment (18). The next step requires reduction and formylation of **21** to the FaPy intermediate, but this step cannot be performed in the presence of the isoxazole. Addition of a water mixture eventually containing urea **5** leads to spontaneous precipitation of **21**. The supernatant containing **4** and **5** can flow away.

The water-insoluble **21**, if brought into contact with dilute formic acid and Zn (found in Earth's crust), reacts immediately to form the compounds **22** and **24** with Zn²⁺ as a side product (Fig. 5A and fig. S10a). These reaction products are water-soluble and can potentially recombine with **4** and **5**. The side product Zn²⁺ can then catalyze the reaction of **4** in the presence of **5** to give *N*-isoxazolyl urea **8** in the presence of **22** and **24** (Fig. 5A and fig. S10b). This leads to the formation of the pyrimidine and purine precursors **8**, **22**, and **24**, which can be transformed into the purine and pyrimidine nucleosides. In this scenario, intermediate **4** of the pyrimidine pathway helps formation of the purine precursor **21**, while Zn²⁺ as a side product of the purine pathway mediates formation of the pyrimidine precursor **8** in a mutually synergistic way, driven by wet-dry cycles.

We combined **8** with different FaPy intermediates and investigated if they reacted in a one-pot scenario with ribose **9** to finally give the purine and pyrimidine nucleosides. To examine this, we dissolved a mixture of **8**, **22**, **25**, ribose **9**, and boric acid and warmed the mixture to 95°C for 14 hours, allowing for slow evaporation of water. The solid material was then taken up with a slightly basic solution containing Fe²⁺ (0.0005 equiv.) and DTT (1.5 equiv.), and we allowed the mixture to warm to 95°C. HPLC-MS analysis proved that these conditions simultaneously provided the purine and pyrimidine nucleosides with cytidine (**11a** to **11d**) and adenosine (**27**) as the

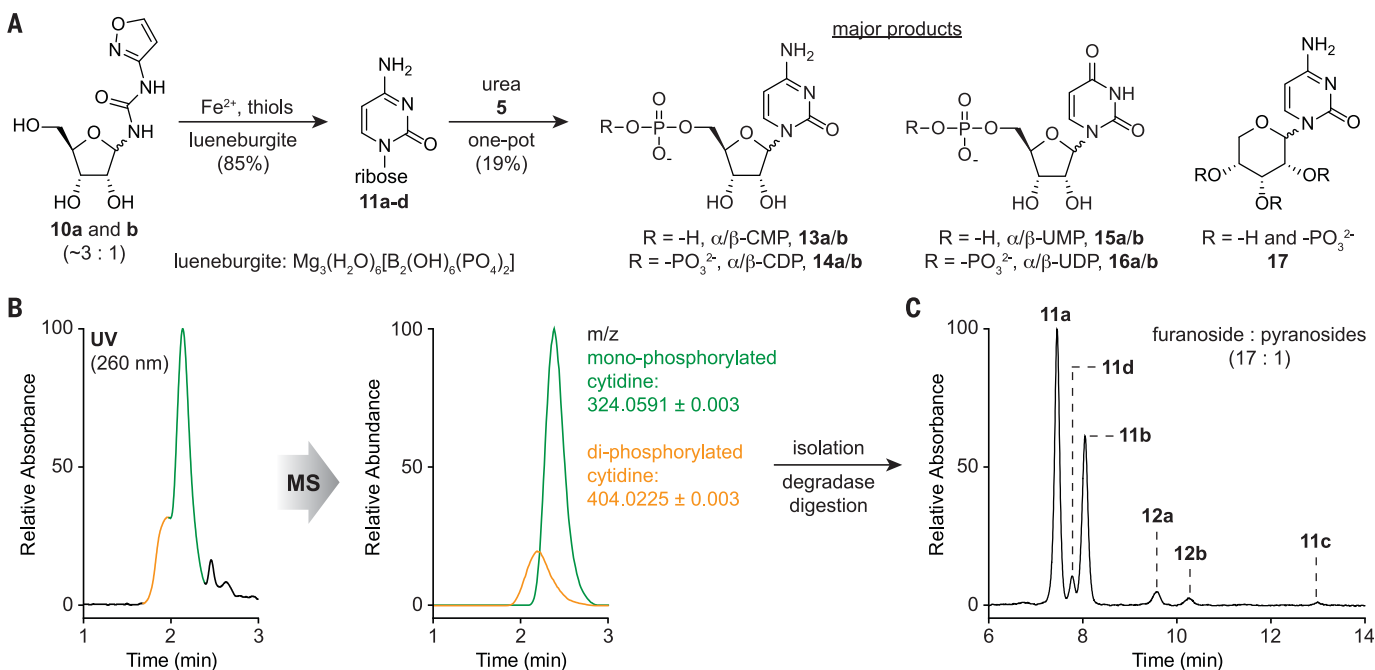


Fig. 3. One-pot nucleotide formation reaction. (A) One-pot synthesis of cytidine and uridine 5'-mono- and 5'-diphosphates (**13a** and **13b** to **16a** and **16b**) after urea addition to the reaction mixture and allowing the mixture to dry-down at 85°C for 20 hours. **a/b** represent the α - and β -anomers, respectively. (B) LC-MS

analysis of the corresponding nucleotide peaks with UV and MS detection and isolation of the formed nucleotides from the prebiotic reaction, followed by an enzymatic removal of the phosphate groups. (C) HPLC analysis of the dephosphorylated product mixture showing predominant formation of α - and β -cytidine **11a** and **11b**.

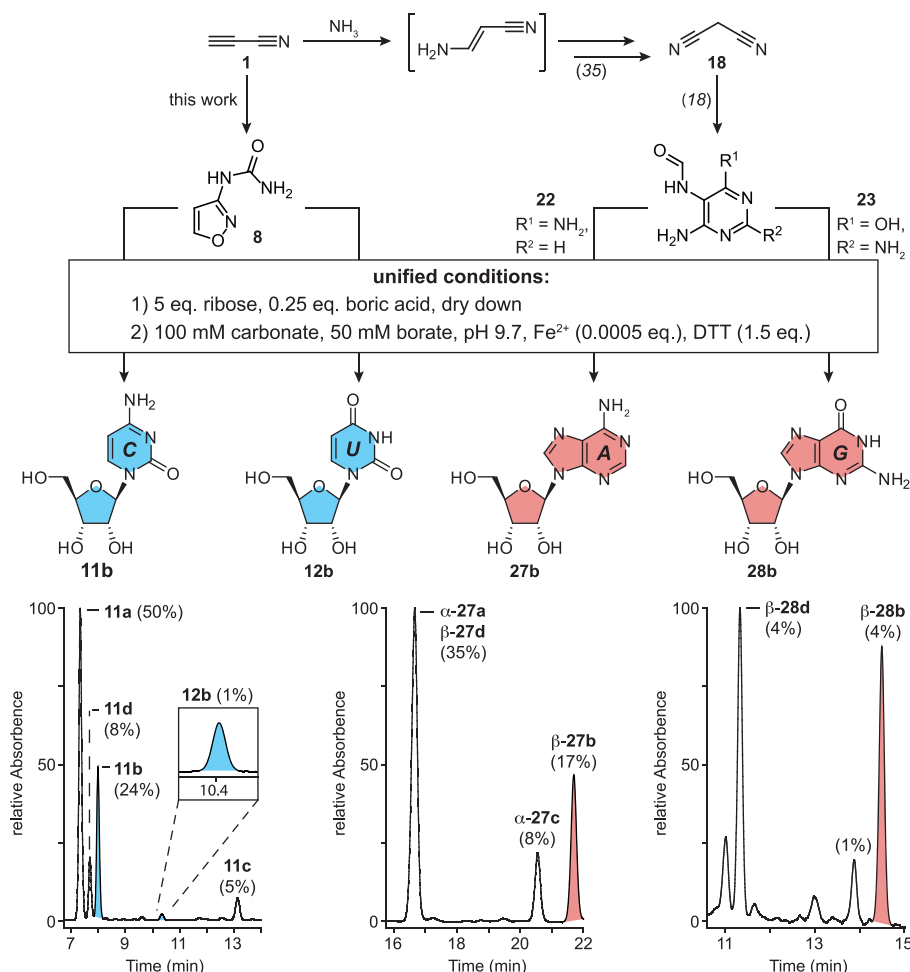


Fig. 4. Formation of all four Watson-Crick RNA building blocks in identical but parallel reactions.

C (**11b**), U (**12b**), A (**27b**), and G (**28b**) are formed under the same conditions separately from **8**, **22**, and **23**. HPLC results are shown with a detection at 260 nm. The nucleosides are labeled as follows: **a** = α -furanosyl; **b** = β -furanosyl; **c** = α -pyranosyl; **d** = β -pyranosyl. Canonical pyrimidine and purine RNA building blocks are labeled in blue and red, respectively.

main products. Diaminopurine nucleosides (DA; **29**), which hydrolyze to guanosine **28**, form in this one-pot reaction as well (Fig. 5A, chromatogram). We noted additional formation of double-ribosylated adenine (rib₂-A). Furthermore, the nucleoside **28** was created in this scenario when we used **23** ($R^1 = \text{OH}$, $R^2 = \text{NH}_2$) as the starting material, but the yields were lower.

Discussion

Ribose-based RNA and the four canonical nucleosides, A, G, C, and U, are central to modern life and to prebiotic hypotheses, such as the “RNA world,” in which RNA strands replicated and evolved to give increasingly complex chemical systems (4). Whether such RNAs were directly assembled from the canonical nucleotides (A, C, G, and U bases) or if they evolved from a simpler proto-RNA system is unclear (36).

Here we show that a reaction network toward the purine and pyrimidine RNA building blocks can be established, starting from simple atmospheric or volcanic molecules. Molecular complexity is generated by wet-dry cycles that can drive the chemical transformations. Therefore, any environment that was able to provide wet-dry phases might have been a suitable place for the origin of RNA building blocks. Our geochemical model assumes that chemistry took place in several basins that were needed to locally separate intermediates. We also needed one or two streams of water in our system to allow exchange of soluble molecules (Fig. 5B). Intermediates might precipitate upon fluctuations of physico-chemical parameters, allowing for the separation of soluble and insoluble materials (e.g., **4** and **21**). After further reactions, which reestablish solubility, the compounds can be recombined (Fig. 5B). For our scenario we need to assume that the early

Earth provided environmental conditions that fluctuated between slightly acidic (pH 3), potentially caused by acidic rain (SO_2 , NO_x), or basic (pH 10) caused by carbonates. Even though most of the chemistry described here was performed at elevated temperatures, the reactions also occur at lower temperatures, but with substantially longer reaction times. We can assume that temperatures fluctuated on the early Earth just like today due to day-night or seasonal cycles. Such fluctuations would certainly have brought about wet-dry cycles, akin to modern droughts and rain. The geophysical requirements needed for the reported chemistry, including elevated temperatures, could have existed in geothermal fields or at surface hydrothermal vents, which are plausible geological environments on early Earth.

Our proposed chemical pathways toward pyrimidines and purines begin with cyanoacetylene **1**, which could have formed in surface hydrothermal vents (13). Reaction of **2**, **3**, or **6** with **1** is the starting point for the pyrimidines, but if **1** reacts instead with ammonia, a pathway to malononitrile **18** as the precursor for purine synthesis is possible (Fig. 4) (35). Another key molecule for the synthesis of purines and pyrimidines is NO_2^- , which is needed to nitrosate malononitrile **18** to **19** (18). NO_2^- is also crucial for the formation of hydroxylamine in the presence of HSO_3^- , which is formed from volcanic SO_2 (27). The concentration of NO_2^- that is reachable in a prebiotic setting is under debate, but it is speculated that the most likely place for its accumulation is in shallow ponds, as needed for our scenario (17). In general, the limited stability of NO_2^- would not be an issue, provided that it is rapidly captured by HSO_3^- upon its formation. Our model assumes a surface environment, where molecules such as NO_2^- , HSO_3^- , or urea **5** could have been delivered by rain after their formation in the atmosphere (Fig. 5B) (25, 37). Our chemistry shows that robust reaction networks can be established that allow all key intermediates to be generated efficiently from relatively complex mixtures, followed by their physical enrichment or separation on the basis of their solubility in water. Wet-dry cycles govern the formation of purine and pyrimidine RNA building blocks in a scenario depicted in Fig. 5B. Of course, we will be unable to definitively prove that the described scenario indeed took place on early Earth, but the reported chemistry shows that, under plausible prebiotic conditions, mutually synergistic reaction pathways can be established in which the intermediates along one pathway help the chemistry of the other. In such a scenario, we show that the key building blocks of life can be created without the need for sophisticated isolation and purification procedures of reaction intermediates that are common in traditional organic chemistry.

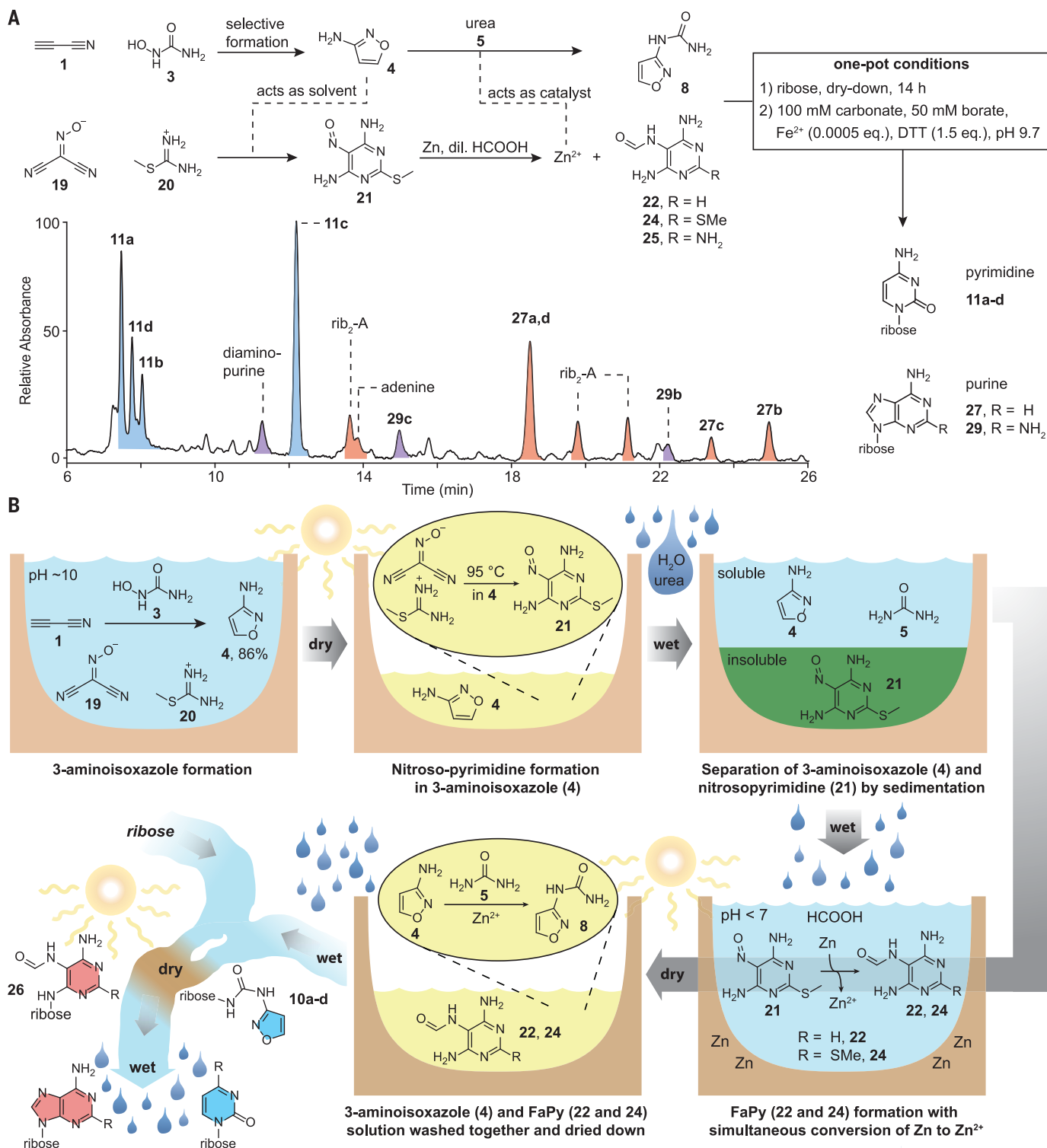


Fig. 5. Unified chemical scenario for the formation of purine and pyrimidine nucleosides. (A) Depiction of the connected reaction pathways to pyrimidine and purine nucleosides, together with the HPLC analysis (260 nm) of the final reaction mixtures. Nucleosides are labeled as follows: **a** = α -furanosyl; **b** = β -furanosyl; **c** = α -pyranosyl; **d** = β -pyranosyl. (B) Proposed geochemical scenario for the simultaneous synthesis of purine and pyrimidine nucleosides,

driven by wet-dry cycles. In yellow, the solvent is 3-aminoisoxazole (**4**), which can be enriched from an aqueous solution due to its high boiling point (228°C). 2-(Methylthio)-5-nitrosopyrimidine-4,6-diamine (**21**) is a general precursor for adenosine and guanosine (**18**). Compounds **8**, **22**, and **24** are accessible in the same pot, and they can react with ribose to the RNA nucleosides in a one-pot reaction.

The concurrent formation of pyrimidine and purine nucleosides in the network can be traced to just a few key starting molecules, such as cyanoacetylene **1**, NH₃, NH₂OH **2** (or the disulfonate **6**), HCN, urea **5**, formic acid, and isocyanate, plus salts such as nitrites, carbonates, and borates. Metals such as Zn or Fe and their ions play an important role in our chemistry, consistent with their proposed involvement in early metabolic cycles (23, 38). In particular, iron-sulfur surfaces needed for pyrimidine formation have been discussed as platforms for early prebiotic chemistry (15, 34, 39). The 5'-(di)phosphorylation is integrated into our pathway if phosphate minerals such as lüneburgite or struvite (figs. S11 to S13) are present. It remains unclear, however, how ribose or any other carbohydrate, such as glycerol or threose, that is needed to form the backbone of RNA or pre-RNA could have formed selectively (29, 40). Sugars such as ribose can be produced nonselectively in a formose-like reaction, which is possible in a variety of different physico-chemical environments (32, 41–43).

REFERENCES AND NOTES

- J. A. Doudna, T. R. Cech, *Nature* **418**, 222–228 (2002).
- D. P. Horning, G. F. Joyce, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9786–9791 (2016).
- J. Attwater, A. Raguram, A. S. Morgunov, E. Gianni, P. Holliger, *eLife* **7**, e35255 (2018).
- W. Gilbert, *Nature* **319**, 618 (1986).
- M. W. Powner, B. Gerland, J. D. Sutherland, *Nature* **459**, 239–242 (2009).
- S. Becker *et al.*, *Science* **352**, 833–836 (2016).
- H.-J. Kim, S. A. Benner, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11315–11320 (2017).
- S. Stairs *et al.*, *Nat. Commun.* **8**, 15270 (2017).
- R. Saladino *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, E2746–E2755 (2015).
- J. F. Kasting, *Science* **259**, 920–926 (1993).
- R. M. Hazen, *Am. J. Sci.* **313**, 807–843 (2013).
- E. D. Swanner *et al.*, *Earth Planet. Sci. Lett.* **390**, 253–263 (2014).
- P. B. Rimmer, O. Shorttle, *Life* **9**, 12 (2019).
- W. Martin, J. Baross, D. Kelley, M. J. Russell, *Nat. Rev. Microbiol.* **6**, 805–814 (2008).
- E. Camprubi, S. F. Jordan, R. Vasiladiou, N. Lane, *IUBMB Life* **69**, 373–381 (2017).
- S. A. Benner, H.-J. Kim, E. Biondi, in *Prebiotic Chemistry and Chemical Evolution of Nucleic Acids*, C. Menor-Salván, Ed. (Springer International Publishing, Cham, 2018), pp. 31–83.
- S. Ranjan, Z. R. Todd, P. B. Rimmer, D. D. Sasselov, A. R. Babbitt, *Geochem. Geophys. Geosyst.* **20**, 2021–2039 (2019).
- S. Becker *et al.*, *Nat. Commun.* **9**, 163 (2018).
- P. Thaddeus, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 1681–1687 (2006).
- R. A. Sanchez, J. P. Ferris, L. E. Orgel, *Science* **154**, 784–785 (1966).
- B. H. Patel, C. Percivalle, D. J. Ritson, C. D. Duffy, J. D. Sutherland, *Nat. Chem.* **7**, 301–307 (2015).
- H. Kofod, B. Wickberg, A. Kjør, *Acta Chem. Scand.* **7**, 274–279 (1953).
- K. B. Muchowska, S. J. Varma, J. Moran, *Nature* **569**, 104–107 (2019).
- V. S. Airapetian, A. Gloer, G. Gronoff, E. Hébrard, W. Danchi, *Nat. Geosci.* **9**, 452–455 (2016).
- H. J. Cleaves, J. H. Chalmers, A. Lazcano, S. L. Miller, J. L. Bada, *Orig. Life Evol. Biosph.* **38**, 105–115 (2008).
- D. P. Summers, S. Chang, *Nature* **365**, 630–633 (1993).
- S. Ranjan, Z. R. Todd, J. D. Sutherland, D. D. Sasselov, *Astrobiology* **18**, 1023–1040 (2018).
- G. K. Rollefson, C. F. Oldershaw, *J. Am. Chem. Soc.* **54**, 977–979 (1932).
- G. F. Joyce, *Nature* **418**, 214–221 (2002).
- D. M. Fialho *et al.*, *Org. Biomol. Chem.* **16**, 1263–1271 (2018).
- H.-J. Kim *et al.*, *Angew. Chem. Int. Ed.* **55**, 15816–15820 (2016).
- A. Ricardo, M. A. Carrigan, A. N. Olcott, S. A. Benner, *Science* **303**, 196–196 (2004).
- M. Kijima, Y. Nambu, T. Endo, *J. Org. Chem.* **50**, 1140–1142 (1985).
- G. Wächtershäuser, *Microbiol. Rev.* **52**, 452–484 (1988).
- U. Trinks, A. Eschenmoser, ETH Zurich (1987); doi: 10.3929/ethz-a-000413538.
- R. Krishnamurthy, *Isr. J. Chem.* **55**, 837–850 (2015).
- J. Liebig, F. Wöhler, *Ann. Phys.* **96**, 369–400 (1830).
- M. Preiner *et al.*, *bioRxiv* 682955 (2019).
- C. Bonfio *et al.*, *Nat. Catal.* **1**, 616–623 (2018).
- J. S. Teichert, F. M. Kruse, O. Trapp, *Angew. Chem. Int. Ed.* **58**, 9944–9947 (2019).
- J. Kofoed, J.-L. Reymond, T. Darbre, *Org. Biomol. Chem.* **3**, 1850–1855 (2005).
- C. Meinert *et al.*, *Science* **352**, 208–212 (2016).
- K. Usami, A. Okamoto, *Org. Biomol. Chem.* **15**, 8888–8893 (2017).

ACKNOWLEDGMENTS

We thank J. Kampmann for x-ray diffraction measurements and S. Balasubramanian for supporting S.B. during the revision of the manuscript. **Funding:** Deutsche Forschungsgemeinschaft (DFG) provided financial support via the programs SFB1309 (TP-A4), SFB749 (TP-A4), SPP-1784, GRK2062/1, and CA275/11-1, the Excellence Cluster EXC114, the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement EPIR 741912), and the Volkswagen Foundation (Initiative "Life": EcoRib). H.O. thanks the European Commission for a Marie Skłodowska-Curie postdoctoral fellowship (PRENUCRNA). **Author contributions:** T.C. designed and supervised research; S.B. helped to design the study, S.B., J.F., S.W., and H.O. performed the experiments. C.S., M.R., and A.C. supported the synthesis and MS quantification, K.I. performed biochemical studies, and T.A. helped to design the synthesis. T.C., S.B., J.F., and S.W. analyzed data. T.C. and S.B. wrote the manuscript and designed the figures. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The x-ray crystallographic data for isoxazoleurea **8** are deposited in the CCDC under accession number 1889652. All other data needed to support the conclusions of this manuscript are included in the main text and supporting material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6461/76/suppl/DC1
Materials and Methods
Figs. S1 to S13
References (44–51)

8 March 2019; resubmitted 21 June 2019
Accepted 21 August 2019
10.1126/science.aax2747

4.2 Studien zur Bildung nicht-kanonischer RNA-Nukleoside unter präbiotisch plausiblen Bedingungen

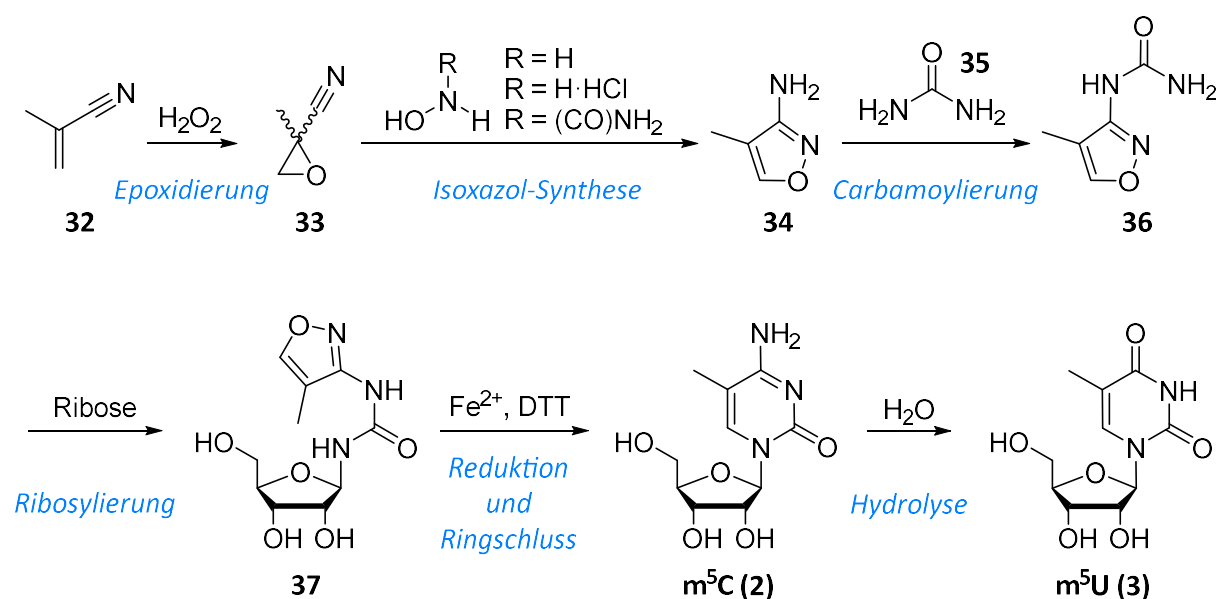
Es scheint plausibel, dass nicht-kanonische RNA-Nukleoside wichtige Funktionen bei der Entstehung des Lebens auf der Erde einnehmen konnten, vgl. Einleitung Abschnitt 1.4. Ziel dieses Teils der Arbeit war es deshalb, präbiotisch plausible Zugänge zu potenziell zentralen nicht-kanonischen RNA-Nukleosiden zu erschließen. Abschnitt 4.2.1 beschreibt Studien zur Bildung der methylierten RNA-Nukleoside m⁵C (5-Methylcytidin, **2**) und m⁵U (5-Methyluridin, **3**) in präbiotischen Modellreaktionen. Im zweiten Abschnitt 4.2.2 werden Untersuchungen zur Bildung des Thio-RNA-Nukleosids s²U (2-Thiouridin, **1**) mittels präbiotischer Modellreaktionen präsentiert.

4.2.1 Studien zur Bildung von m⁵C und m⁵U unter präbiotisch plausiblen Bedingungen

Die Synthese von m⁵U (**3**) gelang in präbiotischen Modellreaktionen bisher nur in Spuren.^[52,56] m⁵C (**2**) wurde unter präbiotisch plausiblen Bedingungen bisher nicht erhalten, vgl. Einleitung Abschnitt 1.4.2. Deshalb sollten in diesem Teil der Arbeit präbiotisch plausible Zugänge zu m⁵C (**2**) und m⁵U (**3**) gefunden werden. Diese sollten idealerweise kompatibel mit der vereinigten Bildung kanonischer Pyrimidin- und Purin-Nukleoside sein, vgl. Abschnitt 4.1.

4.2.1.1 Synthesestrategie

Zunächst wurde ein Synthesepfad um das Schlüsselmolekül 3-Amino-4-methylisoxazol (**34**) entwickelt, siehe Schema 8. Ausgehend von Methacrylnitril (**32**) sollte mittels H₂O₂-Epoxidierung 2-Methyloxiran-2-carbonitril (**33**) gebildet werden. Ein Hydroxylamin-Derivat sollte anschließend die Synthese von 3-Amino-4-methylisoxazol (**34**) ermöglichen, bevor das exozyklische Amin mit Harnstoff (**35**) zu **36** carbamoyliert werden sollte. Anschließende Ribosylierung sollte den m⁵C-Vorläufer **37** liefern. Die Reduktion der Isoxazol-N-O-Bindung mittels Fe²⁺ und DTT sollte daraufhin die Zyklisierung zu m⁵C (**2**) initiieren. m⁵U (**3**) sollte anschließend leicht durch Hydrolyse von m⁵C (**2**) erhalten werden können.^[198-201] Die folgenden Abschnitte beschreiben die Umsetzung dieses Synthesepfades. Die im Folgenden vorgestellten Ergebnisse der Untersuchungen legen nahe, dass die Bildung von m⁵C (**2**) und m⁵U (**3**) in präbiotischen Modellreaktionen möglich ist.



Schema 8 Synthesestrategie zur Bildung von m⁵C (2) und m⁵U (3) in präbiotischen Modellreaktionen, ausgehend von Methacrylnitril (32).

4.2.1.2 präbiotische Modellreaktionen

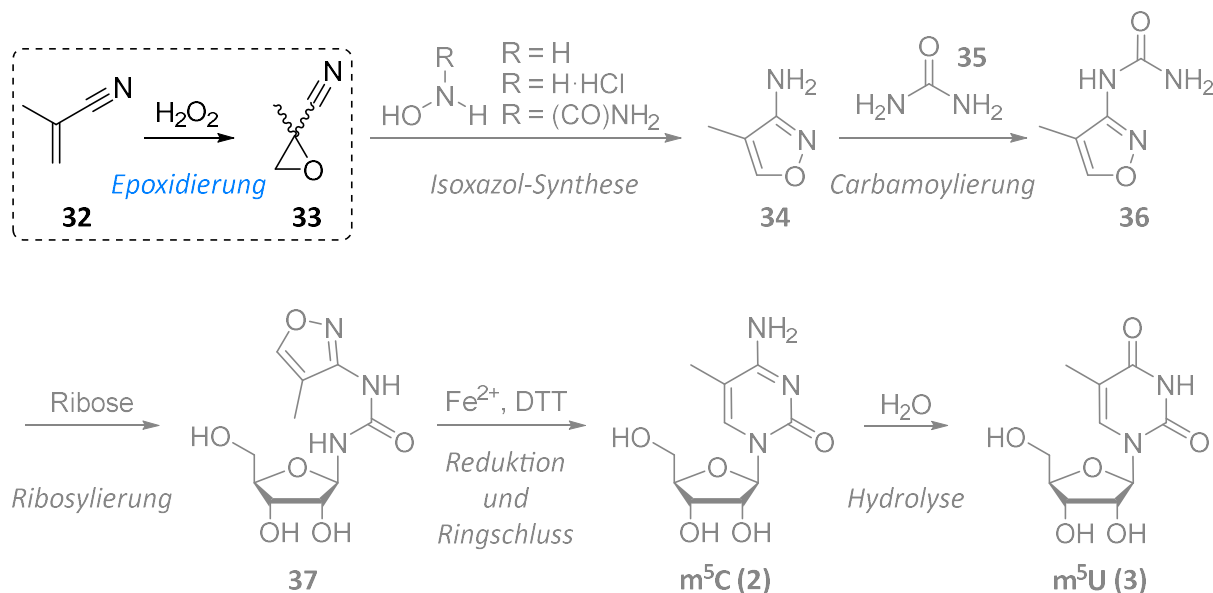
4.2.1.2.1 Methacrylnitril (32)

Die Anwesenheit von Methacrylnitril (32) auf der präbiotischen Erde scheint plausibel, deshalb wurden präbiotische Modellreaktionen zur Bildung von Methacrylnitril (32) nicht durchgeführt. Die Atmosphäre des Saturn-Mondes Titan weist eine komplexe organische Chemie auf, die große Ähnlichkeit zu den präbiotischen Prozessen der frühen Erde annehmen lässt.^[202] Interessanterweise ermöglichten sowohl experimentelle Simulationen als auch Beobachtungen der Titan-Atmosphäre die Detektion von Methacrylnitril (32).^[203,204] Zudem ist die Anwesenheit von Methacrylnitril (32) auch im interstellaren Medium plausibel.^[205] Weitere präbiotische Modellreaktionen zur Bildung von Methacrylnitril (32) wurden deshalb nicht durchgeführt. Für die Darstellung der Referenzverbindung des Zielprodukts 32 mittels herkömmlicher Synthese, siehe Abschnitt 4.2.1.3.1.

4.2.1.2.2 Epoxidierung zu 2-Methyloxiran-2-carbonitril (33)

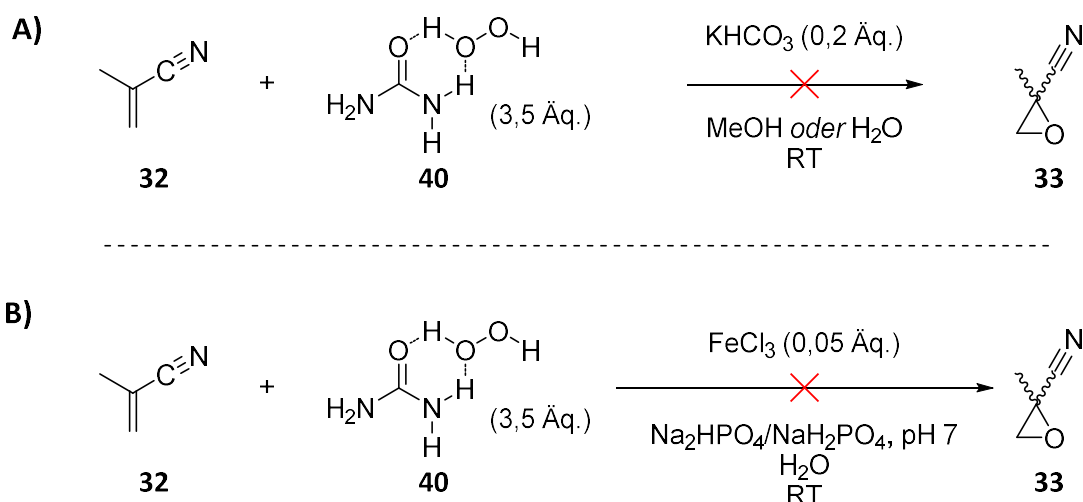
Ziel dieses Abschnitts war die Epoxidierung von Methacrylnitril (32) zu 2-Methyloxiran-2-carbonitril (33) unter präbiotisch plausiblen Bedingungen, siehe Schema 9. Epoxid 33 wird

dabei als Racemat gebildet. Da das Stereozentrum bei der darauffolgenden Isoxazol-Synthese wieder in dessen sp^2 -hybridisierte Form überführt wird, stellt das Racemat keine Einschränkung der Synthesestrategie dar. Für die Darstellung der Referenzverbindung des Zielprodukts **33** mittels herkömmlicher Synthese, siehe Abschnitt 4.2.1.3.2.



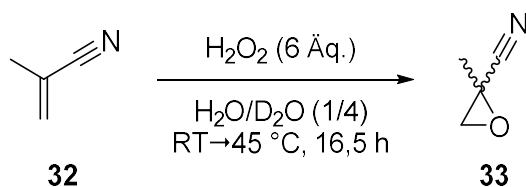
Schema 9 Ziel dieses Abschnitts war die Epoxidierung von Methacrylnitril (**32**) zu 2-Methyloxiran-2-carbonitril (**33**) unter präbiotisch plausiblen Bedingungen.

Weder Carbamidperoxid (**40**) noch Fe^{3+} führten zur Epoxidierung von Methacrylnitril (**32**). Zunächst wurde das Harnstoff- H_2O_2 -Addukt Carbamidperoxid (**40**) zur Epoxidierung von Methacrylnitril (**32**) untersucht, siehe Schema 10. Dabei zeigten Reaktionen in An- sowie in Abwesenheit von Fe^{3+} unter verschiedenen Bedingungen keine Bildung des Zielprodukts 2-Methyloxiran-2-carbonitril (**33**). Auch die Koordination von Fe^{3+} mittels 1-Methylimidazol oder Diaminomaleonitril führten nicht zum Erfolg.^[206]



Schema 10 Umsetzungen von Methacrylnitril (**32**) mit Carbamidperoxid (**40**) führten unter diversen Bedingungen nicht zur Bildung des Zielprodukts 2-Methyloxiran-2-carbonitril (**33**). Umsetzung **B**) führte auch in Anwesenheit von 1-Methylimidazol oder Diaminomaleonitril nicht zur Bildung des Epoxids **33**.

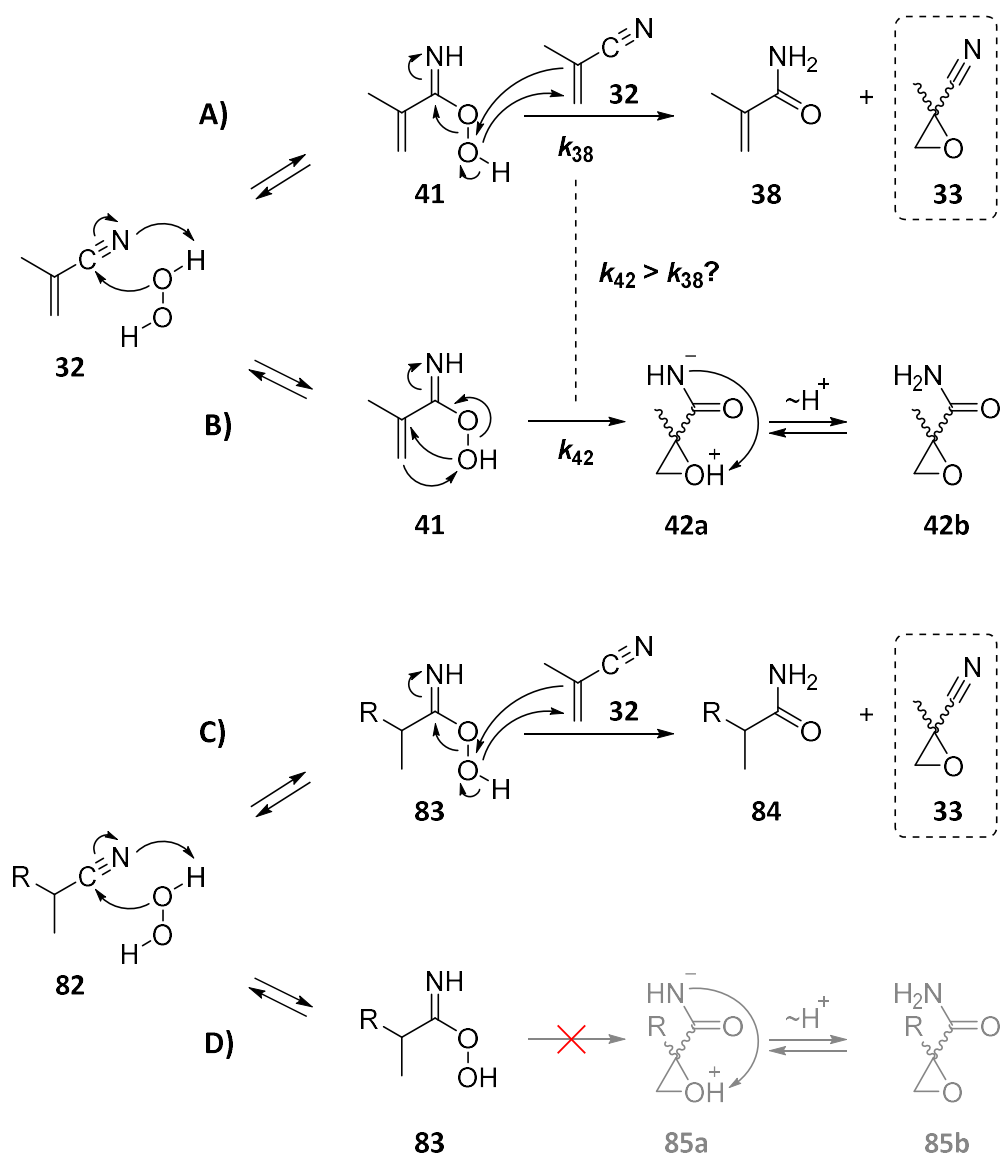
H_2O_2 ermöglichte schließlich die Epoxidierung von Methacrylnitril (**32**) zu Epoxid **33**. Da Methacrylnitril (**32**) lediglich mittels NMR-Spektroskopie detektiert werden konnte, wurde die Reaktion in einem H_2O/D_2O -Gemisch durchgeführt, siehe Schema 11. Nach 16,5 h bei RT gelang die Detektion von 2-Methyloxiran-2-carbonitril (**33**) mittels 1H - und ^{13}C -NMR-Spektroskopie.



Schema 11 Die Bildung von 2-Methyloxiran-2-carbonitril (**33**) gelang durch Epoxidierung von Methacrylnitril (**32**) mittels H_2O_2 in einer präbiotischen Modellreaktion.

Eine Imidoperoxysäure fungiert möglicherweise als Epoxidierungsreagenz. Ein Mechanismus-Vorschlag zur Epoxidierung von Methacrylnitril (**32**) mittels H_2O_2 lässt sich auf Basis der Studien von Payne *et al.*^[207] formulieren: Demnach könnte H_2O_2 die Nitril-Funktionalität von Methacrylnitril (**32**) zunächst zur Imidoperoxysäure **41** umsetzen, siehe Schema 12 A). Diese könnte im nächsten Schritt als reaktive Spezies **41** die Epoxidierung eines weiteren Methacrylnitril-Moleküls bewirken. Als Nebenprodukt entstünde dabei Methacrylamid (**38**). In einer Konkurrenzreaktion könnte es zur intramolekularen Epoxidierung der Imidoperoxysäure **41** zu 2-Methyloxiran-2-carboxamid (**42**) kommen, siehe Schema 12 B). Es kann angenommen werden, dass die Geschwindigkeitskonstante dieser Konkurrenzreaktion (k_{42})

größer ist als die der Zielproduktbildung (k_{38}). Um die Häufigkeit dieser Konkurrenzreaktion zu reduzieren und damit eine mögliche Steigerung der Ausbeute von 2-Methyloxiran-2-carbonitril (**33**) zu erzielen, sollte der Einsatz eines α - β -gesättigten Nitrils (**82**) untersucht werden, siehe Schema 12 C) und D). Dieses könnte die reaktive Spezies **83** bilden, ohne Möglichkeit zur intramolekularen Epoxidierung. Weitere Optimierungen der Reaktion könnten durch Screenings der Reaktionsdauer und -temperatur ermöglicht werden.



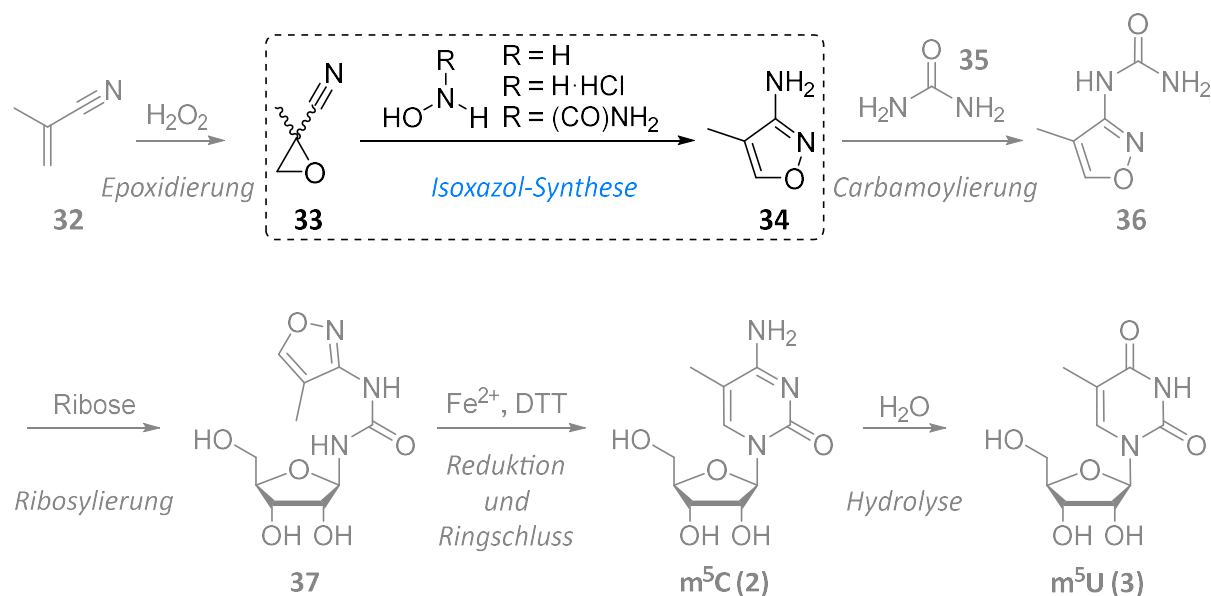
Schema 12 Mechanismus-Vorschlag zur Synthese von 2-Methyloxiran-2-carbonitril (**33**) aus Methacrylnitril (**32**) mittels H_2O_2 . **A)** Bildung der Imidoperoxysäure **41** als reaktive Spezies aus Methacrylnitril (**32**). Anschließend Epoxidierung zum Zielprodukt **33** (umrahmt). **B)** Konkurrenzreaktion zur Bildung des Zielprodukts **33**: Intramolekulare Epoxidierung der reaktiven Spezies **41** zu 2-Methyloxiran-2-carboxamid (**42**). Die Anwesenheit von **42** im Reaktionsgemisch wurde nicht untersucht. k : Geschwindigkeitskonstante **C)** Mechanismus-Vorschlag zur Synthese von 2-Methyloxiran-2-carbonitril (**33**) unter Zugabe eines α - β -gesättigten Nitrils (**82**) zur Bildung der reaktiven

Spezies **83**. Als Nebenprodukt würde das α - β -gesättigte Amid **84** gebildet werden. **D)** Die intramolekulare Epoxidierung der reaktiven Spezies **82** zu **85a** und **85b** ist nicht möglich. $R = C_nH_{2n+1}$, $n = 1, 2, 3...$
Mechanismus in Anlehnung an PAYNE *et al.*^[207]

H₂O₂ könnte auf der präbiotischen Erde als Reagenz fungiert haben. Die Anwesenheit von H₂O₂ auf der präbiotischen Erde scheint möglich. So wurde H₂O₂ sowohl auf dem Mars^[208] als auch auf dem Jupiter-Mond Europa^[209] nachgewiesen. Zwei Himmelskörper, die als die besten Kandidaten für die Entstehung von Leben in unserem Sonnensystem gelten.^[202] Darüber hinaus ist das auf der präbiotischen Erde abundante^[37] Mineral Pyrit (FeS₂) in der Lage H₂O zu H₂O₂ zu oxidieren.^[210] Neuere Erkenntnisse zeigen zudem, dass die photochemische Bildung von H₂O₂ auf der präbiotischen Erde möglich gewesen sein sollte: So wird H₂O unter Einwirkung ultravioletter Strahlung zu Hydroxyl-Radikalen umgesetzt, die anschließend zu H₂O₂ addieren können.^[211] Es liegt also nahe, dass Bildung und Umsetzung von H₂O₂ auf der frühen Erde als präbiotisch plausibel angesehen werden können.

4.2.1.2.3 Isoxazol-Synthese zu 3-Amino-4-methylisoxazol (**34**)

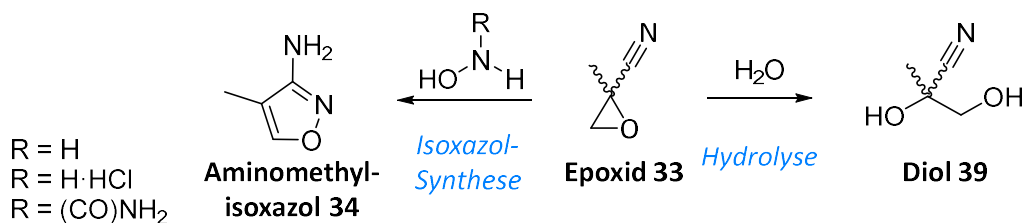
Die Umsetzung von 2-Methyloxiran-2-carbonitril (**33**) zu 3-Amino-4-methylisoxazol (**34**) unter präbiotisch plausiblen Bedingungen, siehe Schema 13, stellte sich als besonders anspruchsvoll heraus. In Vorbereitung der präbiotischen Modellreaktionen wurden Studien zur Stabilität von Epoxid **33** durchgeführt. Daraufhin erfolgte die Untersuchung einer Vielzahl von Bedingungen und Reagenzien zur Etablierung der Isoxazol-Synthese unter präbiotisch plausiblen Bedingungen. Schließlich gelang die Bildung von 3-Amino-4-methylisoxazol (**34**) aus Epoxid **33** mittels Hydroxyharnstoff. Für die Darstellung der Referenzverbindung **34** mittels herkömmlicher Synthese, siehe Abschnitt 4.2.1.3.3.



Schema 13 Ziel dieses Abschnitts war die Umsetzung von 2-Methyloxiran-2-carbonitril (33) zu 3-Amino-4-methylisoxazol (34) unter präbiotisch plausiblen Bedingungen.

4.2.1.2.3.1 Untersuchungen zur Stabilität von 2-Methyloxiran-2-carbonitril (33)

2-Methyloxiran-2-carbonitril (33) ist in wässriger Lösung auch unter erhöhter Temperatur stabil. Die Hydrolyse von 2-Methyloxiran-2-carbonitril (33) zu Diol 39 wurde als zentrale Konkurrenzreaktion zur Bildung von 3-Amino-4-methylisoxazol (34) postuliert, siehe Schema 14. Um Reaktionsbedingungen identifizieren zu können unter denen die Hydrolyse deutlich langsamer verläuft als die Isoxazol-Synthese, wurden zunächst verschiedene Reaktionsbedingungen bezüglich der Geschwindigkeit der Diol-Bildung analysiert. In einer ersten Analyse wurde die Temperatur-Stabilität von 2-Methyloxiran-2-carbonitril (33) untersucht: Epoxid 33 in H₂O/D₂O wurde nach 30 min bei RT für weitere 150 min auf 45 °C erhitzt. ¹H-NMR-Spektroskopie zeigte, dass so lediglich 10 % des Epoxids 33 zu Diol 39 hydrolysierten. Demnach scheint 2-Methyloxiran-2-carbonitril (33) bei neutralem pH keine erhöhte Sensitivität gegenüber Temperatursteigerungen aufzuweisen.



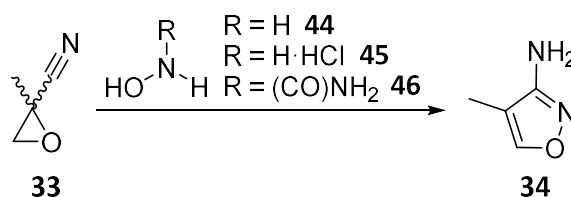
Schema 14 Hydrolyse von 2-Methyloxiran-2-carbonitril (33) zu Diol 39 stellt die zentrale Konkurrenzreaktion zur Bildung des Zielprodukts 3-Amino-4-methylisoxazol (34) dar.

Unter basischen Bedingungen hydrolysiert 2-Methyloxiran-2-carbonitril (33) in kurzer Zeit. Die Umsetzung von Epoxid **33** mit Hydroxylamin-Derivaten sollte bei $\text{pH} \neq 7$ erfolgen. Um ein Verständnis zur Stabilität von Epoxid **33** unter basischen Bedingungen zu erlangen, erfolgten Stabilitätsstudien in $\text{NaHCO}_{3(\text{aq})}$ und $\text{NaOH}_{(\text{aq})}$: Ein Gemisch aus Epoxid **33** und $\text{NaHCO}_{3(\text{aq})}$ (150 mM, 1,5 Äq.) zeigte mittels $^1\text{H-NMR}$ -Spektroskopie nach 30 min bei RT keine Signale des Edukts **33** mehr. In $\text{NaOH}_{(\text{aq})}$ (150 mM, 1,5 Äq.) konnte mittels $^1\text{H-NMR}$ -Spektroskopie nach 60 min bei RT ebenfalls kein Epoxid **33** mehr detektiert werden. Demnach scheint die Hydrolyse-Geschwindigkeit von Epoxid **33** bei $\text{pH} > 7$ vergleichsweise groß zu sein.

In Phosphat-Puffer bei pH 7 hydrolysierte 2-Methyloxiran-2-carbonitril (33) umgehend. Um die Umsetzung von 2-Methyloxiran-2-carbonitril (**33**) in gepufferter Lösung zu untersuchen, erfolgte eine Stabilitäts-Studie in Phosphatpuffer bei pH 7. Bei RT zeigte das Gemisch nach wenigen Minuten mittels $^1\text{H-NMR}$ -Spektroskopie vollständige Hydrolyse von Epoxid **33**. Die Hydrolyse-Geschwindigkeit von 2-Methyloxiran-2-carbonitril (**33**) in einer Phosphat-gepufferten Lösung bei pH 7, scheint damit deutlich geringer zu sein als in einer nicht gepufferten wässrigen Lösung. Für diesen Befund sind verschiedene Ursachen denkbar: So ist es möglich, dass der pH-Wert der Pufferlösung minimal von pH 7 abwich, sodass geringe Mengen H^+ -/ OH^- -Ionen die Hydrolyse des Epoxids **33** ermöglichten. Auch Verunreinigungen der Phosphatpuffer-Substanz könnten eine plausible Erklärung liefern. Überraschend, aber nicht vollständig auszuschließen wäre zudem, dass die geringe Nucleophilie der Phosphat-Ionen die Epoxid-Öffnung katalysierten.

4.2.1.2.3.2 präbiotische Modellreaktionen

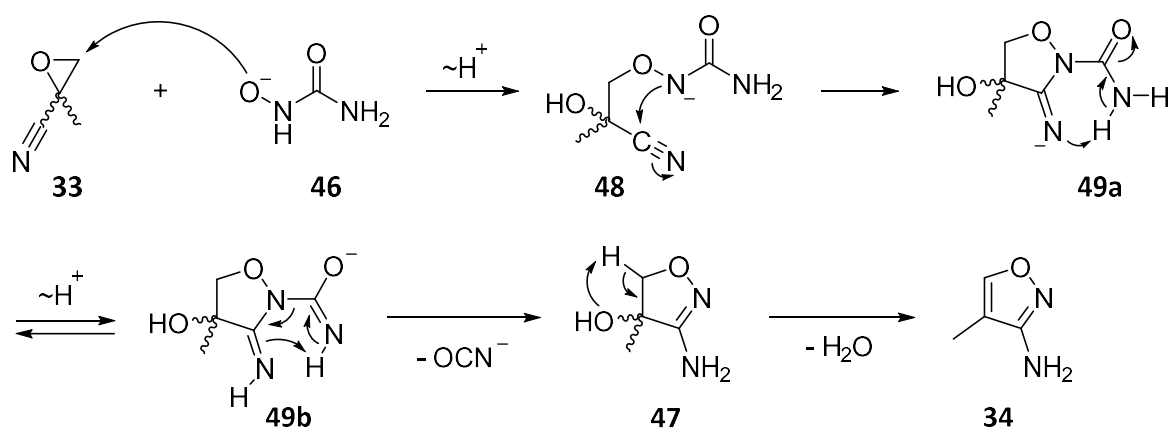
Die Bildung von 3-Amino-4-methylisoxazol (**34**) ausgehend von 2-Methyloxiran-2-carbonitril (**33**) sollte unter verschiedenen präbiotisch plausiblen Bedingungen untersucht werden, siehe Schema 15. So sollten drei verschiedene Hydroxylamin-Derivate eingesetzt werden: Hydroxylamin (**44**), Hydroxylamin-hydrochlorid (**45**) sowie Hydroxyharnstoff (**46**). Um ein fundiertes Verständnis dieser Isoxazol-Synthese zu erlangen, wurde zunächst ein Mechanismus-Vorschlag erarbeitet.



Schema 15 Die Umsetzung von 2-Methyloxiran-2-carbonitril (**33**) zu 3-Amino-4-methylisoxazol (**34**) unter präbiotisch plausiblen Bedingungen sollte mit verschiedenen Hydroxylamin-Derivaten untersucht werden: Hydroxylamin (**44**), Hydroxylamin-hydrochlorid (**45**) sowie Hydroxyharnstoff (**46**).

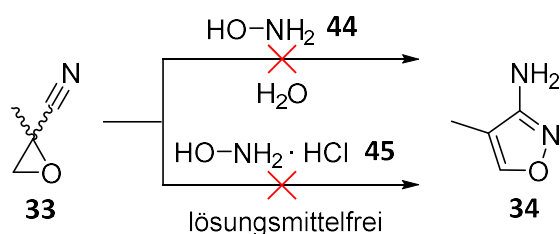
Die Umsetzung von Hydroxyharnstoff (**46**) mit 2-Methyloxiran-2-carbonitril (**33**) zu Methylisoxazolamin **34**, lässt sich in zwei Abschnitte unterteilen: **a**) Bildung der hydratisierten Zwischenstufe **47** sowie **b**) Kondensation zum Zielprodukt **34**. Schema 16 zeigt einen Mechanismus-Vorschlag zur Bildung von Methylisoxazolamin **34** aus Epoxid **33** und Hydroxylamin-Derivaten am Beispiel von Hydroxyharnstoff (**46**):

a) Im ersten Schritt sollte die Addition von Hydroxyharnstoff (**46**) an das Epoxid **33** erfolgen. Die Ringöffnung muss dazu am terminalen Kohlenstoff der Epoxid-Einheit eingeleitet werden. Angriffe an anderen elektrophilen Zentren des Epoxids **33** würden keine Zielproduktbildung erlauben. Zudem muss die Epoxid-Öffnung durch die Hydroxy-Gruppe erfolgen. Ein Angriff durch das benachbarte Stickstoff-Atom würde ebenso keine Zielproduktbildung ermöglichen. Aufgrund dieser speziellen Anforderung besitzt Hydroxyharnstoff (**46**) einen wichtigen Vorteil gegenüber Hydroxylamin **44**: In Hydroxyharnstoff (**46**) weist der Sauerstoff der Hydroxyl-Gruppe eine größere Nucleophilie auf als der benachbarte Stickstoff. Darüber hinaus ist die Hydroxyl-Gruppe des Hydroxyharnstoffs (**46**) weniger sterisch gehindert als das benachbarte Stickstoff-Atom. Beide Effekte erhöhen die Wahrscheinlichkeit eines Angriffs durch die Hydroxyl-Gruppe im Hydroxyharnstoff (**46**) im Vergleich zu Hydroxylamin (**44**). Reagenz-unabhängig ist die Epoxidöffnung am terminalen Kohlenstoff begünstigt, durch die vergleichsweise hohe sterische Abschirmung des quartären Epoxid-Kohlenstoffs. Die Öffnung des Epoxids **33** könnte die Zwischenstufe **48** bilden. Durch Angriff am Nitril-Kohlenstoff könnte anschließend der Ringschluss zum Isoxazol-Derivat **49a** erfolgen. Nach intramolekularem Protonentransfer zu **49b**, könnte über einen sechsgliedrigen Übergangszustand Cyanat eliminiert werden. **b**) Die so gebildete hydratisierte Zwischenstufe **47** könnte anschließend durch Kondensation das Zielprodukt **34** bilden. Dabei würde das Stereozentrum in **47** wieder in ein sp^2 -hybridisiertes Zentrum überführt werden.



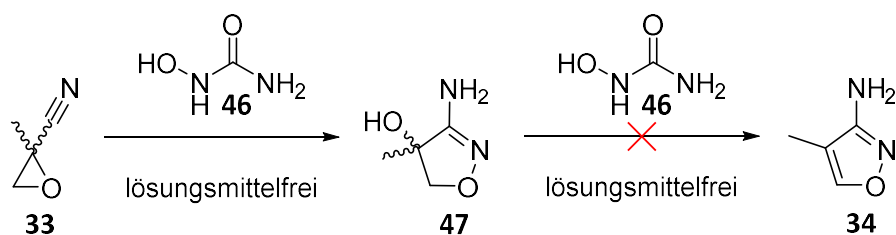
Schema 16 Mechanismus-Vorschlag zur Bildung von 3-Amino-4-methylisoxazol (**34**) durch Umsetzung von 2-Methyloxiran-2-carbonitril (**33**) mit Hydroxylamin-Derivaten, am Beispiel von Hydroxyharnstoff (**46**).

Umsetzungen von 2-Methyloxiran-2-carbonitril (**33**) mit Hydroxylamin (**44**) oder Hydroxylaminhydrochlorid (**45**) führten nicht zur Bildung von Methylisoxazolamin **34**, siehe Schema 17. Wegen dessen geringer Stabilität in reiner Form, wurde Hydroxylamin (**44**) in Lösung (50 %) eingesetzt. Die Umsetzung mit Epoxid **33** bei RT zeigte keine Bildung von Methylisoxazolamin **34**. Ein möglicher Grund dafür könnte in der ambidenten Reaktivität des Hydroxylamins (**44**) liegen: Im Gegensatz zu Hydroxylaminhydrochlorid (**45**) und Hydroxyharnstoff (**46**) besteht im Falle von Hydroxylamin (**44**) ein deutlich geringerer Unterschied zwischen der Nukleophilie von Amino- und Hydroxyl-Gruppe. Um H₂O als zentralen Faktor der Hydrolyse von Epoxid **33** auszuschließen, erfolgte die Umsetzung mit Hydroxylaminhydrochlorid (**45**) lösungsmittelfrei. ¹H-NMR-Spektroskopie nach 40 min zeigte, dass die Hydrolyse zum Diol **39** nicht vermieden werden konnte. Ein Grund dafür liegt vermutlich in der sauren Eigenschaft des Hydroxylaminhydrochlorids (**45**, pK_s 5,8).^[212]



Schema 17 Umsetzungen von 2-Methyloxiran-2-carbonitril (**33**) mit Hydroxylamin (**44**) oder Hydroxylaminhydrochlorid (**45**) führten nicht zur Bildung des Zielprodukts 3-Amino-4-methylisoxazol (**34**).

In Abwesenheit von Lösungsmitteln führten Umsetzungen von Hydroxyharnstoff (**46**) mit Epoxid **33** zur Bildung der hydratisierten Zwischenstufe **47**, siehe Schema 18. Zur Umsetzung von 2-Methyloxiran-2-carbonitril (**33**) mit Hydroxyharnstoff (**46**) ohne Lösungsmittel wurden Stöchiometrie, Temperatur sowie Reaktionsdauer variiert, siehe Tab. 1. In Reaktionen 1, 3 und 4 konnte mittels ¹H-NMR-Spektroskopie die Bildung der hydratisierten Zwischenstufe **47** beobachtet werden. Alle weiteren Reaktionen zeigten mittels LR-(LC-)ESI-MS und ¹H-NMR-Spektroskopie keine Bildung der Zwischenstufe **47** oder des Zielprodukts **34**.



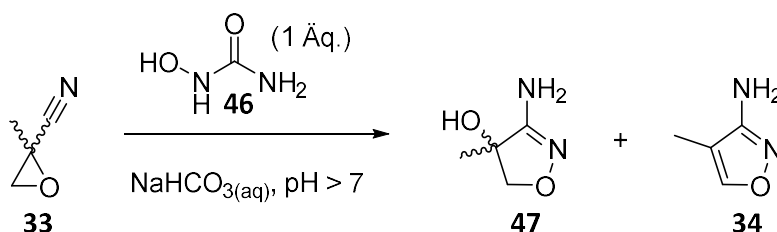
Schema 18 Umsetzung von 2-Methyloxiran-2-carbonitril (**33**) mit Hydroxyharnstoff (**46**) ohne Lösungsmittel: Die in Tab. 1 zusammengefassten Reaktionsbedingungen führten zur Bildung der Zwischenstufe **47**. Die Bildung des Zielprodukts 3-Amino-4-methylisoxazol (**34**) konnte hingegen mittels LR-(LC-)ESI-MS und $^1\text{H-NMR}$ -Spektroskopie nicht detektiert werden.

Tab. 1 Zusammenfassung untersuchter Reaktionsbedingungen zur Umsetzung von 2-Methyloxiran-2-carbonitril (**33**) mit Hydroxyharnstoff (**46**) ohne Lösungsmittel gemäß Schema 18. Detektion der Reaktionsprodukte erfolgte mittels LR-(LC-)ESI-MS und $^1\text{H-NMR}$ -Spektroskopie.

Reaktion	LR-(LC-)ESI-MS und $^1\text{H-NMR}$ -Spektroskopie		$T / ^\circ\text{C}$	t / h	Stöchiometrie / Äq.	
	Zielprodukt 34 ?	Zwischenstufe 47 ?			Epoxid 33	Hydroxyharnstoff (46)
1	-	✓	25 → 40	92	2	1
2	-	-	25	72	1	11
3	-	✓	40	72	1	11
4	-	✓	70	72	1	11
5	-	-	95	72	1	11
6	-	-	140	2	1	11

Im nächsten Schritt sollte die Umsetzung von 2-Methyloxiran-2-carbonitril (**33**) mit Hydroxyharnstoff (**46**) in Lösung und basischem pH untersucht werden, siehe Schema 19. Dazu wurden zunächst Umsetzungen bei pH-Werten zwischen pH 8 und pH 11 untersucht, siehe Tab. 2. Dabei konnte mittels LR-(LC-)ESI-MS in allen Reaktionen die Zwischenstufe **47** detektiert werden. Das Zielprodukt wurde hingegen nur bei $8 \leq \text{pH} \leq 9$ detektiert. Basierend auf diesen Erkenntnissen wurde anschließend der Einfluss der Temperatur auf die Bildung des Zielprodukts **34** untersucht, siehe Tab. 3. Dabei konnte mittels HR-LC-ESI-MS zu allen untersuchten Zeitpunkten sowohl Zielprodukt **34** als auch die Zwischenstufe **47** detektiert werden. Bereits bei der frühesten Messung (21 h bei 25 °C) wurden sowohl Zielprodukt **34** als auch Zwischenstufe **47** detektiert. Das Verhältnis Zielprodukt **34**/Zwischenstufe **47** war stets gering: Es betrug im Mittel über alle Messzeitpunkte etwa 3 %. Ab einer Reaktionszeit > 92 h

konnte die Abnahme von Zielprodukts **34** und Zwischenstufe **47** beobachtet werden. Dabei nahm auch das Verhältnis Zielprodukt **34**/Zwischenstufe **47** stetig ab. Dies weist zum einen auf Degradation hin, zum anderen könnte dies ein Hinweis auf die Rückreaktion des Zielprodukts **34** zur Zwischenstufe **47** sein.



Schema 19 Umsetzung von 2-Methyloxiran-2-carbonitril (**33**) mit Hydroxyharnstoff (**46**) in NaHCO_{3(aq)} ermöglichte die Bildung des Zielprodukts 3-Amino-4-methylisoxazol (**34**). Tab. 2 und Tab. 3 fassen die untersuchten Reaktionsbedingungen zusammen. Die Reaktionskontrolle erfolgte mittels LR-(LC-)ESI-MS. Die Quantifizierung der Ausbeuten steht noch aus.

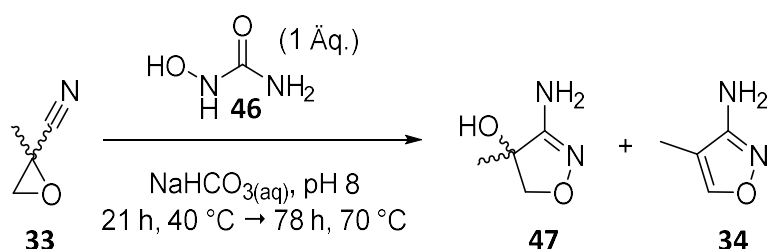
Tab. 2 Zusammenfassung untersuchter Reaktionsbedingungen zur pH-Abhängigkeit der Umsetzung von 2-Methyloxiran-2-carbonitril (**33**) mit Hydroxyharnstoff (**46**) in (Na₂CO₃/NaHCO₃)_(aq), vgl. Schema 19. Detektion mittels LR-(LC-)ESI-MS.

Reaktion	LR-(LC-)ESI-MS		pH	weitere Bedingungen
	Zielprodukt 34 ?	Zwischenstufe 47 ?		
7	✓	✓	8 ≤ pH ≤ 9	21 h, 25 °C, (Na ₂ CO ₃ /NaHCO ₃) _(aq)
8	-	✓	9 ≤ pH ≤ 10	
9	-	✓	10 ≤ pH ≤ 11	

Tab. 3 Zusammenfassung untersuchter Reaktionsbedingungen zur Umsetzung von 2-Methyloxiran-2-carbonitril (**33**) mit Hydroxyharnstoff (**46**) in NaHCO_{3(aq)} bei verschiedenen Temperaturverläufen, vgl. Schema 19. Detektion mittels HR-LC-ESI-MS. Die zentralen Unterschiede zwischen den Temperaturverläufen sind **fett** hervorgehoben.

Reaktion	HR-LC-ESI-MS		Temperaturverlauf	weitere Bedingungen
	Zielprodukt 34 ?	Zwischenstufe 47 ?		
10	✓	✓	21 h, 40 °C → 78 h, 70 °C	pH 8, NaHCO _{3(aq)}
11	✓	✓	21 h, 40 °C → 78 h, 85 °C	
12	✓	✓	21 h, 25 °C → 29 h, 40 °C → 14 d, 70 °C	

Schema 20 fasst die Reaktionsbedingungen zusammen die eine maximale Bildung des Zielprodukts 3-Amino-4-methylisoxazol (**34**) ermöglichten. Kondensationsreaktionen sind in wässriger Lösung aufgrund des Prinzips von LE CHATELIER stark gehemmt. Deshalb stellt die Kondensation zum Zielprodukt **34** die zentrale Hürde zur Steigerung der Ausbeute des Zielprodukts **34** dar. Falls im nächsten Schritt eine weitere Steigerung der Ausbeute angestrebt wird, sollten zunächst die hier etablierten Bedingungen in einem offenen Reaktionssystem angewandt werden. Eine vollständige Evaporation des Lösungsmittels sollte die Kondensation deutlich begünstigen. Zudem sollte die Umsetzung von 2-Methyloxiran-2-carbonitril (**33**) mit Hydroxyharnstoff (**46**) ohne Lösungsmittel genauer untersucht werden. Insbesondere der Einsatz von HR-LC-ESI-MS würde eine bessere Vergleichbarkeit der Ergebnisse ermöglichen. Darüber hinaus sollte im nächsten Schritt eine zusätzliche Validierung der Ergebnisse durch HR-LC-ESI-MS-Koinjektion der Referenzverbindung **34** erfolgen.



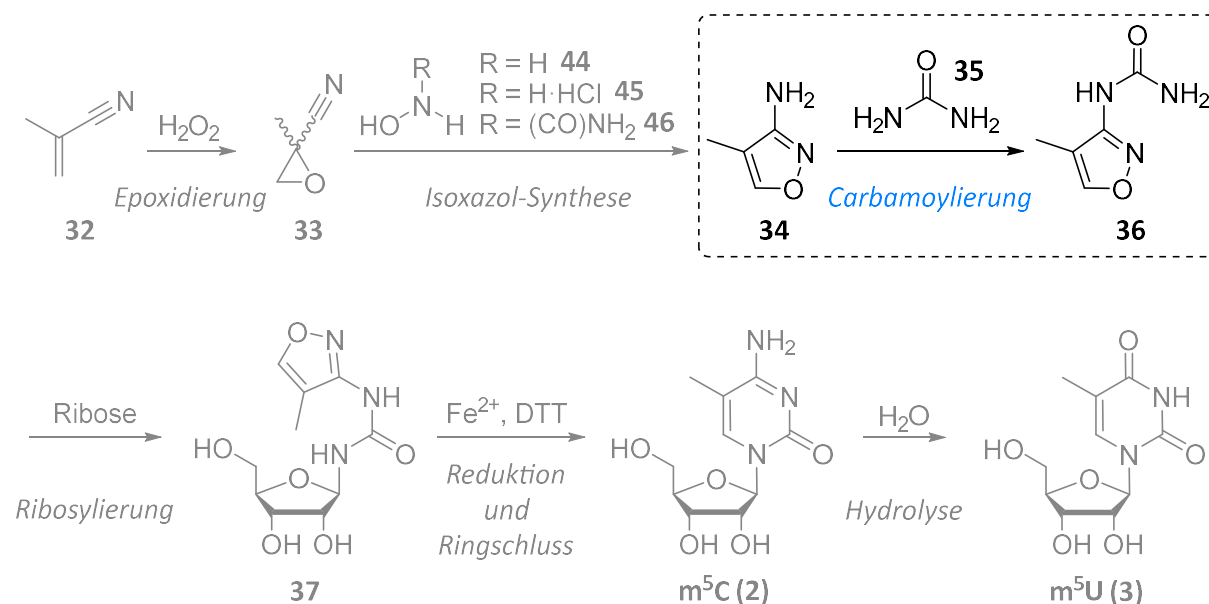
Schema 20 Umsetzung von 2-Methyloxiran-2-carbonitril (**33**) mit Hydroxyharnstoff (**46**) in $\text{Na}_2\text{CO}_3(\text{aq})$ ermöglichte die Bildung des Zielprodukts 3-Amino-4-methylisoxazol (**34**). Tab. 2 und Tab. 3 fassen die untersuchten Reaktionsbedingungen zusammen. Die Reaktionskontrolle erfolgte mittels HR-LC-ESI-MS. Die Quantifizierung der Ausbeuten steht noch aus.

Zusammengenommen konnte in einer Vielzahl präbiotischer Modellreaktionen die robuste Bildung von 3-Amino-4-methylisoxazol (**34**) demonstriert werden. Dies war in einem breiten Temperaturbereich, in An- und Abwesenheit von H_2O sowie im pH-Wert-Bereich zwischen pH 8 und pH 9 möglich. Im Anschluss an die Synthese der Referenzverbindungen, vgl. Abschnitt 4.2.1.3.3, sollte die Quantifizierung der Ausbeute mittels HR-LC-ESI-MS-Koinjektion erfolgen.

4.2.1.2.4 Carbamoylierung zu 4-Methylisoxazol-3-ylharnstoff (**36**)

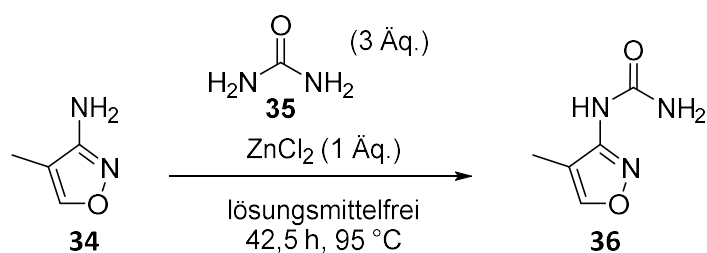
Ziel dieses Abschnitts war die Carbamoylierung von 3-Amino-4-methylisoxazol (**34**) zu 4-Methylisoxazol-3-ylharnstoff (**36**) unter präbiotisch plausiblen Bedingungen. Für die

Darstellung der Referenzverbindung **36** mittels herkömmlicher Synthese, siehe Abschnitt 4.2.1.3.4.



Schema 21 Ziel dieses Abschnitts war die Carbamoylierung von 3-Amino-4-methylisoxazol (**34**) zu 4-Methylisoxazol-3-ylharnstoff (**36**) unter präbiotisch plausiblen Bedingungen.

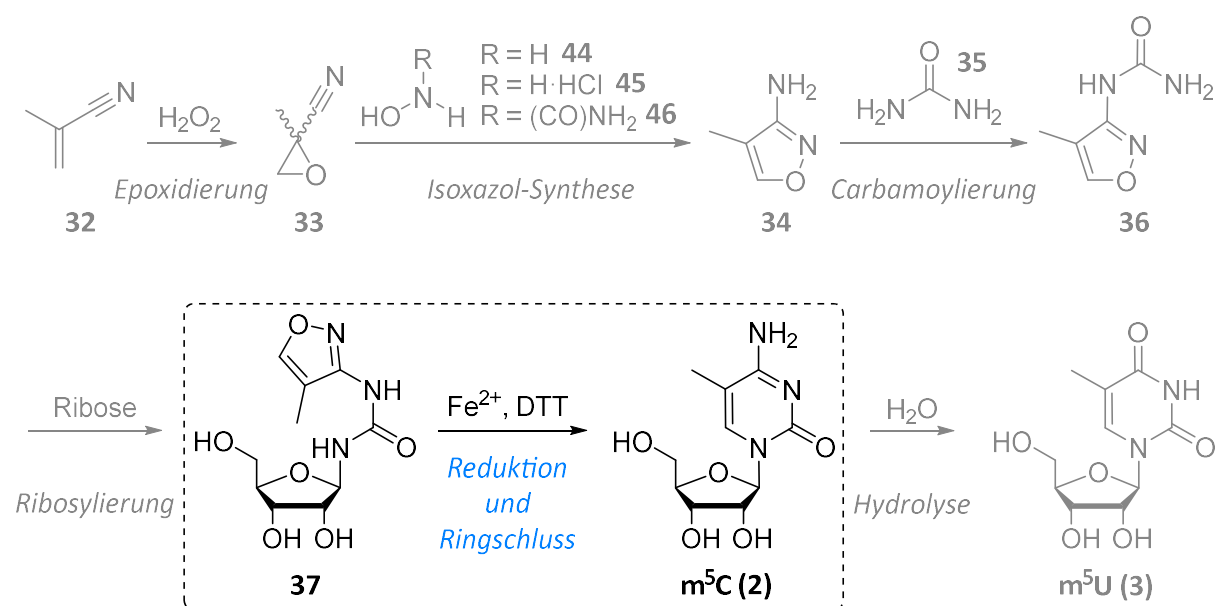
Die Herstellung von 4-Methylisoxazol-3-ylharnstoff (**36**) gelang in einer präbiotischen Modellreaktion lösungsmittelfrei in Anwesenheit von ZnCl_2 . In Anlehnung an die Umsetzung des unmethylierten Edukt-Analogs Isoxazol-3-ylharnstoff, vgl. Abschnitt 4.1, wurde 3-Amino-4-methylisoxazol (**34**) mit Harnstoff (**35**) und ZnCl_2 lösungsmittelfrei bei 95°C umgesetzt. Aufgrund der hygroskopischen Eigenschaften von ZnCl_2 entsteht innerhalb kurzer Zeit ein wässriges Reaktionsgemisch. Untersuchung des Reaktionsgemisches mittels HR-LC-ESI-MS nach 42,5 h zeigte die Bildung des Zielprodukts Methylisoxazolylharnstoff **36**. Darüber hinaus zeigten sowohl das UV-Chromatogramm als auch das Massenspektrum kaum Nebenproduktbildung. Zur Validierung der präbiotischen Synthese sollte im nächsten Schritt eine HR-LC-ESI-MS-Koinjektion der Referenzverbindung **36** erfolgen.



Schema 22 Präbiotische Modellreaktion zur Bildung von 4-Methylisoxazol-3-ylharnstoff (**36**), ausgehend von 3-Amino-4-methylisoxazol (**34**) und Harnstoff (**35**). Die Reaktionskontrolle erfolgte mittels HR-LC-ESI-MS. Die Quantifizierung der Ausbeute ist noch ausstehend.

4.2.1.2.5 Voruntersuchungen zu Reduktion und Ringschluss

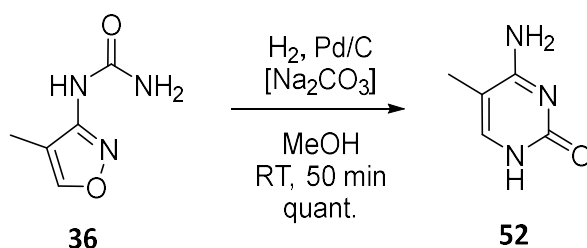
Die in diesem Abschnitt vorgestellte Studie erfolgte als Voruntersuchung des Projekts „4.2.1 Studien zur Bildung von m^5C (**2**) und m^5U (**3**) unter präbiotisch plausiblen Bedingungen“. Ziel war es zunächst zu untersuchen, ob die Reduktion der Isoxazol-N-O-Bindung in Methylisoxazolylharnstoff-Derivaten grundsätzlich selektiv möglich ist und ob so die Zyklisierung initiiert werden kann.



Schema 23 Ziel dieses Abschnitts war die Etablierung von Reduktion und Ringschluss zu m^5C (**2**).

Die selektive Reduktion und Zyklisierung der 4-Methylisoxazol-3-ylharnstoff-Einheit ist möglich. Als Modellsystem für diese Voruntersuchung wurde 4-Methylisoxazol-3-ylharnstoff (**36**) gewählt, siehe Schema 24. Es scheint naheliegend, dass sich 4-Methylisoxazol-3-ylharnstoff (**36**) bezüglich Reduktion und Zyklisierung sehr ähnlich zu N^1 -Ribosyl- N^2 -(4-methylisoxazol-3-yl)harnstoff **37** verhalten sollte. Da 4-Methylisoxazol-3-ylharnstoff (**36**) zudem leicht zugänglich war, stellte es ein ideales Modellsystem dar. Die Voruntersuchung erfolgte durch Reduktion mit H_2 an Pd/C. Nach 50 min konnte die quantitative Bildung von 5-Methylcytosin (**52**) mittels ^{13}C -NMR-Spektroskopie sowie HR-LC-ESI-MS nachgewiesen werden. Eine selektive Reduktion und Zyklisierung der 4-Methylisoxazol-3-ylharnstoff-Einheit ist also möglich. Die Bildung des Zielprodukts **52** mittels H_2 und Pd/C ohne Bildung von Nebenprodukten demonstriert die hohe Beständigkeit des restlichen Moleküls gegenüber vergleichsweise harschen reduktiven Bedingungen. Dieses Ergebnis legte die Basis für den Beginn der chemisch anspruchsvollen Studien zur Bildung von 2-Methyloxiran-2-carbonitril (**33**) sowie Methylisoxazolamin **34**, vgl. Abschnitte 4.2.1.2 und 4.2.1.2.3. Es scheint naheliegend, dass auch die Reduktion der ribosylierten Form **37** in einer präbiotischen Modellreaktion die Bildung von m^5C (**2**) ermöglicht. Im Anschluss

an die Synthese der ribosylierten Form **37** sollte demnach dessen Umsetzung in Anlehnung an die in Abschnitt 4.1 etablierte präbiotische Modellreaktion erfolgen.

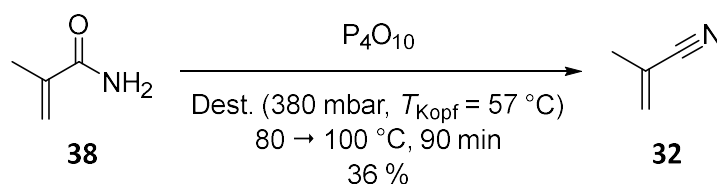


Schema 24 Die Reduktion von 4-Methylisoxazol-3-ylharnstoff (**36**) mit H₂ und Pd/C führte zur quantitativen Bildung von 5-Methylcytosin (**52**).

4.2.1.3 Synthese der Referenzverbindungen

4.2.1.3.1 Methacrylnitril (**32**)

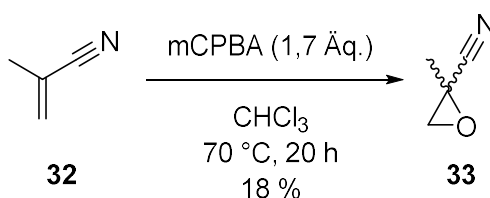
Zur Herstellung von Methacrylnitril (**32**) konnte eine existierende Synthese^[213] weiterentwickelt werden. Da Methacrylnitril (**32**) das Edukt der mehrstufigen Synthese von m⁵C (**2**) darstellt, war dessen Synthese im Gramm-Maßstab notwendig. Dies sollte durch Dehydratisierung von Methacrylformamid (**38**) mittels P₄O₁₀ erreicht werden.^[213] Wiederholte Optimierung der Umsetzung ermöglichte es schließlich Methacrylnitril (**32**) in hoher Reinheit und 36 % Ausbeute zu erhalten, siehe Schema 25. Dies gelang mittels trockener Destillation (Erhitzen des Reaktionsgemisches unter vermindertem Druck). Dabei bildete das unter vermindertem Druck gasförmige Methacrylnitril (**32**) mit dem Reaktionsgemisch einen festen Schaum. Der Einschluss des Zielprodukts **32** im Schaum, trug maßgeblich zur Reduktion der Ausbeute bei. Zudem verhinderte die feste Konsistenz des Schaumes jegliche Durchmischung der Reaktion.



Schema 25 Darstellung von Methacrylnitril (**32**) aus Methacrylformamid (**38**) mittels herkömmlicher Synthese, durch Dehydratisierung mittels P₄O₁₀.

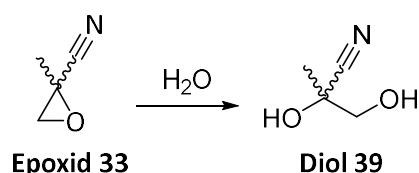
4.2.1.3.2 2-Methyloxiran-2-carbonitril (**33**)

Zur Herstellung des Epoxids **33** konnte eine mCPBA-vermittelte Synthese entwickelt werden. Die Epoxidierung von Methacrylnitril (**32**) mittels $\text{Co}(\text{ClO}_4)_2$ -Katalyse,^[214] führte nicht zur Bildung von Epoxid **33**. Stattdessen gelang die Synthese des Epoxids **33** durch Umsetzung von Methacrylnitril (**32**) mit mCPBA (*meta*-Chlorperbenzoesäure). Filtration des Rohprodukts und wiederholte Vakuumdestillation ergaben 2-Methyloxiran-2-carbonitril (**33**) in hoher Reinheit und 18 % Ausbeute. Gründe für die geringe Ausbeute waren zum einen die anspruchsvolle Aufreinigung. Zum anderen wurde die Destillation vor der vollständigen Trocknung des Rückstandes beendet. Dies geschah, um die Anreicherung möglicherweise gebildeter explosiver Peroxid-Nebenprodukte zu verhindern. Es ist zu erwarten, dass eine weitere Optimierung der Aufreinigung bei Bedarf die Ausbeute erhöhen könnte.



Schema 26 Darstellung von 2-Methyloxiran-2-carbonitril (**33**) aus Methacrylnitril (**32**) mittels herkömmlicher Synthese mit Hilfe von mCPBA.

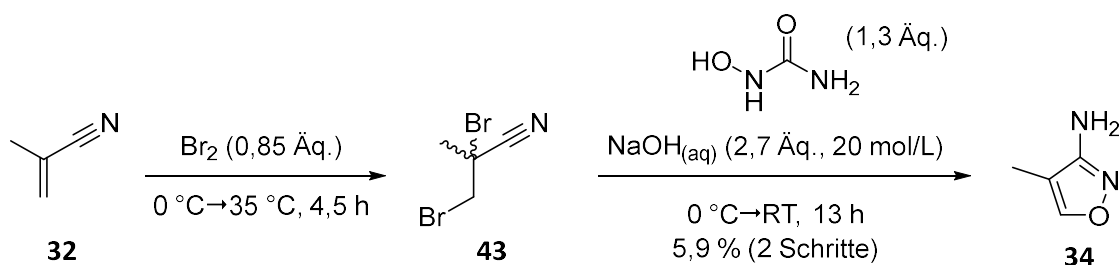
Der Nachweis des Epoxids **33** gelang ausschließlich mittels NMR-Spektroskopie. Der anspruchsvolle Nachweis von Epoxid **33** begründet sich vermutlich in dessen vergleichsweise hoher Reaktivität gegenüber Nukleophilen. So führten weder EI- noch ESI-Massenspektrometrie zu aussagekräftigen Ergebnissen. Auch die zu Epoxid **33** publizierten EI-MS-Daten^[215] enthalten kein Molekül-Ion. Die angegebenen Fragment-Massen deuten lediglich auf den üblichen Hintergrund eines Massenspektrometers hin^[216] und ließen sich nicht reproduzieren. Zudem weist Epoxid **33** keine IR-Schwingungen auf, die eine Unterscheidung von dessen Hydrolyseprodukt Diol **39** erlauben, siehe Schema 27. Der Nachweis des Zielprodukts **33** gelang schließlich mittels NMR-Spektroskopie und Abgleich mit publizierten Daten des Zielprodukts **33** sowie von Verbindungen analog zu Diol **39**.^[217,218] Darüber hinaus existieren in der Literatur keine weiteren aussagekräftige Charakterisierungen des Epoxids **33**. Eine vielversprechende Alternative könnte im nächsten Schritt die Charakterisierung mittels Röntgen-Kristallstrukturanalyse darstellen. Da Epoxid **33** bei RT als Flüssigkeit vorliegt, müssten dazu Kristallisation und Messung unterhalb dessen Schmelzpunktes stattfinden.



Schema 27 Hydrolyse von 2-Methyloxiran-2-carbonitril (**33**) zu Diol **39**.

4.2.1.3.3 3-Amino-4-methylisoxazol (**34**)

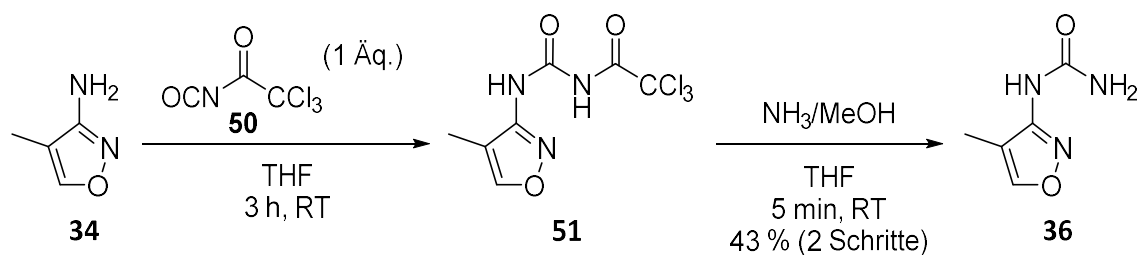
Die Herstellung von 3-Amino-4-methylisoxazol (**34**) gelang mittels herkömmlicher Synthese in Anlehnung an KÖHN et al.,^[219] siehe Schema 28. Zunächst wurde Methacrylnitril (**32**) mit Br₂ für 5 h bei 35 °C umgesetzt. Die Zwischenstufe 2,3-Dibrom-2-methylpropanitril **43** wurde anschließend *in situ* mit NaOH_(aq) und Hydroxyharnstoff reagiert. Nach Aufarbeitung konnte 3-Amino-4-methylisoxazol (**34**) in reiner Form als Referenzverbindung erhalten werden. Die geringe Ausbeute von 5,9 % über zwei Schritte begründet sich in der großen Zahl an Reaktionskontrollen der zweistufigen Synthese. Darüber hinaus sollte in Zukunft bei der Extraktion pH 7 beibehalten werden, um einem möglichen Verlust der deprotonierten Form des Zielprodukts **34** in die wässrige Phase vorzubeugen.



Schema 28 Zweistufige Darstellung von 3-Amino-4-methylisoxazol (**34**) mittels herkömmlicher Synthese, durch Bromierung von Methacrylnitril (**32**).

4.2.1.3.4 4-Methylisoxazol-3-ylharnstoff (**36**)

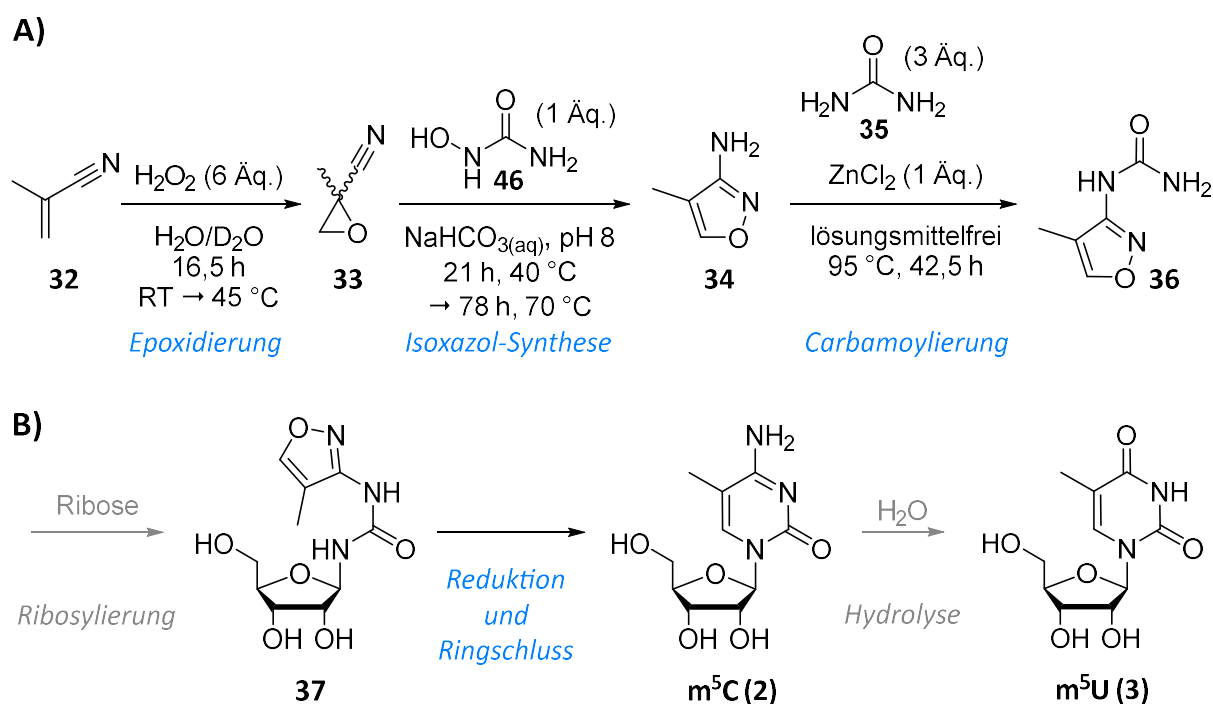
Zur Herstellung von 4-Methylisoxazol-3-ylharnstoff (**36**) konnte eine zweistufige Synthese entwickelt werden, siehe Schema 29. Zunächst wurde 3-Amino-4-methylisoxazol (**34**) mit Trichloracetylisocyanat **50** in THF bei RT carbamoyliert. Nach Bildung der Trichloracetylgeschützten Zwischenstufe **51**, erfolgte dessen Entschützung *in situ* durch Zugabe von NH₃ in MeOH. Aufreinigung lieferte das Zielprodukt 4-Methylisoxazol-3-ylharnstoff (**36**) als Referenzverbindung in einer Ausbeute von 43 % über zwei Schritte.



Schema 29 Darstellung von 4-Methylisoxazol-3-ylharnstoff (**36**) mittels herkömmlicher Synthese in zwei Schritten: Umsetzung von 3-Amino-4-methylisoxazol (**34**) mit 1,1,1-Trichloracetylisocyanat (**50**) und anschließende Entschützung durch NH₃ in MeOH.

4.2.1.4 Zusammenfassung der Studien zur Bildung von m⁵C (**2**) und m⁵U (**3**) und Ausblick

In diesem Teil der Arbeit konnten die zentralen Schritte auf dem Weg zur Synthese von m⁵C (**2**) unter präbiotisch plausiblen Bedingungen etabliert werden. Schema 30 fasst die etablierten präbiotischen Modellreaktionen zusammen (schwarz). Die größten chemischen Unwägbarkeiten befanden sich in Teil A) der für dieses Projekt entwickelten Synthesestrategie: Es gelang aber alle drei Syntheseschritte in präbiotischen Modellreaktionen zu demonstrieren. Darüber hinaus wurden Zugänge zu den Verbindungen **32**, **33**, **34** und **36** mittels herkömmlicher Synthese etabliert. Bezüglich Teil B) der Synthesestrategie wurden Reduktion und Ringschluss anhand von **36** untersucht. Dessen erfolgreiche Reduktion und Zyklisierung unter präbiotisch plausiblen Bedingungen legt nahe, dass auch die Reduktion von **37** in einer präbiotischen Modellreaktion die Bildung von m⁵C (**2**) ermöglicht.



Schema 30 Zusammenfassung der in dieser Arbeit etablierten präbiotischen Modellreaktionen zur Bildung von m^5C (**2**) und m^5U (**3**), unterteilt in Teil **A**) und **B**) des Synthesepfades. In dieser Arbeit nicht adressierte Reaktionen sind grau dargestellt. Zusätzlich wurden mittels herkömmlicher Synthese Zugänge zu den Verbindungen **32**, **33**, **34** und **36** etabliert. Die Quantifizierung der Ausbeuten steht noch aus.

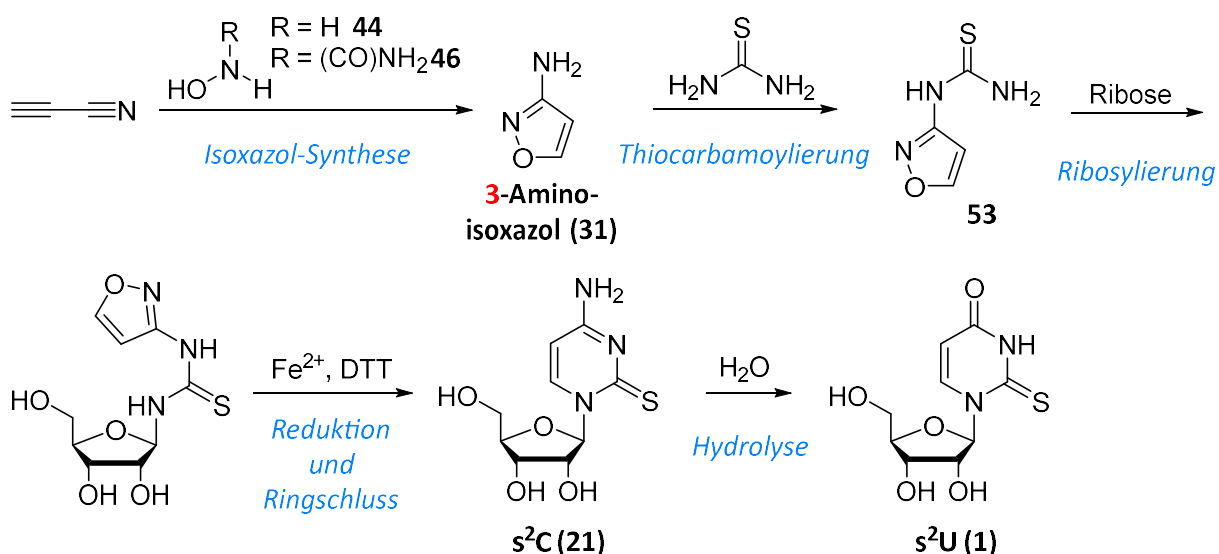
Um dieses Projekt zur Publikationsreife zu führen, sollte die Ribosylierung von **36** unter präbiotisch plausiblen Bedingungen etabliert werden, bevor Reduktion und Ringschluss an **37** demonstriert werden können. Nach Synthese der Referenzverbindungen zu **37** in Anlehnung an Abschnitt 4.1, können Validierungen mittels HR-LC-ESI-MS-Koinjektionen erfolgen. Alle weiteren Referenzverbindungen wurden bereits synthetisiert bzw. sind kommerziell erhältlich. Ausgehend von m^5C (**2**) kann darüber hinaus mittels Hydrolyse m^5U (**3**) generiert werden.^[198–201] Das RNA-Analog und ein möglicher Vorläufer des kanonischen DNA-Nukleosids dT.

4.2.2 Studien zur Bildung von s²U unter präbiotisch plausiblen Bedingungen

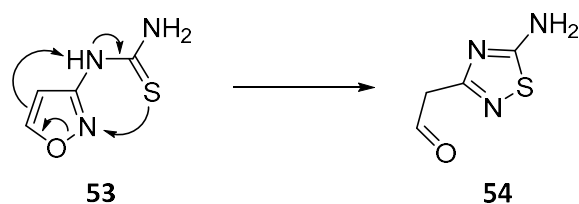
Thio-Nucleoside nahmen womöglich wichtige Funktionen bei der Entstehung des Lebens auf der Erde ein, vgl. Einleitung Abschnitt 1.4.1. Die Bildung des potenziell zentralen Thio-Nucleosids s²U (2-Thiouridin, **1**) konnte bereits in präbiotischen Modellreaktionen demonstriert werden.^[55,57] Die Bedingungen dieses Zugangs zu s²U (**1**) sind aber nicht kompatibel mit einer vereinigten Bildung aller kanonischen RNA-Nucleoside. Demnach sollte in diesem Teil der Arbeit ein präbiotisch plausibler Zugang zur Bildung von s²U (**1**) identifiziert werden. Dieser sollte, wenn möglich, kompatibel mit der vereinigten Bildung der kanonischen Pyrimidin- und Purin-Nucleoside sein, vgl. Abschnitt 4.1.

4.2.2.1 Synthesestrategie

Zunächst wurde die Bildung von s²U (**1**) via 3-Aminoisoxazol (**31**) untersucht. Im ersten Schritt wurde eine Synthesestrategie ausgehend von dem Schlüssel-molekül der vereinigten Pyrimidin- und Purin-Nucleosid-Synthese entworfen: 3-Aminoisoxazol (**31**), siehe Schema 31. Nach der initialen Synthese von 3-Aminoisoxazol (**31**)^[220] sollte dessen Thiocarbamoylierung zu N-(Isioxazol-3-yl)thioharnstoff (**53**) erfolgen. Eine Vielzahl an Voruntersuchungen ermöglichte jedoch keine Isolation des Zielprodukts N-(Isioxazol-3-yl)thioharnstoff (**53**), weder in präbiotischen Modellreaktionen noch mittels herkömmlicher Synthese. Ein Grund dafür könnte die Umlagerung von N-(Isioxazol-3-yl)thioharnstoff (**53**) nach BOULTON-KATRITZKY darstellen, siehe Schema 32.^[221] Somit war die Entwicklung einer alternativen Synthesestrategie nötig.

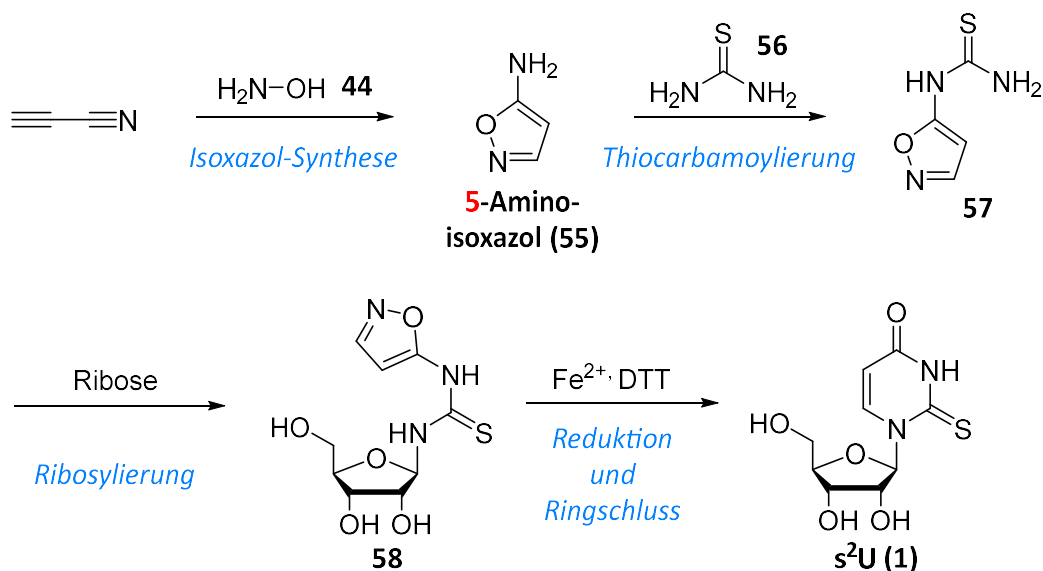


Schema 31 Synthesestrategie zur Bildung von s²C (**21**) und s²U (**1**) in präbiotischen Modellreaktionen, via 3-Aminoisoxazol (**31**).



Schema 32 Mechanismus der BOULTON-KATRITZKY-Umlagerung von *N*-(Isoxazol-3-yl)thioharnstoff (**53**) zum Thiadiazolaldehyd **54**.^[221,222] Die BOULTON-KATRITZKY-Umlagerung führte möglicherweise dazu, dass *N*-(Isoxazol-3-yl)thioharnstoff (**53**) in Voruntersuchungen nicht erhalten werden konnte.

Die Bildung von *s*²U (**1**) via 5-Aminoisoxazol (**55**), umgeht die BOULTON-KATRITZKY-Umlagerung. Zur Vermeidung der BOULTON-KATRITZKY-Umlagerung wurde eine Synthesestrategie ausgehend von 5-Aminoisoxazol (**55**) entwickelt, siehe Schema 33. 5-Aminoisoxazol (**55**) kann aus simplen und präbiotisch plausiblen Vorläufern gebildet werden.^[223] Anschließend sollte die Thiocarbamoylierung des exozyklischen Amins mittels Thioharnstoff (**56**) unter Bildung von *N*-(Isoxazol-5-yl)thioharnstoff (**57**) etabliert werden. Nach Ribosylierung von *N*-(Isoxazol-5-yl)thioharnstoff (**57**) zu **58** sollte der Isoxazol-Ring reaktiv geöffnet werden, um die Zyklisierung zu *s*²U (**1**) zu ermöglichen. Die im Folgenden beschriebenen Ergebnisse legen nahe, dass die Bildung von *s*²U (**1**) in präbiotischen Modellreaktionen möglich ist. Die Durchführung erfolgte größtenteils in Zusammenarbeit mit ANNIKA TÖLKE (AK CARELL, LMU München).

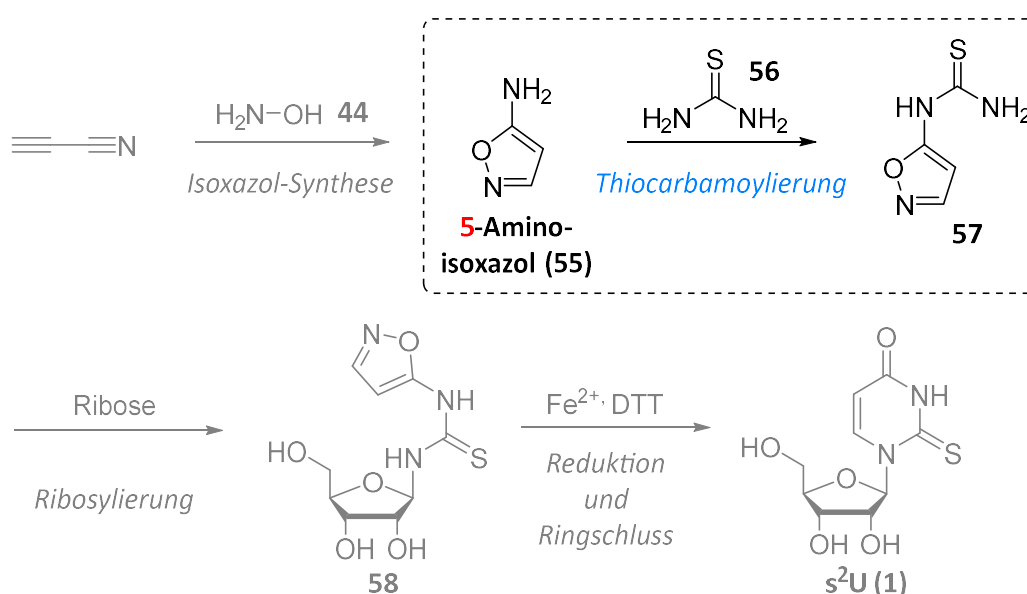


Schema 33 Synthesestrategie zur Bildung von *s*²U (**1**) in präbiotischen Modellreaktionen, via 5-Aminoisoxazol (**55**). Eine BOULTON-KATRITZKY-Umlagerung von *N*-(Isoxazol-5-yl)thioharnstoff (**57**) ist nicht zu erwarten, vgl. Schema 32.

4.2.2.2 präbiotische Modellreaktionen

4.2.2.2.1 Thiocarbamoylierung zu *N*-(Isoxazol-5-yl)thioharnstoff (57)

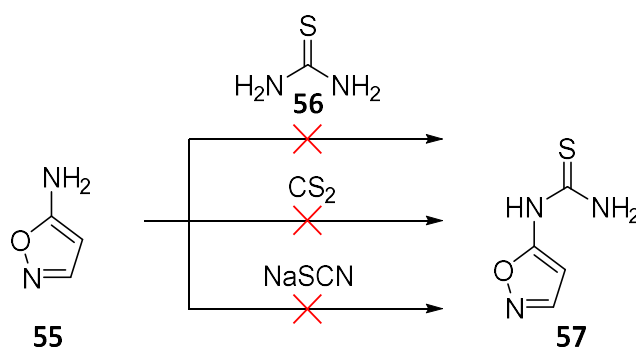
Die Installation der Thioharnstoff-Einheit in präbiotischen Modellreaktionen stellte die größte Herausforderung auf dem Weg zur Bildung von *s*²U (1) dar. Ziel dieses Abschnitts war die Thiocarbamoylierung von 5-Aminoisoxazol (55) unter präbiotisch plausiblen Bedingungen, siehe Schema 34. Zunächst erfolgte die Herstellung von *N*-(Isoxazol-5-yl)thioharnstoff (57) als Referenzverbindung mittels herkömmlicher Synthese. Die Bildung von *N*-(Isoxazol-5-yl)thioharnstoff (57) unter präbiotisch plausiblen Bedingungen stellte sich hingegen als sehr anspruchsvoll heraus. Erst die Untersuchung einer Vielzahl an Reaktionen eröffnete einige vielversprechende Möglichkeiten, die in den folgenden Abschnitten vorgestellt werden. Neben den im Folgenden vorgestellten Zugängen erlaubten auch Umsetzungen von 3-Aminoisoxazol (31) mit Thioharnstoff (56) und Fe-Katalysatoren die Detektion von *N*-(Isoxazol-3-yl)thioharnstoff (53). Aufgrund der beschriebenen Herausforderungen durch die BOULTON-KATRITZKY-Umlagerung, wurden diese Reaktionen hier nicht näher beschrieben. Eine detaillierte Beschreibung dieser Umsetzungen wurde von ANNIKA TÖLKE angefertigt.^[224] Für die Darstellung der Referenzverbindung des Zielprodukts 57 mittels herkömmlicher Synthese, siehe Abschnitt 4.2.2.3.1.



Schema 34 Ziel dieses Abschnitts war die Thiocarbamoylierung von 5-Aminoisoxazol (55) zu *N*-(Isoxazol-5-yl)thioharnstoff (57) unter präbiotisch plausiblen Bedingungen.

4.2.2.2.1.1 Umsetzungen mit Thioharnstoff, CS₂ oder Thiocyanat

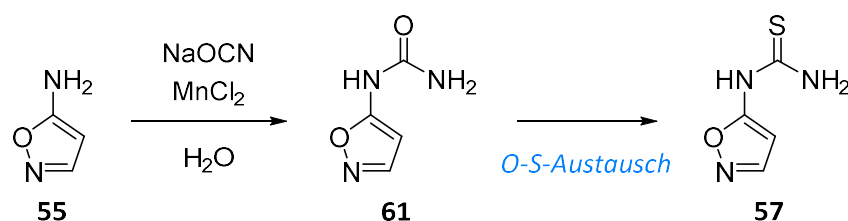
Die Bildung von *N*-(Isoxazol-5-yl)thioharnstoff (**57**) in präbiotischen Modellreaktionen wurde zunächst mit Thioharnstoff (**56**), CS₂ sowie NaSCN unter einer Vielzahl von Bedingungen untersucht. Unter keiner dieser Bedingungen konnte die Bildung des Zielprodukts *N*-(Isoxazol-5-yl)thioharnstoff (**57**) detektiert werden. Die in Schema 35 zusammengefassten Studien wurden im Detail von ANNIKA TÖLKE beschrieben.^[224] In den folgenden beiden Abschnitte werden zwei Strategien präsentiert, die ohne eine *direkte* Thiocarbamylierung von 5-Aminoisoxazol (**55**) die Bildung von *N*-(Isoxazol-5-yl)thioharnstoff (**57**) ermöglichen sollten.



Schema 35 Zusammenfassung untersuchter Umsetzungen zur Bildung von *N*-(Isoxazol-5-yl)thioharnstoff (**57**) aus 5-Aminoisoxazol (**55**). Untersucht wurden eine Vielzahl an Bedingungen in Anwesenheit von Thioharnstoff (**56**), CS₂ oder NaSCN. Keine der untersuchten Bedingungen führte zur Bildung des Zielprodukts **57**. Diese Studien wurden im Detail von ANNIKA TÖLKE beschrieben.^[224]

4.2.2.2.1.2 Umsetzungen mit P₄S₁₀

Ein O-S-Austausch könnte die Synthese von *N*-(Isoxazol-5-yl)thioharnstoff (**57**) ermöglichen. Im Gegensatz zur Thiocarbamylierung von 5-Aminoisoxazol (**55**), konnte dessen Carbamylierung bereits in einer präbiotischen Modellreaktion demonstriert werden.^[223] Um diesen Umstand für die Synthese von *N*-(Isoxazol-5-yl)thioharnstoff (**57**) nutzbar zu machen, sollte ein O-S-Austausch an *N*-(Isoxazol-5-yl)harnstoff (**61**) etabliert werden. Schema 36 zeigt eine Strategie zur Synthese von *N*-(Isoxazol-5-yl)thioharnstoff (**57**) mittels Carbamylierung von 5-Aminoisoxazol (**55**) und anschließendem O-S-Austausch.



Schema 36 Synthesestrategie zur Bildung von *N*-(Isoxazol-5-yl)thioharnstoff (**57**) durch O-S-Austausch an *N*-(Isoxazol-5-yl)harnstoff (**61**). Die Bildung von *N*-(Isoxazol-5-yl)harnstoff (**61**) in einer präbiotischen Modellreaktion wurde außerhalb dieser Arbeit demonstriert.^[223]

Mittels P_4S_{10} gelang der O-S-Austausch an *N*-(Isoxazol-5-yl)harnstoff (**61**). Zunächst erfolgte die Carbamoylierung von 5-Aminoisoxazol (**55**).^[223] Anschließend wurde der O-S-Austausch an *N*-(Isoxazol-5-yl)harnstoff (**61**) mit P_4S_{10} unter verschiedenen Bedingungen untersucht, siehe Tab. 4. Dabei ermöglichte Reaktion 14, eine Umsetzung in CS₂ und NEt₃, die Detektion des Zielprodukts **57** mittels ¹H-NMR-Spektroskopie sowie HR-LC-ESI-MS. Die übrigen Reaktionsbedingungen führten hingegen nicht zum Erfolg.

Tab. 4 Zusammenfassung untersuchter Reaktionsbedingungen des O-S-Austauschs an *N*-(Isoxazol-5-yl)harnstoff (**61**), vgl. Schema 36. Zielprodukt: *N*-(Isoxazol-5-yl)thioharnstoff (**57**). Detektion des Zielprodukts mittels ¹H-NMR-Spektroskopie sowie HR-LC-ESI-MS. Rkt.: Reaktion. CS₂: 787 μ L, NEt₃: 39 μ L. *pressure tube*: geschlossenes Reaktionssystem. Mörser: Vermischung der Edukte mittels Mörser und Pistill.

Rkt.	Zielprodukt 57 detektiert?	P_4S_{10} / Äq.	CS ₂	NEt ₃	Atmosphäre	T / °C	<i>pressure tube</i> ?	weiteres
13	-	0,22	✓	✓	N ₂	RT	-	-
14	✓	0,22	✓	✓	Ar	RT→35	✓	-
15	-	0,22	-	✓	N ₂	RT	✓	-
16	-	0,22	-	✓	N ₂	RT→35	✓	Mörser
17	-	0,44	✓	✓	N ₂	RT→35	✓	-
18	-	0,22	-	-	N ₂	RT	✓	-

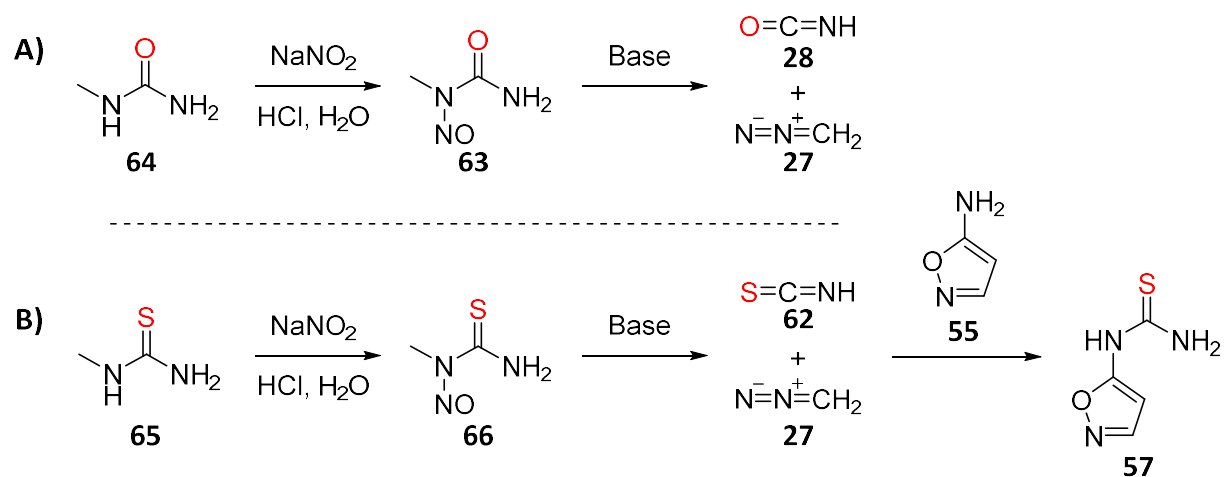
Die Anwesenheit von P_4S_{10} auf der präbiotischen Erde scheint plausibel. P_4S_{10} entsteht bei Umsetzung von Fe^{II}P mit Fe^{II}S₂ oder mit Fe^{II}S und S₈, bei Temperaturen zwischen 700 °C und 900 °C.^[225,226] Aus P₄ und S₈ bildet sich P_4S_{10} sogar bereits bei 300 °C.^[227] Blitzeinschläge, Meteoriteneinschläge sowie Vulkanausbrüche ermöglichten solche Temperaturen vermutlich häufig auf der frühen Erde.^[37] Auch Eisen-Phosphide und -Sulfide scheinen präbiotisch plausibel:

So sind Phosphid-haltige Mineralien wie Schreibersit und Barringerit nach wie vor häufige Bestandteile von Meteoriten.^[228] $\text{Fe}^{\text{II}}\text{S}$ und $\text{Fe}^{\text{II}}\text{S}_2$ waren ebenso wie elementarer Schwefel abundant.^[36,229] Auch die Verfügbarkeit von P^0 im Hadaikum scheint plausibel.^[23,230]

Eine weitere Optimierung der Reaktion sollte möglich sein: So könnte die in Reaktion 14 eingesetzte Menge an P_4S_{10} optimiert werden. Zudem sollte das Lösungsmittel vor der Reaktion gründlich entgast werden. Zusätzlich könnten P^{III} -Schwefel-Spezies in Bezug auf deren Wirksamkeit zum O-S-Austausch untersucht werden. Deren größere Oxophilie gegenüber der eingesetzten P^{V} -Spezies P_4S_{10} , könnte eine weitere Steigerung der Ausbeute ermöglichen. Zusammengenommen konnte gezeigt werden, dass die Bildung von *N*-(Isoxazol-5-yl)thioharnstoff (**57**) mittels P_4S_{10} möglich ist. Im nächsten Schritt sollte die präbiotische Plausibilität der Reaktion gestärkt werden.

4.2.2.2.1.3 Umsetzungen mit *N*¹-Methyl-*N*¹-(nitroso)thioharnstoff (**66**)

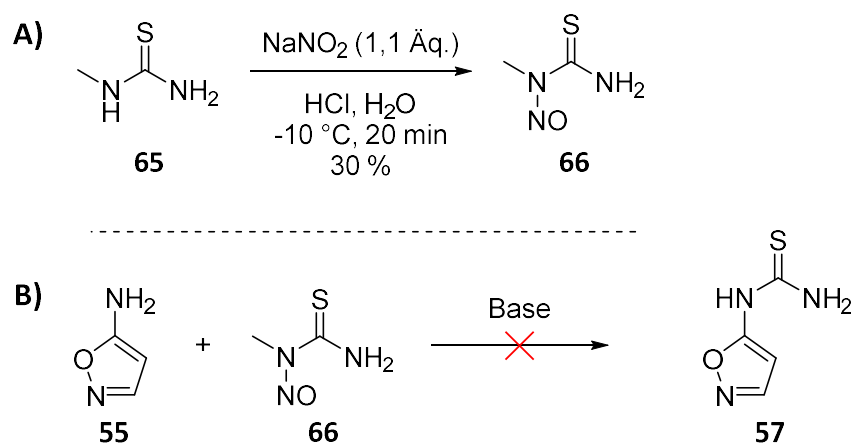
Isothiocyansäure (62) könnte die Synthese von N-(Isoxazol-5-yl)thioharnstoff (57) ermöglichen. In diesem Abschnitt sollte untersucht werden, ob Isothiocyansäure (**62**) die Thiocarbamoylierung von 5-Aminoisoxazol (**55**) ermöglicht. SCHNEIDER *et al.* verwendeten *N*¹-Methyl-*N*¹-nitrosoharnstoff (**63**) als stabilen Speicher der Isocyansäure (**28**) und des Diazomethans (**27**),^[231] siehe Schema 37 A). Unter basischen Bedingungen lassen sich beide Reagenzien, **28** und **27**, gezielt freisetzen. Die Speicher-Verbindung *N*¹-Methyl-*N*¹-nitrosoharnstoff (**63**) kann aus *N*-Methylharnstoff (**64**) und NaNO_2 erzeugt werden. Zur Anwendung für die Thiocarbamoylierung von 5-Aminoisoxazol (**55**), sollte Isothiocyansäure (**62**) aus den entsprechenden Schwefel-Derivaten generiert werden, siehe Schema 37 B). Anschließend sollte Isothiocyansäure (**62**) zur *in situ*-Thiocarbamoylierung von 5-Aminoisoxazol (**55**) eingesetzt werden.



Schema 37 A) Bildung von Isocyan säure (**28**) und Diazomethan (**27**) aus *N*-Methylharnstoff (**64**) via *N*¹-Methyl-*N*¹-nitrosoharnstoff (**63**) nach SCHNEIDER *et al.*^[231] **B)** Synthesestrategie zur Bildung von Isothiocyan säure (**62**) und Diazomethan (**27**) aus *N*-(Methyl)thioharnstoff (**65**) via *N*¹-Methyl-*N*¹-(nitroso)thioharnstoff (**66**). Isothiocyan säure (**62**) sollte anschließend *in situ* zur Thiocarbamylierung von 5-Aminoisoxazol (**55**) eingesetzt werden. **rot:** Unterschiede zwischen A) und B).

Die Anwesenheit der Edukte auf der präbiotischen Erde scheint plausibel. Bezüglich der präbiotischen Plausibilität der Edukte kann angenommen werden, dass *N*-(Methyl)thioharnstoff (**65**) analog zum Harnstoff-Derivat *N*-Methylharnstoff (**64**)^[77], aus Methylamin und Isothiocyanaten in präbiotischen Modellreaktionen gebildet werden kann. NaNO_2 konnte auf der präbiotischen Erde aus NO und NO_2 gebildet werden.^[232] Beide Gase entstanden bei Blitzentladungen durch Reaktion von N_2 mit O_2 , auch wenn O_2 nur in geringem Anteil in der Atmosphäre vorlag.^[232]

Keine der untersuchten Reaktionen führte zur Bildung des Zielprodukts *N*-(Isoxazol-5-yl)thioharnstoff (**57**). Zunächst wurde die Bildung von *N*¹-Methyl-*N*¹-(nitroso)-thioharnstoff (**66**) unter präbiotischen Bedingungen etabliert. Dies gelang durch Umsetzung von *N*-(Methyl)thioharnstoff (**65**) mit NaNO_2 bei $-10\text{ }^\circ\text{C}$ in einer Ausbeute von 30 %, siehe Schema 38 A). Im nächsten Schritt wurde die Umsetzung von 5-Aminoisoxazol (**55**) mit *N*¹-Methyl-*N*¹-(nitroso)thioharnstoff (**66**) unter einer Vielzahl an Bedingungen untersucht, siehe Schema 38 B) und Tab. 5. Unter keiner dieser Reaktionsbedingungen konnte das Zielprodukt *N*-(Isoxazol-5-yl)thioharnstoff (**57**) mittels HR-LC-ESI-MS detektiert werden.

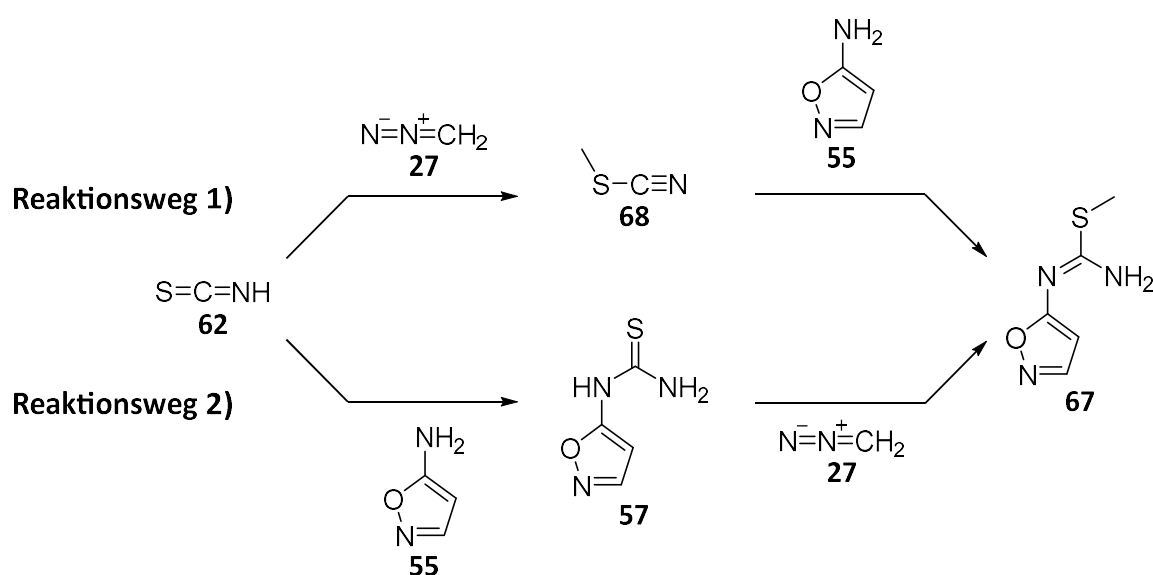


Schema 38 A) Synthese von *N*-Methyl-*N*-Nitroso-Thioharnstoff (**66**) aus *N*-(Methyl)thioharnstoff (**65**) und NaNO_2 in einer präbiotischen Modellreaktion **B)** Die Umsetzung von 5-Aminoisoxazol (**55**) mit *N*-Methyl-*N*-(nitroso)thioharnstoff (**66**) führte unter keiner der untersuchten Bedingungen (siehe Tab. 5) zur Bildung von *N*-(Isoxazol-5-yl)thioharnstoff (**57**).

Tab. 5 Zusammenfassung der untersuchten Reaktionsbedingungen zur Thiocarbamoylierung von *N*-(Isoxazol-5-yl)harnstoff (**61**) gemäß Schema 38 B). Zielprodukt: *N*-(Isoxazol-5-yl)thioharnstoff (**57**). Die Reaktionskontrollen erfolgten mittels HR-LC-ESI-MS.

Reaktion	Zielprodukt 57 detektiert?	<i>N</i> ¹ -Methyl- <i>N</i> ¹ -(nitroso)-thioharnstoff (66) / Äq.	Base	<i>T</i> / °C
19	-	10	NaHCO_3 (10 Äq.)	55
20	-	10	Borax (1 Äq.)	55
21	-	10	$(\text{NH}_4)_2\text{S}$ (3 Äq.)	55
22	-	10	Na_2HPO_4 (10 Äq.)	55
23	-	10	CaCO_3 (5 Äq.)	55
24	-	25	NaHCO_3 (10 Äq.)	8
25	-	25	Borax (1 Äq.)	8
26	-	25	$(\text{NH}_4)_2\text{S}$ (3 Äq.)	8
27	-	25	Na_2HPO_4 (10 Äq.)	8
28	-	25	NaHCO_3 (10 Äq.)	7→70
29	-	25	Borax (1 Äq.)	7→70
30	-	25	$(\text{NH}_4)_2\text{S}$ (3 Äq.)	7→70
31	-	25	Na_2HPO_4 (10 Äq.)	7→70
32	-	25	CaCO_3 (5 Äq.)	7→70

Die untersuchten Reaktionen führten hingegen robust zur Bildung von *S*-Methyl-*N*-(isoxazol-5-yl)thioharnstoff (**67**). HR-LC-ESI-MS zeigte die Bildung des *S*-Methyl-Derivats **67** des ursprünglichen Zielprodukts **57** in allen Reaktionen mit Reaktionstemperaturen über 55 °C. Die Quantifizierung der Ausbeuten sollte im Anschluss an die Synthese der Referenzverbindung erfolgen, vgl. Abschnitt 4.2.2.3.1. In Schema 39 werden zwei mögliche Reaktionswege zur Bildung von *S*-Methyl-*N*-(isoxazol-5-yl)thioharnstoff **67** vorgeschlagen. Aufgrund der vergleichsweise geringen Nucleophilie des exozyklischenamins von 5-Aminoisoxazol (**55**) kann angenommen werden, dass es sich bei der Addition von 5-Aminoisoxazol (**55**) um den jeweils geschwindigkeitsbestimmenden Schritt handelt. So kann postuliert werden, dass Reaktionsweg 1) via Methylthiocyanat (**68**) bevorzugt ablaufen sollte.

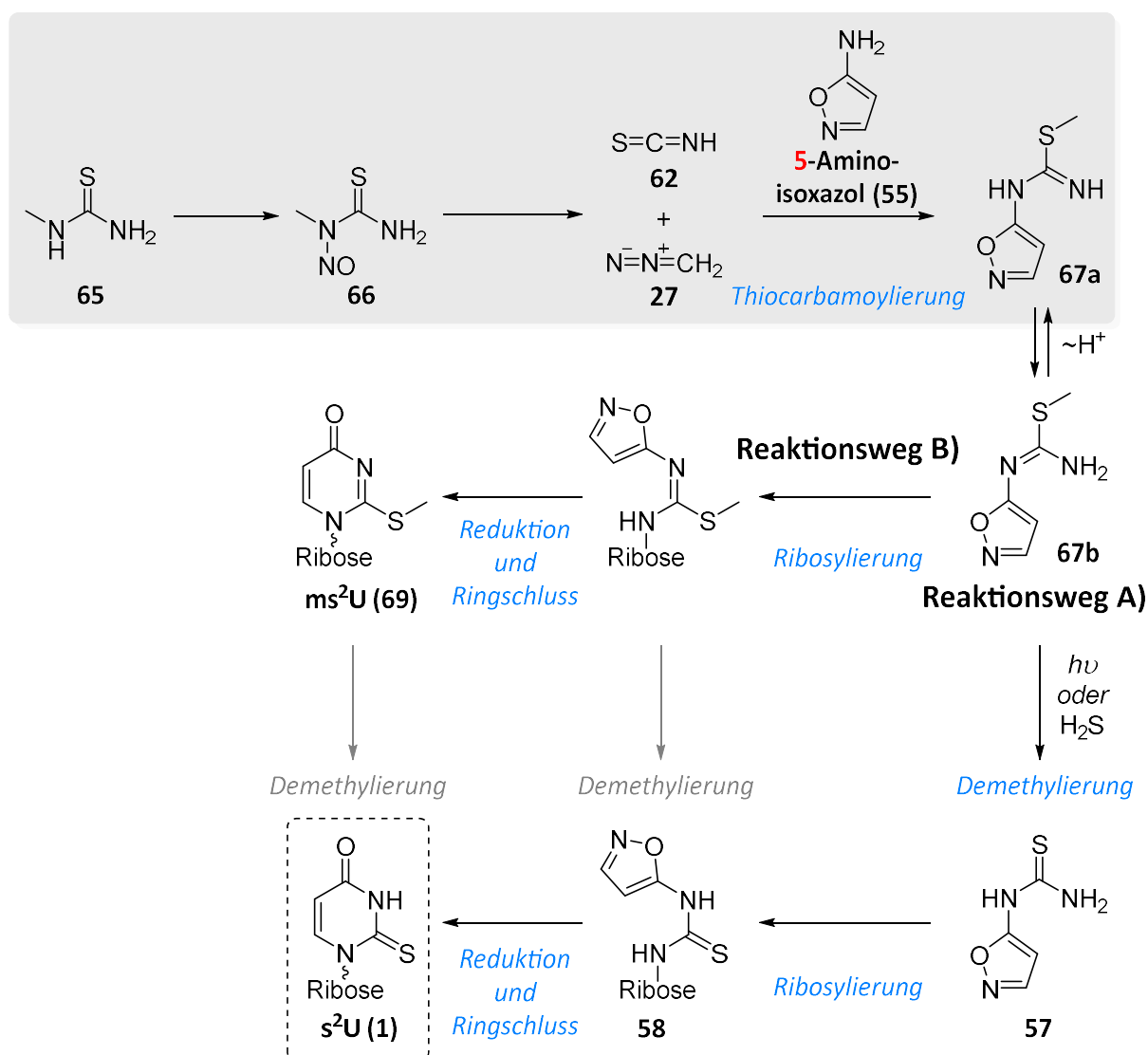


Schema 39 Vorschlag zweier Reaktionswege zur Bildung von *S*-Methyl-*N*-(isoxazol-5-yl)thioharnstoff (**67**) aus 5-Aminoisoxazol (**55**), Isothiocyanäure (**62**) und Diazomethan (**27**).

Ausblick zur Thiocarbamylierung mit *N*¹-Methyl-*N*¹-(nitroso)thioharnstoff (**66**)

Die robuste Bildung von *S*-Methyl-*N*-(isoxazol-5-yl)thioharnstoff (**67**) unter präbiotisch plausiblen Bedingungen eröffnet Möglichkeiten für neue Reaktionswege, siehe Schema 40: Reaktionsweg A) entspricht der bisherigen Strategie zur Synthese von *s*²U (**1**), aufgenommen wurde lediglich die Demethylierung von *S*-Methyl-*N*-(isoxazol-5-yl)thioharnstoff (**67**) mittels Photolyse^[233,234] oder H_2S ^[235,236]. Wie bisher vorgesehen, würde anschließend die Ribosylierung zu **58** erfolgen, bevor Reduktion und Ringschluss *s*²U (**1**) bilden. Darüber hinaus würde eine spätere Demethylierung, im Anschluss an die Ribosylierung oder an Reduktion und Ringschluss (grau dargestellt), weitere Reaktionswege eröffnen. So könnte Reaktionsweg B) einen Zugang zu *ms*²U (**69**) ermöglichen,

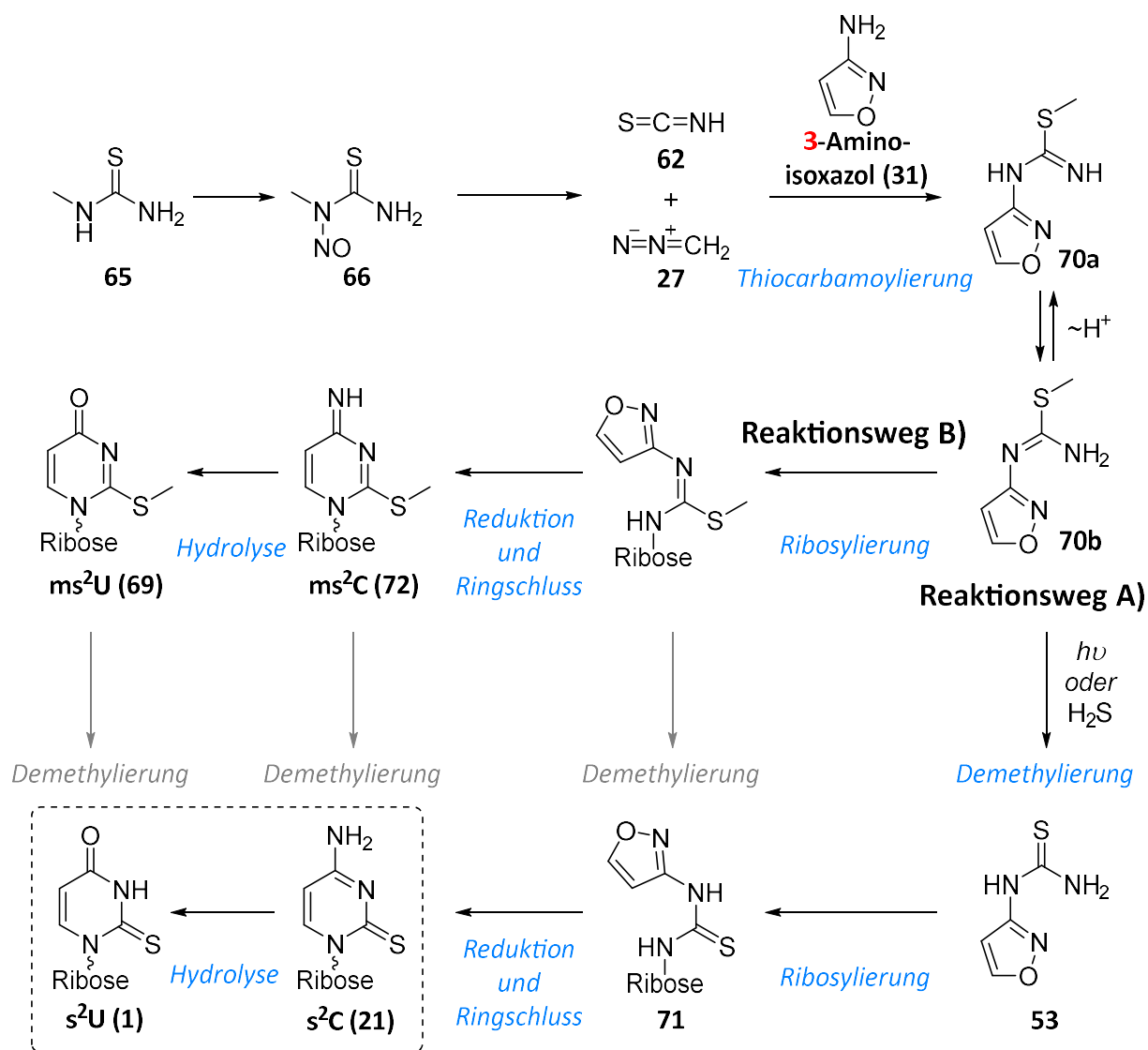
eine potenziell rezente RNA-Modifikation.^[237] Diese diversen Optionen eröffnen die Möglichkeit zur robusten Bildung von s^2U (**1**) über verschiedene Reaktionswege.



Schema 40 Plan zur Synthese von s^2U (**1**) via 5-Aminoisoxazol (**55**), auf Basis der in diesem Abschnitt etablierten Bildung von *S*-Methyl-*N*-(isoxazol-5-yl)thioharnstoff (**67**) (grau hinterlegt). Die Demethylierung könnte mittels Photolyse^[233,234] oder H_2S ^[235,236] erfolgen. Ribosylierung sowie Reduktion und Ringschluss würden analog zum bisherigen Synthesepfad verlaufen, vgl. Schema 33 in Abschnitt 4.2.2. Reaktionsweg B) könnte zusätzlich die Synthese von ms^2U (**69**) ermöglichen.

Der Zugang zu *S*-Methyl-*N*-(isoxazol-5-yl)thioharnstoff (**67**) sollte auch den Einsatz des reaktiveren 3-Aminoisoxazols (**31**) ermöglichen. 3-Aminoisoxazol (**31**) zeigte sich in nukleophilen Additions-Reaktionen deutlich reaktiver als dessen Isomer 5-Aminoisoxazol (**55**). Demnach könnte der Einsatz von 3-Aminoisoxazol (**31**) eine deutliche Steigerung der Ausbeute

ermöglichen; im Vergleich zu der bisher geringen Ausbeute von *S*-Methyl-*N*-(isoxazol-5-yl)thioharnstoff (**67**). Die Methylgruppe des *S*-Methyl-*N*-(isoxazol-3-yl)thioharnstoffs (**70**) sollte dabei als Schutzgruppe des Schwefels fungieren und so die BOULTON-KATRITZKY-Umlagerung unterbinden. Schema 41 zeigt einen Plan zur Synthese von *s*²U (**1**) und *s*²C (**21**) via 3-Aminoisoxazol (**31**). Reduktion des ribosylierten 3-Aminoisoxazol-Derivats **71**, sollte nach Ringschluss unmittelbar zu *s*²C (**21**) führen. Dies kann durch Hydrolyse in *s*²U (**1**) überführt werden.^[57] Darüber hinaus würde diese Strategie via Reaktionsweg B) die Möglichkeit zur Bildung von *ms*²C (**72**) und *ms*²U (**69**) eröffnen. Auch *ms*²C (**72**) stellt eine rezente RNA-Modifikation dar.^[238]



Schema 41 Plan zur Synthese von *s*²U (**1**) und *s*²C (**21**) via 3-Aminoisoxazol (**31**), auf Basis der in diesem Abschnitt etablierten Bildung von *S*-Methyl-*N*-(isoxazol-5-yl)thioharnstoff (**67**), vgl. grau hinterlegte Reaktionen in Schema 40. Die Demethylierung könnte mittels Photolyse^[233,234] oder H₂S^[235,236] erfolgen. Ribosylierung sowie Reduktion und Ringschluss würden analog zum bisherigen Synthesepan verlaufen, vgl. Schema 33 in Abschnitt 4.2.2. Reaktionsweg B) könnte zusätzlich die Synthese von *ms*²C (**72**) und *ms*²U (**69**) ermöglichen.

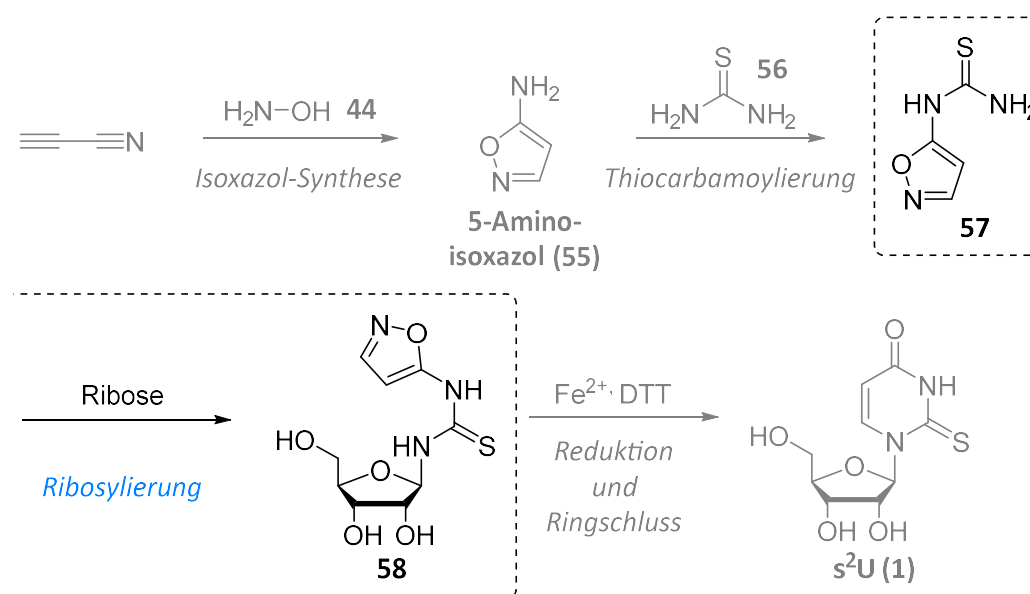
*Im nächsten Schritt ist die Herstellung der S-Methyl-N-(isoxazolyl)thioharnstoff-Derivate **67** und **70** mittels herkömmlicher Synthese nötig.* Als Referenzverbindungen können diese zum einen zur Validierung der gebildeten S-Methyl-N-isoxazolyl-Derivate **67** und **70** eingesetzt werden. Darüber hinaus würden diese Verbindungen die Etablierung der sich anschließenden Syntheseschritte ermöglichen, vgl. Schema 40 und Schema 41. So könnte die Umsetzung von **67** und **70** mit H₂S oder Photolyse untersucht werden. Darüber hinaus sollte eine Optimierung der Bildung von **67** und **70** in präbiotischen Modellreaktionen erfolgen.

4.2.2.2.1.4 Zusammenfassung der Thiocarbamoylierung

Zusammengenommen wurden durch Untersuchung einer Vielzahl an Reaktionsbedingungen zwei unabhängige Zugänge zur Installation der Thioharnstoff-Einheit gefunden. Wir konnten zeigen, dass die direkte Thiocarbamoylierung von 5-Amino- oder 3-Aminoisoxazol mit Thioharnstoff (**56**), Thiocyanaten oder CS₂ unter einer Vielzahl an Bedingungen nicht zum Erfolg führte. Stattdessen gelang der O-S-Austausch an N-(Isoxazol-5-yl)harnstoff (**61**) mit P₄S₁₀. Darüber hinaus wurde ein Zugang zu S-Methyl-N-(isoxazol-5-yl)thioharnstoff (**67**) etabliert. Dieser könnte den Schlüsselschritt einer neuartigen Synthese von s²U (**1**) darstellen.

4.2.2.2 Ribosylierung zu *N*-Ribosyl-*N*-(isoxazol-5-yl)thioharnstoff (**58**)

Das konzeptionelle Herzstück der präbiotischen Synthese von *s*²U (**1**) bildet die regioselektive Ribosylierung, siehe Schema 42. Ziel dieses Abschnittes war die Ribosylierung von *N*-(Isoxazol-5-yl)thioharnstoff (**57**) unter präbiotisch plausiblen Bedingungen. Für die Darstellung der Referenzverbindung des Zielprodukts **58** mittels herkömmlicher Synthese, siehe Abschnitt 4.2.2.3.2.



Schema 42 Ziel dieses Abschnitts war die Ribosylierung von *N*-(Isoxazol-5-yl)thioharnstoff (**57**) zu *N*-Ribosyl-*N*-(isoxazol-5-yl)thioharnstoff (**58**).

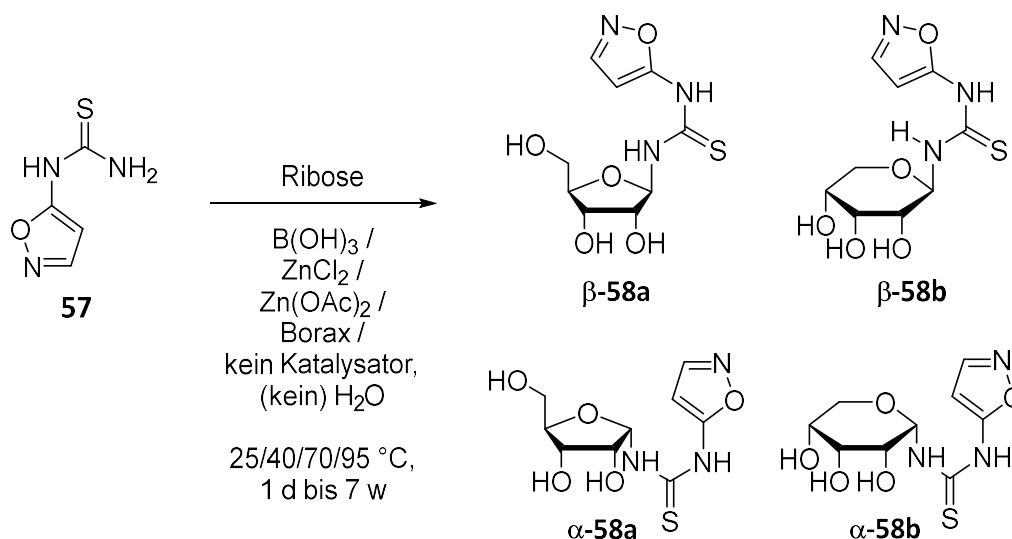
Ribosylierung von N-(Isoxazol-5-yl)thioharnstoff (**57**) lieferte vier Signale, entsprechend den vier Ribosid-Isomeren des Zielprodukts **58**. In Anlehnung an die in Abschnitt 4.1 etablierte Ribosylierung, wurden eine Vielzahl an Reaktionsbedingungen zur Ribosylierung von *N*-(Isoxazol-5-yl)thioharnstoff (**57**) untersucht, siehe Schema 43. Dabei wurden die folgenden drei Parameter variiert: Reaktionstemperatur (25 °C, 40 °C, 70 °C), Katalysator ($\text{B}(\text{OH})_3$, ZnCl_2 , $\text{Zn}(\text{OAc})_2$, Borax, kein Katalysator) und Reaktionsdauer (ein Tag bis 7 Wochen). Zusätzlich wurden Reaktionen bei 95 °C ohne Lösungsmittel durchgeführt. Bereits nach einem Tag bei 70 °C konnten vier Signale detektiert werden, die in ihrer Masse den vier Zielprodukt-Isomeren α - und β -**58a** und -**58b** entsprachen. Dabei konnten folgende Trends beobachtet werden:

Reaktionstemperatur

- das Verhältnis der vier Isomere ist abhängig von Reaktionstemperatur und Katalysator
- niedrige Temperaturen und ZnCl_2 steigern die Bildung von Isomer II und III deutlich
- Umsetzungen unterhalb von 70 °C verliefen sehr langsam
- Umsetzungen bei 95 °C ohne H_2O lieferten kein oder wenig Zielprodukt **58**

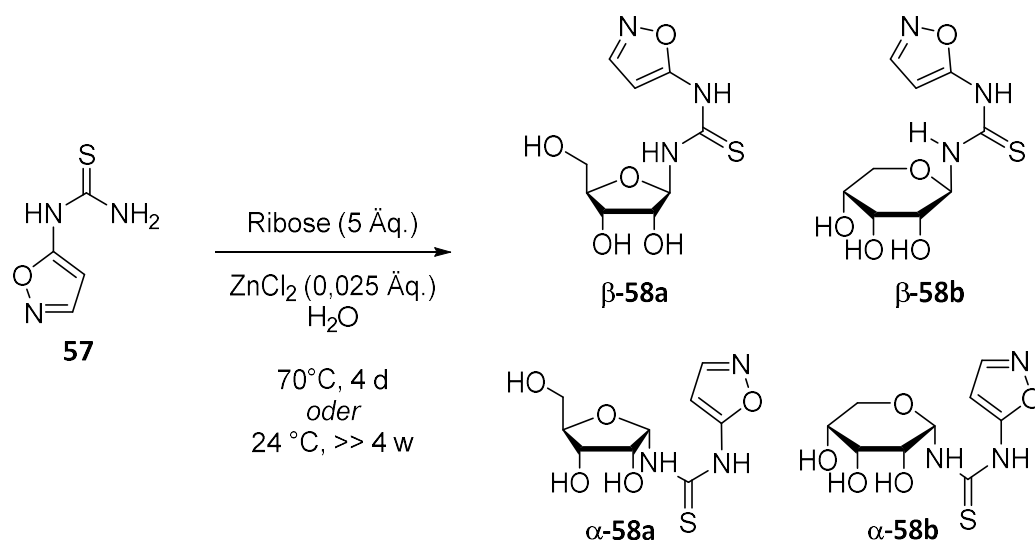
Katalysator

- ZnCl_2 katalysierte die Bildung des Zielprodukts **58** am effizientesten
- $\text{B}(\text{OH})_3$ und kein Katalysator lieferten vergleichbare Ergebnisse
- $\text{Zn}(\text{OAc})_2$ beeinflusste die Bildung des Zielprodukts **58** negativ
- Borax führte vermutlich zur Bildung von Ribose-Borat-Addukten



Schema 43 Zusammenfassung der untersuchten Bedingungen zur Ribosylierung von *N*-(Isoxazol-5-yl)thioharnstoff (**57**). Ein Großteil dieser Bedingungen führte zur Detektion von Signalen mittels HR-LC-ESI-MS, die in ihrer Masse den vier Zielprodukt-Isomeren α - und β -**58a** sowie -**58b** entsprachen. Die Quantifikation der Ausbeuten ist im Anschluss an die Synthese der Referenzverbindungen möglich, vgl. Abschnitt 4.2.2.3.2.

Die Untersuchungen zeigten, dass die Ribosylierung von *N*-(Isoxazol-5-yl)thioharnstoff (**57**) idealerweise in H_2O , in Anwesenheit von ZnCl_2 erfolgt, siehe Schema 44. Eine vollständige Umsetzung kann bei 70 °C in vier Tagen erfolgen. Um eine Steigerung der Isomere II und III zu erzielen, sollte die Ribosylierung hingegen bei verringerter Temperatur über mehrere Wochen bis Monate erfolgen: Abb. 17 zeigt den Einfluss der Reaktionstemperatur auf die Bildung der Isomere II und III. Die Umsetzung bei 70 °C erfolgte für einen Tag, die Umsetzung bei 25 °C für vier Wochen. Bei 25 °C wurde auch nach 7 Wochen nur ein geringer Umsatz des Edukts **55** beobachtet. Diese niedrige Reaktionsgeschwindigkeit erscheint zunächst unpraktikabel. Zur Einordnung dieser Ergebnisse ist es jedoch notwendig die möglichen Zeiträume präbiotischer Reaktionen auf der frühen Erde zu betrachten: So kann angenommen werden, dass diese Zeiträume deutlich größer gewesen sein konnten, als es für diese Arbeit praktikabel gewesen wäre.



Schema 44 Zusammenfassung idealer Reaktionsbedingungen zur Ribosylierung von *N*-(Isoxazol-5-yl)thioharnstoff (**57**) auf Basis der durchgeführten Untersuchungen, vgl. Schema 43. Das Verhältnis der gebildeten Ribosid-Isomere lässt sich durch Wahl der Reaktionstemperatur deutlich beeinflussen, vgl. Abb. 17. Die Quantifikation der Ausbeuten wird durch Synthese der Referenzverbindungen möglich, vgl. Abschnitt 4.2.2.3.2.

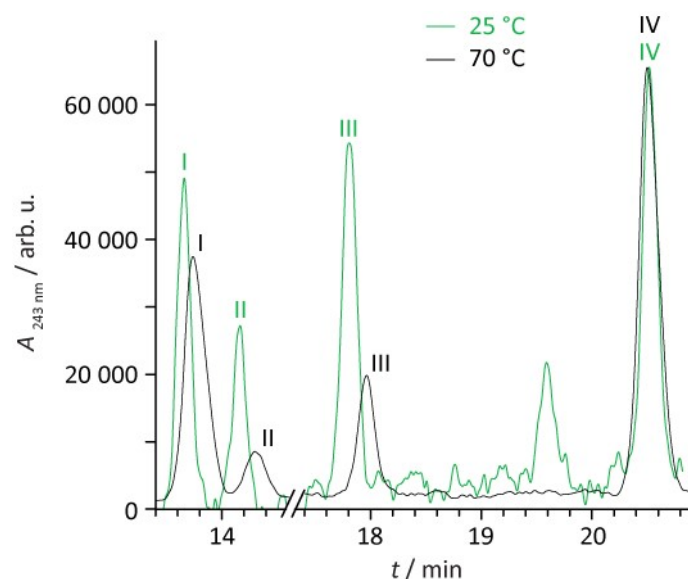


Abb. 17 Dargestellt ist der Temperatureinfluss auf die Bildung der Isomere des Zielprodukts *N*-Ribosyl-*N*-(Isoxazol-5-yl)thioharnstoff (**58**), anhand der HR-LC-ESI-MS-Chromatogramme: Umsetzungen von *N*-(Isoxazol-5-yl)thioharnstoff (**57**) mit Ribose und ZnCl_2 bei 70 °C für einen Tag (grün) sowie bei 25 °C für 4 Wochen (schwarz). Signale, die der Produktmasse (m/z 276,0649) zugeordnet wurden, sind nummeriert und entsprechen mit hoher Wahrscheinlichkeit den Ribosid-Isomeren α -58a und β -58a, sowie α -58b und β -58b, vgl. Schema 44. Zur Steigerung der Übersichtlichkeit ist das Signal des Edukts *N*-(Isoxazol-5-yl)thioharnstoff (**57**) zwischen 15 und 17 min nicht dargestellt. Das Chromatogramm der Umsetzung bei 25 °C ist auf Signal IV der Umsetzung bei 70 °C normiert.

Ein Vergleich zwischen herkömmlicher Synthese und präbiotischer Modellreaktion legt die Bildung der Zielprodukte α - und β -58a sowie -58b nahe. Abb. 18 vergleicht Chromatogramme der präbiotischen Modellreaktion und der herkömmlichen Synthese von *N*-Ribosyl-*N*-(isoxazol-5-yl)thioharnstoff (α - und β -58a sowie -58b), vgl. Abschnitt 4.2.2.3.2. Die große Übereinstimmung der Signal-Verhältnisse liefert einen weiteren Hinweis darauf, dass in der präbiotischen Modellreaktion tatsächlich die vier Zielprodukte α - und β -58a und -58b gebildet wurden. Zur abschließenden Validierung verbleibt damit lediglich die Zuordnung der Signale zu den vier Ribosid-Isomeren.

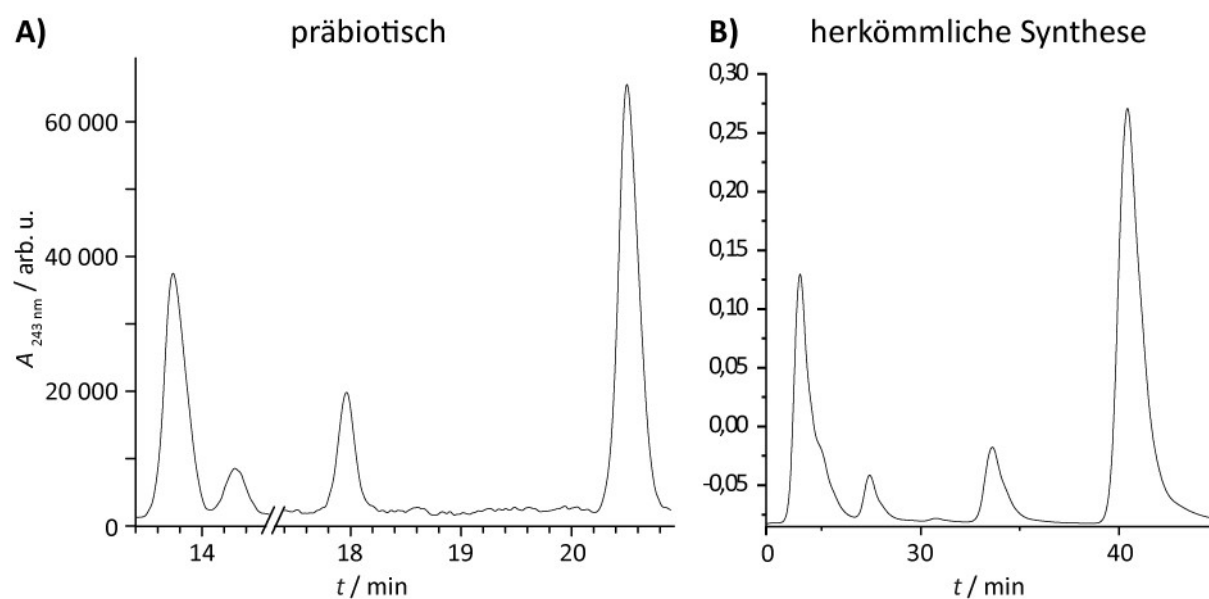
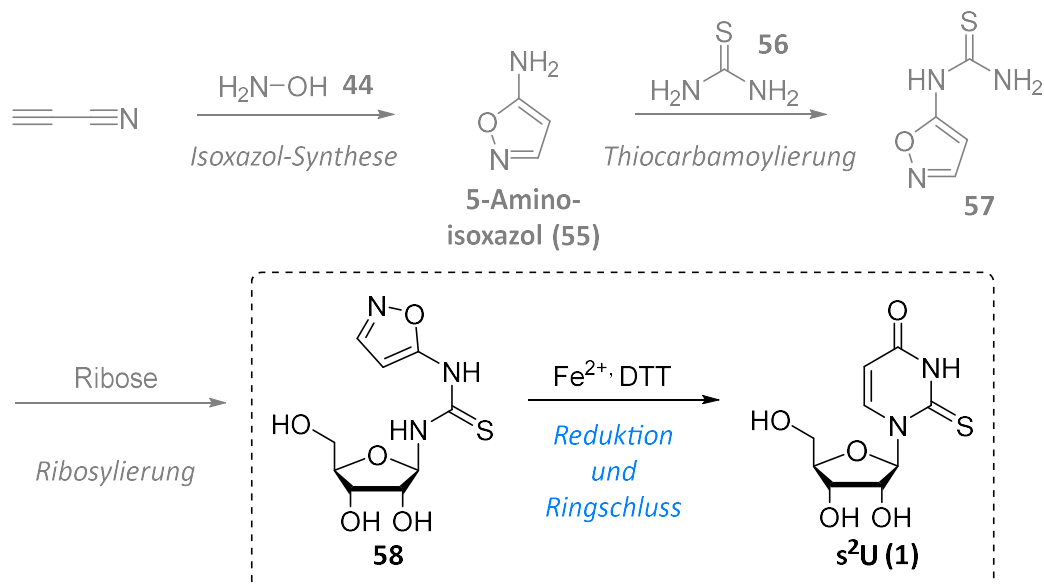


Abb. 18 A) HR-LC-ESI-MS-Chromatogramm der präbiotischen Modellreaktion zur Bildung von *N*-Ribosyl-*N*-(isoxazol-5-yl)thioharnstoff (α - und β -58a sowie -58b) bei 70 °C, vgl. Schema 44 und Abb. 17. **B)** RP-HPLC-Chromatogramm der Darstellung von *N*-Ribosyl-*N*-(isoxazol-5-yl)thioharnstoff (α - und β -58a sowie -58b) mittels herkömmlicher Synthese, vgl. Abb. 22 B) in Abschnitt 4.2.2.3.2. Der Vergleich zwischen A) und B) legt nahe, dass in der präbiotischen Modellreaktion die vier Zielprodukte α - und β -58a und -58b gebildet wurden.

Zusammengenommen wurden Zugänge zu *N*-Ribosyl-*N*-(isoxazol-5-yl)thioharnstoff α - und β -58a sowie -58b mittels präbiotischer Modellreaktionen etabliert. Darüber hinaus wurden Bedingungen gefunden, welche die relative Ausbeute der unterrepräsentierten Isomere II und III deutlich steigern.

4.2.2.2.3 Reduktion und Ringschluss

Im letzten Schritt der Synthese von s^2U (**1**), sollte via Reduktion der Isoxazol-*N*-*O*-Bindung der Ringschluss zum s^2U (**1**) ermöglicht werden, siehe Schema 45. Da s^2U (**1**) kommerziell erhältlich ist, war die Etablierung einer herkömmlichen Synthese nicht notwendig.



Schema 45 Ziel dieses Abschnitts war es Reduktion und Ringschluss zu s^2U (**1**) in einer präbiotischen Modellreaktion zu etablieren.

Zunächst sollten Reduktion und Ringschluss an einem geeigneten Modellsystem etabliert werden. Um die Komplexität der Reaktionskontrolle möglichst gering zu halten, wurde als Modellsystem Bz-geschütztes N^1 - β -Ribofuranosyl- N^2 -(isoxazol-5-yl)thioharnstoff **β -76a** gewählt, siehe Abb. 19. Die Benzoyl-Gruppen sollten die Furanosid-Pyranosid-Isomerisierung inhibieren und so eine simple chromatographische Reaktionskontrolle ermöglichen. Die unterschiedliche chromatographische Retention der vier Zielprodukt-Isomere α - und β -**58a** sowie **58b** hätte die Produkt-Identifikation anderenfalls deutlich erschwert; insbesondere durch die verminderte Intensität von vier statt einer Zielprodukt-Spezies. Zudem kann angenommen werden, dass der Einfluss der Bz-Gruppen auf Reduktion und Ringschluss vernachlässigbar ist. Damit stellte Bz-geschütztes N^1 - β -Ribofuranosyl- N^2 -(isoxazol-5-yl)thioharnstoff **β -76a** ein ideales Modellsystem dar.

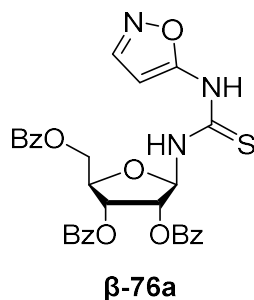
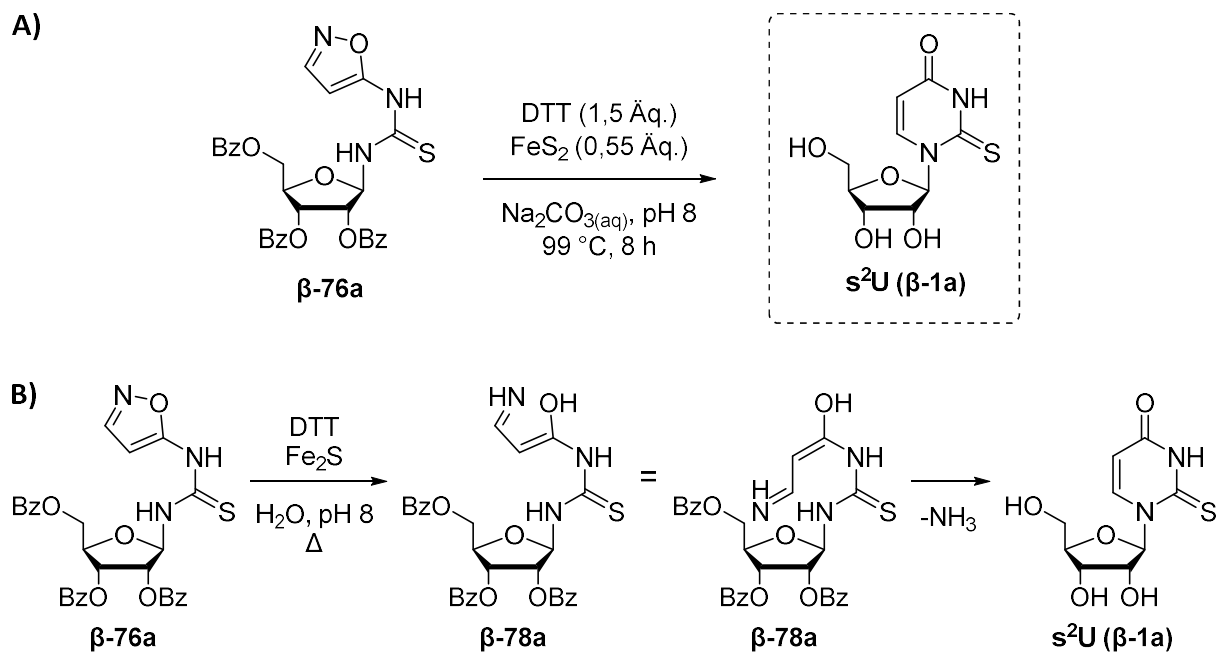


Abb. 19 Bz-geschütztes N^1 -β-Ribofuranosyl- N^2 -(isoxazol-5-yl)thioharnstoff **β-76a** stellte ein ideales Modellsystem zur Etablierung von Reduktion und Ringschluss dar.

*Reduktion von Bz-geschütztem N^1 -β-Ribofuranosyl- N^2 -(isoxazol-5-yl)thioharnstoff **β-76a** ermöglichte die Bildung von s^2U (**1**) in einer präbiotischen Modellreaktion.* Die Umsetzung von **β-76a** erfolgte in Anlehnung an die in Abschnitt 4.1 etablierten Reaktionsbedingungen, siehe Schema 46 A). Interessanterweise erlaubten Reaktionskontrollen via 1H -NMR-Spektroskopie, Dünnschicht-Chromatographie sowie LR-(LC-)ESI-MS keine Detektion des Zielprodukts **1**. Die Detektion des Zielprodukts s^2U (**1**) gelang schließlich mittels HR-LC-ESI-MS. Anschließende HR-LC-ESI-MS-Koinjektion der Referenzverbindung **β-1a** bestätigte die Bildung von s^2U (**β-1a**), siehe Abb. 20. Interessanterweise konnte Bz-geschütztes s^2U nicht detektiert werden. Da stattdessen das Bz-entschützte Produkt s^2U (**β-1a**) detektiert wurde liegt es nahe, dass bei der Umsetzung neben Reduktion und Ringschluss auch die Entschützung der Bz-Gruppen erfolgte. Vermutlich durch Freisetzung von NH_3 bei der Ringschluss-Kondensation zu s^2U (**1**), siehe Schema 46 B).



Schema 46 A) Umsetzung des Bz-geschütztem N^1 - β -Ribofuranosyl- N^2 -(isoxazol-5-yl)thioharnstoffs **β -76a** in einer präbiotischen Modellreaktion ermöglichte die Bildung des Zielproduktes dieses Projekts, s^2U (**β -1a**). **B)** Reaktions-Vorschlag zu Reduktion und Ringschluss von N^1 - β -Ribofuranosyl- N^2 -(isoxazol-5-yl)thioharnstoff (**β -76a**) mittels FeS_2 und DTT. Die Ringschluss-Kondensation von **β -78a** im zweiten Schritt setzt NH_3 frei, welches vermutlich die Bz-Entschützung bewirkte.

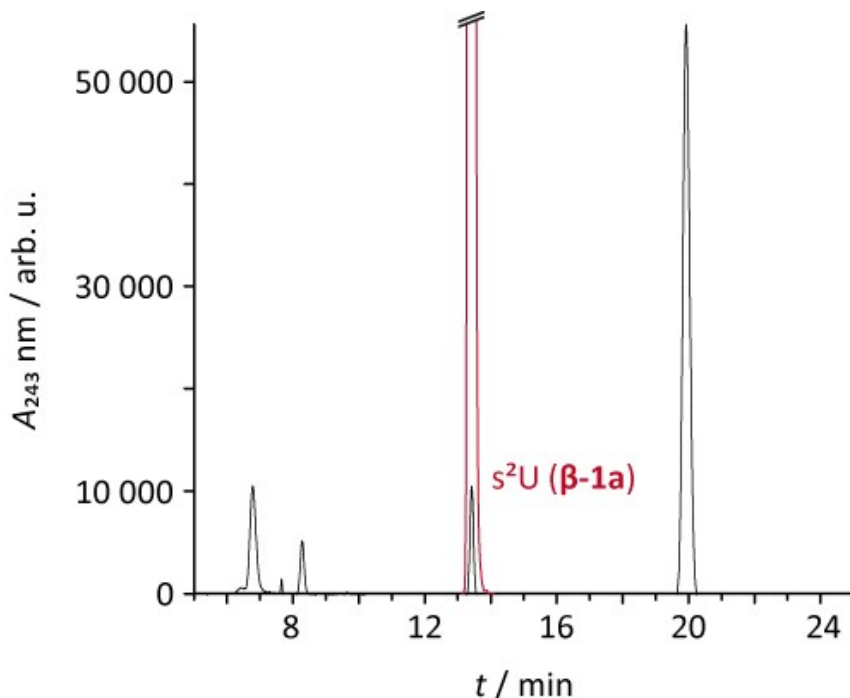
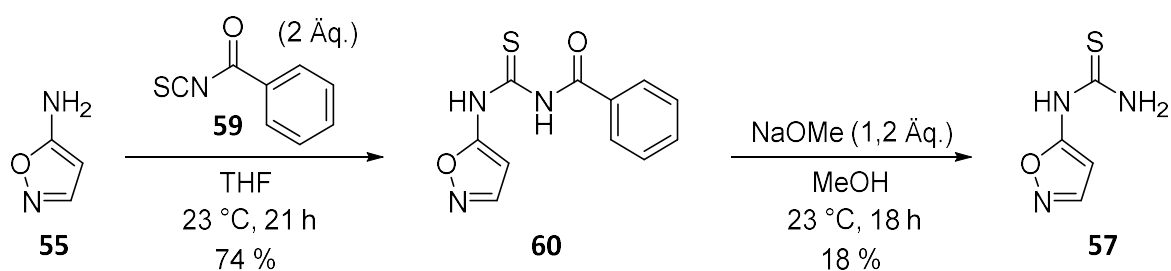


Abb. 20 HR-LC-ESI-MS-Koinjektion bestätigte die Bildung von s^2U (**β -1a**) in der präbiotischen Modellreaktion, vgl. Schema 46 A). Dargestellt ist eine Überlagerung zweier Spektren: schwarz: Reaktionsgemisch der Reduktion von Bz-geschütztem N^1 -Ribosyl- N^2 -(isoxazol-5-yl)thioharnstoff **β -76a** gemäß Schema 46. rot: s^2U -Signal der Koinjektion von s^2U (**β -1a**) und dem Reaktionsgemisch.

4.2.2.3 Synthese der Referenzverbindungen

4.2.2.3.1 *N*-(Isoxazol-5-yl)thioharnstoff (57)

Zur Herstellung von *N*-(Isoxazol-5-yl)thioharnstoff (57) wurde eine zweistufige Synthese entwickelt. Zunächst wurde 5-Aminoisoxazol (55)^[223] mit Benzoylisothiocyanat (59) in THF umgesetzt. Nach 21 h bei 23 °C konnte *N*¹-Benzoyl-*N*³-(isoxazol-5-yl)thioharnstoff (60) in einer Ausbeute von 74 % erhalten werden. Eine weitere Steigerung der Ausbeute könnte bei Bedarf durch wiederholte Filtrationen der Mutterlauge erreicht werden. Im zweiten Schritt erfolgte die Entschützung mit NaOMe in MeOH. Diese Reaktion lieferte nach 18 h bei 23 °C *N*-(Isoxazol-5-yl)thioharnstoff (57) in einer Ausbeute von 18 %. Studien zur Optimierung der Reaktionsdauer zeigten, dass eine verkürzte Reaktionsdauer zu einer Steigerung der Ausbeute führen könnte. Weitere Optimierungen der Reaktionsbedingungen könnten in Anlehnung an ähnliche Umsetzungen erfolgen.^[239–244] Dabei sollte beachtet werden, dass bei Temperaturen über 70 °C eine schnelle Degradation des Edukts 5-Aminoisoxazol (55) beobachtet wurde.^[223]

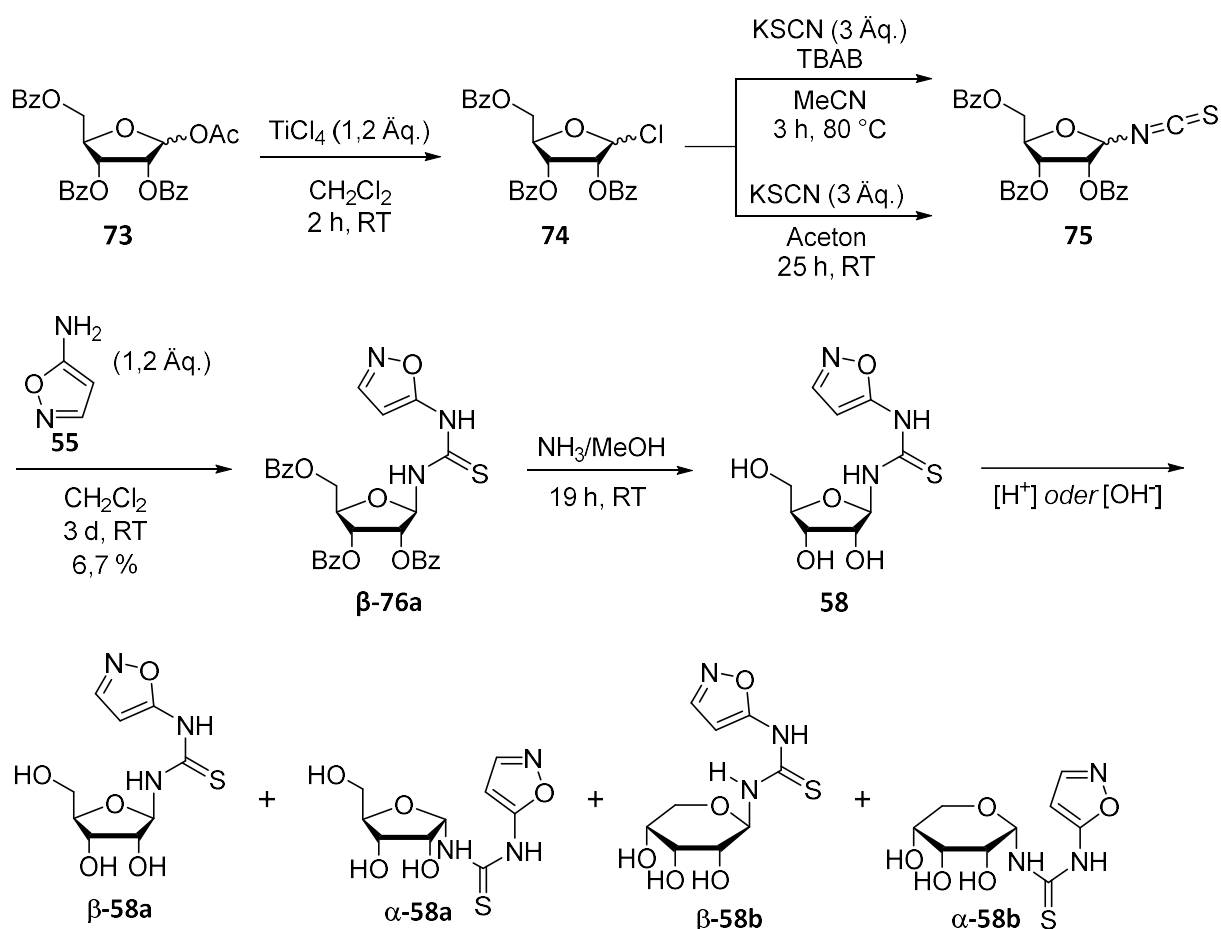


Schema 47 Zweistufige Darstellung von *N*-(Isoxazol-5-yl)thioharnstoff (57) als Referenzverbindung mittels herkömmlicher Synthese. Im ersten Schritt erfolgte die Thiocarbamoylierung von 5-Aminoisoxazol (55) mit Benzoylisothiocyanat (59) zu 60. Im zweiten Schritt erfolgte dessen Entschützung zu *N*-(Isoxazol-5-yl)thioharnstoff (57) mittels NaOMe.

4.2.2.3.2 *N*-Ribosyl-*N*-(isoxazol-5-yl)thioharnstoff (58)

In Anlehnung an die in Abschnitt 4.1 etablierte Ribosylierung, wurde eine Strategie zur Synthese der vier Ribosid-Isomere von *N*-Ribosyl-*N*-(isoxazol-5-yl)thioharnstoff α - und β -58a und -58b entwickelt und umgesetzt, siehe Schema 48. Zunächst wurde 1'-*O*-Acetyl-2',3',5'-tri-*O*-benzoyl-D-ribose (73) mit TiCl₂ in den 1'-Chlorozucker 74 überführt.^[220] Nach Aufarbeitung wurde ein Teil des Rohprodukts 74 in Aceton mit KSCN bei RT für 25 h umgesetzt. Eine zweite Umsetzung des Rohprodukts 74 erfolgte mit KSCN und dem Phasen-Transfer-Katalysator TBAB (Tetrabutylammoniumbromid) in MeCN bei 80 °C für 3 h. Beide Strategien lieferten das

Zielprodukt 1'-Isothiocyanato-2',3',5'-tri-*O*-benzoyl-D-ribose (**75**). Da sich die Abtrennung von TBAB als schwierig herausstellte, sollte in Zukunft die Umsetzung von **74** ohne TBAB erfolgen.



Schema 48 Zusammenfassung der Darstellung der α - und β -Furanosid- sowie α - und β -Pyranosid-Isomere von *N*-Ribosyl-*N*-(isoxazol-5-yl)thioharnstoff α - und β -**58a** und **58b** als Referenzverbindungen mittels herkömmlicher Synthese.

Für die Ausbeute der Addition von 5-Aminoisoxazol (**55**) an den SCN-Zucker **75** könnte der Einsatz von Pyridin vorteilhaft sein. Im dritten Schritt der Synthese erfolgte die Addition von 5-Aminoisoxazol (**55**) an das Isothiocyanat **75**. Dazu wurde **75** in CH_2Cl_2 bei RT für drei Tage mit 5-Aminoisoxazol (**55**) umgesetzt. Aufreinigung lieferte das Benzoyl-geschützte N^1 - β -Ribofuranosyl- N^2 -(isoxazol-5-yl)thioharnstoff (β -**76a**) in einer Ausbeute von 6,7 %. Zur Optimierung dieser Reaktion führte ANNIKA TÖLKE vorläufige Untersuchungen in verschiedenen Lösungsmitteln (CH_2Cl_2 , Toluol, THF und Pyridin) durch.^[224] Die Ergebnisse dieser Untersuchungen könnten darauf hindeuten, dass die Addition von 5-Aminoisoxazol (**55**) an das Isothiocyanat **75** in Pyridin mit größerer Ausbeute verlaufen könnte als in den anderen Lösungsmitteln der Untersuchung. Eine mögliche Erklärung für die deutlich höhere Ausbeute in Pyridin liegt möglicherweise in der Bildung einer aktivierten Zwischenstufe. So könnte Pyridin in

der Lage sein durch nukleophilen Angriff eine Pyridinium-Isothiocyanato-Zwischenstufe **77** zu bilden, siehe Abb. 21. Diese könnte eine geringere Aktivierungsenergie des Angriffs von 5-Aminoisoxazol (**55**) ermöglichen und so die Ausbeute steigern.

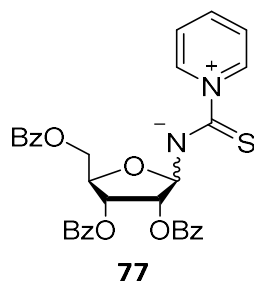


Abb. 21 Mögliche Pyridinium-Isothiocyanato-Zwischenstufe **77**.

Die Zielprodukt-Isomere α - und β -**58a** sowie **58b** neigen zur *Re*-Isomerisierung. Im letzten Schritt der Synthese erfolgte die Entschützung von *N*¹-(2',3',5'-tri-*O*-Benzoyl- β -D-ribofuranosyl)-*N*²-(isoxazol-5-yl)thioharnstoff (**β -76a**) in methanolischer NH₃-Lösung. Abb. 22 A) zeigt das RP-HPLC-Chromatogramm nach Vor-Aufreinigung mittels *Flash*-Säulenchromatographie. Durch RP-HPLC-Fraktionierung und LR-(LC-)ESI-MS konnten vier Signale dem Zielprodukt *N*-Ribosyl-*N*-(isoxazol-5-yl)thioharnstoff (**58**) zugeordnet werden, siehe Nummerierung in Abb. 22 A). Die finale Aufreinigung erfolgte mittels semi-präparativer RP-HPLC in H₂O/MeCN. Abb. 22 B) zeigt das RP-HPLC-Chromatogramm 2 h nach der Aufreinigung des Ribosid-Isomers IV (blau hinterlegt in Abb. 22 A)). Zusätzlich zur aufgereinigten Spezies IV sind drei weitere Signale zu erkennen (rot). Diese Signale konnten mittels LR-(LC-)ESI-MS dem Zielprodukt **58** zugeordnet werden. Dementsprechend erfolgte im Anschluss an die Aufreinigung eine *Re*-Isomerisierung. Es kann angenommen das dies durch den leicht sauren pH \approx 6 des Puffersystems (H₂O/MeCN) begünstigt wurde. Interessanterweise wurde bei der Aufreinigung der analogen Harnstoff-Verbindungen *N*-Ribosyl-*N*-(isoxazol-5-yl)harnstoff im gleichen Puffersystem keine unmittelbare Isomerisierung beobachtet.^[223] Demnach bewirkt die Anwesenheit des Schwefels eine drastische Steigerung der Isomerisierungsneigung der Zielprodukte α - und β -**58a** und **58b**. Vermutlich durch die deutlich erhöhte Elektronendichte des Schwefels im Vergleich zum Sauerstoff. Dies begünstigt vermutlich die Ringöffnung der Ribose.

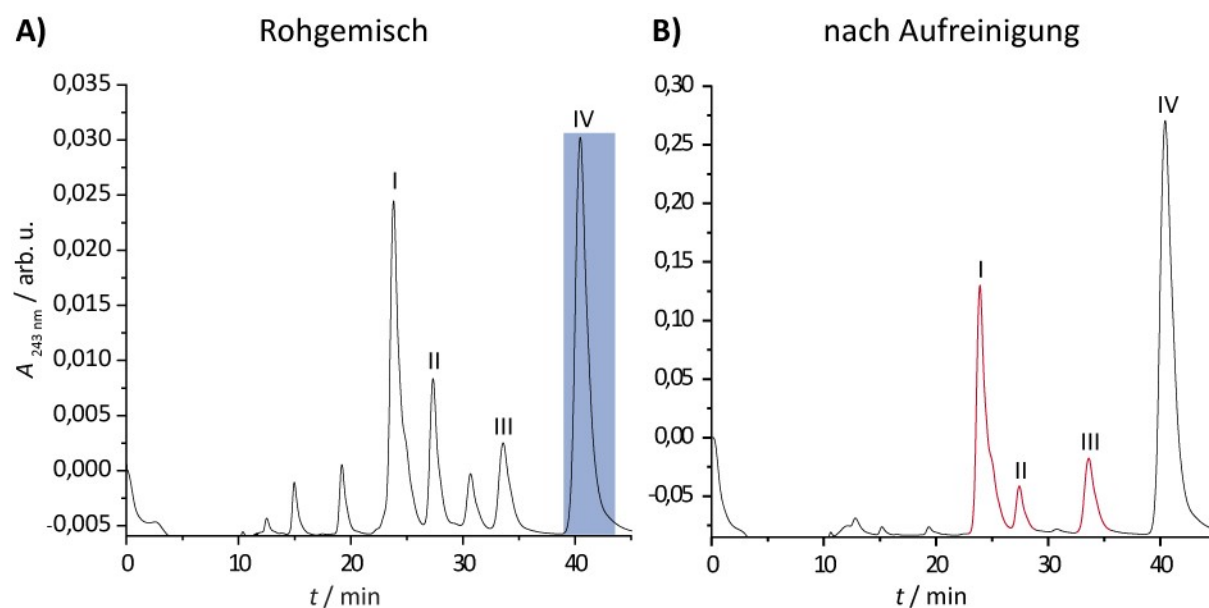


Abb. 22 A) RP-HPLC-Chromatogramm des Rohgemischs der Ribosid-Isomere von *N*-Ribosyl-*N*-(isoxazol-5-yl)thioharnstoff α - und β -**58a** und **-58b**, nach Benzoyl-Entschützung von β -**76a**, vgl. Schema 48. **blau:** Signal IV, welches mittels RP-HPLC aufgereinigt wurde. **B)** Chromatogramm der RP-HPLC-Aufreinigung des in A) markierten Signals IV, 2 h nach der Aufreinigung. Die zusätzlich gebildeten Signale (**rot**) legen nahe, dass im Anschluss an die Aufreinigung eine Re-Isomerisierung erfolgte.

Ein neutrales Puffersystem und umgehendes Gefrieren sollten eine Re-Isomerisierung verhindern können. Ribosid-Isomerisierungen können sowohl sauer als auch basisch katalysiert werden. Deshalb sollte bei der nächsten RP-HPLC-Aufreinigung der *N*-Ribosyl-*N*-(isoxazol-5-yl)thioharnstoff-Isomere (α - und β -**58a** und **-58b**) ein pH-neutrales Puffersystem gewählt werden, beispielsweise HNEt₃OAc in H₂O/MeCN. Darüber hinaus könnten die Fraktionen direkt im Anschluss an die Aufreinigung mittels N_{2(l)} gefroren werden. Anschließende Gefriertrocknung würde die Entfernung des flüchtigen Puffersystems in gefrorenem Zustand ermöglichen. Die Entfernung des Puffersystems mittels Vakuumzentrifuge ist hingegen nicht zu empfehlen. Diese Methode verursacht häufig erhöhte Proben-Temperaturen sowie eine ungleichmäßige Verdampfung der Puffer-Komponenten. Letzteres kann zum Teil drastische Änderungen des neutralen pH-Werts der Probe bedingen.

Zusammenfassung und Ausblick zur Synthese von *N*-Ribosyl-*N*-(isoxazol-5-yl)thioharnstoff (**58**)

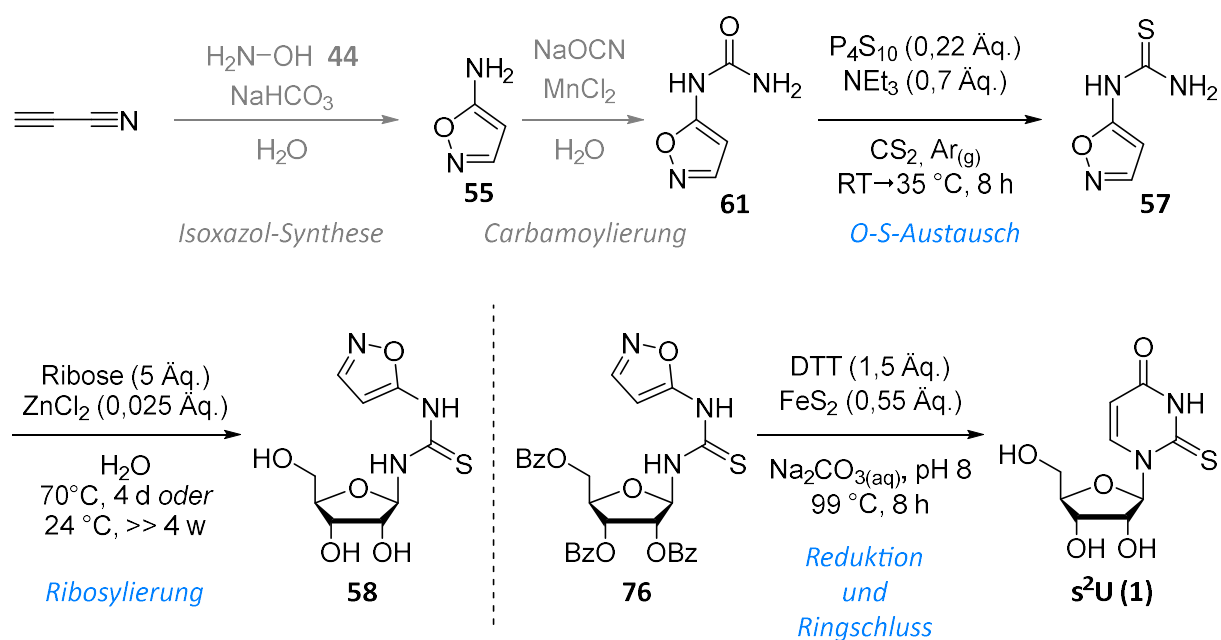
*Zusammengenommen wurden Zugänge zu *N*-Ribosyl-*N*-(isoxazol-5-yl)thioharnstoff α - und β -**58a** sowie **-58b** sowohl mittels präbiotischer Modellreaktionen als auch durch herkömmliche Synthese etabliert.* Die Re-Isomerisierung der Referenzverbindungen nach deren HPLC-Aufreinigung

verhinderte die Validierung der präbiotischen Modellreaktion mittels HR-LC-ESI-MS-Koinjektion. Die nächsten Schritte umfassen demnach die erneute Aufreinigung der Referenzverbindungen α - und β -**58a** sowie **58b**. Anschließend kann die präbiotische Modellreaktion mittels HR-LC-ESI-MS-Koinjektion validiert bzw. die Signale den Ribosid-Isomeren zugeordnet werden. Alternativ könnte eine Isolation der Referenzverbindungen auch durch Skalierung der präbiotischen Modellreaktion auf präparativen Maßstab und anschließende HPLC-Aufreinigung gelingen.

4.2.2.4 Zusammenfassung und Ausblick der Studien zur Bildung von s^2U (**1**)

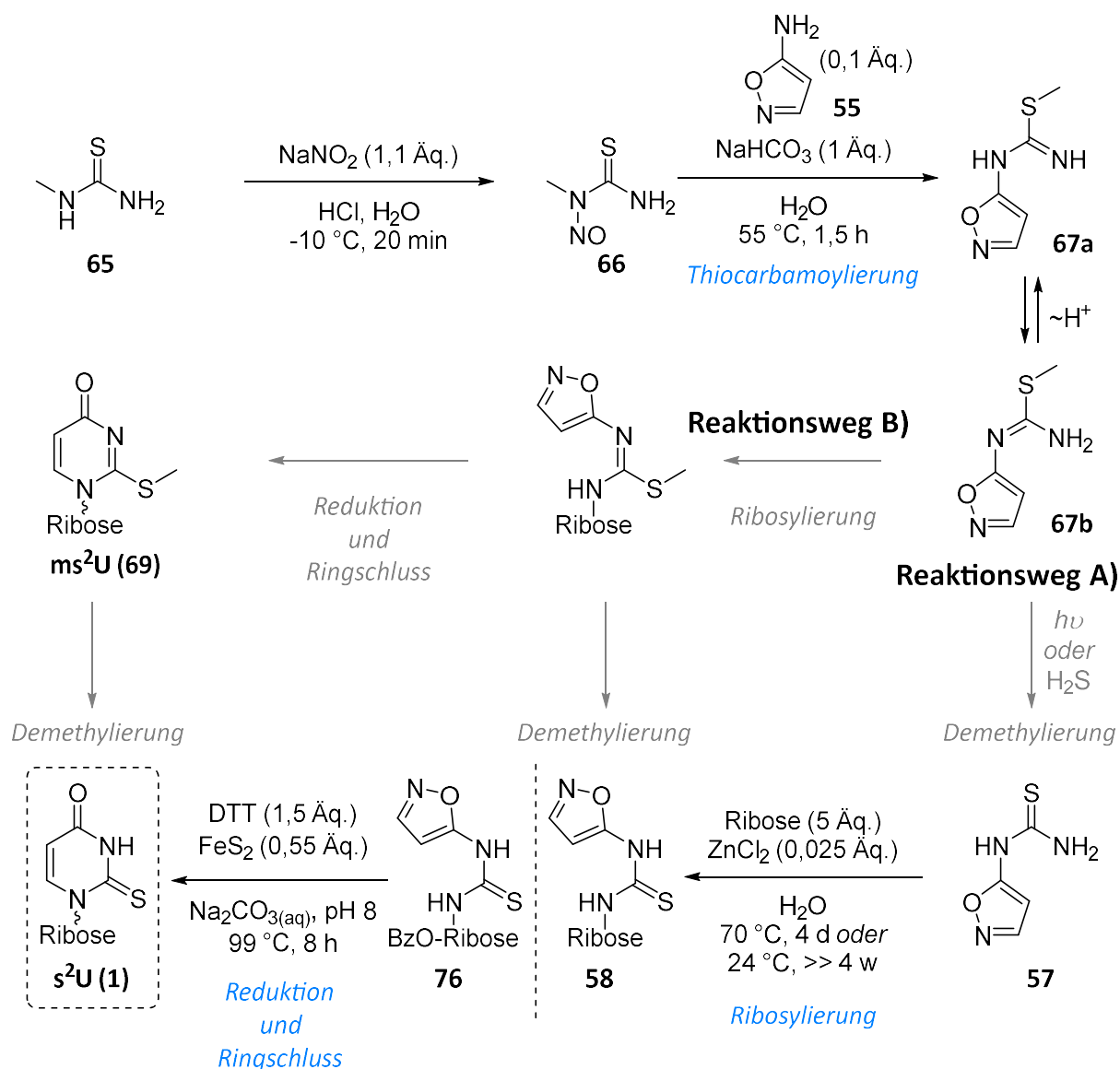
Zusammengenommen konnten in diesem Projekt die zentralen Schritte auf dem Weg zur Synthese von s^2U (**1**) in präbiotischen Modellreaktionen etabliert werden. Im Laufe dieser Arbeit wurden zwei Synthesestrategien entwickelt, die sich in der Art der Thiocarbamoylierung unterscheiden. Im Folgenden werden Zusammenfassungen beider Strategien präsentiert.

Schema 49 fasst die präbiotischen Modellreaktionen zur Synthese von s^2U (**1**) via P_4S_{10} und 5-Aminoisoxazol (**55**) zusammen. In dieser Arbeit etablierte Schritte sind schwarz dargestellt. Alle weiteren Schritte wurden bereits außerhalb dieser Arbeit demonstriert und benötigen keiner weiteren Etablierung (grau).^[223] Zusätzlich zu den präbiotischen Modellreaktionen gelang es Zugänge zu den Referenzverbindungen **55** sowie **58** mittels herkömmlicher Synthese zu etablieren. Damit wurden Zugänge zu allen notwendigen Referenzverbindungen entwickelt. Alle weiteren Referenzverbindungen sind kommerziell erhältlich oder wurden außerhalb dieser Arbeit etabliert.^[223]



Schema 49 Zusammenfassung der in dieser Arbeit etablierten präbiotischen Modellreaktionen zur Bildung von s²U (**1**) via P₄S₁₀ und 5-Aminoisoxazol (**55**). Die Quantifizierung der Ausbeuten steht noch aus. Alle grau dargestellten Schritte wurden bereits außerhalb dieser Arbeit etabliert. Alle in dieser Arbeit etablierten Schritte sind schwarz dargestellt.

Schema 50 fasst die präbiotischen Modellreaktionen zur Synthese von s²U (**1**) via N¹-Methyl-N¹-(nitroso)thioharnstoff (**66**) und 5-Aminoisoxazol (**55**) zusammen. In dieser Arbeit etablierte Schritte sind schwarz dargestellt. Diese Strategie beginnt unter Bedingungen analog zur Synthese methylierter RNA-Nukleoside und Aminosäure-RNA-Nukleoside nach SCHNEIDER *et al.* Via Reaktionsweg A) würde diese Strategie die Möglichkeit zur Bildung von s²U (**1**) durch Demethylierung mittels Photolyse^[233,234] oder H₂S^[235,236] eröffnen. Reaktionsweg B) würde darüber hinaus die Möglichkeit bieten einen Zugang zu ms²U (**69**) zu etablieren. In Bezug auf Reaktionsweg A) wurden darüber hinaus Zugänge zu allen relevanten Referenzverbindungen **66**, **67**, **55** und **58** etabliert.



Schema 50 Zusammenfassung der in dieser Arbeit etablierten präbiotischen Modellreaktionen zur Bildung von s^2U (1) via N^1 -Methyl- N^1 -(nitroso)thioharnstoff (66) und 5-Aminoisoxazol (55). BzO-Ribose: 2',3',5'-tri-*O*-benzoyl-D-ribosyl. Die Quantifizierung der Ausbeuten steht noch aus. Alle in dieser Arbeit etablierten Schritte sind schwarz dargestellt. Die grau dargestellten Schritte stehen noch zur Etablierung aus.

Um dieses Projekt zur Publikationsreife zu führen verbleiben demnach unabhängig von der Synthesestrategie die folgenden Schritte: Zuordnung der Ribosid-Isomere mittels HR-LC-ESI-MS-Koinjektion sowie Anwendung der in dieser Arbeit für Reduktion und Ringschluss an 57 etablierten Bedingungen auf 58. Quantifizierung der Reaktionsausbeuten mittels HR-LC-ESI-MS-Koinjektion. Bezüglich der P_4S_{10} -Strategie sollte die Ausbeute des O-S-Austauschs optimiert werden. Bezüglich der N^1 -Methyl- N^1 -(nitroso)thioharnstoff-Strategie sollten im nächsten Schritt die Demethylierung via Photolyse^[233,234] oder H_2S ^[235,236] etabliert sowie die Ausbeute der Thiocarbamoylierung optimiert werden.

5 Protein-Interaktoren der mRNA-Modifikation m⁶A

„N⁶-methyladenosine (m⁶A) recruits and repels proteins to regulate mRNA homeostasis”
R. R. Edupuganti, S. Geiger, R. G. H. Lindeboom, H. Shi, P. J. Hsu, Z. Lu, S. Y. Wang,
M. P. A. Baltissen, P. W. T. C. Jansen, M. Rossa, M. Müller, H. G. Stunnenberg, C. He,
T. Carell, M. Vermeulen, *Nat. Struct. Mol. Biol.* **2017**, *24*, 870–878.

Prolog

Die hier vorgestellte Studie stellt die erste systematische Identifizierung von m⁶A-Protein-Interaktoren dar. Durch umfassende massenspektrometrische Untersuchungen in verschiedenen Zelllinien und Sequenzkontexten konnte eine Vielzahl an m⁶A-Protein-Interaktoren identifiziert werden. Durch statistische Klassifizierung der Affinitäten wurde bestimmt, welche Proteine m⁶A spezifisch binden (*reader*-Proteine) und welche Proteine von m⁶A spezifisch abgestoßen werden. Dabei sind zwei Entdeckungen besonders hervorzuheben: FMR1 und G3BP1. FMR1 konnte als m⁶A-bindendes Protein identifiziert werden. Dies stellt die erste bekannte Verbindung zwischen einer mRNA-Modifikation und einer Krankheit des Autismus-Spektrums dar. Darüber hinaus konnte gezeigt werden, dass G3BP1 von m⁶A stark abgestoßen wird. G3BP1 ist ein bekanntes Stressgranulat-Protein und erhöht in dieser Funktion die Stabilität von mRNA.^[245] Die Anwesenheit von m⁶A in mRNA führt zur Abstoßung von G3BP1 und damit zur Verringerung der Lebensdauer von m⁶A-modifizierter mRNA. Durch diese Studie konnten zudem bekannte m⁶A-Interaktoren bestätigt werden. Nicht zuletzt stellt diese Arbeit eine Ressource dar: Zu einer Vielzahl mit m⁶A-interagierenden Proteinen (sowohl bindende als auch abgestoßene) sowie zu Proteinen die nicht mit m⁶A-/A-interagieren („Hintergrund-Proteine“).




Autorenbeitrag

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

Lizenz

Kopie der Publikation mit Erlaubnis des Verlags

N^6 -methyladenosine (m^6A) recruits and repels proteins to regulate mRNA homeostasis

Raghu R Edupuganti^{1,4}, Simon Geiger^{2,4}, Rik G H Lindeboom^{1,4}, Hailing Shi^{3,4} , Phillip J Hsu³, Zhike Lu³, Shuang-Yin Wang¹, Marijke P A Baltissen¹, Pascal W T C Jansen¹, Martin Rossa², Markus Müller², Hendrik G Stunnenberg¹, Chuan He³, Thomas Carell²  & Michiel Vermeulen¹ 

RNA modifications are integral to the regulation of RNA metabolism. One abundant mRNA modification is N^6 -methyladenosine (m^6A), which affects various aspects of RNA metabolism, including splicing, translation and degradation. Current knowledge about the proteins recruited to m^6A to carry out these molecular processes is still limited. Here we describe comprehensive and systematic mass-spectrometry-based screening of m^6A interactors in various cell types and sequence contexts. Among the main findings, we identified G3BP1 as a protein that is repelled by m^6A and positively regulates mRNA stability in an m^6A -regulated manner. Furthermore, we identified FMR1 as a sequence-context-dependent m^6A reader, thus revealing a connection between an mRNA modification and an autism spectrum disorder. Collectively, our data represent a rich resource and shed further light on the complex interplay among m^6A , m^6A interactors and mRNA homeostasis.

m^6A was first recognized as an abundant eukaryotic mRNA modification in the 1970s^{1,2}, but the functional significance of this modification remained unclear for years. Recently, it was shown that m^6A is enzymatically added to and removed from mRNA molecules^{3–5}. A heterodimer of METTL3–METTL14, in complex with WTAP, KIAA1429 and ZFP217, catalyzes the addition of a methyl group to adenosine in the context of a conserved consensus sequence (ACU)^{5–8}. Two demethylases, FTO and ALKBH5, are known to remove the m^6A modification^{3,4}. Furthermore, researchers have used antibody-based enrichment methods to determine the transcriptome-wide distribution of m^6A ^{9,10}. m^6A marks several thousand mRNAs in mammalian cells and is enriched at the translation-termination region^{9,10}. In recent years, various biological phenomena have been associated with the m^6A modification on mRNA, such as obesity¹¹, plant development¹², yeast meiosis¹³, cancer^{14,15}, cell fate transitions¹⁶, fertilization¹⁷ and pluripotency¹⁸. At the molecular level, m^6A is known to affect mRNA stability¹⁹, translation^{20–22}, microRNA biogenesis²³, splicing^{11,24}, X-chromosome inactivation²⁵ and other biological processes²⁶.

One important question is how m^6A affects mRNA homeostasis. Previous work has shown that the YTH domain is a specific ' m^6A -reader' domain⁹, suggesting that YTH-domain-containing proteins act as effectors of m^6A in cells. Indeed, recent studies have revealed a role for YTH-domain-containing proteins in the regulation of various aspects of mRNA homeostasis: YTHDF2 regulates mRNA stability¹⁹, YTHDC1 regulates splicing²⁴, and YTHDF1 and YTHDF3 regulate translation^{22,27}. In addition to YTH-domain-containing

proteins, HNRNP proteins have also been shown to recognize m^6A ^{28–30}. Whether additional m^6A -reader proteins exist is currently unclear. Furthermore, it is currently unknown whether certain proteins are specifically repelled by m^6A -containing mRNA sequences, a phenomenon that is observed, for example, for methylated DNA³¹.

Here we describe global mass-spectrometry-based interaction screening for m^6A readers in various cell types and mRNA sequence contexts. Our data show that YTH proteins are conserved, cell-type-independent m^6A readers. In addition, we identified sequence-context-dependent m^6A readers, including FMR1, thus linking m^6A to the regulation of mRNA translation and fragile X-linked mental retardation³². Strikingly, our data also show that many proteins have a preference for unmodified versus m^6A -containing mRNA. One such repelled protein is G3BP1, a known stress granule protein³³. Our RNA-seq and pulsed-SILAC experiments showed that G3BP1 and FMR1 regulate m^6A -linked mRNA stability and translation rates, respectively. In summary, our data reveal intricate interplay between m^6A readers and repelled proteins and various aspects of mRNA homeostasis.

RESULTS

The m^6A interactome in mammalian cells

To identify readers for m^6A , we carried out RNA pulldowns in mammalian cell lysates followed by quantitative mass-spectrometry-based proteomics (Fig. 1 and Supplementary Fig. 1a). To this end, we designed RNA probes containing four repeats of the m^6A consensus

¹Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, Nijmegen, the Netherlands.

²Center for Integrated Protein Science at the Fakultät für Chemie und Pharmazie, Ludwig-Maximilians-Universität München, Munich, Germany. ³Department of Chemistry, Department of Biochemistry and Molecular Biology, and Institute for Biophysical Dynamics, Howard Hughes Medical Institute, The University of Chicago, Chicago, Illinois, USA. ⁴These authors contributed equally to this work. Correspondence should be addressed to C.H. (chuanhe@uchicago.edu), T.C. (thomas.carell@lmu.de) or M.V. (Michiel.Vermeulen@science.ru.nl).

Received 16 May; accepted 9 August; published online 4 September 2017; doi:10.1038/nsmb.3462

sequence GGACU^{9,10}. As a control, we used an unmethylated strand (Supplementary Fig. 1a). We incubated immobilized probes (Supplementary Fig. 1b–d, Supplementary Table 1) with ‘light’ (L) and ‘heavy’ (H) SILAC-labeled lysates. Each experiment consisted of a

‘forward’ and a ‘reverse’ (label-swap) pull-down. After incubation and washes, bound proteins were subjected to on-bead trypsin digestion and analyzed by liquid chromatography–tandem mass spectrometry (LC-MS/MS) (Supplementary Fig. 1a). Supplementary Figure 1a,

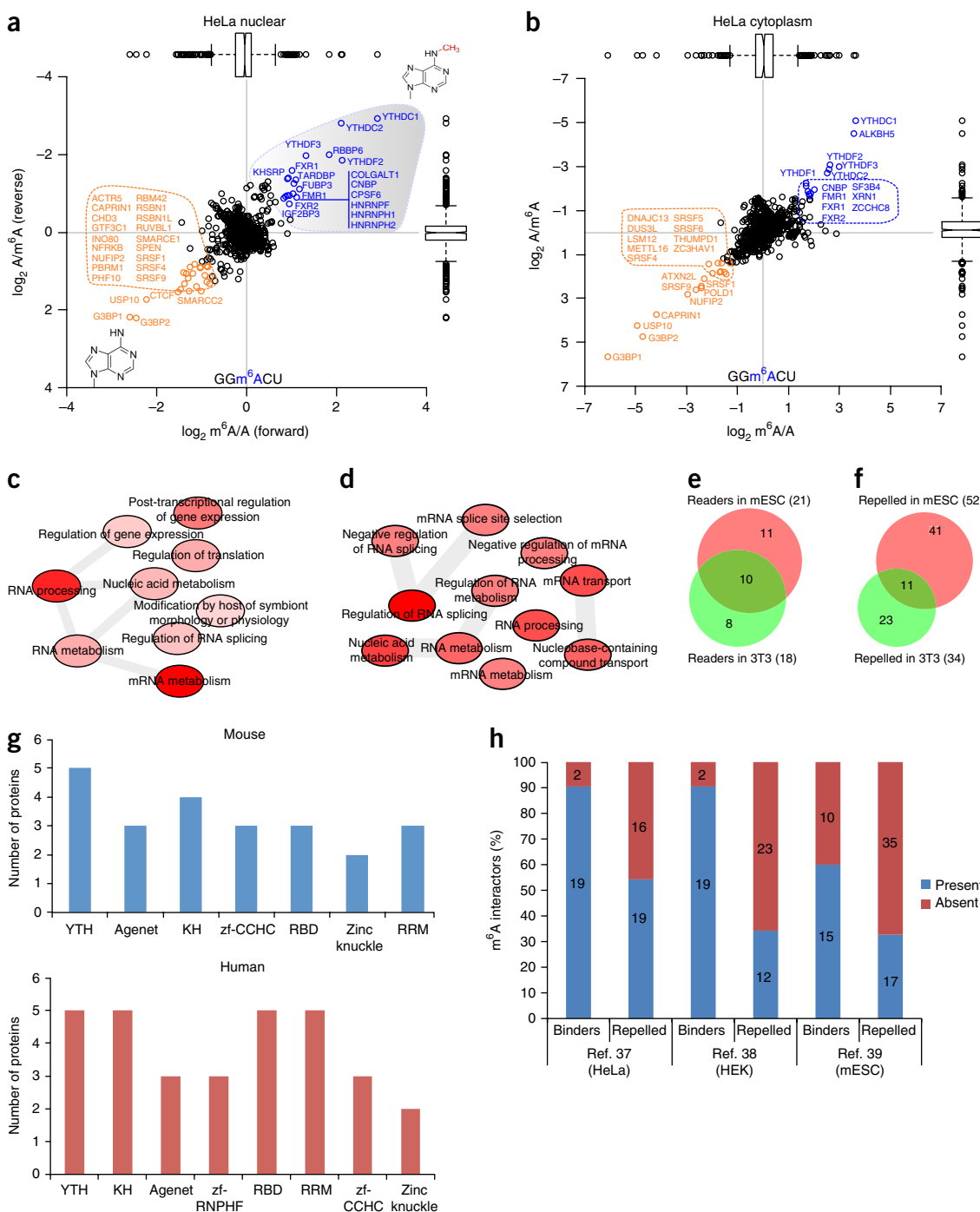


Figure 1 A global m⁶A interactome. (a,b) Results of a SILAC-based m⁶A RNA pull-down in HeLa nuclear extract (a) and cytoplasmic lysates (b). m⁶A-interacting proteins are depicted in blue, repelled proteins are in orange, and background proteins are in black. We used a threshold of 1.5× the interquartile range to identify significant interactors. In the box plots, center lines indicate medians, box edges represent the interquartile range (IQR), and whiskers extend to ±1.5× the IQR; outlier values are shown as black circles. (c,d) GO-term enrichment analysis (Online Methods) of m⁶A readers (c) and repelled proteins (d) in HeLa cells. The relationships between various GO terms are depicted by gray bars. The intensity of the red shading indicates the significance of the relationship (darker shading denotes greater significance). (e) The overlap between m⁶A readers identified in mESCs and differentiated cells (NIH 3T3). (f) The overlap between m⁶A-repelled proteins in mESCs and NIH 3T3 cells. The Venn diagrams summarize data obtained in three independent experiments. (g) Protein domains present in m⁶A interactors in mouse (top) and human (bottom) cells. (h) Many m⁶A interactors have also been identified in previously published *in vivo* mRNA–protein interaction studies^{37–39}. Source data for g and h are available online. Proteomics data and GO enrichment analysis are presented in Supplementary Data Set 2.

which depicts a typical RNA-pulldown interaction plot, shows readers in the upper right quadrant (high forward H/L ratio; low H/L ratio in the reverse experiment). Proteins appearing in the lower left quadrant bound preferentially to the unmodified probe. Background proteins are clustered together around the origin of the plot, and showed ~1:1 ratios in both experiments.

We identified 21 m⁶A readers in HeLa nuclear (Fig. 1a) and cytoplasmic (Fig. 1b) lysates. YTH-domain-containing proteins were among the most prominent m⁶A readers (Fig. 1a,b, Supplementary Data Set 2). We identified ALKBH5, an m⁶A demethylase³, as a reader for m⁶A-containing RNA, which indicates that our pulldown approach was able to capture putative enzymes involved in m⁶A metabolism. Other prominent m⁶A readers are FMR1 and its paralogs FXR1 and FXR2. We also identified three HNRNP proteins as m⁶A readers in HeLa nuclear lysates. Other HNRNP proteins—namely, HNRNPG³⁰, HNRNPC²⁹ and HNRNPA2B1 (ref. 28)—have previously been linked to m⁶A. These proteins recognize m⁶A in a structure- and sequence-dependent context, which is probably why they were missing in our pulldowns. An analysis of Gene Ontology (GO)-term enrichment in m⁶A readers in HeLa cells showed, as expected, that the most enriched biological pathways among m⁶A readers are related to mRNA metabolism and processing (Fig. 1c).

Epigenetic modifications such as m⁵C not only recruit but also repel proteins³¹. We identified a number of proteins that preferentially interacted with the unmodified RNA sequence and thus are repelled by m⁶A. Prominent among these were two RNA-binding proteins, G3BP1 and G3BP2, both of which are known stress granule proteins³⁴, as well as their known interaction partners USP10 and CAPRIN1 (ref. 35). This indicates that in our RNA pulldowns we identified both direct RNA-protein interactors and indirect RNA-protein interactors, which are mediated by protein-protein interactions. Another repelled protein is METTL16, which was very recently identified as an adenosine methyltransferase for small nuclear RNA³⁶. GO-term enrichment analysis of proteins that are repelled by m⁶A revealed terms related to mRNA splicing and transport (Fig. 1d).

Although m⁶A levels are mostly stable in different cell types, the m⁶A interactome may vary between different cell types. To address this, we carried out m⁶A pulldowns in mouse embryonic stem cells (mESCs), neuronal progenitor cells and NIH 3T3 cell lysates. In mESCs, Yth proteins, Alkbh5 and Fmr1 proteins represented the most prominent m⁶A readers, which is consistent with data obtained for HeLa cells (Supplementary Fig. 1e,f). G3bp1 and G3bp2, along with Usp10, again represented the most prominent repelled proteins. RNA pulldowns in mouse neuronal progenitor cell lysates combined with a chemical stable-isotope labeling strategy (dimethyl labeling) once again showed that Yth-domain-containing proteins were conserved, cell-type-independent m⁶A interactors (Supplementary Fig. 1g,h). m⁶A interactors in NIH 3T3 cells were similar to those identified in other cell types. Repelled proteins, however, tended to be more diverse and numerous between different cell types (Supplementary Fig. 1i, Fig. 1e,f).

To discriminate between m⁶A readers that are dependent on or independent of the RNA sequence context, we generated two different probes with degenerate sequences surrounding m⁶A and then used these baits for RNA pulldowns in mouse whole cell extracts. From the results, it is clear that YTH-domain-containing proteins largely interacted with m⁶A in an RNA-sequence-context-independent manner, whereas other readers, such as FMR1, showed RNA-sequence-context-dependent binding. Repelled proteins showed limited overlap between different degenerate strands (Supplementary Fig. 1j,k). A global analysis of RNA-binding domains present among m⁶A

readers and repelled proteins showed that apart from the YTH domain, the KH, RRM and RBD domains were predominant, which indicates that some of these domains may be regulated by m⁶A directly, albeit in most cases in an RNA-sequence-context-dependent manner (Fig. 1g, Supplementary Fig. 1l).

Finally, we reasoned that certain proteins identified in our screens may have been identified previously in large-scale mRNA interactome studies, which have been conducted via mRNA-protein cross-linking methodologies in HeLa³⁷, HEK293T³⁸ and mESC³⁹ mRNA-protein binding resources (Fig. 1h). A substantial number of the m⁶A readers that we identified in our pulldowns are also covered in these large-scale mRNA-protein cross-linking studies, which indicates that at least in some cases, these detected *in vivo* mRNA-protein interactions are driven or regulated by mRNA modifications such as m⁶A.

Stress granule protein G3BP1 is repelled by m⁶A and regulates mRNA stability

One important observation from our pulldown experiments is that many proteins are repelled by the m⁶A modification. The two most consistently and strongly repelled proteins were G3BP1 and G3BP2 (G3BPs). These proteins are known to have critical roles in stress granule assembly³³ and are important for embryonic development. G3BPs have very similar domain structures and contain canonical RNA-binding domains⁴⁰ (Fig. 2a). First, we used recombinant protein to validate preferential G3BP1 binding to the unmodified RNA. We carried out pulldown experiments in two predominant m⁶A sequence contexts: GGm⁶ACU and GAm⁶ACU⁴¹ (Fig. 2b). In agreement with our quantitative mass spectrometry results (Fig. 1), G3BP1 clearly interacted with the GGACU probe, but not with the GGm⁶ACU probe. This methylation-sensitive binding was less clear in the GAACU sequence context (Fig. 2b). We further verified these results via quantitative mass spectrometry (Supplementary Fig. 2a). Thus, we conclude that G3BP1 is repelled by m⁶A in an RNA-sequence-context-dependent manner.

To identify global mRNA binding sites for G3BP1 and G3BP2, we carried out PAR-CLIP (photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation) analyses in HEK293T cells with Flag-tagged constructs (Fig. 2c,d, Supplementary Fig. 2b). We identified ~3,800 and ~3,300 mRNA binding sites for G3BP1 and G3BP2, respectively; ~2,300 of these were shared between both proteins (Fig. 2e, Supplementary Data Set 3). At the transcript level, G3BP1 bound to 2,019 genes, whereas G3BP2 bound to 1,672 genes (Fig. 2e); 1,377 target genes were shared between G3BP1 and G3BP2. Most G3BP peaks were found in the 3'-UTR region, followed by intergenic and coding regions (Supplementary Fig. 2c). As has been observed for m⁶A^{10,42}, G3BP-binding sites were enriched at 3'-UTR regions (Fig. 2f). However, unlike for m⁶A, 5'-UTR regions showed less G3BP binding. The most predominant mRNA-binding motif of G3BP1 was CAACUC ($P = 1 \times 10^{-81}$), whereas CAACUCU was the most predominant binding motif of G3BP2 ($P = 1 \times 10^{-76}$) (Fig. 2g). These motifs were enriched in 31% and 35% of G3BP1 and G3BP2 binding sites, respectively. Thus, the central part of the known m⁶A consensus sequence (ACU) was conserved among G3BP-binding sites. GO-term enrichment analysis showed that G3BP1-bound mRNAs were significantly enriched for intracellular protein trafficking, mRNA splicing and pre-mRNA processing (Fig. 2h), whereas G3BP2-bound mRNAs were enriched for intracellular protein trafficking, chromatin packaging and the cell cycle (Supplementary Fig. 2d). These results indicate that G3BPs regulate mRNAs with both diverse and overlapping biological functions.

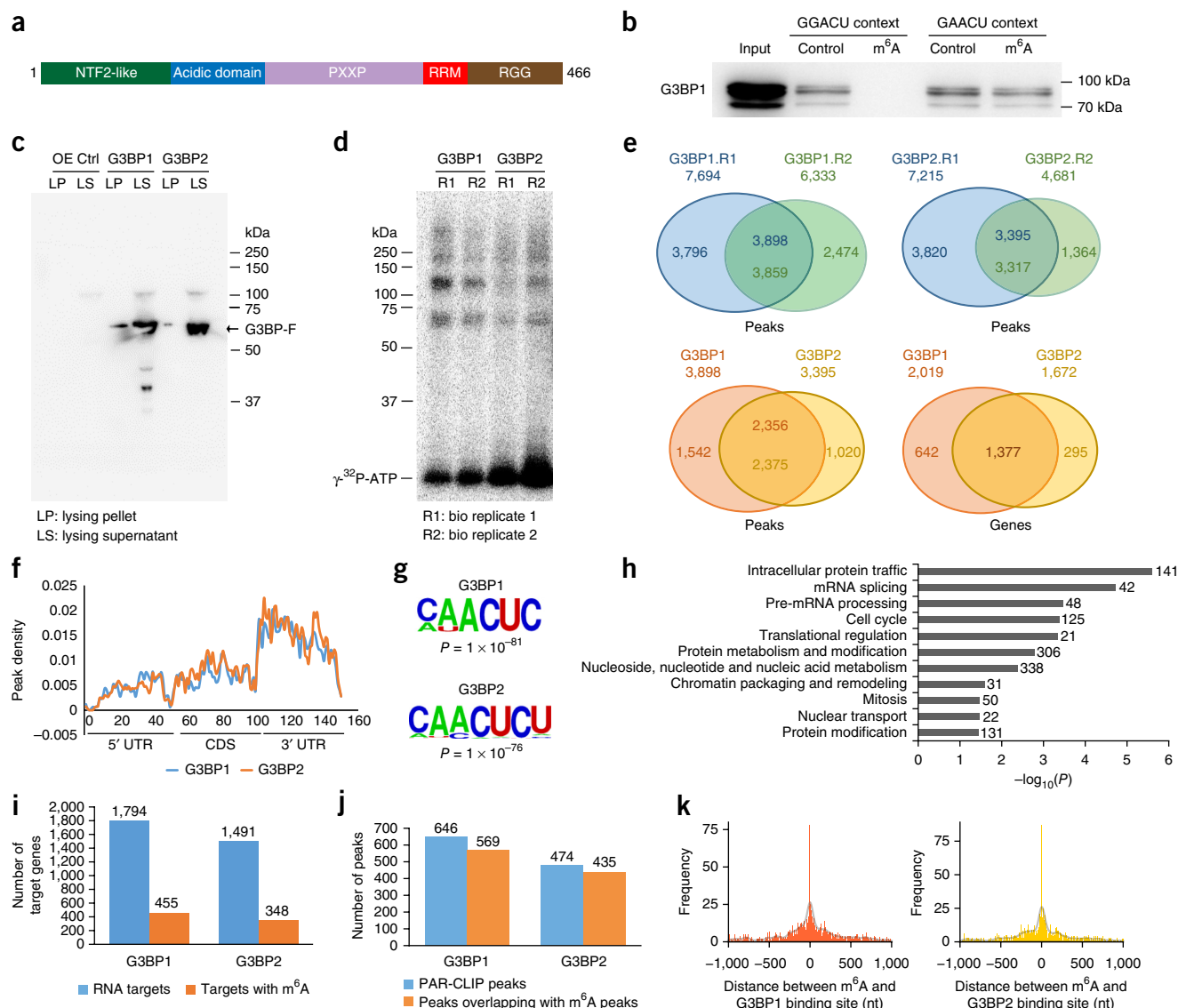


Figure 2 G3BP1 and G3BP2 interact with thousands of transcripts. **(a)** The G3BP1 domain structure. **(b)** GST western blot showing RNA-sequence-context-dependent inhibition of GST-G3BP1 binding to RNA by m⁶A. Molecular weights are indicated on the right. **(c, d)** RNA cross-link immunoprecipitation in HEK293 cells expressing either Flag-G3BP1 or Flag-G3BP2. Flag western blots **(c)** and radioactively labeled RNA-protein complexes **(d)** are shown. OE, overexpression; Ctrl, control. **(e)** Common peaks between PAR-CLIP biological replicates 1 and 2 for G3BP1 (top left) and G3BP2 (top right), and comparisons between G3BP1 and G3BP2 peaks (bottom left) and target genes (bottom right). **(f)** Metagene profiles of G3BP1 and G3BP2 distribution across the transcriptome. **(g)** The most enriched consensus sequences of G3BP1 (top) and G3BP2 (bottom) on mRNA; *P* values were determined as described in the Online Methods. **(h)** The most significantly enriched GO terms among G3BP1 target genes. The numbers to the right of the bars represent the number of genes in that GO term. **(i)** The number of G3BP1 and G3BP2 peaks that overlap with m⁶A peaks among m⁶A-containing mRNAs. **(j)** The number of G3BP1 and G3BP2 peaks on m⁶A-containing mRNAs that overlap with m⁶A. **(k)** The distribution of G3BP1 (left) and G3BP2 (right) relative to m⁶A sites at single-nucleotide resolution⁴³. The distance between PAR-CLIP peaks and m⁶A sites was counted in 10-nucleotide bins. A Gaussian kernel smoothing over the histogram is plotted as a transparent black line. The uncropped blot image for **b** is shown in **Supplementary Data Set 1**. Source data for **i** and **j** are available online. G3BP1 and G3BP2 PAR-CLIP data and related GO term enrichment analysis are presented in **Supplementary Data Set 3**.

Next, we set out to investigate the potential interplay between m⁶A and G3BP1. Approximately 25% of G3BP1 target genes and 23% of G3BP2 target genes are known to carry m⁶A (Fig. 2i). On these target genes, the m⁶A site frequently overlaps with the binding site for G3BP1/G3BP2 (88% and 92%, respectively) (Fig. 2j). This held true in comparisons of G3BP PAR-CLIP data with recently published base-resolution m⁶A data⁴³ (Fig. 2k). This indicates that at least a fraction of mRNA G3BP1- and G3BP2-binding sites are direct targets for m⁶A. In certain cases—for example, in a GGACU context—m⁶A would inhibit G3BP binding. To further support this observation,

we plotted the distance between G3BP peaks relative to previously reported YTHDF2 and YTHDC1 PAR-CLIP data^{19,24}. This analysis showed that a substantial amount of G3BP and YTHDF2/YTHDC1 peaks overlap, which further suggests that mRNA binding of G3BPs could be regulated by m⁶A (Supplementary Fig. 2e–h). However, it should once again be noted that many G3BP1-binding sites occur in a sequence context (AACU) in which methylation may not negatively affect G3BP1 binding.

Given that m⁶A negatively correlates with mRNA stability and that G3BP1 binding to mRNA is negatively affected by m⁶A in certain

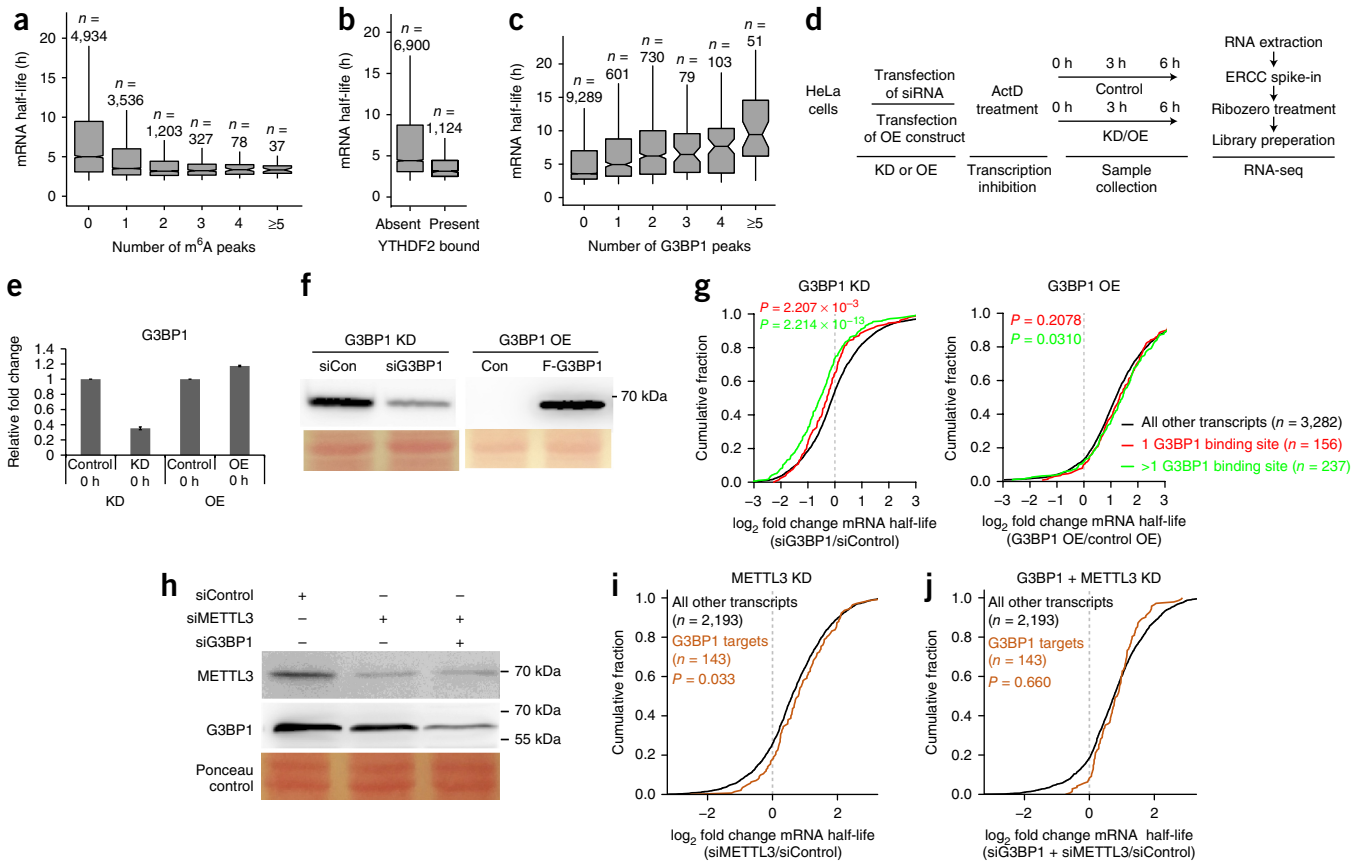


Figure 3 G3BP1 protects target mRNAs from degradation. (a) The correlation between the number of m⁶A peaks¹⁹ on mRNA molecules and mRNA half-life⁴⁴. (b) The correlation between YTHDF2 mRNA binding¹⁹ and mRNA stability. (c) The correlation between G3BP1 mRNA binding and target gene transcript stability⁴⁴. In a–c, *n* represents the number of genes, center lines in box plots indicate medians, box edges represent the interquartile range, and whiskers extend to ±1.5× the interquartile range; outlier values are not shown. (d) mRNA half-life measurements after G3BP1 knockdown (KD) or overexpression (OE). ActD, actinomycin D. (e) qPCR analysis of G3BP1 levels after protein knockdown or overexpression. GAPDH was used to normalize expression. Data are shown as the mean ± s.e.m.; *n* = 3 independent experiments. (f) Western blots depicting knockdown and overexpression of G3BP1. Ponceau-stained protein bands are shown as loading controls. Con, control plasmid expressing Flag tag. (g) Cumulative distribution function plots showing the effect of G3BP1 knockdown on target genes. The log₂ fold changes in mRNA half-life were grouped and analyzed on the basis of the number of G3BP1-binding sites on each transcript. *P* values were calculated by two-sided Mann–Whitney *U* test. (h) Western blots depicting knockdown of METTL3 and G3BP1. Ponceau-stained protein bands are shown as loading controls. (i) A cumulative distribution function plot showing the effect of METTL3 knockdown on G3BP1 target genes. (j) A cumulative distribution function plot showing the effect of combined G3BP1 and METTL3 knockdown on G3BP1 target genes. log₂ fold changes were grouped and analyzed as described for g. Uncropped blot images for f and h are shown in **Supplementary Data Set 1**. A summary of mRNA half-lives as determined by RNA-seq analysis is presented in **Supplementary Data Set 4**.

sequence contexts, we reasoned that G3BP1 binding might have a positive effect on the stability of target mRNAs. To address this, we integrated G3BP1 PAR-CLIP data with global mRNA half-life data⁴⁴. First, we assessed the effect of m⁶A on mRNA half-life. As reported previously¹⁹, we observed a negative correlation between mRNA half-life and m⁶A (Fig. 3a). Because YTHDF2 is the main effector of m⁶A-induced mRNA degradation, we also correlated YTHDF2 PAR-CLIP data¹⁹ to mRNA half-life. As was observed for m⁶A, YTHDF2 binding was negatively correlated with mRNA half-life (Fig. 3b). The effect of YTHDF2 binding on mRNA half-life can be completely explained by the amount of m⁶A on the protein’s mRNA targets (Supplementary Fig. 3a). Next, we correlated the effect of G3BP1 binding to the half-life of mRNA targets. Here we observed a significant increase in mRNA half-life concordant with an increasing number of G3BP1 binding sites on target mRNAs (Fig. 3c; *P* < 2.2 × 10⁻¹⁶).

To further investigate the role of G3BP1 in regulating mRNA half-life, we carried out G3BP1 expression perturbation experiments combined with RNA-seq (Fig. 3d–f, Supplementary Data Set 4).

We added actinomycin D to cells that contained either reduced or increased amounts of G3BP1 in order to inhibit new mRNA synthesis. Next, we carried out RNA-seq on equal numbers of cells at fixed time points after the addition of actinomycin D. We added an ERCC spike-in to the purified RNA, which, combined with accurate quantification of cell numbers, facilitated absolute mRNA quantification. As expected, G3BP1 knockdown resulted in a significant decrease in the stability of G3BP1 target mRNAs (Fig. 3g). In agreement with the *in silico* analysis (Fig. 3c), the number of G3BP1 binding sites on target mRNA molecules correlated with the stability of those mRNAs. mRNA targets with one G3BP1-binding site showed decreased stability after G3BP1 knockdown, but this effect was more pronounced for targets that contained more than one binding site. In contrast, after G3BP1 overexpression, we observed a small but consistent increase in the stability of target transcripts (Fig. 3g). We also validated some of these results by qPCR (Supplementary Fig. 3b,c). Together, these results strongly suggest that G3BP1 binding to target mRNAs results in stabilization of the targets.

Finally, to investigate whether the stabilizing effects of G3BP1 binding to target mRNA molecules can be regulated by m⁶A, we determined mRNA half-lives after METTL3 knockdown alone or in combination with G3BP1 depletion (Fig. 3h). When m⁶A levels were reduced by METTL3 knockdown, we observed a significant increase in the stability of G3BP1 target mRNAs (Fig. 3i). This phenocopied the observed effects of G3BP1 overexpression (Fig. 3g). Combined knockdown of METTL3 and G3BP1 negated this effect (Fig. 3j). We thus conclude that the increased mRNA stability observed after METTL3 knockdown is at least partially caused by increased G3BP1 binding to these transcripts. In summary, these results reveal the intricate interplay among m⁶A, m⁶A readers, and proteins that are repelled by m⁶A to regulate mRNA stability and turnover.

FMR1 interacts with and regulates translation of m⁶A-containing transcripts

In both mouse and human cells, we consistently identified FMR1 as an RNA-sequence-context-dependent m⁶A reader (Fig. 1, Supplementary Fig. 1). We first set out to verify this observation by using recombinant protein. As shown in Figure 4a,b, recombinant FMR1 preferentially bound to the methylated GGACU RNA probe. As a negative control, we used an FMR1 point mutant (I304N) that is known to be defective in mRNA binding. FMR1 contains two N-terminal Aget domains, three KH domains and one RGG domain, which are known RNA-binding domains⁴⁵. Very recently, it was shown that under certain conditions RGG domains recognize m⁶A-modified RNA³⁰. To identify the domain(s) responsible for m⁶A binding, we generated several FMR1 GST-deletion constructs (Supplementary Fig. 4a). RNA pulldowns with these deletion constructs showed that all four RNA-binding domains were needed for preferential m⁶A binding (Supplementary Fig. 4b). These data thus validate our mass spectrometry data and confirm that FMR1 preferentially interacts with m⁶A-containing RNA *in vitro* in certain sequence contexts.

FMR1 binds to hundreds of mRNAs to negatively regulate their translation, and loss of FMR1 leads to fragile X-linked mental retardation syndrome^{32,46,47}. Given the overlap between the m⁶A and FMR1 consensus mRNA sequences (Fig. 4c)⁴⁸ and our *in vitro* interaction data, we asked whether FMR1 preferentially interacts with m⁶A-containing mRNA *in vivo*. To address this, we first integrated previously published m⁶A-sequencing and FMR1 PAR-CLIP data^{10,48}. We observed clear enrichment of mRNA binding by FMR1 isoforms 1 and 7 on m⁶A sites, suggesting that FMR1 binding may co-occur with m⁶A on mRNA (Fig. 4d). However, the observed spatial proximity between FMR1 and m⁶A could have been due to overlapping FMR1 and m⁶A consensus motifs, rather than (partially) driven by direct m⁶A binding. To determine whether FMR1 preferentially binds to m⁶A-containing mRNA *in vivo*, we carried out cross-linking immunoprecipitation followed by quantitative ultra-high-performance liquid chromatography–tandem mass spectrometry (CLIP-UPLC-MS/MS) to directly measure levels of Flag-FMR1-bound m⁶A. We subjected HEK293T cells with stable expression of Flag-FMR1 to UV cross-linking, after which we carried out immunoprecipitation of FMR1 from whole cell extracts. As negative controls, we carried out Flag pulldowns from Flag and Flag-FMR1-I304N cells (Fig. 4e,f, Supplementary Fig. 4c). Next, we analyzed immunopurified mRNA samples by mass spectrometry to determine the amount of m⁶A in the samples. We observed specific enrichment of m⁶A (more than two-fold) compared with the input in cells that expressed Flag-FMR1 (Fig. 4g). Measurements of m⁶A levels in the immunoprecipitated RNA without RiboMinus treatment did not show any significant differences in m⁶A levels among the three samples (Supplementary Fig. 4d),

which indicates that the enriched m⁶A signal was indeed derived from mRNA. Together, these results show that FMR1 colocalizes with m⁶A on mRNA and preferentially interacts with m⁶A-containing mRNA molecules *in vivo*.

Because FMR1 binds to m⁶A and is known to repress translation by stalling ribosome translocation⁴⁷, we asked what the biological consequence of FMR1 binding to m⁶A-containing mRNA might be. Recent studies demonstrated a positive correlation between m⁶A and translation mediated by YTHDF1 (ref. 22) and YTHDF3 (ref. 27) in HeLa cells. First, to investigate whether YTHDF1 and FMR1 bind to overlapping mRNA targets, we mined published PAR-CLIP data sets for both proteins in HeLa cells^{22,48}. This analysis revealed significant overlap between the two binding profiles, suggesting that a large subset of target mRNAs are shared ($P < 1.259 \times 10^{-38}$; Fig. 4h). We even observed significant overlap between FMR1 targets in mouse brain⁴⁷ and YTHDF1 targets in HeLa cells ($P < 7.120 \times 10^{-112}$) (Supplementary Fig. 4e), which indicated that many of the overlapping targets are cell-type independent. Common targets included HUWE1 and KDM5C, which are also implicated in X-linked mental retardation disorders. GO-term enrichment analysis of common FMR1/YTHDF1 target genes between brain and HeLa cells highlighted biological processes related to neurogenesis (Supplementary Fig. 4f).

To investigate the potential interplay among FMR1, m⁶A and translation rates, we made use of a pulsed-SILAC-based method⁴⁹ (Fig. 4i, Supplementary Data Set 5). Briefly, we grew HeLa cells that expressed either inducible FMR1 or FMR1-I304N in light (Lys0Arg0 (K0R0)) medium. We induced transgene expression for 24 h (Fig. 4j), after which we transferred the induced cells to medium-heavy (M) (K4R6) SILAC medium, whereas we transferred non-induced control cells to heavy (K8R10) SILAC medium. We harvested different batches of cells at 2, 6, 10 and 24 h after transfer to medium-heavy or heavy SILAC medium. The incorporation of SILAC labels during this incubation provided a proxy for the translation rates. Furthermore, comparison of the M/L and H/L ratio for each protein provided a readout for the effect of transgene expression on translation rates (Fig. 4k). The translation rate was defined as the time needed for half of each quantified protein to be labeled with the given SILAC label⁵⁰. At the global level, FMR1 expression resulted in global translational inhibition compared with the expression of FMR1-I304N (Fig. 4l). To investigate the potential interplay between the regulation of translation by FMR1 and m⁶A, we carried out a second pulsed-SILAC experiment in which we combined FMR1 or FMR1-I304N expression with simultaneous METTL3 knockdown (Fig. 4j). As expected, a reduction of m⁶A levels in cells expressing the FMR1-I304N mutant led to decreased translation rates (Fig. 4l). Notably, the negative effect of wild-type FMR1 expression on translation was aggravated by a simultaneous reduction in the amount of m⁶A (Fig. 4l, Supplementary Fig. 4g). We interpret this as follows: YTH-domain-containing proteins are strictly dependent on m⁶A for mRNA binding. Their positive effects on translation are therefore mitigated by METTL3 knockdown and the concomitant reduction of m⁶A levels. FMR1 preferentially binds m⁶A-containing mRNA, but it is not strictly dependent on m⁶A for mRNA binding (Fig. 4a). On the basis of our results, we propose that YTHDF1 and FMR1 compete for binding to m⁶A sites on mRNA. Thus, after METTL3 knockdown, the inhibitory effects of FMR1 on translation rates become even more pronounced, because many putative ACU-containing FMR1-binding sites become available in the absence of YTHDF1/3. These results further illustrate the complex interplay among m⁶A, m⁶A readers and mRNA homeostasis.

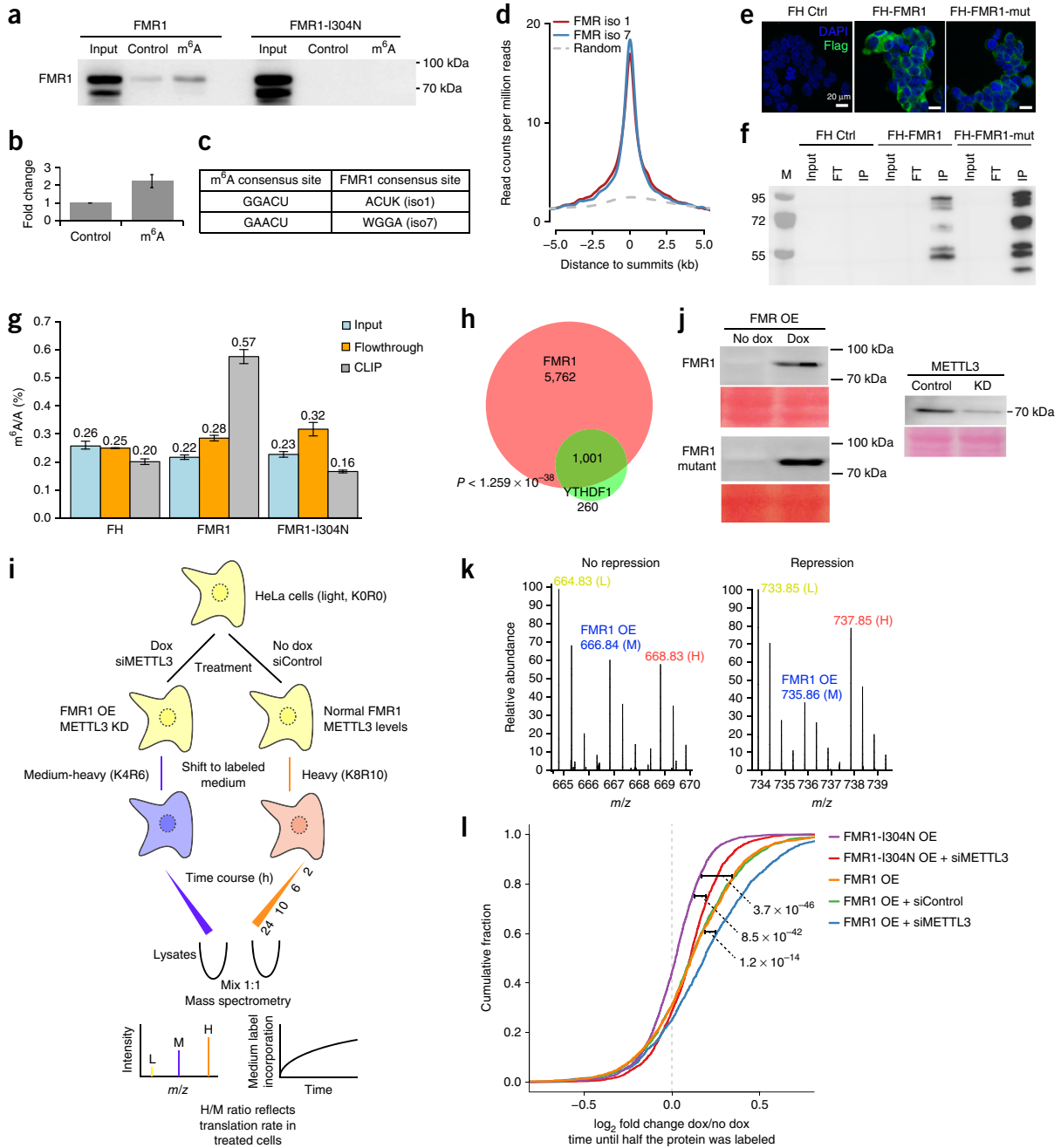


Figure 4 FMR1 preferentially binds to m⁶A-containing mRNA *in vitro* and *in vivo* and affects the translation of its targets. **(a)** Western blots showing preferential m⁶A binding by recombinant GST-FMR1 *in vitro*. **(b)** Quantification of FMR1 binding to control and m⁶A probes (data are shown as mean ± s.e.m.; n = 3 independent experiments). **(c)** Predominant consensus sequences for m⁶A^{9,10} and consensus mRNA-binding sites for FMR1 isoforms 1 and 7 (ref. 48). **(d)** Visualization of binding of FMR1 isoforms 1 and 7 to mRNA relative to known m⁶A sites on mRNA^{10,48}. Randomized peaks were generated from the transcriptome and used as a control as described in the Online Methods. **(e)** Immunofluorescence analysis of the expression of Flag-HA (FH)-tagged FMR1 isoform 1 and Flag-HA-tagged FMR1 isoform 1 I304N mutant (mut) in HEK293T cells. **(f)** RNA cross-linking and immunoprecipitation in lysates from HEK293 cells expressing either Flag-HA-tagged FMR1 or Flag-HA-tagged FMR1-I304N, using anti-Flag. IP, immunoprecipitate; FT, flow-through; M, molecular weight ladder. **(g)** Representative LC-MS quantification showing enrichment of m⁶A in Flag-HA-tagged FMR1-bound mRNA. Error bars represent the range; n = 2 independent experiments. **(h)** The overlap between FMR1 (ref. 48) and YTHDF1 target transcripts²². **(i)** A schematic representation of the pulsed SILAC workflow. Dox, doxycycline. **(j)** Western blots showing the results of METTL3 knockdown and inducible FMR1/FMR1 mutant protein expression in HeLa cells. Ponceau-stained protein bands are shown as loading controls. **(k)** Representative peptide spectra showing SILAC label incorporation for a peptide belonging to a protein whose translation rate is not affected by FMR1 overexpression (left) and a peptide belonging to a protein that is translationally repressed by FMR1 overexpression (right) (compare the incorporation of medium-heavy (M) and heavy (H) label in the two plots). **(l)** A cumulative distribution function plot showing the effect of METTL3 knockdown (siMETTL3), FMR1 overexpression, and FMR1-I304N overexpression on translation rates in HeLa cells. We used the time until half the protein was labeled as a proxy for the translation rate (Online Methods). A summary of the proteomics data from the pulsed SILAC analyses is provided in **Supplementary Data Set 5**. Uncropped blot images for **a**, **f** and **j** are shown in **Supplementary Data Set 1**. Source data for **g** are available online.

DISCUSSION

Here we describe a systematic mass-spectrometry-based proteomics screen to identify m⁶A readers in a variety of cell types and different RNA sequence contexts. Our data show that YTH-domain-containing proteins are evolutionarily conserved cell-type-independent m⁶A readers. In addition, we identified a number of previously unrecognized m⁶A readers, although we emphasize that almost all of these newly identified readers, including FMR1, interact with m⁶A in a manner that is dependent on RNA sequence, and possibly also the secondary structure context. Multiple RNA-binding domains may be responsible for these RNA-sequence-dependent m⁶A interactions, such as the KH and RGG domains. Additional biochemical studies are clearly needed to further investigate this. These observations are analogous to what has been observed for DNA-methylation readers. DNA methylation is read by a few DNA-sequence-independent domains³¹. In addition, other domains, such as the Forkhead box, Krueppel-like zinc fingers and the homeobox, interact with methylated DNA sequences in a DNA-sequence-dependent manner.

One important conclusion from our work is that m⁶A could serve to repel proteins from binding to m⁶A-containing mRNA sequences. This is again analogous to what has been observed in the context of DNA methylation³¹. For RNA sequences that contained ACU motifs, we observed a lot of interacting proteins, and for a substantial number of those, binding was inhibited when the adenosine base in the sequence was methylated. The most strongly repelled proteins in our RNA pulldown experiments were G3BP1 and G3BP2. However, this phenomenon seems to be dependent on the RNA sequence context. In a GGACU context, m⁶A strongly repelled G3BP1 binding, whereas inhibition of binding was not obvious when methylation occurred in a GAACU context (Fig. 2b). The results of our functional experiments for G3BP1 (Figs. 2 and 3) suggest that G3BP1 binding to mRNA is positively correlated with mRNA stability, and that this positive correlation can be affected by manipulation of m⁶A levels. mRNA stability in *cis* is mediated by three major factors: (1) the extent of m⁶A presence, (2) polyadenylation and (3) codon optimization⁵¹. Of these three factors, m⁶A is the only one that can be dynamically altered through methylation and demethylation. The *cis* factors are aided by *trans* factors such as protein readers, which regulate the stability of mRNA. It was previously reported¹⁹ that YTHDF2 predisposes mRNA to degradation in an m⁶A-dependent manner. Our data suggest that G3BP1 is one additional m⁶A-regulated factor that affects RNA stability. G3BP1 could stabilize mRNA molecules in several ways. G3BP1 is found in the nucleus in resting cells⁵², and it could bind to newly transcribed mRNA molecules in the nucleus and prevent methylation by competing with the m⁶A methylation machinery. Alternatively, G3BP1 could bind to its target mRNAs in the cytoplasm and form ribonucleoprotein granules (RNPs). These RNPs could assemble into higher-order stress granules consisting of other proteins and translation factors⁵³. The formation of G3BP1-mRNA RNPs or stress granules could prevent access to the RNA degradation machinery, thereby protecting mRNA. Another possible mechanism is that under certain conditions, cells could demethylate mRNA, thereby allowing G3BP1 to bind to GGACU-containing mRNA sequences, resulting in increased stability of those mRNA molecules. Interestingly, G3BP1 is also known to bind inosine⁵⁴, a modified adenosine base, which indicates a broader role for G3BPs in RNA-modification-dependent modulation of mRNA homeostasis. In any case, our findings reveal intricate interplay between RNA-binding proteins and the regulation of mRNA homeostasis. Dynamic m⁶A levels in cells, which affect many RNA-binding proteins, seem to represent a regulatory switch via which mRNA stability and translation can be quickly fine-tuned to

regulate gene expression and cell function in the presence of cellular stimuli and perturbations.

The work presented here is also relevant from a clinical perspective. It is well established that m⁶A has an important role in the regulation of mRNA homeostasis. It is therefore not surprising that disturbances of readers, writers and erasers of m⁶A are implicated in a number of diseases, including obesity and cancer^{14,15}. We have identified FMR1, an important constituent of neuronal granules and stress granules⁵⁵, as a clinically relevant, RNA-sequence-context-dependent reader for m⁶A. FMR1 binds to ribosomes and represses the translation of target mRNAs. Loss of this repression in neurons due to the absence of functional FMR1 leads to fragile X-linked mental retardation⁴⁷. Our work suggests that a reduction of m⁶A levels or inhibition of YTHDF1, both of which inhibit mRNA translation, may result in the alleviation of disease symptoms in fragile X-linked mental retardation syndrome. Further work, however, is clearly needed to test this hypothesis. In any case, our work here has uncovered what is, to the best of our knowledge, the first reported link between an mRNA modification and an autism spectrum disorder, thus paving the way toward a better molecular understanding, and possibly treatment, of this disease.

METHODS

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

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank members of the Vermeulen lab for fruitful discussions. Work in the Vermeulen lab is supported by the NWO Gravitation program Cancer Genomics Netherlands. Work in the He lab is supported by NHGRI, NIH (HG008688). C.H. is an Investigator of the Howard Hughes Medical Institute. Work in the Carell lab is supported by Deutsche Forschungsgemeinschaft (grants SFB749, SFB1032 and SPP1784) and Bundesministerium für Bildung und Forschung (EXC114).

AUTHOR CONTRIBUTIONS

R.R.E. and M.V. conceived the project and wrote the manuscript, with input from all other authors. R.R.E. performed most wet-lab experiments. S.G. and M.R. prepared the m⁶A probes. R.G.H.L. analyzed RNA-seq and whole cell proteome data. H.S. and P.J.H. performed CLIP and PAR-CLIP experiments. Z.L. analyzed PAR-CLIP data. S.-Y.W. performed bioinformatics analysis of FMR1 PAR-CLIP data. M.P.A.B. provided technical support. P.W.T.C.J. measured mass spectrometry samples. M.M. and H.G.S. provided scientific input. C.H., T.C. and M.V. supervised the project.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Dimock, K. & Stoltzfus, C.M. Sequence specificity of internal methylation in B77 avian sarcoma virus RNA subunits. *Biochemistry* **16**, 471–478 (1977).
2. Desrosiers, R.C., Friderici, K.H. & Rottman, F.M. Characterization of Novikoff hepatoma mRNA methylation and heterogeneity in the methylated 5' terminus. *Biochemistry* **14**, 4367–4374 (1975).
3. Zheng, G. *et al.* ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **49**, 18–29 (2013).
4. Jia, G. *et al.* N⁶-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7**, 885–887 (2011).
5. Liu, J. *et al.* A METTL3-METTL14 complex mediates mammalian nuclear RNA N⁶-adenosine methylation. *Nat. Chem. Biol.* **10**, 93–95 (2014).
6. Ping, X.L. *et al.* Mammalian WTAP is a regulatory subunit of the RNA N⁶-methyladenosine methyltransferase. *Cell Res.* **24**, 177–189 (2014).
7. Schwartz, S. *et al.* Perturbation of m⁶A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Rep.* **8**, 284–296 (2014).
8. Aguiló, F. *et al.* Coordination of m⁶A mRNA methylation and gene transcription by ZFP217 regulates pluripotency and reprogramming. *Cell Stem Cell* **17**, 689–704 (2015).

9. Dominissini, D. *et al.* Topology of the human and mouse m⁶A RNA methylomes revealed by m⁶A-seq. *Nature* **485**, 201–206 (2012).
10. Meyer, K.D. *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).
11. Zhao, X. *et al.* FTO-dependent demethylation of N⁶-methyladenosine regulates mRNA splicing and is required for adipogenesis. *Cell Res.* **24**, 1403–1419 (2014).
12. Zhong, S. *et al.* MTA is an *Arabidopsis* messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell* **20**, 1278–1288 (2008).
13. Agarwala, S.D., Blitzblau, H.G., Hochwagen, A. & Fink, G.R. RNA methylation by the MIS complex regulates a cell fate decision in yeast. *PLoS Genet.* **8**, e1002732 (2012).
14. Lin, S., Choe, J., Du, P., Triboulet, R. & Gregory, R.I. The m⁶A methyltransferase METTL3 promotes translation in human cancer cells. *Mol. Cell* **62**, 335–345 (2016).
15. Li, Z. *et al.* FTO plays an oncogenic role in acute myeloid leukemia as a N⁶-methyladenosine RNA demethylase. *Cancer Cell* **31**, 127–141 (2017).
16. Batista, P.J. *et al.* m⁶A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **15**, 707–719 (2014).
17. Zhao, B.S. *et al.* m⁶A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. *Nature* **542**, 475–478 (2017).
18. Geula, S. *et al.* m⁶A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* **347**, 1002–1006 (2015).
19. Wang, X. *et al.* N⁶-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117–120 (2014).
20. Meyer, K.D. *et al.* 5' UTR m⁶A promotes cap-independent translation. *Cell* **163**, 999–1010 (2015).
21. Zhou, J. *et al.* Dynamic m⁶A mRNA methylation directs translational control of heat shock response. *Nature* **526**, 591–594 (2015).
22. Wang, X. *et al.* N⁶-methyladenosine modulates messenger RNA translation efficiency. *Cell* **161**, 1388–1399 (2015).
23. Alarcón, C.R., Lee, H., Goodarzi, H., Halberg, N. & Tavazoie, S.F. N⁶-methyladenosine marks primary microRNAs for processing. *Nature* **519**, 482–485 (2015).
24. Xiao, W. *et al.* Nuclear m⁶A reader YTHDC1 regulates mRNA splicing. *Mol. Cell* **61**, 507–519 (2016).
25. Patil, D.P. *et al.* m⁶A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537**, 369–373 (2016).
26. Cao, G., Li, H.B., Yin, Z. & Flavell, R.A. Recent advances in dynamic m⁶A RNA modification. *Open Biol.* **6**, 160003 (2016).
27. Shi, H. *et al.* YTHDF3 facilitates translation and decay of N⁶-methyladenosine-modified RNA. *Cell Res.* **27**, 315–328 (2017).
28. Alarcón, C.R. *et al.* HNRNPA2B1 is a mediator of m⁶A-dependent nuclear RNA processing events. *Cell* **162**, 1299–1308 (2015).
29. Liu, N. *et al.* N⁶-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518**, 560–564 (2015).
30. Liu, N. *et al.* N⁶-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. *Nucleic Acids Res.* **45**, 6051–6063 (2017).
31. Spruijt, C.G. & Vermeulen, M. DNA methylation: old dog, new tricks? *Nat. Struct. Mol. Biol.* **21**, 949–954 (2014).
32. Penagarikano, O., Mulle, J.G. & Warren, S.T. The pathophysiology of fragile X syndrome. *Annu. Rev. Genomics Hum. Genet.* **8**, 109–129 (2007).
33. Matsuki, H. *et al.* Both G3BP1 and G3BP2 contribute to stress granule formation. *Genes Cells* **18**, 135–146 (2013).
34. Tourrière, H. *et al.* The RasGAP-associated endoribonuclease G3BP assembles stress granules. *J. Cell Biol.* **160**, 823–831 (2003).
35. Kedersha, N. *et al.* G3BP-Caprin1-USP10 complexes mediate stress granule condensation and associate with 40S subunits. *J. Cell Biol.* **212**, 845–860 (2016).
36. Pendleton, K.E. *et al.* The U6 snRNA m⁶A methyltransferase METTL16 regulates SAM synthetase intron retention. *Cell* **169**, 824–835 (2017).
37. Castello, A. *et al.* Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406 (2012).
38. Baltz, A.G. *et al.* The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **46**, 674–690 (2012).
39. Kwon, S.C. *et al.* The RNA-binding protein repertoire of embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1122–1130 (2013).
40. Irvine, K., Stirling, R., Hume, D. & Kennedy, D. Rasputin, more promiscuous than ever: a review of G3BP. *Int. J. Dev. Biol.* **48**, 1065–1077 (2004).
41. Wei, C.M. & Moss, B. Nucleotide sequences at the N⁶-methyladenosine sites of HeLa cell messenger ribonucleic acid. *Biochemistry* **16**, 1672–1676 (1977).
42. Ke, S. *et al.* A majority of m⁶A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes Dev.* **29**, 2037–2053 (2015).
43. Linder, B. *et al.* Single-nucleotide-resolution mapping of m⁶A and m⁶Am throughout the transcriptome. *Nat. Methods* **12**, 767–772 (2015).
44. Tani, H. *et al.* Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res.* **22**, 947–956 (2012).
45. Myrick, L.K., Hashimoto, H., Cheng, X. & Warren, S.T. Human FMRP contains an integral tandem Agenet (Tudor) and KH motif in the amino terminal domain. *Hum. Mol. Genet.* **24**, 1733–1740 (2015).
46. Brown, V. *et al.* Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell* **107**, 477–487 (2001).
47. Darnell, J.C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011).
48. Ascano, M. Jr. *et al.* FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* **492**, 382–386 (2012).
49. Schwanhäusser, B., Gossen, M., Dittmar, G. & Selbach, M. Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* **9**, 205–209 (2009).
50. Visscher, M. *et al.* Proteome-wide changes in protein turnover rates in *C. elegans* models of longevity and age-related disease. *Cell Rep.* **16**, 3041–3051 (2016).
51. Chen, C.A. & Shyu, A.B. Emerging themes in regulation of global mRNA turnover in cis. *Trends Biochem. Sci.* **42**, 16–27 (2017).
52. Costa, M., Ochem, A., Staub, A. & Falaschi, A. Human DNA helicase VIII: a DNA and RNA helicase corresponding to the G3BP protein, an element of the ras transduction pathway. *Nucleic Acids Res.* **27**, 817–821 (1999).
53. Anderson, P. & Kedersha, N. Stress granules: the Tao of RNA triage. *Trends Biochem. Sci.* **33**, 141–150 (2008).
54. Scadden, A.D. Inosine-containing dsRNA binds a stress-granule-like complex and downregulates gene expression in trans. *Mol. Cell* **28**, 491–500 (2007).
55. Kim, S.H., Dong, W.K., Weiler, I.J. & Greenough, W.T. Fragile X mental retardation protein shifts between polyribosomes and stress granules after neuronal injury by arsenite stress or in vivo hippocampal electrode insertion. *J. Neurosci.* **26**, 2413–2418 (2006).

ONLINE METHODS

Cell culture, SILAC labeling and cell-extract preparation. HeLa, HEK293T and NIH 3T3 cells were cultured in Dulbecco's modified Eagle's medium (DMEM; Thermo) supplemented with 10% FBS (Gibco), 1× penicillin–streptomycin (15140-122; Thermo) and 1 mM glutamine (Invitrogen). The cells were regularly tested in-house for the presence of mycoplasma. HeLa cells were SILAC-labeled by culture in SILAC DMEM (88420; Thermo) supplemented with 10% dialyzed FBS (DS-1003; DundeeCell), 1× Glutamax (35050-061; Thermo), 1× penicillin–streptomycin, 73 mg/mL L-lysine (light/K0, Sigma, A6969; or heavy/K8, Sigma, 608041) and 29.4 mg/mL arginine (light/R0, Sigma, A6969; or heavy/R10, Sigma, 608033). ESCs were grown in serum-free 2i medium. Briefly, we prepared serum-free ESC medium by mixing the following components: Neurobasal medium (Gibco; 21103-049), 250 mL; DMEM/F12 (Gibco; 11320-033), 250 mL; N2-Supplement (Gibco; 17502-048), 2.5 mL; B27 minus RA (Gibco; 12587001), 5 mL; 7.5% BSA (Gibco; 15260-037), 3.33 mL; penicillin–streptomycin, 5 mL; 1 mM glutamine (Invitrogen); 1% nonessential amino acids (Invitrogen); 6.3 μl monothioglycerol (M6145; Sigma); LIF (1,000 U/mL); and CHIR99021 and PD0325901 (3 and 1 μM, respectively). For SILAC labeling of ESCs, we used custom-made Neurobasal and DMEM/F12 (Life technologies) without lysine and arginine. Mouse neuronal progenitor cells were cultured as described previously⁵⁶.

Whole cell lysates for RNA pulldowns were prepared according to the following protocol: Cells were harvested with trypsin and washed twice with chilled PBS. Cells were then resuspended in five pellet volumes of whole cell extract buffer (50 mM Tris-HCl, pH 8, 150 mM NaCl, 1% NP-40, 0.5 mM DTT, 10% glycerol, and protease inhibitor cocktail) and incubated for 90 min at 4 °C on a rotating wheel. Crude lysates were then centrifuged for 30 min at 14,000 r.p.m. at 4 °C. Soluble whole cell extracts were finally aliquoted and snap-frozen in liquid nitrogen until further usage. Nuclear and cytoplasmic extracts were prepared as described previously⁵⁶. Cell lysates for western blots were prepared with RIPA buffer (10 mM Tris-Cl, pH 8.0, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 150 mM NaCl, 0.1% SDS and protease inhibitor cocktail). Protein quantification was performed with a Bradford or BCA kit (Thermo) according to the manufacturer's protocol.

Plasmids and constructs. G3BP1 cDNA was cloned into the *Hind*III and *Bam*HI sites of pcDNA3-c-Flag vector (Plasmid #20011; Addgene) to create an expression construct with a C-terminal Flag tag. For the creation of N-terminally GST-tagged constructs, G3BP1 cDNA was amplified from HeLa cDNA and cloned into pGEX-5X1 vector (GE Healthcare). FMR1 isoform 1 (pFRT-TODestFLAGHAhFMRPiso1; 48690) and FMR1-I304N isoform 1 (pFRT-TODestFLAGHAhFMRPiso1I304N; 48692) vectors were obtained from Addgene. FMR1, FMR1-I304N and various domains of FMR1 were amplified from these vectors. We created FMR1 overexpression constructs for the creation of doxycycline-inducible HeLa cell lines by subcloning FMR1 and FMR1-I304N in AAVS1-TRE3G-EGFP (Addgene; 52343). A Kozac sequence and a Flag-HA tag were added to the N terminus of FMR1 and FMR1-I304N by PCR. AAVS1-TRE3G-EGFP⁵⁷ contains an inducible promoter and AAVS1 homology arms for correct integration. By cloning FMR1 in *Sall* and *Mlu*I sites, we generated an AAVS1-TRE3G-FMR1 construct. Primer sequences used to amplify specific genes and constructs are listed in **Supplementary Table 1**.

Generation of cell lines. We generated HeLa cells expressing inducible FMR1 and FMR1-I304N by knocking Flag-HA-FMR1 or the I304N mutant into the AAVS1 locus. To this end, HeLa cells were transfected with a donor vector (AAVS1-TRE3G-FMR1) and a CRISPR-Cas9 vector with guide RNAs targeting the AAVS1 locus⁵⁸, using Lipofectamine 3000 (Invitrogen). In this system, cells with correct integration at the AAVS1 locus become puromycin resistant. Cells with stably integrated gene products were selected by puromycin (1 μg/mL) selection for several days. Individual colonies were picked, and doxycycline-induced expression was tested for several clones. Overexpression of FMR1 and FMR1-I304N was induced by the addition of 1 and 0.5 μg/mL doxycycline, respectively.

Recombinant protein expression and purification. Rosetta BL-21 bacteria (Novagen) were used for GST-fusion protein expression. Cells were grown in LB medium supplemented with 100 μg/mL ampicillin and 34 μg/mL chloramphenicol. IPTG was added to a final concentration of 0.5 mM, after which

protein expression was induced for 4 h or overnight at 16 °C. Cells were then harvested, washed with PBS and resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 15% glycerol, 1 mM EDTA, 0.5 mM PMSF, 1 mM DTT, 100 mM NaCl, 1% Triton X-100, 0.25% NP-40, protease inhibitor cocktail, and lysozyme to a final concentration of 0.5 mg/ml). Cells were lysed by repeated (five times) freeze-thaw cycles. Bacterial debris was removed by centrifugation at 14,000 r.p.m. for 20 min at 4 °C, after which soluble extracts were aliquoted and snap-frozen in liquid nitrogen until further usage.

SILAC-based RNA pulldowns. All buffer solutions were prepared with high-quality, RNase-free reagents. Each SILAC-based RNA pulldown experiment consisted of four individual pulldowns. 10 μl of streptavidin Sepharose high-performance beads (GE Healthcare) were used for each pulldown (20 μl of 50% slurry). Beads were first washed twice in 1 mL of RNA binding buffer (50 mM HEPES-HCl, pH 7.5, 150 mM NaCl, 0.5% NP40 (v/v), 10 mM MgCl₂). After each wash step, samples were centrifuged for 3 min at 4,000 r.p.m. in a pre-cooled (4 °C) tabletop centrifuge to spin down the beads. To inactivate and remove RNases, we incubated the beads with RNase inhibitor RNasin plus (Promega) in RNA binding buffer (100 μl buffer with 0.8 units of RNasin/μl) for 30 min on ice. After centrifugation and removal of RNasin–RNase complexes, beads were preblocked with yeast tRNA (50 μg/mL; AM7119; Life Technologies) in RNA binding buffer overnight at 4 °C on a rotation wheel. The preblocked beads were washed twice with RNA binding buffer and then incubated with 5 μg of biotinylated RNA probe (per pulldown) diluted with RNA binding buffer to a final volume of 600 μl. Beads were incubated for 30 min at 4 °C in a rotation wheel to allow binding of biotinylated probes to the streptavidin beads. The beads were washed once with 1 mL of RNA wash buffer (50 mM HEPES-HCl, pH 7.5, 250 mM NaCl, 0.5% NP-40 and 10 mM MgCl₂) and twice with protein incubation buffer (10 mM Tris-HCl, pH 7.5, 150 mM KCl, 1.5 mM MgCl₂, 0.1% (v/v) NP-40, 0.5 mM DTT, and complete protease inhibitors –/– EDTA (Roche)). Beads containing immobilized RNA were then incubated with 600 μg of nuclear extract or 1,000 μg of cytoplasmic extracts in a total volume of 600 μl of protein binding buffer. The incubation reaction also contained 30 μg of yeast tRNA to prevent nonspecific binding, and RNasin. In the forward experiment, the control probe was incubated with light-labeled (R0K0) lysates, whereas the m⁶A probe was incubated with heavy (R10K8) lysates. The reverse experiment represented a biological-replicate label swap. The reactions were incubated at room temperature for 30 min and then for 90 min on a rotation wheel at 4 °C. The beads were then washed three times with protein incubation buffer and twice with ice-cold PBS to remove detergent from the beads. The last PBS wash was used to combine beads from the forward and the reverse pulldowns as follows:

Forward reaction: control probe, R0K0 (light); m⁶A probe, R10K8 (heavy)
Reverse reaction: control probe, R10K8 (heavy); m⁶A probe, R0K0 (light)

Proteins were on-bead digested with trypsin. Briefly, beads were resuspended in 100 μL of elution buffer (50 mM Tris, pH 8.5, 2 M urea and 10 mM DTT) and then incubated for 20 min at room temperature in a thermoshaker at 1,100 r.p.m. Iodoacetamide was then added to a final concentration of 55 mM, and the mixture was incubated for 10 min in a thermoshaker (1,100 r.p.m.) at room temperature in the dark. Proteins were then partially digested from the beads by the addition of 250 ng of trypsin for 2 h at room temperature in a thermoshaker in the dark. After incubation, the supernatant was collected in a separate tube. The beads were then incubated with 50 μL of elution buffer for 5 min at room temperature in a thermoshaker (1,100 r.p.m.). 100 ng of fresh trypsin was added to the pooled eluates, and proteins were digested overnight at room temperature. Finally, tryptic peptides were acidified to pH <2 with TFA (10%) and desalted with C18 Stage tips before MS analyses.

Dimethyl-labeling-based RNA pulldowns. Dimethyl-labeling-based RNA pulldowns were performed with unlabeled lysates. Tryptic peptides from individual pulldowns obtained after on-bead digestion were differentially labeled with dimethyl isotopes (CH₂O or CD₂O) essentially as described⁵⁹. We created forward and reverse reactions by mixing labeled peptides as follows:

Forward reaction: control probe, CH₂O (light); m⁶A probe, CD₂O (medium)
Reverse reaction: control probe, CD₂O (medium); m⁶A probe, CH₂O (light)

Labeled reactions were acidified to pH <2 with TFA (10%) and desalted with C18 Stage tips before MS analyses. RNA probes used in SILAC and dimethyl-labeling-based pulldowns are listed in **Supplementary Table 1**.

siRNA knockdown. siRNAs targeting human G3BP1 (SR306845) or METTL3 (SR311159) and control siRNAs (SR30004) were obtained from Origene. siRNAs were transfected at a final concentration of 25 nM using Lipofectamine 3000 (Invitrogen) according to the manufacturer's instructions.

Quantitative RT-PCR. qRT-PCR reactions were done with iQ SYBR Green Supermix. Oligonucleotides used in this study are listed in **Supplementary Table 1**.

Western blot analyses and antibodies. SDS-PAGE gels were blotted onto nitrocellulose membrane with the Trans-Blot Turbo transfer system (Bio-Rad). Blots were stained with Ponceau stain to confirm equal loading of proteins. Blots were blocked in 5% skim milk in PBS. Primary antibodies to the following proteins were used: G3BP1 (Santa Cruz Biotechnology; sc-98561), Flag (Sigma-Aldrich; F1804), GST (Thermo; MA4-004) and METTL3 (Bethyl Laboratories; A301-567-A). Images were processed in ImageJ. Western blot quantifications were done with ImageJ.

FMR1 CLIP. 2.4×10^8 HEK293T cells stably expressing Flag-HA, Flag-HA-FMR1, or Flag-HA-FMR1-I304N were UV-cross-linked twice with 150 mJ/cm² at 254 nm and harvested with a cell scraper. Cells were lysed in two volumes of lysis buffer containing 150 mM KCl, 2 mM EDTA, 10 mM HEPES, 0.5% NP-40, 0.5 mM DTT, 1:100 cOmplete protease inhibitor (Roche; 04693116001), and 1:100 SUPERase In RNase inhibitor (20 U/μL; Ambion; AM2694) for 30 min in a rotation wheel at 4 °C. Lysate was cleared by centrifugation at 14,000 r.p.m. for 15 min at 4 °C followed by syringe filtration through a 0.45-μm filter. 50 μL of lysate was saved as input. 200 μL of anti-Flag magnetic beads (Sigma; M8823) were washed with 500 μL of ice-cold wash buffer four times (200 mM NaCl, 2 mM EDTA, 0.05% NP-40, 50 mM Tris-HCl, pH 7.6, 0.5 mM DTT, 1:1,000 SUPERase In RNase inhibitor, 1:1,000 protease inhibitor). Lysate was added to the beads and incubated for 2 h on a rotation wheel at 4 °C. Supernatant was removed and saved as flow-through, and beads were washed six times with 1 mL of ice-cold wash buffer. 250 μL of elution solution (500 ng μl⁻¹ 3× Flag peptide (Sigma; F4799) in wash buffer) was added to each sample, and the mixture was rotated for 1 h at 4 °C to elute. We made proteinase K solution by adding 10% proteinase K solution (20 mg/mL; Invitrogen; 25530049) to 2× proteinase K buffer (100 mM Tris, pH 7.7, 150 mM NaCl, 12.5 mM EDTA, 2% (w/v) SDS). The eluent, 50 μL of input, and 50 μL of flow-through were each incubated in one volume of proteinase K solution for 45 min at 55 °C to release RNA. Three volumes of TRIzol were added to each sample, and total RNA was extracted on a Direct-zol microprep spin column according to the manufacturer's protocol for RNA extraction and DNase I digestion. rRNA was removed with the RiboMinus Eukaryote Kit v2. 50 ng of input, flow-through, and immunoprecipitated mRNA were analyzed by LC-MS/MS as previously reported¹⁹.

PAR-CLIP. PAR-CLIP was performed as previously reported¹⁹. Briefly, five 15-cm plates of HEK293T cells were seeded on day 1. After 18 h, the cells were transfected with Flag-tagged G3BP1 or G3BP2 plasmid at 80% confluency. After another 6 h, the media was changed and 200 μM 4SU was added. After 14 h, the media was aspirated and the cells were washed twice with PBS for each plate. The plates were kept on ice, and cross-linking was carried out with 0.15 J/cm² 365-nm UV light. 2 ml of PBS was added, and the cells were collected with a cell lifter. RNA was purified as previously reported¹⁹. The purified bound RNA was dissolved in 12 μL of RNase-free water, of which 6 μL was subjected to small RNA library preparation with the NEBNext Multiplex Small RNA Library Prep Set for Illumina (E7300S; NEB). Two biological replicates were carried out for G3BP1 and G3BP2 each. All samples were sequenced on an Illumina HiSeq 2000 with single-end 50-bp read length. For data analysis, the adaptors were trimmed with the FASTX-Toolkit⁶⁰. Deep-sequencing data were mapped to human genome version hg38 with TopHat version 2.0 (ref. 61), without any gaps and allowing for at most two mismatches. PAR-CLIP data were analyzed by PARalyzer v1.1 with default settings⁶². HOMER⁶³ was used for *de novo* motif discovery. For the comparison of PAR-CLIP and m⁶A peaks from Liu *et al.*⁵ (**Fig. 2i,j**), at least 1-bp

overlap was applied as the criterion for defining overlapping peaks. To assess the distribution of m⁶A with respect to PAR-CLIP peaks, we used publicly available single-nucleotide-resolution m⁶A data from Linder *et al.*⁴³ and converted the reported m⁶A sites to hg38 with liftOver⁶⁴. For FMR1 and m⁶A colocalization analysis (**Fig. 4d**), we generated random peaks by randomly permuting the location of m⁶A peaks on the transcriptome.

GO analysis. GO-term enrichment analysis was performed with DAVID⁶⁵. The top 10–15 most enriched GO terms with a Bonferroni or Benjamini–Hochberg corrected *P* value of <0.05 were used. To reduce redundancy in the GO terms involved and summarize them, we used the online tool REVIGO⁶⁶. We used Cytoscape⁶⁷ to make figures from REVIGO output.

RNA half-life determination. RNA half-life was determined by RNA-sequencing as described previously¹⁹. Briefly, six 10-cm plates were seeded with HeLa cells at 25% confluency. After 8 h, two plates were transfected with siControl, two with siG3BP1, and two with siG3BP1 + siMETTL3. For overexpression experiments, a Flag-control plasmid and a Flag-G3BP1 construct were transfected into two 10-cm plates of HeLa cells each. 24 h after transfection, cells were trypsinized and live cells were counted. Cells were then equally divided over three fresh plates. 48 h after transfection, actinomycin D was added to a final concentration of 5 μg/mL (Santa Cruz Biotechnology; sc-200906). Cells were then harvested 0, 3 and 6 h after the addition of actinomycin D, respectively. The harvested cells were counted once again and snap-frozen. RNA was purified with the Qiagen RNeasy mini kit. rRNA was removed with the Ribo-Zero rRNA removal kit (Illumina). RNA fragmentation was performed with the NEBNext Magnesium RNA Fragmentation Module (NEB). The first cDNA strand was synthesized with Superscript III in the presence of actinomycin D and random hexamers. The second strand was synthesized with *Escherichia coli* DNA polymerase in the presence of dUTP. USER enzyme was added before library amplification to maintain strand identity. NEXTflex 8-bp adaptors were used for adaptor ligation. Size selection was performed with E-gel, and quality was assessed on a Bioanalyzer.

RNA-seq libraries were quantified with kallisto version 0.43.0 (ref. 68) by pseudo-alignment to the RefSeq transcript database downloaded 28 June 2016. Absolute transcript concentrations per cell were computed by linear-fitting the RNA spike-in TPM values to attomoles. RNA half-life measures were calculated as described by Wang *et al.*¹⁹. In short, for every transcript, the RNA-degradation constant was determined as the log₂ fold change in RNA abundance 3 or 6 h after transcription was inhibited, and the average degradation constant was used to calculate the RNA half-life per transcript. To ensure accurate RNA half-life estimations, we considered only transcripts with finite RNA half-lives of <24 h for further analyses. The RNA half-life measures in **Figure 3a–c** were taken from Tani *et al.*⁴⁴.

Mass spectrometry. Tryptic peptides were separated with an Easy-nLC 1000 (Thermo) connected online to a Q Exactive or Orbitrap-Fusion mass spectrometer (Thermo). RNA pulldown samples were separated with a 94-min acetonitrile gradient (7% to 32%) followed by washes at 50% and then 95% acetonitrile for a total of 120 min of data acquisition. For measurements on the Q Exactive, the top ten most abundant peptides were fragmented for every full scan. For RNA pulldowns measured on the LTQ-Orbitrap Fusion Tribrid mass spectrometer (Thermo), a similar LC gradient was used, and samples were measured in top-speed mode. For the whole-cell proteome experiment measured on the LTQ-Orbitrap Fusion Tribrid mass spectrometer (Thermo), peptides were separated with a 214-min acetonitrile gradient (7% to 30%) followed by washes at 60% and then 95% acetonitrile for a total of 240 min of data acquisition. Scans were collected in data-dependent top-speed mode in cycles of 3 s with dynamic exclusion enabled and set to 60 s.

MaxQuant analysis. Data processing was done with the MaxQuant software. For SILAC-labeling-based experiments, raw data were analyzed with MaxQuant version 1.5.0.1 with the default settings and K0R0, K4R6 or K8R10 (light, medium-heavy or heavy) labels, and matching between runs enabled⁶⁹. For dimethyl-labeling-based experiments, raw data were analyzed with DimethylLys0 and DimethylLys4 (light and medium) labels, and matching between runs was enabled. We used the UniProt database downloaded on 13 July 2014 for identification. To identify significant interactors, we plotted normalized ratios from

MaxQuant output tables for the forward and reverse pulldowns. Outliers were identified independently in the forward and reverse pulldowns by means of box plot statistics (threshold: 1.5× the interquartile range). Proteins were considered significant if they were identified as outliers in both experiments. Proteins identified in MaxQuant output tables were classified into four groups: readers, repelled proteins, background proteins, and contaminants. For whole-cell proteome experiments, we used a Swiss-Prot curated database for identification (downloaded on 7 February 2016). Scatter plots were generated in R.

Pulsed SILAC. Pulsed SILAC is a variant SILAC method used to measure translation rates in cells⁴⁹. HeLa cells expressing FMR1 under doxycycline control were SILAC-labeled in light medium (K0R0) for a week to allow them to adapt to the culture conditions. Doxycycline was added for 24 h to induce expression of FMR1 or the FMR1 mutant. Cells without doxycycline induction served as a control. In the case of FMR1 overexpression combined with METTL3 knockdown, METTL3 siRNAs were transfected 24 h before the addition of doxycycline. Control siRNA was transfected into control cells that were not treated with doxycycline. 24 h after doxycycline induction, cells were washed with PBS to remove any light amino acid-containing medium. Cells with FMR1 overexpression alone or in combination with METTL3 knockdown were transferred to medium-heavy (K4R6) medium. Non-doxycycline-treated control cells were transferred to heavy (K8R10) medium. Cells were subsequently harvested 2, 6, 10 and 24 h after transfer into medium-heavy or heavy SILAC medium. The harvested cells were then lysed on ice by the addition of 150 µl of lysis buffer (100 mM Tris-HCl, pH 7.6, 4% SDS) and heated for 5 min at 95 °C. The samples were sonicated with a Diagenode Bioruptor to reduce viscosity. Crude lysates were then centrifuged at 14,000 r.p.m. for 20 min, and supernatants were recovered. Lysate protein concentrations were measured with the BCA kit (Thermo). Equal amounts of protein from control and treated samples at each time point were then mixed. DTT was added to a 100 mM final concentration, and the samples were incubated for 15 min. Samples were then further processed via the FASP method⁷⁰. A trypsin-LysC mix (Promega) was used for digestion. Finally, tryptic peptides were acidified to pH <2 with TFA (10%) and desalted with C18 Stage tips before MS analyses.

The time it took for half the protein to be labeled (THPL) was calculated for all identified proteins as described by Visscher *et al.*⁵⁰. In short, we performed a linear regression between the log-transformed fraction of labeled protein and the pulse time. We then extrapolated the THPL by dividing the slope of the regression by log(0.5). In addition to the four pulsed time points, we forced the regression lines through zero as described previously⁵⁰. To ensure high confidence of our individual THPL measurements, we used only the proteins with an identified label ratio in all time points and a Pearson correlation coefficient of 0.95 or higher for downstream analyses.

Statistical analysis. Various statistical methods were used as appropriate and are indicated in the figure legends. The interquartile range (1.5×) was used to identify outliers in affinity-purification mass spectrometry experiments. The *P* values of the

identified RBP motifs were determined by HOMER⁶³. Significantly enriched gene ontologies were identified with a modified Fisher exact test, after multiple testing correction by the Bonferroni or Benjamini–Hochberg method. A Fisher exact test was used to analyze overlap between PAR-CLIP peaks. Two-sided nonparametric Mann–Whitney *U* tests were used to test for differences in protein half-lives and in mRNA half-lives.

A Life Sciences Reporting Summary for this article is available.

Data availability. All mass spectrometric data generated in this study have been deposited to the ProteomeXchange Consortium via the PRIDE repository under project identifier [PXD006498](https://doi.org/10.60027/201701000006498). RNA-sequencing and PAR-CLIP data have been deposited at Gene Expression Omnibus (GEO) under accession code [GSE98856](https://doi.org/10.1101/109856). Uncropped images of western blots, agarose gels and PAR-CLIP images are in **Supplementary Data Set 1**. Source data for **Figures 1g,h, 2i,j** and **4g** are available online. Any other data are available from the corresponding author upon request.

56. Spruijt, C.G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
57. Qian, K. *et al.* A simple and efficient system for regulating gene expression in human pluripotent stem cells and derivatives. *Stem Cells* **32**, 1230–1238 (2014).
58. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
59. Boersema, P.J., Raijmakers, R., Lemeer, S., Mohammed, S. & Heck, A.J. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.* **4**, 484–494 (2009).
60. Pearson, W.R., Wood, T., Zhang, Z. & Miller, W. Comparison of DNA sequences with protein sequences. *Genomics* **46**, 24–36 (1997).
61. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
62. Corcoran, D.L. *et al.* PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol.* **12**, R79 (2011).
63. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
64. Hinrichs, A.S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).
65. Huang, W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
66. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).
67. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
68. Bray, N.L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
69. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
70. Wiśniewski, J.R., Zougman, A. & Mann, M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J. Proteome Res.* **8**, 5674–5678 (2009).

6 Cytosin-Modifikationen der DNA: Funktionsstudien und Verbesserung der Sequenzierung

Neben den etablierteren Funktionen von mdC, mehren sich Hinweise auf epigenetische Eigenschaften von hmdC, fdC und cadC.^[154] Selbst hemi-modifizierte Positionen besitzen womöglich eigenständige epigenetische Funktionen.^[171,175,176,178] In diesem Abschnitt werden Beiträge zur Aufklärung dieses dynamischen Feldes beschrieben: In Abschnitt 6.1 wird eine im Rahmen dieser Arbeit publizierte Studie zur Funktion hemi-modifizierter CpG-Dyaden vorgestellt. Ein zentraler Zugang zu Funktionen der Cytosin-Modifikationen stellt deren exakte quantitative Sequenzierung dar. Um diese zu ermöglichen, wurden darüber hinaus drei unabhängige Methoden zur Verbesserung der Sequenzierung von mdC, hmdC, fdC und cadC entwickelt. In Abschnitt 6.2 werden diese Methoden vorgestellt.

6.1 *in vitro*-Funktionsänderung eines Promotors, durch hmdC, fdC oder cadC im hemi-CpG-Kontext

„Functional impacts of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine at a single hemi-modified CpG dinucleotide in a gene promoter” N. Kitsera, J. Allgayer, E. Parsa, N. Geier, M. Rossa, T. Carell, A. Khobta, *Nucleic Acids Res.* **2017**, *45*, 11033–11042.

Prolog

Diese Arbeit sollte Licht auf die komplexe Genregulation durch mdC, hmdC, fdC und cadC im hemi-CpG-Kontext werfen. Dazu wurde ein Genexpressions-Konstrukt hergestellt, welches eine hemi-CpG-Position in einem eigens entworfenen Promotor enthält. Durch die Modifikation dieser Position mit mdC, hmdC, fdC oder cadC konnte deren Einfluss auf die Expression eines Reportergens quantitativ untersucht werden. So ermöglicht unser System einen normierten Vergleich beliebiger CpG-Modifikationen.

Wir konnten zeigen, dass die dC-Modifikationen mdC, hmdC, fdC und cadC, im hemi-CpG-Kontext eines Promotors die Genexpression reduzieren. Dies erfolgt über die direkte Beeinflussung der Transkriptionsfaktor-Bindung. Im Falle von mdC und hmdC blieb die Abnahme der Genexpression konstant und war proportional zur Reduktion der Transkriptionsfaktor-Bindung. Hingegen führte der Einbau von fdC und cadC zu einer fortschreitenden Abnahme der Genexpression, bis zum nahezu vollständigen Stillstand. Wir zeigten, dass dieser Effekt ausschließlich bei aktivem TDG auftritt. Durch das Ausschalten von TDG, welches fdC und cadC aus der DNA entfernt, findet hingegen keine weitere Reduktion der Genexpression statt.

Autorenbeitrag

Für diese Studie habe ich die Synthese der Phosphoramidite von 2'-F-fdC sowie 2'-F-cadC übernommen. Anschließend habe ich diese Phosphoramidite zur Festphasensynthese der entsprechenden DNA-Oligonukleotide eingesetzt. Die so synthetisierten Oligonukleotide wurden anschließend von mir mittels Umkehrphasen-HPLC aufgereinigt und via MALDI-TOF-Massenspektrometrie charakterisiert.

Lizenz

Kopie der Publikation mit Erlaubnis des Verlags

Functional impacts of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine at a single hemi-modified CpG dinucleotide in a gene promoter

Nataliya Kitsera¹, Julia Allgayer², Edris Parsa³, Nadine Geier¹, Martin Rossa³, Thomas Carell³ and Andriy Khobta^{1,2,*}

¹Institute of Toxicology, University Medical Center of the Johannes Gutenberg University Mainz, Mainz 55131, Germany, ²Institute of Pharmacy and Biochemistry, Johannes Gutenberg University of Mainz, Mainz 55128, Germany and ³Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians-Universität München, Munich 81377, Germany

Received February 28, 2017; Revised July 21, 2017; Editorial Decision August 05, 2017; Accepted August 08, 2017

ABSTRACT

Enzymatic oxidation of 5-methylcytosine (5-mC) in the CpG dinucleotides to 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC) and 5-carboxycytosine (5-caC) has central role in the process of active DNA demethylation and epigenetic reprogramming in mammals. However, it is not known whether the 5-mC oxidation products have autonomous epigenetic or regulatory functions in the genome. We used an artificial upstream promoter constituted of one cAMP response element (CRE) to measure the impact of 5-mC in a hemi-methylated CpG on the promoter activity and further explored the consequences of 5-hmC, 5-fC, and 5-caC in the same system. All modifications induced mild impairment of the CREB transcription factor binding to the consensus 5'-TGACGTCA-3' CRE sequence. The decrease of the gene expression by 5-mC or 5-hmC was proportional to the impairment of CREB binding and had a steady character over at least 48 h. In contrast, promoters containing single 5-fC or 5-caC underwent further progressive loss of activity, up to an almost complete repression. This decline was dependent on the thymine-DNA glycosylase (TDG). The results thus indicate that 5-fC and 5-caC can provide a signal for perpetuation and enhancement of the repressed transcriptional state by a mechanism that requires base excision repair.

INTRODUCTION

DNA methylation at cytosine residues plays essential roles in transcriptional regulation and genome maintenance in

vertebrates (1–3). Most abundant and best studied context in which 5-methylcytosine is present in the genomes is symmetric methylation of CpG dinucleotides. CpG methylation can be copied to daughter DNA strands during replication and thus constitutes an epigenetic mark propagated through cell division (4,5). CpG methylation in the promoter regions regulates transcriptional activity of genes by at least two mechanisms. One is direct modulation of binding of the methylation-sensitive transcription factors to the regulatory motifs containing methylated CGs (6–8). Another is reinforcement and long-term propagation of chromatin-mediated gene silencing by recruitment of the methyl-CpG-binding domain (MBD) proteins and, in turn, histone deacetylases (1,2). Both these modes of action appear to be important for silencing and transcriptional repression of genes whose promoters contain long-range arrays rich in CpG dinucleotides, known as CpG islands (CGIs). In the non-CGI promoters, functions of CpG methylation are even more complex and the outcomes for transcriptional activity may be very variable for genes controlled by different transcription factors and in different cell lineages (9–11). DNA demethylation mechanisms add another complexity level to the regulation of gene transcription by erasing 5-methylcytosine or converting it to other types of base modifications whose regulatory functions still have to be explored (12–15). Thus, currently the best characterized DNA demethylation pathway in vertebrates mediated by the ten eleven translocation (TET) family oxygenases (16–21) generates 5-hmC, 5-fC and 5-caC, which may have independent functions in the regulation of gene expression (13,22). However, understanding of functional consequences of cytosine demethylation products at specific regulatory elements is hindered by the extremely low levels of these modifications in the genome (17,19,23) and by the dynamic character of 5-fC and 5-caC which are efficiently removed by DNA repair mechanisms (15,18,24).

*To whom correspondence should be addressed. Tel: +49 6131 17 9271; Fax: +49 6131 17 8499; Email: khobta@uni-mainz.de

Transitions between the methylated and unmethylated states of CpG dinucleotides require hemi-modified states in which 5-mC or its oxidation products are present in only one DNA strand (25). Functional impact of such modifications on the regulation of gene expression is unknown because deliberate hemi-modification of chosen cytosine residues in DNA cannot be generated enzymatically. To overcome this hurdle, we have previously proposed an efficient procedure for a site-specific incorporation of synthetic oligonucleotides containing various base modifications into plasmid-borne reporter genes with the help of sequence-specific nicking endonucleases (26). With these tools, we have already obtained important insights into the mechanisms of excision repair of several types of structurally defined DNA lesions in functional reporter genes (27–29). Here, we applied an analogous approach to measure the functional impacts of 5-mC, 5-hmC, 5-fC and 5-caC on the regulation of gene transcription. Our strategy was to generate expression constructs containing each of these cytosine modifications in a CpG dinucleotide within an artificial promoter composed of a cyclic adenosine monophosphate (cAMP) response element (CRE) upstream from a TATA box-containing RNA polymerase II basal promoter. We chose CRE because it contains a CpG dinucleotide in the consensus sequence and because it was known that binding of the specific transcription factors (e.g., CREB and C/EBP alpha) and the induced transcription can be either negatively or positively regulated by the symmetric CpG methylation (8,30). Moreover the presence of 5-mC and 5-hmC in just one DNA strand significantly reduce, whereas the C>U substitution in either one or both DNA strands increase the affinity of CREB binding (31,32). The aim of present work was to compare the functional significance of CpG hemi-methylation and hemi-hydroxymethylation for the CRE-driven gene expression and to investigate the impacts of the DNA demethylation intermediates 5-fC and 5-caC on the promoter activity.

MATERIALS AND METHODS

Cell lines and reporter vectors

Isogenic HeLa-derived cell lines in which TDG, SMUG1 or UNG1/2 DNA glycosylases were knocked down by stable expression of the specific shRNA and the corresponding control HeLa/pEpS cell line were generated and characterised by western blotting and the specific DNA glycosylase activity testing previously (29). The sequence cloned into the pENTR/pSuper+ vector (Addgene, Cambridge, MA, USA) for expression of shRNA which efficiently targets human TDG was 5' GATCCGGGAACGAAATATGGACGTTCAACTCG AGTTGAACGTCATATTTCTTTTGGAAA. The parental vector pENTR/pSuper+ was used as a scrambled transcript control. Vectors pCRE-uno with a unique full-length 5'-TGACGTCA-3' consensus CRE sequence upstream from the RNA polymerase II transcription initiation region of the reporter EGFP gene and pCMV1111 with four CRE sequences were described previously (33). The vectors contain tandem sites for the Nb.BsrDI nicking endonuclease spaced with the 18-nucleotide interval around the 5'-TGACGTCA-3' consensus CRE sequence (Figure

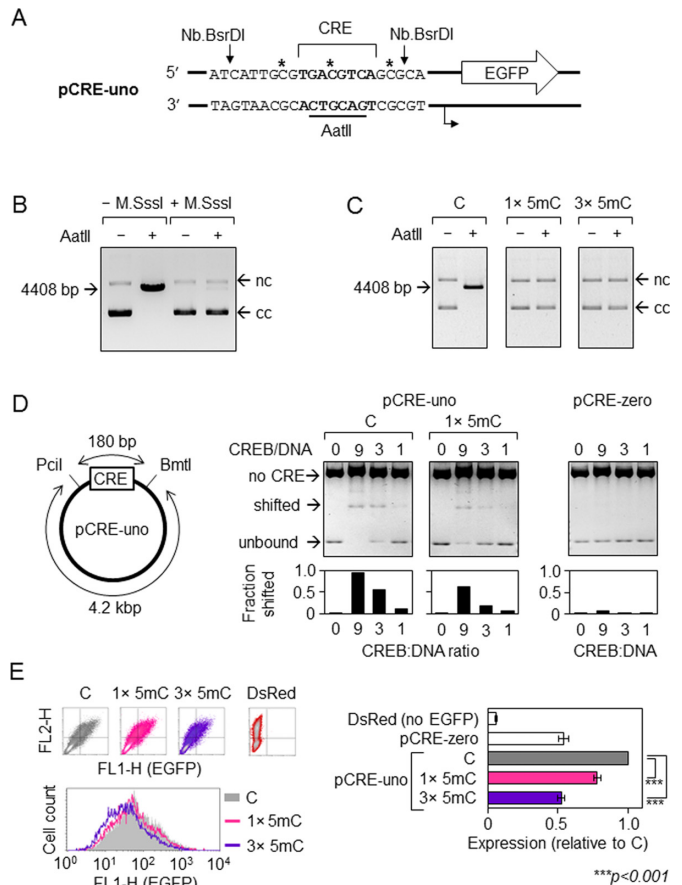


Figure 1. Impact of CpG hemi-methylation in the minimal CRE promoter on the CREB binding and the reporter EGFP gene expression. (A) Artificial CRE-uno promoter: transcription start (broken arrow), CRE sequence (bold), AatII site (underlined), Nb.BsrDI nicking positions (vertical arrows), and positions of 5-mC/5-hmC/5-fC/5-caC in the incorporated synthetic oligonucleotides (asterisks). (B) Inhibition of the AatII restriction endonuclease by the M.SssI methylation of the central CG-dinucleotide in CRE. Agarose gel electrophoresis of pCRE-uno incubated with AatII. Arrows indicate migration positions of the linearized vector (4408 bp) and of the covalently closed (cc) and nicked (nc) forms of circular vector DNA. (C) Verification of the incorporation of synthetic oligonucleotides containing one or three 5-mC (1 × 5mC, 3 × 5mC) by inhibition of the AatII cleavage. (D) Scheme of the PciI/BmtI digest of the pCRE-uno vector and electrophoretic mobility shift assay (EMSA) of the PciI/BmtI fragments under the indicated CREB:DNA molar ratios. The shorter (180 bp) fragment contains CRE with no or one 5-mC in the central CG-dinucleotide. The right panel shows control CRE-less vector (pCRE-zero) incubated with CREB in parallel. (E) EGFP expression in HeLa cells transfected with pCRE-uno containing none, one (in the central CG-dinucleotide of CRE) or three 5mC. Representative FACS data and relative EGFP expression values (mean ± SD, $n = 7$). Expression in the absence of an EGFP-coding vector (DsRed) denotes the lower detection limit. CRE-less vector (pCRE-zero) was included in two experiments as a reference for the basal expression level. P -values calculated by the Student's t -test.

1A). The derived vectors pCRE-uno-C (with inverse orientation of the Nb.BsrDI sites but otherwise identical to pCRE-uno) and pCRE-zero (in which CRE was mutated to produce an inert 5'-ACTGACTG-3' sequence) were generated by PCR with the overlapping primers using PfuTurbo[®] polymerase (Agilent Technologies, Waldbronn, Germany). The template DNA was digested with DpnI (Thermo Fischer Scientific, Darmstadt, Germany)

followed by direct repair in the ultracompetent *Escherichia coli* scs-8 (Agilent Technologies).

Synthetic oligonucleotides

DNA CE-phosphoramidites Bz-dA, Bz-dC, iBu-dG, T and Bz-mC were obtained from Glen Research (Sterling, VA) or Link Technologies (Bellshill, Scotland). 5-HmC, 5-fC and 5-caC phosphoramidites were synthesized and solid-phase synthesis of the 18-mer 5'-CATTGCGTGACGTCAGCG deoxyribo-oligonucleotides was performed using the standard protocols described previously (22). Underlined cytosines show positions in which 5-mC/5-hmC/5-fC/5-caC were incorporated, as specified in the text. Synthesized DNA chains were HPLC-purified with VP 250/10 Nucleosil 100-7 C 18 columns (Macherey-Nagel, Düren, Germany) and the quality determined by MALDI-MS (22). 2'-(*R*)-fluorinated derivatives of 5-fC and 5-caC were synthesized and handled as described previously (34). The sequence for incorporation of the 2'-fluorinated nucleotides was 5'-CATTGCGTGACGTCAGCG. Oligonucleotides 5'-CATTGCGTGA[THF/S-THF]GTCAGCG containing the abasic site analog tetrahydrofuran with either phosphodiester (THF) or phosphorothioate (S-THF) 5'-linkage as well as 5'-CATTGCGTGAC[8-oxoG]TCAGCG containing 8-oxo-7,8-dihydroguanine were purchased from BioSpring GmbH (Frankfurt am Main, Germany).

Generation of the reporter constructs with 5-mC/5-hmC/5-fC/5-caC specifically positioned in the CRE sequence

Site-specific double nicks on both sides of the CRE sequence in the vectors pCRE-uno ('top' strand), pCRE-uno-C ('bottom' strand) or pCMV1111 were generated by the Nb.BsrDI endonuclease (NEB GmbH, Frankfurt am Main, Germany). Synthetic 18-mer oligonucleotides containing 5-mC/5-hmC/5-fC/5-caC, the 2'-(*R*)-fluorinated derivatives of 5-fC/5-caC or THF/S-THF in the specified positions were used to displace the excised native DNA strand fragment and seamlessly ligated into vector DNA by the strand exchange protocol described previously (26). The respective control oligonucleotide without modifications was always ligated in parallel for every independent vector preparation. The presence of 5-mC/5-hmC/5-fC/5-caC in the covalently closed vector DNA was verified by the inhibition of cleavage by the AatII restriction endonuclease (NEB).

Electrophoretic mobility shift assays

Vectors pCRE-uno and pCRE-zero were digested with PciI and BmtI (both NEB) to produce the CRE-less 4228 bp fragment and the 180 bp fragment with one (pCRE-uno) or no CRE (pCRE-zero) and cleaned up. Binding reactions contained 10 nM DNA (400 ng in 15 μ l reaction volume) and proportional variable amounts of the purified CREB protein (BioCat, Heidelberg, Germany) in the binding buffer composed of 20 mM Tris-HCl (pH 8.0), 50 mM KCl, 5 mM MgCl₂, 1 mM dithiothreitol and 5% glycerol (all reagents from Sigma-Aldrich, Seelze, Germany) supplemented with 1 mg/ml bovine serum albumin (NEB). After 30-min incubation at 37°C reactions were chilled on ice and

fragments separated by agarose gel electrophoresis in the 0.5 \times TBE buffer (Bio-Rad Laboratories, Dreieich, Germany) under cooling conditions, as described previously (35). Band intensities were quantified following the ethidium bromide staining with the help of GelDoc™ XR+ molecular imager and the Image Lab™ software (Bio-Rad).

Transfections and gene expression analyses

Exponentially growing in six-well plates cells were transfected with the help of Effectene (QIAGEN, Hilden, Germany) with the EGFP reporter constructs (with or without the specified modifications) in combination with the tracer pDsRed-Monomer-N1 vector (Clontech, Saint-Germain-en-Laye, France), 400 ng each vector. The method for quantitative determination of EGFP expression in transiently transfected cells by flow cytometry was validated by titration of the EGFP-encoding vector and described in detail previously (36). Briefly, formaldehyde-fixed cells were equilibrated in phosphate buffered saline (PBS) and analysed using FACSCalibur™ and the CellQuest™ Pro software (Beckton Dickinson GmbH, Heidelberg, Germany). FSC/SSC gating was routinely used to exclude fragmented and aggregated cells and DsRed signal (FL2-H) was applied as additional gating marker to select for the effectively transfected cells. After the exclusion of untransfected cells, EGFP fluorescence (FL1-H) distribution plots were generated and average EGFP expression per cell determined as the median of the distribution. To assess variability between independent experiments, expression levels of constructs with cytosine modifications were calculated in each experiment relative to the expression of the control construct (without modifications) as a ratio of the respective FL1-H values.

RESULTS

Generation of expression constructs containing the C5 atom modifications of defined cytosines

We have previously generated the EGFP-encoding pCRE-uno reporter vector in which the basal RNA polymerase II promoter is coupled with a single CRE as a minimal upstream regulatory element. The CRE-motif is flanked by strand-specific nicking sites for the endonuclease Nb.BsrDI (Figure 1A), which are neutral for the regulation of gene expression, whereas CRE itself measurably activates transcription of the reporter EGFP gene (33). We sought to use these properties for the investigation of consequences of defined cytosine modifications at a single CpG dinucleotide for the regulation of the CRE-driven gene expression. For this purpose, we have introduced 5-mC (or 5-hmC/5-fC/5-caC) into the vector DNA sequence by substitution of the Nb.BsrDI-excised native DNA strand fragment for the synthetic strands containing the respective modifications at the central CG-dinucleotide of the CRE sequence.

As a proof of principle, we first generated expression constructs containing one or three 5-mC residues in the substituted DNA strand (Figure 1A). The pCRE-uno CRE sequence accommodates a unique recognition sequence for the AatII restriction endonuclease, whose activity is inhibited by the CpG methylation (Figure 1B), which en-

ables easy verification of the incorporation of 5-mC with the synthetic oligonucleotide (Figure 1C). Subsequently, we used the same approach for the incorporation of oligonucleotides containing 5-hmC, 5-fC or 5-caC and demonstrated that all the modifications were efficiently integrated into vector DNA (Supplementary Figure S1).

CREB binding and gene expression in the presence of the CpG hemi-methylation

We have analysed by the electrophoretic mobility shift assay the influence of single 5-mC in the central CpG dinucleotide of the CRE consensus sequence on binding of the transcription factor CREB (Figure 1D). Comparison with the nearly identical vector pCRE-zero in which the CRE site has been deleted showed that CREB binds specifically to the restriction fragment containing the CRE-motif and that single 5-mC in the central CpG dinucleotide significantly reduces the band shifting. The equivalent levels of band shifting in the presence of 5-mC, were achieved by increasing the CREB:DNA ratio by the factor of ~ 2.5 (Figure 1D and not shown data), suggesting that single methyl group causes a 2- to 3-fold reduction of CREB affinity to its binding site.

To check whether the impaired CREB binding has a functional significance for the gene transcription in cells, we next measured the impact of CpG hemi-methylation on the EGFP expression levels in transfected human cells. Analyses in HeLa cells showed that single 5-mC incorporated into the central CpG nucleotide decreased the expression by approximately one quarter. Hemi-methylation of all three CpG dinucleotides within the exchanged DNA fragment resulted in the halved expression level with respect to the expression of unmethylated control templates (Figure 1E). This quantitatively corresponded to the expression level of the CRE-less vector, suggesting that hemi-methylation at these three positions reduces gene transcription to the basal levels and, hence, nullifies the activatory effect of CREB. Accordingly, we derive that hemi-methylation of the central CpG alone caused an ~ 2 -fold reduction of the CREB-dependent fraction of transcription. For comparison, total methylation of the pCRE-uno vector by M.SssI almost completely obliterated the EGFP expression, thus indicating a very efficient repression of both induced and basal transcription (Supplementary Figure S2). Together, the results indicate that even a single methyl group—if present at the critical CpG dinucleotide—can measurably reduce the level of gene transcription in cells by direct reduction of the affinity of the transcription factor binding.

The impact of 5-hmC on the gene expression: comparison with 5-mC

We have next generated expression constructs containing synthetic 5-hmC at the same CpG dinucleotides as described above for 5-mC and determined the expression levels in HeLa cells under the same conditions. As 5-mC, 5-hmC caused a significant reduction in the EGFP expression levels if present in the central CpG dinucleotide of the CRE sequence and even stronger reduction when all three CpG dinucleotides of the incorporated synthetic DNA were modified by the hydroxymethyl group (Figure 2A). Quanti-

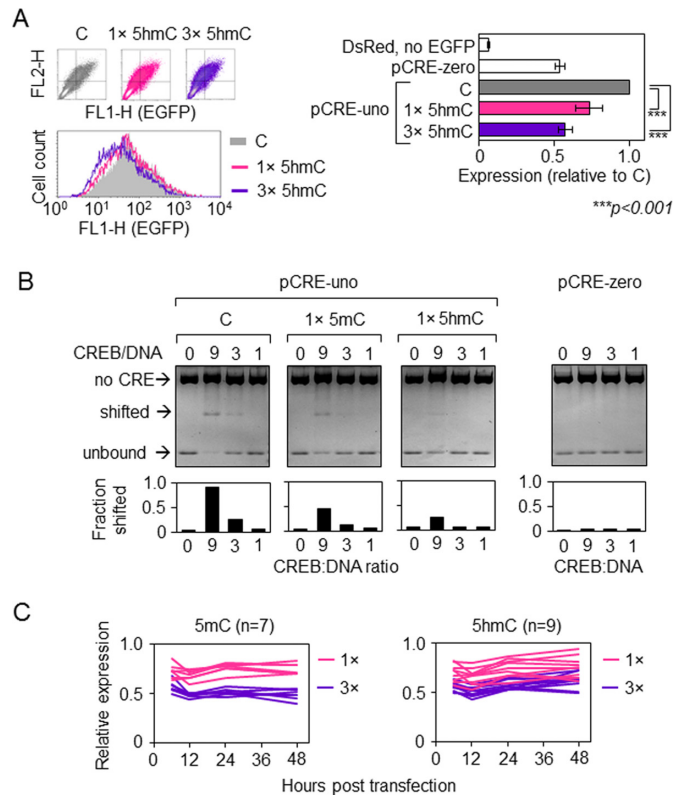


Figure 2. Impacts of one or three 5-hmC in the minimal CRE promoter on the EGFP gene expression and comparison with the respective effects of 5-mC. (A) EGFP expression in HeLa cells transfected with pCRE-uno containing none, one or three 5-hmC. Representative fluorescence scatter plots and the correspondent overlaid fluorescence distribution plots of HeLa cells 24 h after transfection (on the left) as well as mean EGFP expression values determined relative to the construct containing unmodified oligonucleotide ('C'). Mean of nine independent experiments \pm SD; *P*-values calculated by the Student's *t*-test. (B) CREB binding to CRE containing one 5-mC or 5-hmC in the central CG-dinucleotide detected by EMSA of the pCRE-uno PciI/BmtI fragment. (C) EGFP expression in HeLa cells transfected with pCRE-uno containing either one (1 \times) or three (3 \times) of the indicated CpG modifications in the promoter fragment (relative to pCRE-uno without CpG modifications). Lines represent independent transfections.

tatively, these effects were similar to the magnitude of the inhibitory effects of 5-mC at the same positions. CREB binding to the modified CRE was also impaired by single 5-hmC to at least the same extent as by 5-mC (Figure 2B), which suggests that the observed decrease in the gene transcription levels in cells was caused primarily by the impaired binding of the transcriptional activator. It has to be noted, however, that the presence of additional methyl or hydroxymethyl groups at the CpG sites outside from the minimal consensus CRE sequence did not cause further prominent decrease of CREB binding in the band-shifting assays (Supplementary Figures S3 and S4), thus indicating that the binding mode and the impact of 5-hmC on the gene expression in cells may be additionally influenced by other factors.

Because 5-hmC has been proposed to play a functional role in DNA demethylation (by either TDG-dependent (16,17,21) or -independent (37,38) pathways, we wondered whether the magnitudes of the repressory effects of both CpG modifications (5-mC and 5-hmC) would vary in the

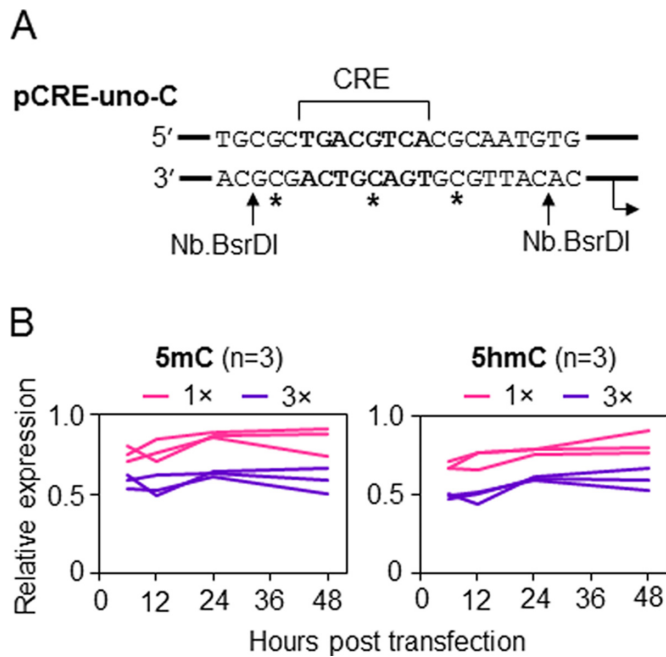


Figure 3. Impacts of 5-mC and 5-hmC in the 'bottom' DNA strand of the minimal CRE promoter on the EGFP gene expression. (A) Scheme of the promoter fragment of the pCRE-uno-C vector used for modification of CG-dinucleotides in the bottom strand (asterisks). (B) EGFP expression in HeLa cells transfected with pCRE-uno containing either one (1×) or three (3×) of the indicated modifications in the promoter fragment (relative to pCRE-uno without modifications). Lines represent independent transfections.

course of time. However, the results showed that the effects of both modifications persisted for at least 48 hours post transfection. During the whole time course, constructs with 5-mC and 5-hmC showed on average equal extents of transcriptional inhibition (Figure 2C), suggesting that 5-hmC has not been removed from CRE. The values for 5-hmC had a somewhat higher interexperimental noise; however, there was no clear tendency to recovery of the gene expression levels over the time course. The results thus indicate that, at least in the cell model used, hemi-methylated and hemi-hydroxymethylated CpG dinucleotides have overall very similar negative impacts on CREB binding and the CRE-activated transcription.

The effects of 5-mC and 5-hmC are manifested in both DNA strands

It is not known, whether biological consequences of hemi-methylation (or hemi-hydroxymethylation) depend on the DNA strand which is methylated. Because only one strand of the pCRE-uno plasmid (Figure 1A, top strand) can be substituted for a synthetic strand, we have generated an analogous pCRE-uno-C plasmid with flipped BsrDI sites, in which the complementary DNA strand can be replaced (Figure 3A, bottom strand). With CRE sequence as a reference point, we introduced 5-mC and 5-hmC at exactly the same CpG dinucleotides as previously in the top strand and performed time course expression analyses in HeLa cells (Figure 3B). Analogously to the previous experiments (Figure 2C), the results showed that either 5-mC or 5-hmC in

the central CpG of the CRE sequence resulted in losses of approximately one fourth of the total gene expression level. Additional modifications at the neighbouring CpG dinucleotides caused ulterior decrease of the gene expression resulting in the loss of approximately half of the total transcriptional activity, which corresponds to residual activity equivalent to the basal transcription level of the CRE-less vector described above. Once again, the effects of both modifications persisted for at least 48 hours. Overall, we conclude that the effects of 5-mC and 5-hmC at the same hemi-modified CpG dyads on the transcriptional activation/repression of the CRE-regulated promoter are qualitatively and quantitatively very similar, regardless of the DNA strand affected by the modification.

Potent and largely indirect repression by 5-fC and 5-caC

The TET-induced mechanism of enzymatic DNA demethylation described previously (16–21) foresees transient generation of 5-fC and perhaps 5-caC. Although active DNA demethylation is regarded as a potential mechanism for re-expression of genes originally silenced by DNA methylation, functional impact of these oxidation products on gene expression has not been sufficiently characterized to date. We have used the experimental setup described above to characterize the impacts of these cytosine modifications on the CRE-driven gene transcription. Because of the variable band-shifting activity between different batches of the recombinant CREB protein, we had to adjust the protein amounts in these experiments with respect to the conditions used in Figures 1D and 2B in order to achieve measurable band shifting with unmodified ('C') DNA (Figures 4A and 5A). We found that both 5-fC and 5-caC to some extent interfere with CREB-binding *in vitro*, however, do not prevent the binding completely, i.e. CREB binding properties to its target sequence were altered by these modifications to similar extents as earlier observed for hemi-methylated and hemi-hydroxymethylated CpG.

Expression analyses of the constructs containing one or three 5-fC in and around the CRE sequence showed that this modification leads to decreased reporter gene expression in the transfected HeLa cells (Figure 4B and C). The impairment of the EGFP expression was initially very mild (as judged from the fluorescence measured at the earliest time point of 6 h); however, the magnitude of the negative impact of 5-fC on the gene expression greatly increased by the 24 h time point after transfection (Figure 4B). At 24 h post-transfections, the degrees of the inhibition of gene expression by 5-fC and 5-caC clearly exceeded the respective effects of 5-mC and 5-hmC reported above (Figures 1–3). In-depth monitoring of the expression of constructs containing one or three hemi-modified CpG sites over a time span of 6–48 h post-transfection revealed a very strong and steady decline of the EGFP expression with no signs of recovery over the whole observation period, suggesting that 5-fC induced persistent repression of the gene transcription (Figure 4C). Thereby, the gene repression could not be accounted solely to the inhibition of CREB binding, as concluded from both strength and kinetics of the observed effects.

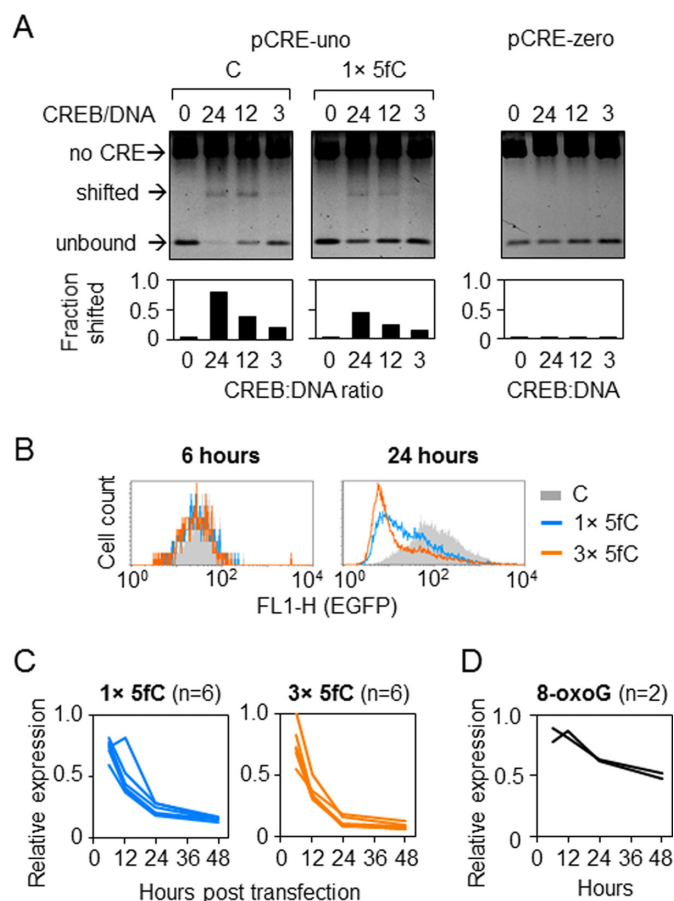


Figure 4. Effects of 5-fC in the minimal CRE promoter. (A) CREB binding to CRE containing one 5-fC in the central CG-dinucleotide detected by EMSA. (B) Representative fluorescence distribution plots of HeLa cells 24 h after transfection with constructs containing 5-fC in one (1×) or three (3×) CG-dinucleotides (overlaid with the reference ‘C’ construct obtained with unmodified synthetic oligonucleotide). (C) Time-course of the EGFP expression (relative to ‘C’) in HeLa cells transfected with constructs containing 1× or 3× 5-fC. Lines show independent transfections. Mean values for each time point are highly significantly different from 1 ($P < 0.001$, heteroscedastic Student’s *t*-test). (D) Time-course of the EGFP expression in HeLa cells transfected with constructs containing single 8-oxoG in the CRE (two independent experiments).

It is interesting to note that a similar gradual transcriptional repression was previously reported for reporter constructs containing various nucleobase modifications and has been identified as a hallmark of DNA lesions undergoing base excision repair (BER) (27,29,33). The role of BER in the gene repression has been inferred based on the requirements of the lesion-specific DNA glycosylases and of the strand-cleaved reaction intermediate generated by the apurinic/apyrimidinic site endonuclease APE1 (33). To test whether BER of a modification situated in the CRE sequence would affect the gene expression in a comparable way, we have placed single 8-oxo-7,8-dihydroguanine (8-oxoG) as a representative BER substrate into the CRE sequence. The results indeed showed a time-dependent decline of the gene expression (Figure 4D). Thus, 8-oxoG and 5-fC in the CRE sequence induced qualitatively similar tran-

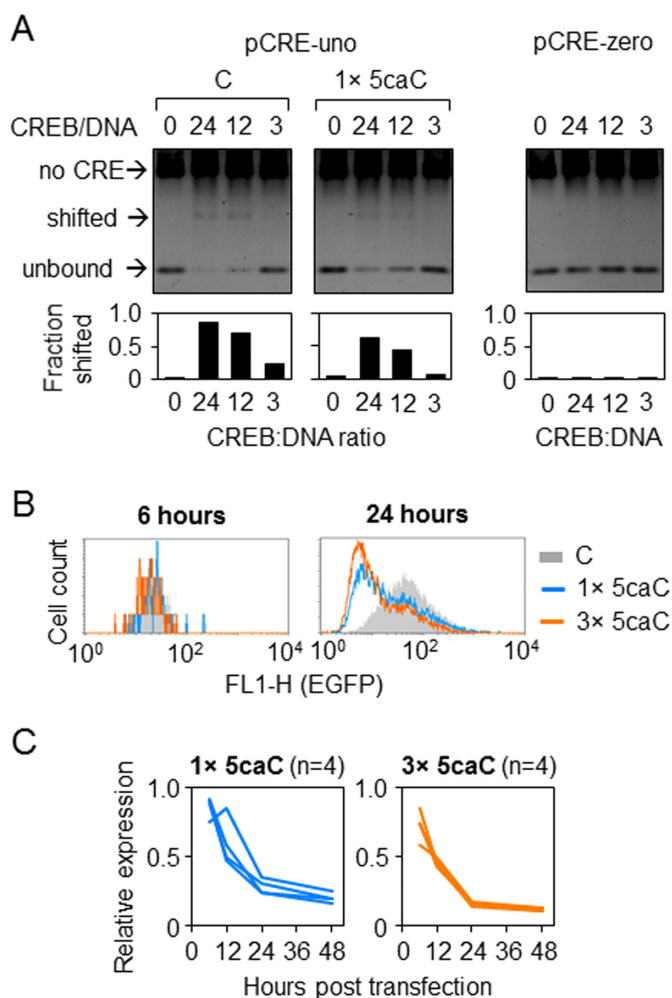


Figure 5. Effects of 5-caC in the minimal CRE promoter. (A) CREB binding to CRE containing one 5-caC in the central CG-dinucleotide detected by EMSA. (B) Representative fluorescence distribution plots of HeLa cells 24 h after transfection with constructs containing 5-caC in one (1×) or three (3×) CG-dinucleotides overlaid with the reference ‘C’ construct. (C) Time-course of the EGFP expression (relative to ‘C’) in HeLa cells transfected with constructs containing 1× or 3× 5-caC. Lines show independent transfections. Mean values for each time point are highly significantly different from 1 ($P < 0.001$, heteroscedastic Student’s *t*-test).

scriptional responses; however, the negative impact of 5-fC on the gene expression was much more powerful.

Further, we have analysed the CREB binding (Figure 5A) as well as the EGFP expression (Figure 5B and C) with 5-caC incorporated into the pCRE-uno reporter vector. The results showed that responses to 5-fC essentially recapitulated the response to 5-fC, i.e. mild inhibition of CREB binding and time-dependent progressive repression of transcription in transfected HeLa cells. Together, the results indicate that transcriptional responses to 5-fC and 5-caC largely resemble the response to DNA lesions, as exemplified by 8-oxoG. The observed transcriptional repression of the reporter constructs containing 5-fC or 5-caC further suggests that also oxidation of 5-mC/5-hmC to 5-fC/5-caC in the process of TET-induced DNA demethylation may not necessarily lead to transcriptional re-activation.

TDG mediates BER-dependent transcriptional repression by 5-fC and 5-caC

Because BER intermediate has been previously implicated in transcriptional repression or silencing induced by various DNA lesions (27,29,33,36), we questioned whether the progressive inhibition of expression of the reporter vectors containing 5-fC or 5-caC is similarly mediated by BER. Both 5-fC and 5-caC are efficiently removed from DNA *in vivo* and *in vitro* by TDG (16,18); therefore we performed gene expression analyses of vectors containing these modifications in cells with knocked down TDG expression (Supplementary Figure S5). The negative effects of 5-caC in pCRE-uno on the reporter gene expression were greatly reduced in the cell line stably expressing the TDG-specific short hairpin (sh) RNA but not in the isogenic cell lines with knocked down UNG1/2 or SMUG1 DNA glycosylases and not in the cell line stably transfected with non-specific shRNA (Figure 6A, left panel). The results thus indicate that excision of 5-caC by TDG is a prerequisite of ulterior transcriptional silencing of the CRE-controlled expression constructs. The magnitude of the inhibitory effects of 5-fC on the gene expression was also significantly moderated by the TDG knockdown, however to a lesser extent than it was observed for 5-caC.

To verify whether the effects of TDG knockdown were specific to 5-fC/5-caC, we further generated expression constructs carrying AP site (THF) in the same position in CRE (Supplementary Figure S6). As expected, the expression analyses showed a strong decrease of EGFP expression induced by THF but not by the APE1-resistant AP site (S-THF). Importantly, the magnitude of the effect of THF was not at all alleviated by the TDG knockdown. We thus conclude that BER reactions downstream from the base excisions step were not affected.

Because TDG knockdown had quantitatively only modest impact on the expression of constructs containing 5-caC and especially 5-fC, we sought to increase the dynamic range for measurements of the EGFP expression. Therefore, we performed analogous experiments in the pCMV1111 vector which contains a strong promoter with multiple regulatory elements including four consensus CRE sequences, one of which was chosen for targeted incorporation of synthetic C/5-fC/5-caC (Supplementary Figure S1B). Remarkably, also in the context of the strong promoter, both 5-caC and 5-fC induced dynamic decrease of the gene expression and in both cases the effects were significantly moderated by the TDG knockdown (Figure 6A, right panel). Also in this promoter the effect of TDG knockdown was somewhat smaller for 5-fC than 5-caC, but the difference was far less pronounced than in pCRE-uno.

The results obtained in the TDG knockdown cells underpinned the role of BER of 5-fC and 5-caC in the transcriptional repression by these cytosine modifications (Figure 6A); however, since the recovery of gene expression was not complete, they did not exclude a potential contribution of other mechanisms. To eliminate the base excision activity completely, we have applied a chemical approach. It was previously reported that sugar 2'-fluorination of the 5-fC and 5-caC nucleotides prevents the TDG excision activity completely while having only minimal impact

on the physical and functional properties of the DNA helix (34). Therefore, we next generated the pCMV1111 expression constructs carrying single 5-fC/5-caC deoxyribonucleotides or their 2'-fluorinated analogs in the specified position in CRE (Supplementary Figure S7). The expression analyses in HeLa cells explicitly showed that 2'-fluorination of deoxyribose fully reversed the gene repression by 5-fC and 5-caC (Figure 6B and C). This effect was exactly reproduced in the pCRE-uno vector (Supplementary Figure S8). The results thus show that regulatory promoter elements retain their activatory functions and that 5-fC and 5-caC do not lead to transcriptional repression when the base excision is inhibited, e.g. by the deoxyribose 2'-fluorination.

DISCUSSION

There are at least two mechanisms for regulation of gene transcription by 5-mC in DNA. It was previously suggested that selective methylation of CpG dinucleotides present in upstream promoter elements can interfere with recruitment of the specific transcription factors (6-8). On the other hand, CpG methylation can recruit general co-repressors via the MBD binding (1,2). Using a minimal CRE-driven promoter with an artificially introduced hemi-methylated CpGs in defined positions, we show here that a single methyl group in the central CpG of the CRE sequence causes a 2- to 3-fold reduction of the specific CREB binding affinity and an equivalent decrease in the CRE-dependent gene expression in human cells (Figure 1). This proportionality suggests that impaired CREB binding has direct functional relationship with the gene expression. Nevertheless, hemi-methylation of the neighbored CpGs clearly added to the decreased gene expression in cells (Figures 2 and 3), even though CREB binding was not further impaired by the modifications outside of the CRE consensus sequence (Figures 2 and 3, Supplementary Figures S3 and S4), which indicates that impaired CREB binding is not the only mechanism for the declined gene expression if more than one 5-mC is present. Indeed, massive methylation of the reporter vector by M.SssI resulted in a complete transcriptional silencing of the reporter constructs (Supplementary Figure S2). This mechanism appears to dominate over the CREB-induced transcriptional activation, since the enzymatic methylation eliminated both induced and basal expression. This notion is supported by further observation that selective partial demethylation of the CRE region did not rescue the expression of the M.SssI methylated reporter constructs (data not shown). In summary, the results demonstrate that methylation and even hemi-methylation of single CpG dinucleotides in critical gene regulatory elements can have substantial negative impacts on the gene expression by directly impeding the transcription factor binding. It is necessary to stress, however, that this effect can only be functionally significant when the surrounding DNA sequence is unmethylated.

The outcome of CRE hydroxymethylation on the gene expression was essentially the same as of methylation (Figures 1-3), thus suggesting that the functional modes of inhibition of the activator-induced gene transcription by these two modifications are similar or common. As for 5-mC, the magnitude of negative effect of single 5-hmC on the gene ex-

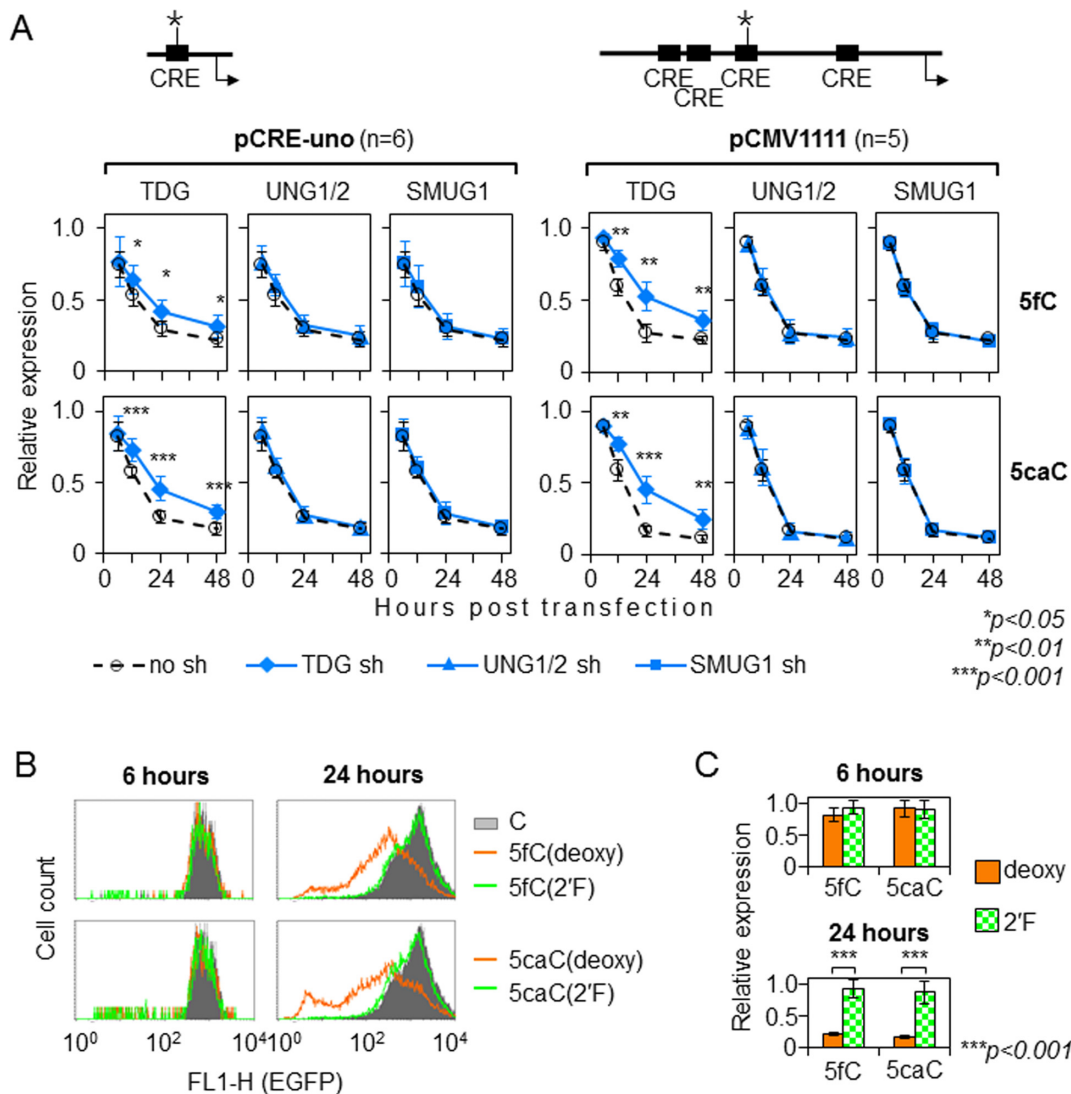


Figure 6. Impact of BER on the expression of reporter constructs containing single 5-fC/5-caC in the CRE sequence. (A) Effects of 5-fC and 5-caC on EGFP expression in HeLa-derived cell lines with stable shRNA knockdown of the UNG1/2, TDG or SMUG1 DNA glycosylases as compared to the control isogenic cell line ('no sh'). Expression of the constructs containing 5-fC or 5-caC in the pCRE-uno (left panel) and pCMV1111 (right panel) vectors at each time point is presented relative to the respective reference constructs containing cytosine (mean \pm SD; P -values calculated by the Student's t -test). (B, C) EGFP expression in HeLa cells at the specified times after transfection with pCMV1111 containing single 5-fC or 5-caC as deoxyribonucleotides (deoxy) or the respective 2'-(R)-fluorinated derivatives (2'F). Representative FACS data (B) and relative EGFP expression values (mean \pm SD, $n = 5$, P -values calculated by the Student's t -test) (C).

pression is consistent with the impediment of CREB binding to CRE sequence by the modification. This result suggests that the effects of both 5-mC and 5-hmC are likely mediated by prevention of binding of activatory proteins specific for unmodified CpG rather than by recruitment of co-repressors via a specific interaction with 5-mC or 5-hmC. Such a mechanism is in agreement with proteomics data which identified little overlap between the spectra of specific interactors of 5-mC and 5-hmC, whereas binding of a much larger set of proteins, including several families of transcriptional activators, was disrupted by either modification (22).

The degrees of inhibition of gene expression by single 5-mC and 5-hmC stayed steady over the whole time-course of the gene expression analyses (Figures 2 and 3), suggesting that epigenetic state of the promoter region remained sta-

ble over at least 48 hours. In a marked contrast, the expression of constructs containing 5-fC or 5-caC showed strong negative dynamics, indicating a progressive silencing of the affected promoter (Figures 4 and 5). Intriguingly, the spectrum of putative 5-fC and 5-caC readers previously identified by the proteomics approaches includes candidates with potential gene silencing functions such as DNMT1 and the components of the Mi-2/NuRD nucleosome remodeling deacetylase complex (22,39). On the other hand, the enrichment of DNA repair proteins including TDG and MPG in the sets of putative 5-fC and 5-caC readers raises the possibility that proteins bound to products or the DNA repair reactions have been also classified as interactors in these experimental setups.

We have previously reported that excision of various base modifications by the specific DNA glycosylases in concert with the apurinic/aprimidinic site endonuclease commonly initiates transcriptional silencing (27,29,33). Now we show that TDG clearly contributes to progressive transcriptional silencing induced by 5-caC and 5-fC (Figure 6A), which strongly suggests the key role of BER in this process. Incomplete recovery of the gene expression levels under the conditions of TDG knockdown should likely be attributed to low residual TDG activity in cells, since complete abolition of excision of 5-fC/5-caC by the deoxyribose 2'-fluorination fully restored the expression (Figure 6B and C). Nevertheless, based on these results it still cannot be ruled out that some other DNA glycosylase (or endonuclease) could also contribute to processing of 5-fC and/or 5-caC. Thus, NEIL1, NEIL2 and NEIL3 interact with TET1 and have been implicated in transcriptional activation of TET-oxidised M.SssI-methylated vector DNA in the absence of TDG (40). Even though the evidence is missing that NEIL proteins autonomously excise 5-fC, another study showed that NEIL1 and NEIL2 accelerate the TDG turnover and thus may stimulate excision when TDG is limited (15). Since our findings implicate a BER intermediate in transcriptional silencing induced by 5-fC and 5-caC, it is intriguing to suggest that balance between the activity of individual enzymes within BER or coordination between the BER steps in a given cell type or during a particular differentiation phase may have profound consequences for the regulation of gene expression by these DNA modifications. In our system, both generation of AP site and subsequent strand cleavage by APE1 are required for the gene silencing (Figure 6, Supplementary Figure S6). Taking into account intrinsically low processivity of TDG, it would be important to investigate, whether the switches exist which would halt one of these steps thereby possibly preventing the gene silencing. Such mechanisms could be critical to safeguard the proper gene function following the restoration of unmethylated CpG in the genome.

In summary, our findings have implications for understanding of transcription regulatory functions of hemimethylated (or hemi-modified by 5-hmC, 5-fC or 5-caC) CpG dinucleotides. Such CpG modifications can arise in cells during DNA demethylation or de novo DNA methylation reactions. The results indicate that 5-mC and 5-hmC act as stable regulatory marks predominantly by decreasing binding of transcriptional activator CREB to the modified target motif. The results further suggest that 5-fC and 5-caC trigger dynamic changes in the promoter activity. Strikingly, in the minimal CRE-regulated promoter employed here, both 5-fC and 5-caC functioned primarily as repressory marks. The impacts of 5-fC and 5-caC on the promoter activity are only partly attributable to the direct impairment of CREB binding and the negative effects on the gene expression strongly advance with time in a TDG-dependent fashion. Extrapolating present results to the situation during active DNA demethylation in cells, we suggest that generation of 5-fC and 5-caC in the regulatory promoter elements and further processing of these modifications by BER by default leads to perpetuation and enhancement of the repressed transcriptional state and that transcriptional reactivation would require additional signals.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Bork Lühnsdorf, Bernd Epe, Alexander Ishchenko, Lars Schomacher and Michael Musheev for discussions of the experimental data. The authors are grateful to Bernd Epe and his group members for free access to the FACS instrument and other laboratory facilities.

FUNDING

Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [KH 263/1, KH263/2, Heisenberg Fellowship KH 263/3 to A.K.; GRK 2062 Fellowship to M.R.]. Funding for open access charge: DFG, German Research Foundation.

Conflict of interest statement. None declared.

REFERENCES

- Schubeler, D. (2015) Function and information content of DNA methylation. *Nature*, **517**, 321–326.
- Jones, P.A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.*, **13**, 484–492.
- Deaton, A.M. and Bird, A. (2011) CpG islands and the regulation of transcription. *Genes Dev.*, **25**, 1010–1022.
- Wigler, M., Levy, D. and Perucho, M. (1981) The somatic replication of DNA methylation. *Cell*, **24**, 33–40.
- Gruenbaum, Y., Cedar, H. and Razin, A. (1982) Substrate and sequence specificity of a eukaryotic DNA methylase. *Nature*, **295**, 620–622.
- Ben-Hattar, J. and Jiricny, J. (1988) Methylation of single CpG dinucleotides within a promoter element of the Herpes simplex virus tk gene reduces its transcription in vivo. *Gene*, **65**, 219–227.
- Watt, F. and Molloy, P.L. (1988) Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev.*, **2**, 1136–1143.
- Iguchi-Ariga, S.M. and Schaffner, W. (1989) CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes Dev.*, **3**, 612–619.
- Hodges, E., Molaro, A., Dos Santos, C.O., Thekkat, P., Song, Q., Uren, P.J., Park, J., Butler, J., Rafii, S., McCombie, W.R. *et al.* (2011) Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol. Cell*, **44**, 17–28.
- Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D. *et al.* (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, **480**, 490–495.
- Ziller, M.J., Gu, H., Muller, F., Donaghey, J., Tsai, L.T., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E. *et al.* (2013) Charting a dynamic DNA methylation landscape of the human genome. *Nature*, **500**, 477–481.
- Cadet, J. and Wagner, J.R. (2014) TET enzymatic oxidation of 5-methylcytosine, 5-hydroxymethylcytosine and 5-formylcytosine. *Mut. Res.*, **764–765**, 18–35.
- Hahn, M.A., Szabo, P.E. and Pfeifer, G.P. (2014) 5-Hydroxymethylcytosine: a stable or transient DNA modification? *Genomics*, **104**, 314–323.
- Klungland, A. and Robertson, A.B. (2016) Oxidized C5-methyl cytosine bases in DNA: 5-hydroxymethylcytosine; 5-formylcytosine; and 5-carboxycytosine. *Free Radic. Biol. Med.*, **107**, 62–68.
- Schomacher, L., Han, D., Musheev, M.U., Arab, K., Kienhofer, S., von Seggern, A. and Niehrs, C. (2016) Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. *Nat. Struct. Mol. Biol.*, **23**, 116–124.

16. He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L. *et al.* (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, **333**, 1303–1307.
17. Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C. and Zhang, Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, **333**, 1300–1303.
18. Maiti, A. and Drohat, A.C. (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.*, **286**, 35334–35338.
19. Pfaffeneder, T., Hackner, B., Truss, M., Munzel, M., Muller, M., Deiml, C.A., Hagemeyer, C. and Carell, T. (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem.*, **50**, 7008–7012.
20. Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D. *et al.* (2013) Global epigenomic reconfiguration during mammalian brain development. *Science*, **341**, 1237905.
21. Weber, A.R., Krawczyk, C., Robertson, A.B., Kusnierczyk, A., Vagbo, C.B., Schuermann, D., Klungland, A. and Schar, P. (2016) Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nat. Commun.*, **7**, 10806.
22. Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W., Bauer, C., Munzel, M., Wagner, M., Muller, M., Khan, F. *et al.* (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell*, **152**, 1146–1159.
23. Kriaucionis, S. and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, **324**, 929–930.
24. Zhang, L., Lu, X., Lu, J., Liang, H., Dai, Q., Xu, G.L., Luo, C., Jiang, H. and He, C. (2012) Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat. Chem. Biol.*, **8**, 328–330.
25. Kress, C., Thomassin, H. and Grange, T. (2006) Active cytosine demethylation triggered by a nuclear receptor involves DNA strand breaks. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 11112–11117.
26. Luhnsdorf, B., Kitsera, N., Warken, D., Lingg, T., Epe, B. and Khobta, A. (2012) Generation of reporter plasmids containing defined base modifications in the DNA strand of choice. *Anal. Biochem.*, **425**, 47–53.
27. Kitsera, N., Stathis, D., Luhnsdorf, B., Muller, H., Carell, T., Epe, B. and Khobta, A. (2011) 8-Oxo-7,8-dihydroguanine in DNA does not constitute a barrier to transcription, but is converted into transcription-blocking damage by OGG1. *Nucleic Acids Res.*, **39**, 5926–5934.
28. Kitsera, N., Gasteiger, K., Luhnsdorf, B., Allgayer, J., Epe, B., Carell, T. and Khobta, A. (2014) Cockayne syndrome: varied requirement of transcription-coupled nucleotide excision repair for the removal of three structurally different adducts from transcribed DNA. *PLoS One*, **9**, e94405.
29. Luhnsdorf, B., Epe, B. and Khobta, A. (2014) Excision of uracil from transcribed DNA negatively affects gene expression. *J. Biol. Chem.*, **289**, 22008–22018.
30. Rishi, V., Bhattacharya, P., Chatterjee, R., Rozenberg, J., Zhao, J., Glass, K., Fitzgerald, P. and Vinson, C. (2010) CpG methylation of half-CRE sequences creates C/EBPalpha binding sites that activate some tissue-specific genes. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 20311–20316.
31. Moore, S.P., Toomire, K.J. and Strauss, P.R. (2013) DNA modifications repaired by base excision repair are epigenetic. *DNA Repair*, **12**, 1152–1158.
32. Verri, A., Mazzarello, P., Biamonti, G., Spadari, S. and Foche, F. (1990) The specific binding of nuclear protein(s) to the cAMP responsive element (CRE) sequence (TGACGTCA) is reduced by the misincorporation of U and increased by the deamination of C. *Nucleic Acids Res.*, **18**, 5775–5780.
33. Allgayer, J., Kitsera, N., Bartelt, S., Epe, B. and Khobta, A. (2016) Widespread transcriptional gene inactivation initiated by a repair intermediate of 8-oxoguanine. *Nucleic Acids Res.*, **44**, 7267–7280.
34. Schroder, A.S., Parsa, E., Iwan, K., Traube, F.R., Wallner, M., Serdjukow, S. and Carell, T. (2016) 2'-(R)-Fluorinated mC, hmC, fC and caC triphosphates are substrates for DNA polymerases and TET-enzymes. *Chem. Commun.*, **52**, 14361–14364.
35. Alessandri, M., Beretta, G.L., Ferretti, E., Mancia, A., Khobta, A. and Capranico, G. (2004) Enhanced CPT sensitivity of yeast cells and selective relaxation of Gal4 motif-containing DNA by novel Gal4-topoisomerase I fusion proteins. *J. Mol. Biol.*, **337**, 295–305.
36. Khobta, A., Anderhub, S., Kitsera, N. and Epe, B. (2010) Gene silencing induced by oxidative DNA base damage: association with local decrease of histone H4 acetylation in the promoter region. *Nucleic Acids Res.*, **38**, 4285–4295.
37. Grin, I. and Ishchenko, A.A. (2016) An interplay of the base excision repair and mismatch repair pathways in active DNA demethylation. *Nucleic Acids Res.*, **44**, 3713–3727.
38. Xue, J.H., Xu, G.F., Gu, T.P., Chen, G.D., Han, B.B., Xu, Z.M., Bjaras, M., Krokan, H.E., Xu, G.L. and Du, Y.R. (2016) Uracil-DNA glycosylase UNG promotes Tet-mediated DNA demethylation. *J. Biol. Chem.*, **291**, 731–738.
39. Iurlaro, M., Ficiz, G., Oxley, D., Raiber, E.A., Bachman, M., Booth, M.J., Andrews, S., Balasubramanian, S. and Reik, W. (2013) A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.*, **14**, R119.
40. Muller, U., Bauer, C., Siegl, M., Rottach, A. and Leonhardt, H. (2014) TET-mediated oxidation of methylcytosine causes TDG or NEIL1 glycosylase dependent gene reactivation. *Nucleic Acids Res.*, **42**, 8592–8604.

6.2 Methoden zur Verbesserung der Sequenzierung von mdC, hmdC, fdC und cadC

Obwohl Möglichkeiten zur Sequenzierung von mdC, hmdC, fdC und cadC seit Jahren existieren, ist deren exakte Sequenzierung keine gängige Praxis. Insbesondere die Sequenzierung der seltenen Modifikationen fdC und cadC ist noch immer nicht verlässlich möglich.^[194,195] Um dies zu adressieren, sollten Methoden zur Verbesserung der Sequenzierung von mdC, hmdC, fdC und cadC entwickelt werden. Im ersten Abschnitt 6.2.1 wird die Entwicklung von Haarnadel-Oligonukleotiden beschrieben. Diese ermöglichen die Strang-spezifische Sequenzierung von mdC, hmdC, fdC sowie cadC in höchster Genauigkeit; also die Unterscheidung zwischen hemi-, hybrid- und vollständig modifizierten Positionen in doppelsträngiger DNA. Der zweite Abschnitt 6.2.2 beschreibt die Entwicklung von Referenz-Oligonukleotiden, sogenannten *spike-ins*. Diese ermöglichen die Erfassung und ggf. Steigerung der Genauigkeit von Sequenzierungen. Der dritte Abschnitt 6.2.3 beschreibt die Etablierung einer Methode zur verlässlichen Sequenzierung von fdC. Dazu wurde ein im AK Carell entwickeltes Hydroxylamin-Reagenz auf die Bisulfit-Sequenzierung (BS-Seq) von fdC angewendet. Dabei zeigte sich, dass so deutlich verlässlichere fdC-Sequenzierungen möglich sind als mit der aktuell gebräuchlichsten Methode fCAB. Zusätzlich wird die naheliegende Anwendung der hier entwickelten Methoden auf Sequenzierungen der 3. Generation besprochen.

6.2.1 Haarnadel-Oligonukleotide: Referenzen zur Sequenzierung komplementärer DNA-Einzelstränge

Dieses Projekt adressiert den Bedarf für Referenznukleoside bei der Haarnadel-Sequenzierung von mdC, hmdC, fdC und cadC. Die Sequenzierung komplementärer DNA-Einzelstränge ist über die Ligation von Haarnadel-Oligonukleotiden (HPs) möglich, vgl. Einleitung Abschnitt 3.2. Um eine verlässliche Sequenzierung der seltenen dC-Modifikationen hmdC, fdC oder cadC zu ermöglichen, ist der Einsatz von Referenznukleosiden dringend angeraten. In diesem Kooperationsprojekt zwischen dem AK WALTER (*Universität des Saarlandes*) und dem AK CARELL (*LMU München*) wurden HPs hergestellt, die eine verlässliche Haarnadel-Sequenzierung von mdC, hmdC, fdC und cadC ermöglichen. So entwarf PASCAL GIEHR (AK WALTER) die HPs auf Basis der HP-BS-Seq von mdC und hmdC, dem HPoxBS.^[180] Trotz anspruchsvoller Synthesen und Aufreinigungen gelang die Herstellung fünf individueller HP-Konstrukte sowie deren Anwendung mittels BS-Tiefensequenzierung.

6.2.1.1 Konzept

Die HPs lassen sich in drei Bereiche unterteilen: (1) Referenznukleoside, (2) Identifikationssequenz und (3) Ligationsüberhang, siehe Abb. 23 A):

(1) *Im HP-Stamm dienen die Referenznukleoside dC, mdC, hmdC, fdC und cadC als interne Standards.* Alle Referenznukleoside liegen sowohl im GpC- als auch im CpG-Kontext vor. Als einzige Ausnahme liegt cadC im DdeI HP nur im CpG-Kontext vor. Alle weiteren HPs enthalten cadC im CpC-Kontext, siehe Abb. 23 B). Lokale Effekte wie Sekundärstrukturen und GC-Gehalt können die Zugänglichkeit einer Modifikation bei der Sequenzierung beeinflussen. Durch die kovalente Verknüpfung der Referenznukleoside mit dem DNA-Fragment kann erwartet werden, dass die Referenznukleoside ähnlich beeinflusst werden, wie die Modifikationen im zu untersuchenden DNA-Fragment. So kann z. B. die ausbleibende Desaminierung von Referenznukleosiden wichtige Hinweise darauf geben, dass ggf. vorhandene Modifikationen im zu untersuchenden DNA-Abschnitt nicht detektiert werden konnten.

(2) *Die Identifikationssequenz ermöglicht die eindeutige Identifikation jedes DNA-Abschnitts.* Die Identifikationssequenz oder UMI, für *unique molecular identifier*, befindet sich in der Haarnadelschleife des HPs. Sie besteht aus einer zufälligen Sequenz von acht W-Nukleosiden (dA, dT, dG). Auf dC wurde verzichtet, um die bioinformatische Auswertung der BS-Seq zu erleichtern: Da dC durch BS desaminiert wird, diese Umwandlung aber nicht immer quantitativ verläuft (vgl. Einleitung Abschnitt 3.2), wäre die Zuordnung der UMI aufgrund erhöhter Sequenz-Komplexität schwierig. Die acht D-Nukleoside der UMI ermöglichen bis zu 6561 individuelle Identifikationssequenzen. Da die HPs an beide Enden des zu untersuchenden DNA-Fragments ligiert werden, ergeben sich daraus etwa 43 Mio. individuelle Kombinationen.

(3) *Fünf verschiedene Ligations-Überhänge ermöglichen die Kombination unserer HPs mit einer Vielzahl an Restriktionsenzymen.* Die Fragmentierung der zu sequenzierenden DNA erfolgt bei lokaler Tiefensequenzierung und RRBS üblicherweise durch Restriktionsenzyme. DNA die mittels WGBS untersucht werden soll, wird häufig mittels Ultraschall fragmentiert. Um eine möglichst hohe Flexibilität zu ermöglichen, entwarf PASCAL GIEHR fünf verschiedene Ligations-Überhänge, siehe Abb. 23 B). Damit sind unsere HPs mit den Restriktionsenzymen MspI, DdeI, BsaWI sowie Eco47I kompatibel. Daneben erzeugen Ultraschall-Fragmentierung sowie eine Vielzahl von Restriktionsenzymen *blunt ends*. Diese sind mit unserem *A-Tail*-HP kompatibel. Da Ligationen von *blunt ends* nur mit geringer Effizienz verlaufen, erfolgt zunächst das sogenannte *A-Tailing* der zu untersuchenden DNA. Dabei wird das *blunt end* am 3'-Ende mit einem einzelnen dA-Nukleotid versehen. Dieser *A-Tail* ist komplementär zum 3'-dT-Überhang des *A-Tail*-HPs und ermöglicht so dessen Ligation in hoher Effizienz.

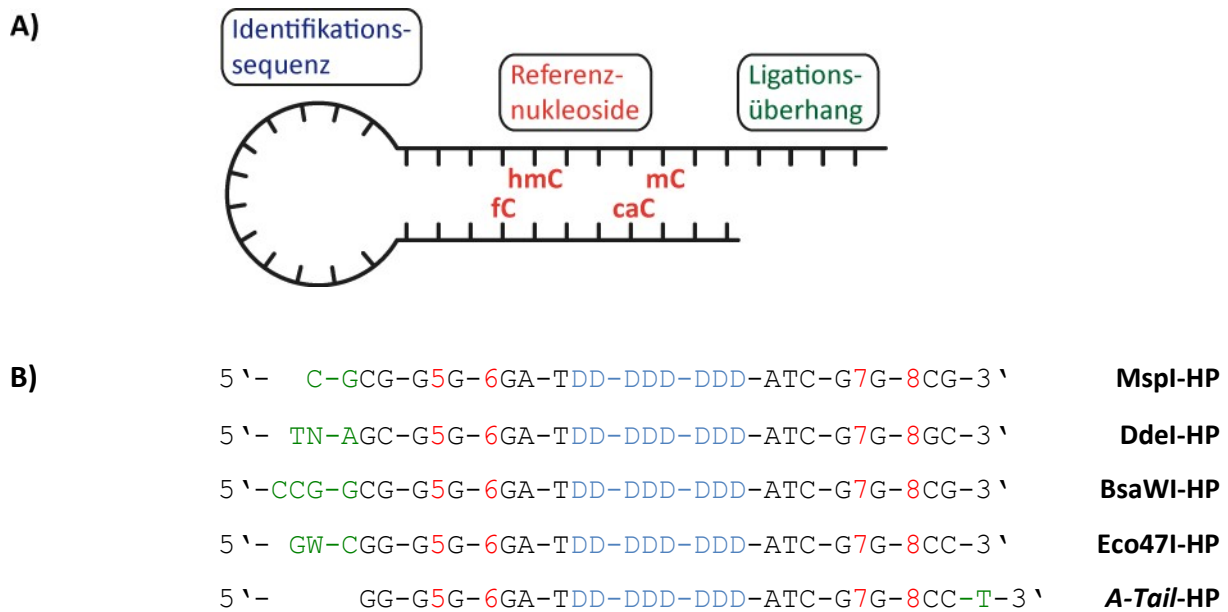


Abb. 23 **A)** Schematische Darstellung der Haarnadel-Oligonukleotide (HPs). **B)** HP-Sequenzen und Benennung, basierend auf deren Ligationsüberhängen. N = A/T/G/C, D = A/T/G, W = A/T. 5: mdC, 6: hmdC, 7: fdC, 8: cadC, grün: Ligationsüberhang, rot: Referenznukleoside, blau: Identifikationssequenz.

Die Anwendung unserer HPs besteht in deren Ligation an die Fragmente der Ziel-DNA. Alle weiteren Schritte entsprechen der regulären BS-Tiefensequenzierung. Die sechs Schritte der HP-BS-Seq sind in Abb. 24 schematisch dargestellt:

1. Fragmentierung der DNA-Probe mittels Restriktionsenzym-Verdau
 - bei *blunt ended* DNA: zusätzlich *A-Tailing*
2. Ligation von DNA-Fragmenten und HPs
3. BS-Behandlung
4. PCR-Amplifikation
5. Tiefensequenzierung
6. bioinformatische Auswertung

Die Anwendung unserer HPs erfolgte im Rahmen der in dieser Arbeit entwickelten Methode zur fdC-Sequenzierung und wird in Abschnitt 6.2.3.4 beschrieben, vgl. Abb. 44.

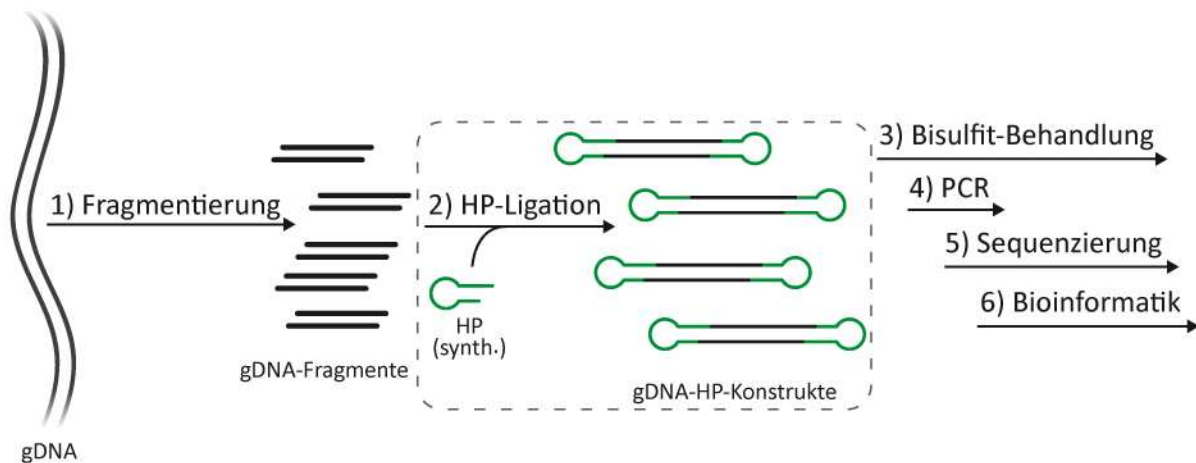


Abb. 24 Ablauf der BS-Seq unter Verwendung von Haarnadel-Oligonukleotiden (HPs): HP-BS-Seq. Die Anwendung der HPs besteht in deren Ligation mit den Fragmenten der DNA-Probe (umrahmt). Alle weiteren Schritte entsprechen der regulären BS-Sequenzierung. Abbildung adaptiert aus.^[246] gDNA: genomische DNA; HP (synth.): synthetische Haarnadel-Oligonukleotide.

6.2.1.2 Einordnung, Zusammenfassung und Ausblick

Zusammengenommen ermöglichen unsere HPs die Einzelstrang-spezifische Sequenzierung von mdC, hmdC, fdC sowie cadC mit höchster Genauigkeit. Dies wird durch Markierung jedes DNA-Fragments sowie durch den Einsatz der Referenznucleoside erreicht. Mit den aktuell gängigen Verfahren kann dies hingegen nicht erreicht werden. Unsere HPs sind dabei nicht auf die Anwendung in der BS-Seq beschränkt. Stattdessen ermöglichen sie die Sequenzierung von mdC, hmdC, fdC sowie cadC mit einer Vielzahl von Methoden. Deren Entwicklung adressiert insbesondere zwei Anwendungen: Die Sequenzierung von fdC und cadC sowie die Sequenzierung von hemi- und hybrid-modifizierten Positionen. Da fdC und cadC sehr selten im Genom auftreten, ist deren aussagekräftige Sequenzierung ohne spezielle Maßnahmen kaum möglich. Unsere HPs ermöglichen nun eine verlässliche Sequenzierung von fdC und cadC. Erste Hinweise auf die epigenetischen Funktionen von hemi- und hybrid-modifizierten Positionen rücken diese bisher kaum untersuchte Klasse seit kurzem verstärkt in den Fokus.^[171,175,176,178] Unsere HPs ermöglichen es auch dieses neue Feld aufzuklären. Die Synthesen und Aufreinigungen der HPs gelangen für alle fünf Konstrukte, trotz anspruchsvoller Haarnadel-Sekundärstrukturen sowie Randomisierungen.

Unsere HPs sollten sich ideal mit SMRT-Sequenzierungen kombinieren lassen. SMRT gehört zu den Sequenzierungs-Methoden der 3. Generation. Sequenzierungen mittels SMRT erfolgen an kreisförmiger DNA, sogenannten *dumbbells*. Diese werden durch Ligation der DNA-Fragmente an Haarnadel-Oligonukleotide erhalten. Die einzig nötige Anpassung zur

Verwendung unserer HPs, sollte im Hinzufügen der Polymerase-spezifischen Sequenzen liegen: Da SMRT-Sequenzierungen durch Polymerase-vermitteltes *sequencing by synthesis* erfolgen, sind Sequenzen für die Assoziation der Polymerase nötig sowie eine Primer-Bindestelle. Beide Sequenzen befinden sich bei SMRT-Sequenzierungen üblicherweise im Haarnadel-Oligonukleotid. Durch simple Ligation sollten unsere HPs um diese Sequenzen erweitert werden können, so dass die Vorteile unserer HPs direkt auf SMRT-Sequenzierungen angewendet werden können. Der Einsatz unserer HPs bei SMRT-Sequenzierungen wird aktuell von PASCAL GIEHR untersucht. Da auch *Oxford Nanopore*-Sequenzierungen auf Haarnadel-Oligonukleotide angewiesen sind, wäre der Einsatz unserer HPs auch für diese vielversprechende Sequenzierungsmethode der 3. Generation denkbar.

6.2.1.3 Experimentelle Details

Dieser Abschnitt gibt einen Überblick der experimentellen Details. Dies umfasst die Oligonukleotid-Synthese der HPs sowie deren Entschützung und Aufreinigung.

Die Herstellung der HPs gliederte sich in fünf Schritte: 1) Phosphoramidit-Synthese, 2) Oligonukleotid-Synthese, 3) Entschützung, 4) Aufreinigung und 5) enzymatische Phosphorylierung. Die HP-Herstellung erfolgte im AK Carell als Kooperations-Projekt: Die Synthese der nicht-kommerziellen Phosphoramidite von hmdC, fdC und cadC erfolgte durch KRISTOF HUFNAGEL in Anlehnung an SCHRÖDER *et al.*^[247]. Die Etablierung der Schritte 2), 3) und 4) erfolgte in Zusammenarbeit mit MATTHIAS KURZ. Die enzymatische Phosphorylierung der HPs erfolgte durch PASCAL GIEHR (AK Walter).

6.2.1.3.1 Oligonukleotid-Synthese

Die Bedingungen der Oligonukleotid-Synthese wurden zur Erhöhung der Ausbeute angepasst. Die Synthese der Oligonukleotide erfolgte mittels Festphasen-Synthese (FPS). Haarnadel-Strukturen stellen für die FPS eine besondere Herausforderung dar. Die Ausbildung einer Haarnadel-Sekundärstruktur während der Synthese kann den Zugang der Reagenzien zum reaktiven Ende der DNA deutlich erschweren. Um einer geringen Ausbeute vorzubeugen, wurden die Phosphoramidite von mdC, hmdC, fdC und cadC in einer um 25-50 % erhöhten Konzentration eingesetzt. Zudem wurden die Reaktionszeiten für alle modifizierten Phosphoramidite auf 3 min vervierfacht. Eine zusätzliche Durchmischung der Reaktionsgemische während der Kupplung wurde erreicht, indem die Reaktionsgemische abschließend acht Mal für je 0,1 s bewegt wurde. Die Reaktionszeit zwischen diesen Durchmischungs-Schritten betrug jeweils 4 s. Zur Steigerung

der Gesamtmenge erfolgte die Synthese im 1 μmol -Maßstab, an CPG- (*controlled pore glass*-) Festphasenträgern.

Die Synthese randomisierter Sequenzen stellte eine Herausforderung dar. Der Einbau randomisierter Positionen wurde durch manuelle äquimolare Mischung der entsprechenden Phosphoramidit-Bausteine erzielt. So wurden etwaige Einflüsse des Geräts zur automatisierten FPS, wie unterschiedliche Flussraten für verschiedene Phosphoramidit-Lösungen, ausgeschlossen. Die randomisierten Positionen sind in Abb. 25 blau hervorgehoben. Randomisierungen stellen in vielerlei Hinsicht eine Herausforderung dar. Bei der Oligonukleotid-Synthese führt die Randomisierung von Positionen dazu, dass der Einbau dieser Phosphoramidite nicht überwacht werden kann. Da jedes Phosphoramidit eine DMT-Schutzgruppe trägt, erlaubt die Detektion der abgespaltenen DMT-Schutzgruppe eine Aussage über den Erfolg des Kupplungs-Schrittes. Dazu wird das Konzentrationsmaximum der DMT-Schutzgruppe während der DMT-Entschützung bestimmt. Das DMT-Konzentrationsmaximum hängt von der Kinetik der Entschütungsreaktion ab, welche sich stark zwischen verschiedenen Phosphoramiditen unterscheidet. Im Falle randomisierter Positionen wird das DMT-Konzentrationsmaximum lediglich als Mittelwert der randomisierten Phosphoramidite erhalten. Dadurch konnte keine Überwachung der Synthese randomisierter Sequenzen erfolgen



Abb. 25 HP-Sequenzen und Benennung. N = A/T/G/C, D = A/T/G, W = A/T. 5: mdC, 6: hmdC, 7: fdC, 8: cadC, grün: Ligationsüberhang, rot: Referenznukleoside, blau: Identifikationssequenz.

6.2.1.3.2 Entschützung

Randomisierte Sequenzen erschwerten die Überwachung der sauren Entschützung. Die Abspaltung der Oligonukleotide von der Festphase erfolgte gemeinsam mit der Entschützung durch $\text{NaOH}_{(\text{aq})}$. Zur Entschützung der Säure-labilen Acetal-Schutzgruppe des fdC-Phosphoramidit-Bausteins erfolgte anschließend eine Entschützung mittels $\text{HOAc}_{(\text{aq})}$. Durch Randomisierung wurden je Festphasen-Träger im Idealfall tausende unterschiedliche Oligonukleotide erzeugt. Für das DdeI HP z.B. bis zu 26 244 Oligonukleotide. Deren

unterschiedliche Zusammensetzung führte zur Verteilung über einen Massenbereich von etwa $\Delta m/z$ 1000, mit Maxima zwischen m/z 8500 und m/z 9500, siehe Abb. 26. Da die Entschützung der Acetal-Schutzgruppe lediglich zu einer Massendifferenz von m/z 59 führt, war eine Überwachung des Entschützens-Fortschritts sowie möglicher Depurinierungen nicht möglich. Auch die Charakterisierung der Zielprodukte war so nur sehr eingeschränkt möglich.

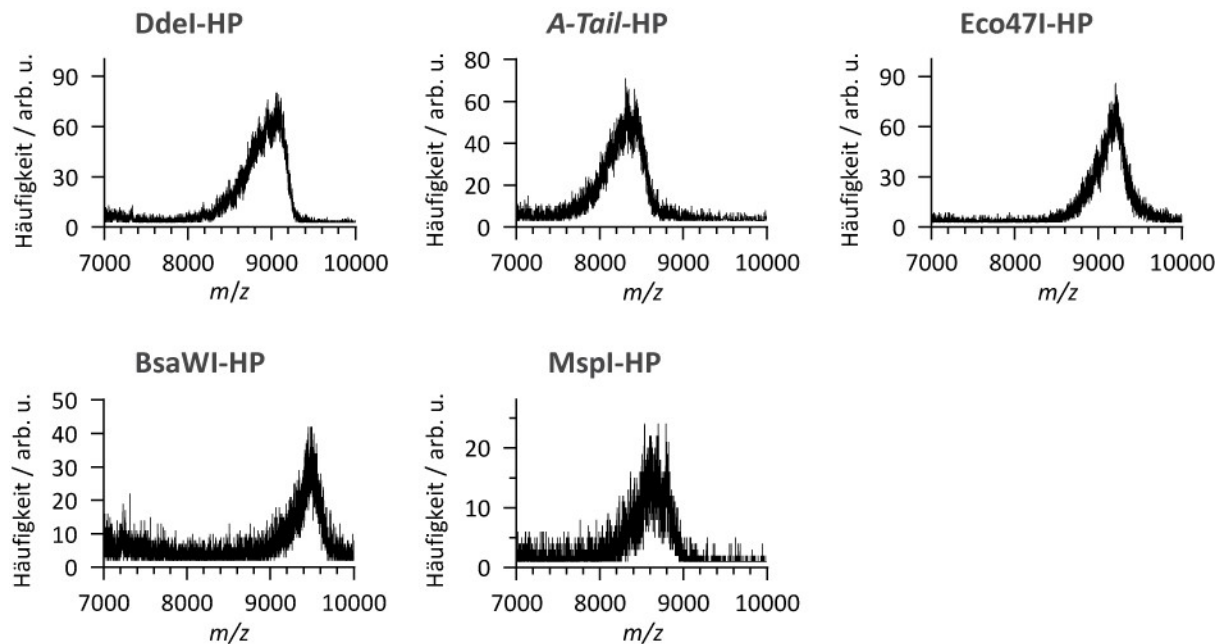


Abb. 26 Die randomisierten Positionen der HPs erschwerten aufgrund der Massen-Verteilung die Analytik mittels MALDI-TOF-MS: Dargestellt sind die MALDI-TOF-MS-Spektren der HPs nach Aufreinigung mittels RP-HPLC.

6.2.1.3.3 Aufreinigung

Die randomisierten Sequenzen erschwerten die Aufreinigung der HPs mittels RP-HPLC. Im Anschluss an die Entschützung der Oligonukleotide, wurden die Rohgemische zunächst mittels analytischer RP-HPLC (Umkehrphasen-HPLC) und MALDI-TOF-MS charakterisiert. Dabei zeigte sich, dass die Vielzahl verschiedener Oligonukleotide je Festphasen-Träger ein breites Spektrum verschiedener Retentionen besaß. Durch die Verteilung der Retentionszeiten über einen Bereich von etwa 10 min, war eine gezielte Abtrennung von Abbruchfragmenten und Nebenprodukten der Oligonukleotid-Synthese kaum möglich. Stattdessen wurde ein Kernbereich identifiziert, der den Großteil der Ziel-Oligonukleotide enthielt. Dazu erfolgte zunächst eine kleinteilige Fraktionierung des Rohgemischs mittels präparativer RP-HPLC. Anschließend wurden alle relevanten Fraktionen mittels MALDI-TOF-MS analysiert. Dies ermöglichte eine Abschätzung des Anteils an Abbruchfragmenten und Nebenprodukten gegenüber Ziel-Oligonukleotiden. Kombiniert wurden daraufhin lediglich die Fraktionen, die einen hohen Anteil an

Ziel-Oligonukleotiden enthielten. Die Aufreinigungen der HPs erfolgte in Zusammenarbeit mit MATTHIAS KURZ und wurden durch ihn näher beschrieben.^[246]

Ionenaustausch-Chromatographie könnte die Aufreinigung der HPs in Zukunft vereinfachen. Bei der Aufreinigung mittels Ionenaustausch-Chromatographie erfolgt die Auftrennung relativ zum Ladungszustand der Substanzen. Der Ladungszustand aller Ziel-Oligonukleotide einer Synthese ist identisch und unterscheidet sich von allen relevanten Verunreinigungen. Im Gegensatz zur RP-HPLC erfolgt die Auftrennung mittels Ionenaustausch-Chromatographie unter denaturierenden Bedingungen. Somit haben randomisierten Sequenzen und Sekundärstrukturen keinen Einfluss auf die Retention der Oligonukleotide. Ionenaustausch-Chromatographie wurde jedoch nicht eingesetzt, um die teuren HPLC-Geräte zu schützen. Die Ionenaustausch-Chromatographie benötigt konzentrierte Salz-Lösungen als mobile Phase. Diese Salzlösungen lassen sich häufig nicht restlos aus dem Pumpensystem der HPLC-Geräte entfernen. Dies kann sowohl zu irreversiblen Schäden am Gerät als auch zur Verunreinigung folgender Auftrennungen führen. Sollte ein geeignetes Gerät zur Verfügung stehen, könnte Ionenaustausch-Chromatographie die Aufreinigung der HPs in Zukunft stark vereinfachen.

6.2.2 *Spike-ins*: Referenz-Oligonukleotide für die Sequenzierung von mdC, hmdC, fdC und cadC

Dieses Projekt adressiert das Fehlen klar definierter Referenz-Oligonukleotide bei der Sequenzierung von mdC, hmdC, fdC und cadC. BS-Seq ist auf Referenznukleoside angewiesen, um eine verlässliche Interpretation der Ergebnisse zu ermöglichen, vgl. Einleitung Abschnitt 3.2. Üblicherweise werden dazu kurze Oligonukleotide eingesetzt, sogenannte *spike-ins*. Da *spike-ins* in der Regel mittels PCR hergestellt werden, ist der gezielte Einbau von dC und dessen Modifikationen nicht möglich. Um dieses Problem zu adressieren, wurden in einem Kooperationsprojekt zwischen dem AK WALTER (Universität des Saarlandes) und dem AK CARELL (LMU München) *spike-ins* entwickelt, die sowohl dC als auch mdC, hmdC, fdC und cadC in definierten Sequenzkontexten enthalten. Dies war durch deren Darstellung mittels Oligonukleotid-Festphasensynthese möglich. Trotz sehr anspruchsvoller Synthesen und Aufreinigungen gelang die Herstellung von neun *spike-in*-Oligonukleotiden in hochreiner Form. Abschließend konnte die Eignung der *spike-ins* als Referenz für die BS-Seq bestätigt werden. Dies erfolgte durch BS-Behandlung und anschließende Tiefensequenzierung.

6.2.2.1 Konzept

Das Design unserer *spike-ins* ermöglicht die gleichzeitige Kontrolle von mdC-, hmdC-, fdC- und cadC-Sequenzierungen. Um trotz der Länge und des hohen Modifikationsgrades unserer *spike-ins* praktikable Ausbeuten der Festphasensynthese zu ermöglichen, wurde jedes *spike-in* aus vier kürzeren *spike-in*-Oligonukleotiden hergestellt. SASCHA TIERLING (AK WALTER) entwarf 9 *spike-in*-Oligonukleotide, deren flexible Kombinierbarkeit die Erzeugung 24 doppelsträngiger *spike-ins* ermöglicht. Abb. 27 A) zeigt schematisch den Aufbau der *spike-ins*: Einzelsträngige *spike-ins* setzen sich aus *spike-in*-Oligonukleotiden gleichen Typs (A oder B) sowie gerader und ungerader Zahl zusammen: Siehe z. B. *spike-in* **A1-A2+B1-B2** in Abb. 27 B). Die Derivatisierung der dC-Nukleotide zur Sequenzierung gelingt selten quantitativ. In Folge dessen führen nicht umgesetzte Nukleotide zur Verzerrung des Sequenzier-Ergebnisses. Unsere *spike-ins* adressieren dieses Problem: Sie enthalten neben dC auch mdC, hmdC, fdC sowie cadC als Referenznukleoside. So ist die parallele Bestimmung der Umwandlungsraten aller dC-Nukleotide möglich. Unsere *spike-ins* ermöglichen es damit diese bedeutende Verzerrung des Sequenzier-Ergebnisses statistisch zu korrigieren. Darüber hinaus decken unsere *spike-ins* diverse Sequenz-Kontexte ab: Vgl. Abschnitt 6.2.2.4.1 der experimentellen Details für einen Überblick aller *spike-in*-Sequenzen sowie deren Sequenz-Kontexte.

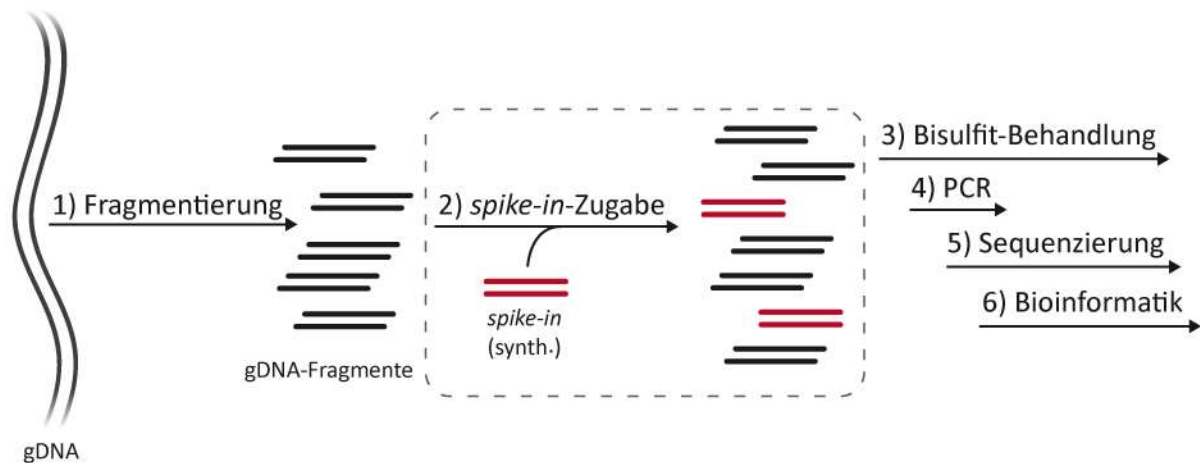


Abb. 28 Ablauf der BS-Seq unter Verwendung von *spike-ins*. Die Anwendung besteht in der Zugabe der *spike-ins* zur fragmentierten DNA-Probe (umrahmt). Alle weiteren Schritte entsprechen der regulären BS-Sequenzierung. gDNA: genomische DNA; *spike-in* (synth.): synthetische *spike-ins*.

6.2.2.2 Sequenzierung

Spike-in A3-A4+B3-B4 wurde mittels BS-Tiefensequenzierung erfolgreich validiert. Zur Validierung unseres *spike-in*-Konzepts, behandelte SASCHA TIERLING *spike-in A3-A4+B3-B4* mit Bisulfit unter *OPTI-Bisulfite*-Bedingungen^[184]. Anschließend erfolgte dessen Tiefensequenzierung. Die Auswertung des *spike-in*-Einzelstrangs **A3-A4** ist in Abb. 29 als gestapeltes Säulendiagramm dargestellt. Gelb kennzeichnet den Anteil der nicht desaminierten Positionen (als C sequenziert), violett den Anteil der desaminierten Positionen (als T sequenziert). dC wurde in beiden Sequenzkontexten nahezu quantitativ desaminiert, ebenso wie cadC. mdC und hmdC reagieren unter *OPTI-Bisulfite*-Bedingungen nicht mit BS.^[184] Entsprechend erscheinen beide Modifikationen nahezu quantitativ gelb, also nicht desaminiert. fdC zeigt ein differenziertes Bild: 95,3 % dieser Position wurden als T sequenziert, also desaminiert. Dieses Ergebnis ist repräsentativ für die geringe BS-Reaktivität von fdC und bildet damit die bekannten Verhältnisse ab.^[184] Ohne Derivatisierung von fdC mit einem geeigneten Reagenz ist dessen Sequenzierung demnach nicht verlässlich möglich.

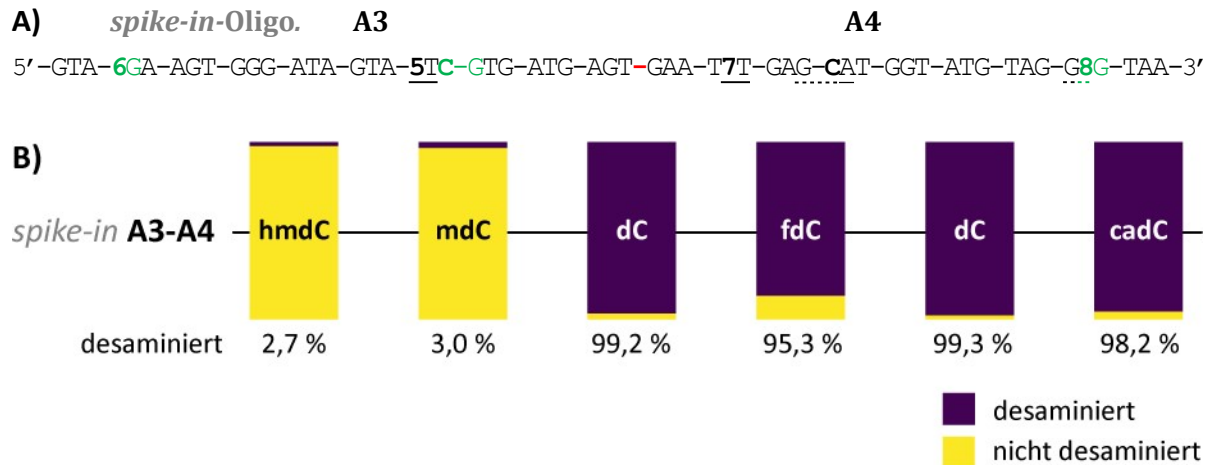


Abb. 29 A) Sequenz des *spike-ins* A3-A4 und Hervorhebung der Sequenz-Kontexte: CpG, CpW, GpC. 5: mdC, 6: hmdC, 7: fdC, 8: cadC. - Ligations-Stelle. **B)** Desaminierungsraten der BS-Tiefensequenzierung von *spike-in* A3-A4, als gestapeltes Säulendiagramm. violett: desaminiert (als T sequenziert), gelb: nicht desaminiert (als C sequenziert). Die Behandlung erfolgte durch Zugabe von 200 pg des doppelsträngigen *spike-ins* A3-A4+B3-B4 zu genomischer DNA. BS-Seq erfolgte unter OPTI-Bisulfite-Bedingungen.^[184] Die Auswertung von B3-B4 ist nicht dargestellt.

Nach BS-Behandlung, Adapter-Ligation und PCR-Amplifikation wurde Concatemer-Bildung beobachtet. Abb. 30 zeigt die Auftrennung der PCR-Produkte nach BS-Behandlung, Adapter-Ligation und PCR-Amplifikation von *spike-in* A3-A4+B3-B4 mittels Gelelektrophorese. Abb. 30 dient als repräsentatives Gelbild: Unterhalb von 300 bp ist eine Bande erkennbar, die sowohl das *spike-in* als auch die PCR-Primer-Dimere enthält. Direkt unterhalb von 400 bp sowie 500 bp ist je eine Bande zu beobachten, die auf die Concatemer-Bildung hindeuten. Concatemer-Bildung bezeichnet eine Zusammenlagerung zu repetitiven Sequenzen. Effekte dieser Concatemer-Bildung auf die nachfolgende Sequenzierung wurden nicht beobachtet. Demnach kann angenommen werden, dass die Bildung von Concatemeren durch *spike-in* A3-A4+B3-B4 deren Einsatz nicht beeinträchtigt.

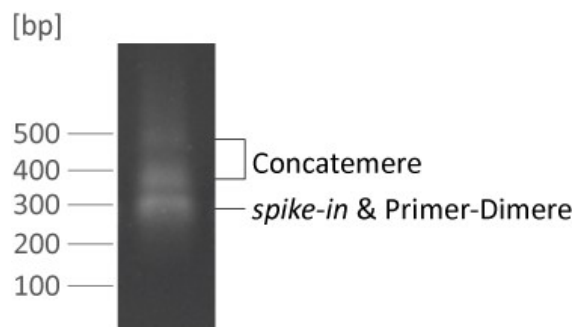


Abb. 30 Gelelektrophorese zeigt Hinweise auf Concatemer-Bildung von *spike-in* A3-A4+B3-B4 nach Adapter-Ligation, BS-Behandlung und PCR.

6.2.2.3 Zusammenfassung, Einordnung und Ausblick

Zusammengenommen ermöglichen unsere *spike-ins* die gleichzeitige Kontrolle von mdC-, hmdC-, fdC- und cadC-Umsetzungen in verschiedenen definierten Sequenz-Kontexten. Unsere 9 *spike-in*-Oligonukleotide können flexibel zu 24 doppelsträngigen *spike-ins* kombiniert werden. Deren Einsatz ermöglicht sowohl die Kontrolle von BS-basierten Sequenzierungen, als auch beliebiger anderer Reagenzien. Das *spike-in*-Design adressiert dabei die typischen Nachteile herkömmlicher *spike-ins*. Unsere *spike-ins* ermöglichen damit eine wirkungsvolle und reproduzierbare Kontrolle von dC-Modifikations-Sequenzierungen bei simpler Anwendung.

Unsere *spike-ins* könnten einen wichtigen Beitrag zur Etablierung von Sequenzierungs-Methoden der 3. Generation leisten. Sequenzierungen mittels SMRT und insbesondere mittels Oxford Nanopore sind in großem Maße vom jeweiligen Sequenz-Kontext abhängig. Unsere *spike-ins* enthalten nicht nur dC sondern auch mdC, hmdC, fdC sowie cadC in diversen Sequenz-Kontexten. Damit ließen sich die Algorithmen zur Detektion dieser dC-Modifikationen ideal trainieren. Dies könnte einen wichtigen Beitrag zur verlässlichen Detektion der dC-Modifikation wie auch der kanonischen DNA-Nukleotide leisten. Zur Anwendung bei Oxford Nanopore-Sequenzierungen können unsere *spike-ins* an einen geeigneten DNA-Abschnitt ligiert werden. Um die für SMRT-Sequenzierungen notwendigen *dumbbells* zu erzeugen, können unsere *spike-ins* mit den in dieser Arbeit entwickelten Haarnadel-Oligonukleotiden kombiniert werden, vgl. Abschnitt 6.2.1.2.

6.2.2.4 Experimentelle Details

Dieser Abschnitt gibt einen Überblick der experimentellen Details dieses Projekts. Die im Abschnitt 6.2.2.4.1 vorgestellten Sequenzen und Sequenz-Kontexte ermöglichen den Anwendern unserer *spike-ins* die Kombination der *spike-in*-Oligonukleotide gemäß den Anforderungen des individuellen Projekts. Anschließend wird die Herstellung der *spike-in*-Oligonukleotide beschrieben, bevor in Abschnitt 6.2.2.4.4 eine günstige und schnelle Methode zur Auswertung unserer *spike-ins* vorgestellt wird.

6.2.2.4.1 Sequenz-Kontexte der *spike-ins*

Unsere *spike-ins* decken diverse Sequenz-Kontexte ab. Cytosin-Modifikationen treten häufig im CpG-Kontext auf, aber auch weitere Sequenz-Kontexte sind relevant, vgl. Einleitung Abschnitt 3.1. Zur besseren Übersicht wurden die Sequenz-Kontexte in Abb. 31 und Tab. 6

Tab. 6 Aufschlüsselung der Sequenz-Kontexte aller *spike-in*-Oligonukleotide. ✓ = vorhanden. W = T oder A. Die Markierungen **CpG**, CpW, GpC dienen der Zuordnung zu **Abb. 31**. C: dC, mC: mdC, hmC: hmdC, fC: fdC, caC: cadC.

spike-in-Oligo.	Modifikationen im Sequenzkontext														
	CpG					<u>CpW</u>					<u>GpC</u>				
	C	mC	hmC	fC	caC	C	mC	hmC	fC	caC	C	mC	hmC	fC	caC
A1	✓		✓			✓	✓		✓		✓			✓	
A2		✓			✓	✓			✓		✓				✓
A3	✓		✓				✓								
A4					✓	✓			✓		✓				✓
B1	✓		✓			✓	✓			✓	✓		✓		
B2	✓			✓	✓	✓		✓			✓				
B3				✓		✓				✓	✓		✓		
B4	✓					✓									
B5	✓				✓	✓	✓	✓			✓				✓

Das modulare Konzept unserer *spike-ins* ermöglicht es Referenznukleoside und Sequenz-Kontexte gezielt für die jeweilige Anwendung zusammenzustellen. Tab. 7 präsentiert *spike-ins* für diverse Anwendungen: Spike-in **A1-A2+B1-B2** deckt die maximale Anzahl an Sequenz-Kontexten ab. Sollen lediglich Modifikationen im CpG-Kontext untersucht werden, so stellt **A3-A2+B0-B2** das geeignete *spike-in* dar. Geeignet bezieht sich dabei auf die Vermeidung nicht benötigter Sequenz-Kontexte. Sollen nur Modifikationen im CpW-Kontext untersucht werden, so stellen die *spike-ins* **A0-A4+B1-B2** sowie **A1-A0+B3-B2** die geeigneten Varianten dar. Zur Sequenzierung von Modifikationen im GpC-Kontext ist **A1-A4+B3-B0** die geeignete Variante.

Tab. 7 Aufschlüsselung doppelsträngiger *spike-ins* zur Analyse bestimmter Sequenz-Kontexte. Adressierte Sequenz-Kontexte sind fett hervorgehoben. ✓ = vorhanden. W = T oder A. Die Markierungen CpG, CpW, GpC dienen der Zuordnung zu **Abb. 31**. C: dC, mC: mdC, hmC: hmdC, fC: fdC, caC: cadC. **0** kennzeichnet *spike-in*-Oligonukleotide ohne mdC, hmdC, fdC, cadC, die kommerziell erworben werden können. – kennzeichnet Ligationen.

<i>spike-in</i> (ds)	Modifikationen im Sequenzkontext														
	CpG					CpW					GpC				
	C	mC	hmC	fC	caC	C	mC	hmC	fC	caC	C	mC	hmC	fC	caC
A1-A2 B1-B2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
A3-A2 B0-B2	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓				✓
A0-A4 B1-B2	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓		✓		✓
A1-A0 B3-B2	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	
A1-A4 B3-B0	✓		✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓

6.2.2.4.2 Oligonukleotid-Synthese

Die Herstellung der *spike-ins* gliederte sich in sechs Schritte: 1) Phosphoramidit-Synthese, 2) Oligonukleotid-Synthese, 3) Entschützung, 4) Aufreinigung, 5) Hybridisierung und Ligation. Die Etablierung der *spike-in*-Herstellung wurde als Kooperations-Prozess geplant: Die Synthese der nicht kommerziellen Phosphoramidite von hmdC, fdC und cadC erfolgte durch KRISTOF HUFNAGEL (AK CARELL, LMU München) in Anlehnung an SCHRÖDER *et al.*^[247]. Die Etablierung der Schritte 2), 3) und 4) erfolgte durch mich in Zusammenarbeit mit MATTHIAS KURZ (AK CARELL). Im Anschluss an die Herstellung der *spike-in*-Oligonukleotide erfolgte deren Hybridisierung und Ligation durch SASCHA TIERLING (AK WALTER, Universität des Saarlandes).

Alle Oligonukleotide wurden mittels FPS mit 5'-terminalen Phosphatgruppen versehen. Zur Herstellung der *spike-ins* ist eine Ligation der *spike-in*-Oligonukleotide nötig. Um die flexible Ligation aller Oligonukleotide zu ermöglichen und enzymatische Phosphorylierungen zu vermeiden, wurden alle Oligonukleotide mittels FPS 5'-terminal phosphoryliert. Dies erfolgte unter Verwendung des Phosphorylierungsreagenzes *solid CPR II 80* der Firma LinkTech, siehe Abb. 32. Die Kupplung des Phosphorylierungsreagenzes erfolgte analog zur Kupplung der Phosphoramidite aller dC-Modifikationen. 5'-Phosphorylierungen mittels FPS sind aus zwei

Gründen anspruchsvoll: Zum einen erfolgen sie am Ende der FPS mit vergleichsweise geringer Ausbeute. Jede fehlgeschlagene Phosphorylierung führt demnach zur Veresterung des prä-finalen Oligonukleotids (*capping*). Veresterte Oligonukleotide können nicht verwendet werden und führen somit zum Verlust des Oligonukleotids und aller eingebauten Phosphoramidite. Dies ist insbesondere relevant für die aufwändig synthetisierten Phosphoramidite von hmdC, fdC und cadC. Darüber hinaus stellen 5'-terminale Phosphatgruppen auch für die Aufreinigung der Oligonukleotide eine Herausforderung dar, siehe folgender Abschnitt 6.2.2.4.3.

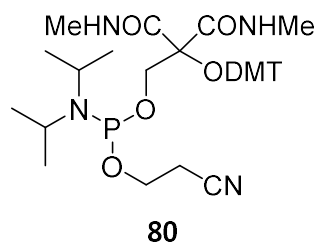


Abb. 32 Phosphorylierungsreagenz *solid CPR II 80* – Die 5'-Phosphorylierung aller *spike-in*-Oligonukleotide erfolgte via Festphasen-Synthese unter Verwendung von *solid CPR II 80*.

Die Synthese des *spike-in*-Oligonukleotids **B4** erforderte die Verwendung eines universellen Festphasenträgers (*universal support*). Die Oligonukleotid-Festphasen-Synthese erfolgt unter Verwendung von Festphasenträgern in 3'>5'-Richtung. Konventionelle Festphasenträger sind dazu bereits mit dem 3'-terminalen Nucleosid beladen. So können die finalen Oligonukleotide im Anschluss an die FPS unter milden kompatiblen Bedingungen von der Festphase abgespalten werden. Da das *spike-in*-Oligonukleotid **B4** ein 3'-terminales mdC aufweist, war eine Verwendung konventioneller Festphasenträger nicht möglich, siehe Abb. 33 A). Stattdessen wurden zwei universelle Polystyrol-Q-Festphasenträger der Firma *LinkTech* im 0,2 µmol-Maßstab eingesetzt. Diese enthalten lediglich die Festphase, aber kein 3'-terminales Nucleosid. Aufgrund der veränderten Detritylierungs-Kinetik des universellen Festphasenträgers, war eine Anpassung der Detritylierungs-Dauer notwendig. Dazu wurde die Detritylierungs-Dauer von 73 s für modifizierte Phosphoramidite auf 126 s für die Detritylierung des universellen Festphasenträgers gesteigert. So gelang die Synthese des 3'-mdC-modifizierten *spike-in*-Oligonukleotids **B4** in sehr hoher Ausbeute (nach Aufreinigung: 93,6 nmol, 23,4 %).

3'-5AT-GCT-TCA-CCC-TAT-CAT-GAG-CAA-TCC-5' *spike-in*-Oligonukleotid **B4**

Abb. 33 Sequenz des *spike-in*-Oligonukleotids **B4**. Die 3'-terminale Position des nicht-kanonischen Nucleosids mdC verlangte den Einsatz eines universellen Festphasenträgers.

Im Anschluss an die FPS erfolgte die Abspaltung der *spike-ins* von der Festphase sowie deren Entschützung. Eine detaillierte Beschreibung der Entschützung-Bedingungen findet sich in Abschnitt 8.2.

6.2.2.4.3 Aufreinigung

Die terminale Phosphatgruppe der *spike-in*-Oligonukleotide gestaltete die Aufreinigung mittels RP-HPLC besonders anspruchsvoll. Nach der Entschützung wurde das *spike-in*-Rohgemisch zunächst mittels RP-HPLC und MALDI-TOF-MS charakterisiert. Dabei zeigte sich, dass die terminale 5'-Phosphatgruppe der Oligonukleotide die Aufreinigung besonders anspruchsvoll gestaltete. Bei der Oligonukleotid-Aufreinigung mittels RP-HPLC, führt jedes zusätzliche kanonische Nukleotid zu einer *Erhöhung* der Retention, so dass das Ziel-Oligonukleotid üblicherweise die größte Retention aufweist. Die terminale 5'-Phosphorylierung führte hingegen zur *Verringerung* der Retention. Somit wiesen das Ziel-Oligonukleotid und diverse Nebenprodukte eine nahezu identische Retention auf. Abb. 34 zeigt beispielhaft das Chromatogramm des *spike-in*-Oligonukleotids **B2**. Es zeigt Verunreinigungen sowohl mit geringerer als auch größerer Retention als das Zielprodukt (größtes Signal). Dies erschwerte sowohl die Identifikation des Ziel-Oligonukleotids im HPLC-Chromatogramm als auch dessen Aufreinigung deutlich.

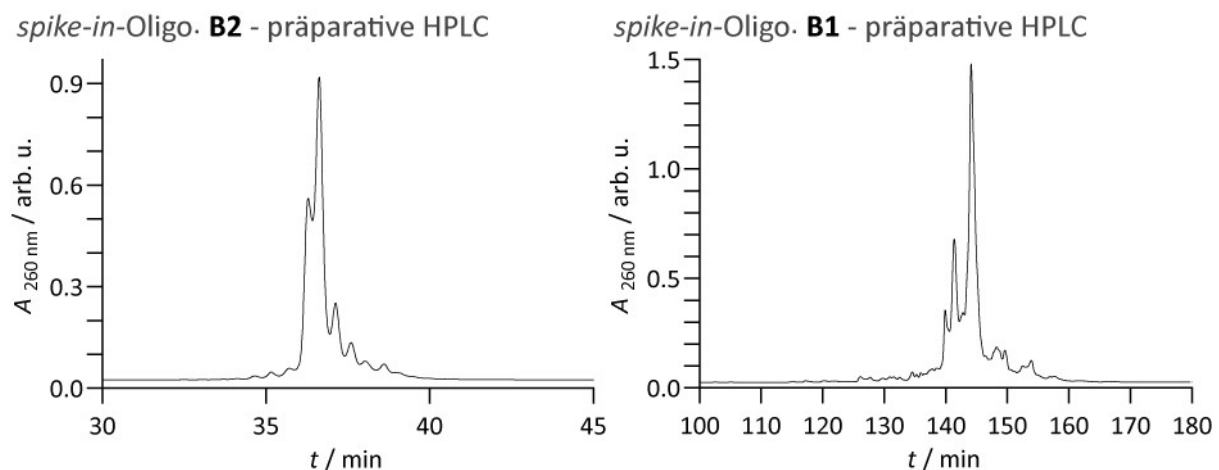


Abb. 34 RP-HPLC-Chromatogramme der *spike-in*-Oligonukleotide **B2** und **B1** vor der Aufreinigung.

Gestreckte Gradienten ermöglichten die Auftrennung komplexer Gemische. Um die ungewöhnlich geringen Unterschiede in der Retention zwischen Ziel-Oligonukleotid und Nebenprodukten zu adressieren, wurden bei Bedarf stark gestreckte RP-HPLC-Gradienten entwickelt. Dazu wurde die Länge der Gradienten von 45 min auf 120-180 min erhöht. Abb. 34 zeigt beispielhaft zwei

RP-HPLC-Chromatogramme vor der Aufreinigung: Die Retentions-Unterschiede im Rohgemisch von *spike-in*-Oligonukleotid **B2** ließ eine Aufreinigung ohne Streckung zu. Die geringen Retentions-Unterschiede im Rohgemisch von **B1** erforderten hingegen eine deutliche Streckung des RP-HPLC-Gradienten, um eine Auftrennung zu ermöglichen. Darüber hinaus wurden einige Aufreinigungen mehrfach wiederholt, um eine hohe Reinheit zu erreichen.

Die besonders anspruchsvolle Aufreinigung von *spike-in*-Oligo. **A2** erfolgte mittels analytischer RP-HPLC. Üblicherweise werden Oligonukleotid-Synthesen im 400 nmol-Maßstab mittels semi-präparativer HPLC aufgereinigt. Semi-präparative HPLC weist jedoch eine deutlich geringere Trennleistung auf als ein analytisches HPLC-System. Um auch die besonders anspruchsvolle Aufreinigung von *spike-in*-Oligonukleotid **A2** adäquat adressieren zu können, erfolgte dessen Aufreinigung mittels analytischer HPLC. Abb. 35 zeigt beispielhaft ein Chromatogramm dieser Aufreinigung. Die Abtrennung der Verunreinigungen wäre mittels semi-präparativer HPLC nicht möglich gewesen. Interessanterweise limitierte nicht die Säulen-Kapazität die Injektionsmenge, sondern das Detektionslimit des analytischen HPLC-Systems. So konnte der obere Teil des Signals nicht vom Detektor erfasst werden. Selbst die Aufreinigung von 10 nmol führte dabei nicht zu einer verringerten Trennleistung. Um trotz der geringen Detektions-Obergrenze auch Signal-Schultern detektieren zu können, wurde das aufzureinigende Reaktionsgemisch aliquotiert und separat aufgereinigt. Die Aufreinigungen wurden mehrfach wiederholt, um eine hohe Reinheit erzielen zu können.

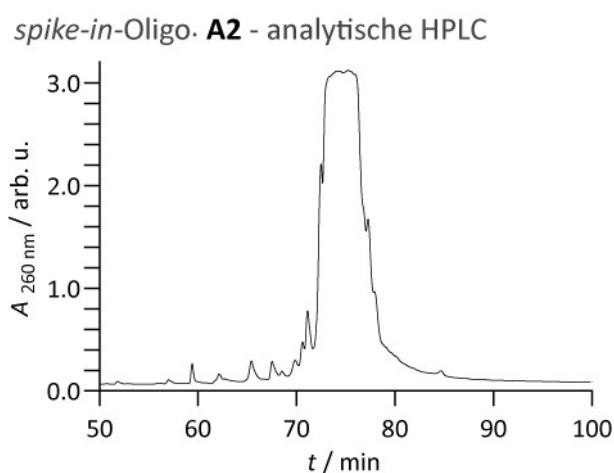


Abb. 35 RP-HPLC-Chromatogramm der Aufreinigung von *spike-in*-Oligonukleotid **A2**. Aufreinigung mittels analytischer RP-HPLC war in diesem Fall nötig. Der obere Teil des Signals überschritt das Detektions-Limit des Systems.

Alle *spike-in*-Oligonukleotide wurden in guter Ausbeute und hoher Reinheit erhalten. Tab. 8 fasst die Aufreinigungs-Strategien sowie die Ausbeuten aller *spike-in*-Oligonukleotide

zusammen. Es konnten Ausbeuten von 10 bis 94 nmol erzielt werden. Der Großteil der Oligonukleotide wurde trotz hoher Reinheit und zum Teil sehr anspruchsvoller Aufreinigung in Ausbeuten um 8 % erhalten. *spike-in*-Oligonukleotid **B4** konnte sogar in einer sehr guten Ausbeute von 23 % erhalten werden. Dies begründet sich vermutlich in dessen geringer Anzahl an Nukleosid-Modifikationen: **B4** enthält als einziges *spike-in*-Oligonukleotid ausschließlich die Modifikation mdC. Alle anderen *spike-in*-Oligonukleotide enthalten hingegen zwei bis drei Nukleosid-Modifikationen. Die geringere Effizienz des Einbaus von Phosphoramiditen modifizierter Nukleoside scheint demnach eine plausible Erklärung für die hohe Ausbeute von **B4** darzustellen. Lediglich zwei Oligonukleotide mit besonders anspruchsvoller Aufreinigung, **A2** und **A4**, wurden in Ausbeuten < 6 % erhalten. Abschließend erfolgte die finale Charakterisierung der Oligonukleotide mittels analytischer RP-HPLC und MALDI-TOF-MS. Abb. 36 A) und B) zeigen die entsprechenden Absorptions- und Massenspektren aller neun *spike-in*-Oligonukleotide.

Tab. 8 Aufreinigungs-Strategien und Ausbeuten aller *spike-in*-Oligonukleotide.

<i>spike-in</i> - Oligonukleotid	HPLC- System	gestreckt?	mehrfach aufgereinigt?	Ausbeute /	
				nmol	%
A1	semi-präparativ	-	✓	25,3	6,3
A2	analytisch	✓	✓	10,0	2,5
A3	semi-präparativ	✓	-	17,5	8,8
A4	semi-präparativ	-	✓	17,0	4,3
B1	semi-präparativ	✓	-	27,7	6,9
B2	semi-präparativ	-	-	28,1	7,0
B3	semi-präparativ	-	-	23,8	6,0
B4	semi-präparativ	-	-	93,6	23
B5	semi-präparativ	✓	✓	17,5	8,8

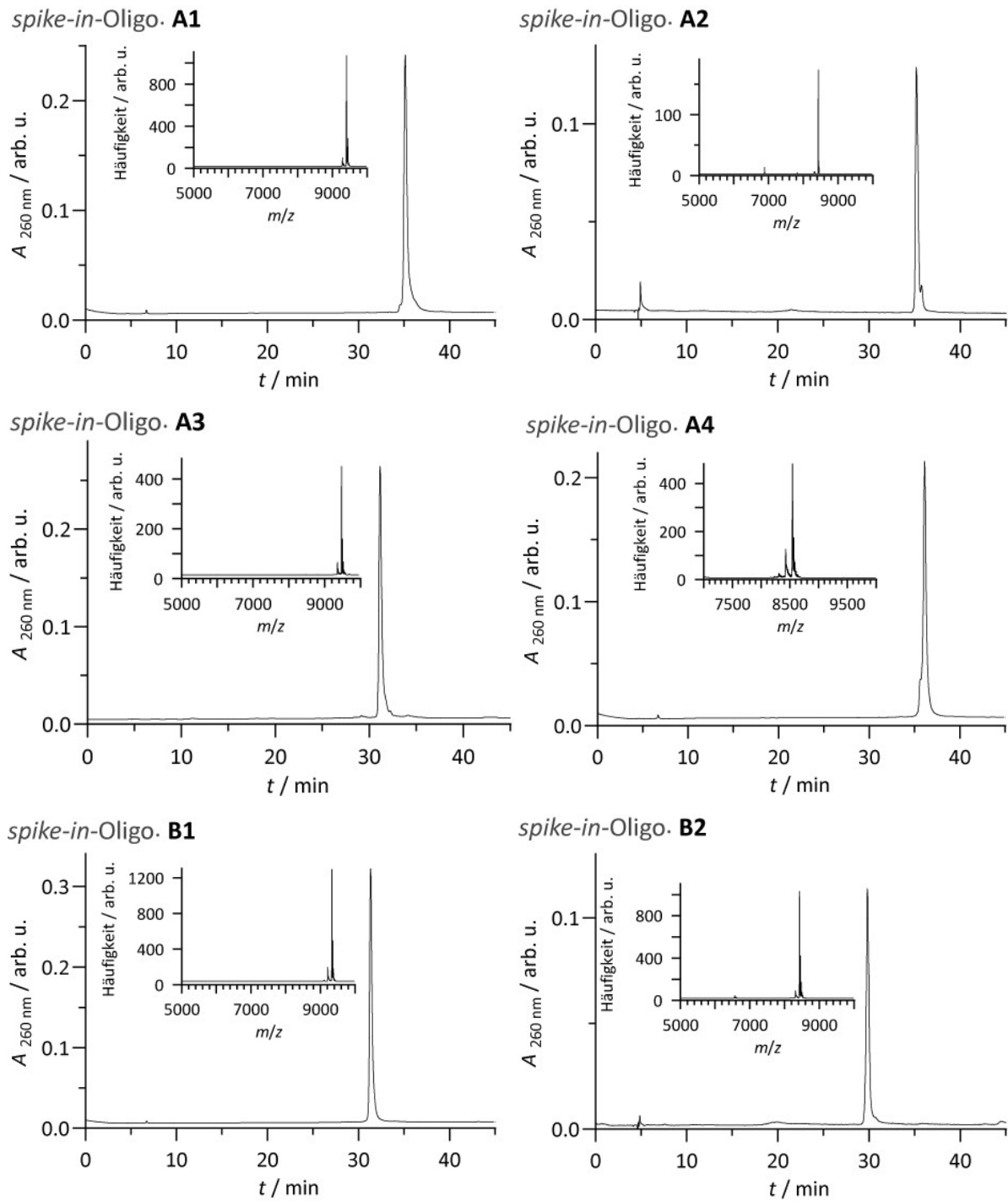


Abb. 36 Teil 1 – RP-HPLC-Chromatogramme und MALDI-TOF-Massenspektren der neun *spike-in*-Oligonukleotide nach deren Aufreinigung. $A_{260\text{ nm}}$: Absorption bei 260 nm. arb. u.: willkürliche Einheiten.

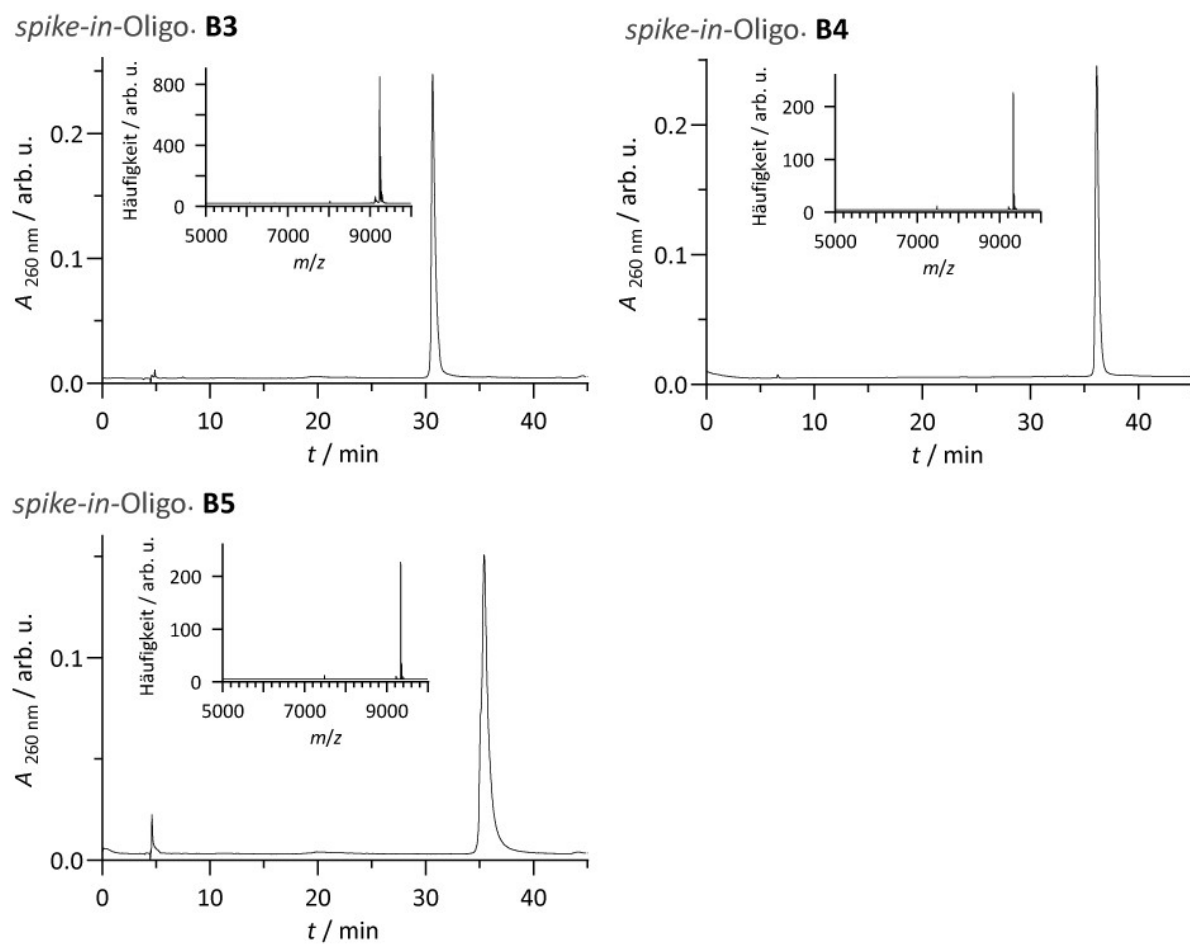


Abb. 37 Teil 2 – RP-HPLC-Chromatogramme und MALDI-TOF-Massenspektren der neun *spike-in*-Oligonukleotide nach deren Aufreinigung. $A_{260 \text{ nm}}$: Absorption bei 260 nm. arb. u.: willkürliche Einheiten.

In Zukunft könnte die *spike-in*-Aufreinigung durch enzymatische Phosphorylierung oder Ionenaustausch-Chromatographie vereinfacht werden. Ionenaustausch-Chromatographie ermöglicht die Auftrennung relativ zum Ladungszustand der Substanzen. Die terminale 5'-Phosphatgruppe der *spike-ins* führt zwar zu einer erhöhten Polarität und damit zu einer erschwerten Aufreinigung mittels RP-HPLC. Der Ladungszustand des Ziel-Oligonukleotids hingegen, unterscheidet sich klar von nahezu allen Verunreinigungen. Somit könnte die Ionenaustausch-Chromatographie eine geeignete Methode zur Aufreinigung der 5'-phosphorylierten *spike-ins* darstellen. Ionenaustausch-Chromatographie wurde in dieser Arbeit jedoch nicht eingesetzt, um die teuren HPLC-Geräte zu schützen. So benötigt die Ionenaustausch-Chromatographie konzentrierte Salz-Lösungen als mobile Phase. Diese Salzlösungen lassen sich häufig nicht restlos aus dem Pumpensystem der HPLC-Geräte entfernen. Dies kann sowohl zu irreversiblen Schäden am Gerät als auch zur Verunreinigung folgender Auftrennungen führen. Sollte ein geeignetes Gerät zur Verfügung stehen, könnte Ionenaustausch-Chromatographie die Aufreinigung der 5'-phosphorylierten *spike-ins* jedoch stark vereinfachen. Eine andere

zielführende Anpassung der *spike-in*-Herstellung könnte die enzymatische Phosphorylierung der *spike-in*-Oligonukleotide darstellen. Durch Synthese der *spike-ins* ohne terminale 5'-Phosphatgruppe wäre eine vergleichsweise simple RP-HPLC-Aufreinigung zu erwarten.

Zusammengenommen basieren die Vorteile unserer *spike-ins* auf deren Herstellung mittels Oligonukleotid-Festphasen-Synthese (FPS), die den gezielten Einbau verschiedener dC-Modifikationen erlaubt. Alle neun *spike-in*-Oligonukleotide konnten nach FPS, Entschützung und Aufreinigung in hoher Reinheit und guter Ausbeute erhalten werden. Der hohe Modifikationsgrad der Oligonukleotide sowie die via FPS installierte 5'-Phosphatgruppe, führten dabei zu größtenteils sehr anspruchsvollen Synthesen und Aufreinigungen. Die Synthesen im 2 x 200 µmol-Maßstab gelangen durch spezielle Kupplungsbedingungen sowie teilweise unter Verwendung universeller Festphasenträger. Die hohe Reinheit konnte durch Streckung der Gradienten, Aufreinigung mittels analytischer RP-HPLC sowie Mehrfach-Aufreinigungen erreicht werden. In Zukunft könnten eine enzymatische Phosphorylierung oder die Aufreinigung mittels Ionenaustausch-Chromatographie die Herstellung der *spike-ins* deutlich erleichtern.

6.2.2.4.4 *spike-in*-Auswertung mittels SNuPE

Primer ermöglichen die schnelle und günstige Auswertung unserer spike-ins. Für *spike-in A3-A4+B3-B4* wurden spezielle Primer entworfen, siehe Abb. 38 A). Diese Primer binden nach der PCR-Amplifikation der *spike-ins* unmittelbar vor diversen Referenznukleosiden. Da die BS-Seq zu einer Umwandlung von C zu T führt lässt sich, über den Einbau des komplementären Nukleotids (G oder A), die Identität dieser Referenz-Positionen auslesen. Diese Technik wird als *single nucleotide primer extension* (SNuPE) bezeichnet. Dabei werden die Primer enzymatisch jeweils um ein Nukleotid verlängert und mittels Gelelektrophorese oder HPLC analysiert. So können zentrale Informationen bereits vor der vergleichsweise aufwändigen und teuren Sequenzierung analysiert werden. Damit ist zum einen eine schnelle und günstige Qualitätskontrolle vor der Sequenzierung möglich. Darüber hinaus liefert die SNuPE für viele Anwendungen bereits verwertbare Ergebnisse. Abb. 38 B) zeigt den Ablauf der SNuPE anhand von *spike-in A3-A4*.

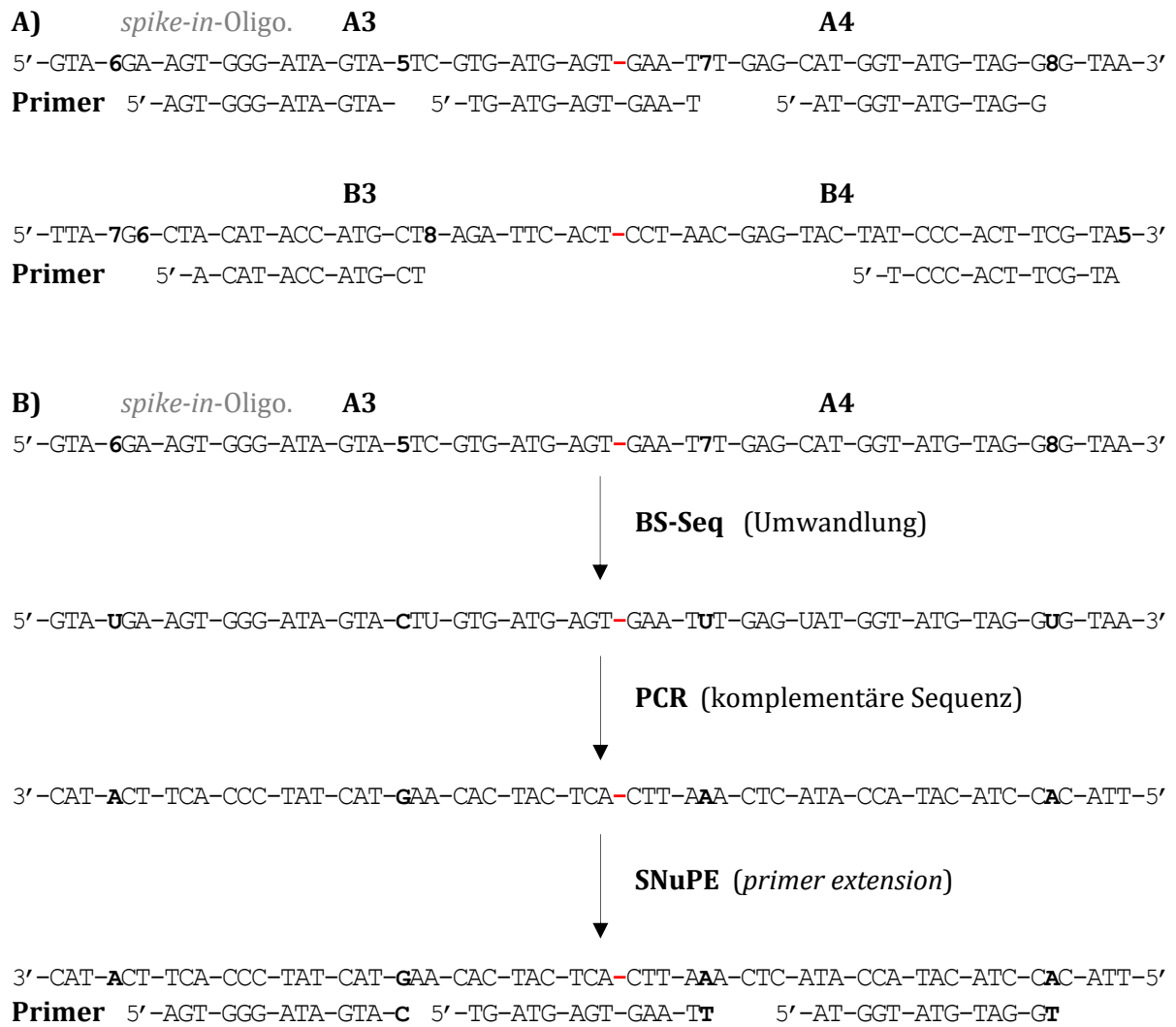


Abb. 38 A) Primer zur Analyse der *spike-ins* mittels SNuPE. Die so analysierbaren Referenznukleoside sind grün hinterlegt. **B)** Ablauf der *spike-in*-Analyse mittels SNuPE. Beispielhaft anhand von *spike-in* A3-A4. SNuPE: *single nucleotide primer extension*, BS-Seq: Bisulfit-Sequenzierung, **5**: mdC, **6**: hmdC, **7**: fdC, **8**: cadC, U: dU. - Ligation-Stelle.

6.2.3 Eine überlegene Methode zur fdC-Sequenzierung: fdC-HA-Seq

Die Sequenzierung von fdC birgt Herausforderungen, die in diesem Projekt adressiert wurden. Seit einigen Jahren existieren verschiedene Methoden zur fdC-BS-Sequenzierung. Die wohl gebräuchlichste dieser Methoden ist fCAB, eine fdC-Sequenzierung mittels *O*-Ethyl-Hydroxylamin. Entscheidend für die zuverlässige Sequenzierung mit Hilfe eines Reagenzes ist die effiziente und robuste Derivatisierung des Ziel-Nukleosids in genomischer DNA. Robust bezieht sich dabei auf die zuverlässige Derivatisierung in allen Sequenz-Kontexten. Der Anteil an derivatisiertem fdC liegt nach fCAB-Behandlung üblicherweise deutlich unterhalb einer quantitativen Umsetzung.^[194,195] Um eine verlässliche Sequenzierung von fdC zu etablieren, wurde in diesem Projekt ein im AK CARELL entwickeltes Hydroxylamin-Reagenz^[168] auf die BS-Seq von fdC angewendet. Die so entwickelte Methode wird hier als fdC-Hydroxylamin-Reagenz-Sequenzierung (fdC-HA-Seq) bezeichnet. Die Entwicklung erfolgte in Kooperation zwischen dem AK WALTER (*Universität des Saarlandes*) und dem AK CARELL (*LMU München*). Federführend involviert waren PASCAL GIEHR und MARTIN ROSSA. Substanziell beigetragen hat darüber hinaus MATTHIAS KURZ. Des Weiteren involviert waren KRISTOF HUFNAGEL, FABIO SPADA, RENÉ RAHIMOFF, EVA KORYTIAKOVÁ, YINGQIAN ZHANG und SASCHA TIERLING.

6.2.3.1 Hydroxylamin-Reagenz

Formyl-Gruppen in DNA können mit einem im AK CARELL entwickelten Reagenz detektiert werden. Hydroxylamine besitzen im Allgemeinen eine hohe Reaktivität gegenüber Formyl-Gruppen. Die Umsetzung vereinzelter Formyl-Gruppen in genomischer DNA stellt hingegen besondere Anforderungen: Insbesondere die geringe Quantität der Formyl-Gruppen sowie die physikochemischen Eigenschaften der genomischen DNA machen eine verlässliche Umsetzung in allen Sequenz-Kontexten schwierig. Um dies zu adressieren, wurde im AK CARELL ein Hydroxylamin-Reagenz (HA) **81** mit hoher DNA-Affinität entwickelt, siehe Abb. 39. Dieses Reagenz wurde bisher zur genomweiten Detektion von abasischen Stellen eingesetzt.^[168] Dessen Anwendungsbereich sollte nun auch auf fdC erweitert werden, siehe Schema 51.

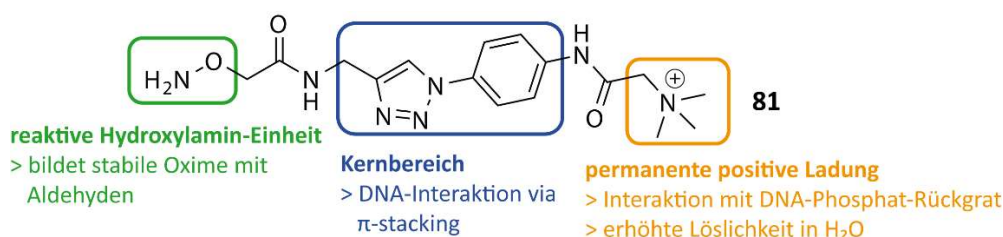
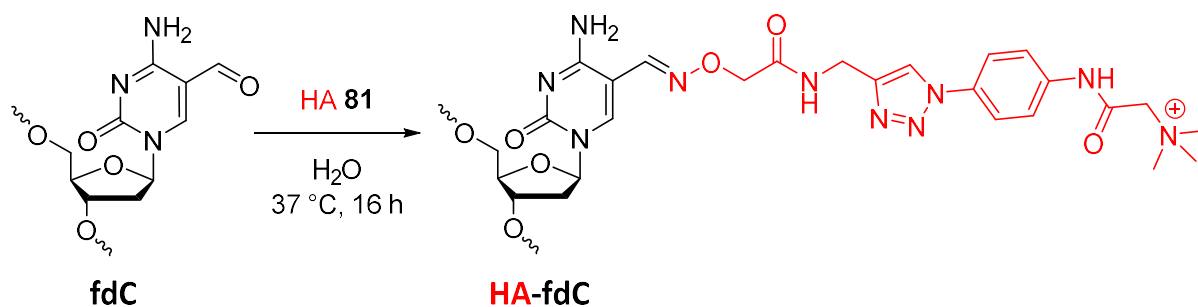


Abb. 39 Im AK CARELL entwickeltes^[168] Hydroxylamin-Reagenz **81** mit hoher DNA-Affinität zur Umsetzung von Formyl-Gruppen in genomischer DNA. Die drei funktionellen Einheiten sind farblich markiert.



Schema 51 Derivatisierung von fdC in DNA mit dem im AK Carell entwickelten^[168] Hydroxylamin-Reagenz **81**, unter milden Bedingungen. HA: Hydroxylamin-Reagenz.

6.2.3.2 Konzept

Die *fdC*-Sequenzierung mit Hydroxylamin-Derivaten basiert auf dem Blockieren der BS-vermittelten Desaminierung. Dass die Sequenzierung von fdC mit Hydroxylamin-Derivaten möglich ist, wurde bereits mittels fCAB demonstriert.^[191] So wurde gezeigt, dass fdC nach Hydroxylamin-Derivatisierung nicht mehr BS-vermittelt desaminiert wird, siehe Abb. 40. Demnach ermöglicht die vergleichende Sequenzierung einer Hydroxylamin-derivatisierten DNA sowie einer unbehandelten Referenzreaktion die Detektion von fdC.

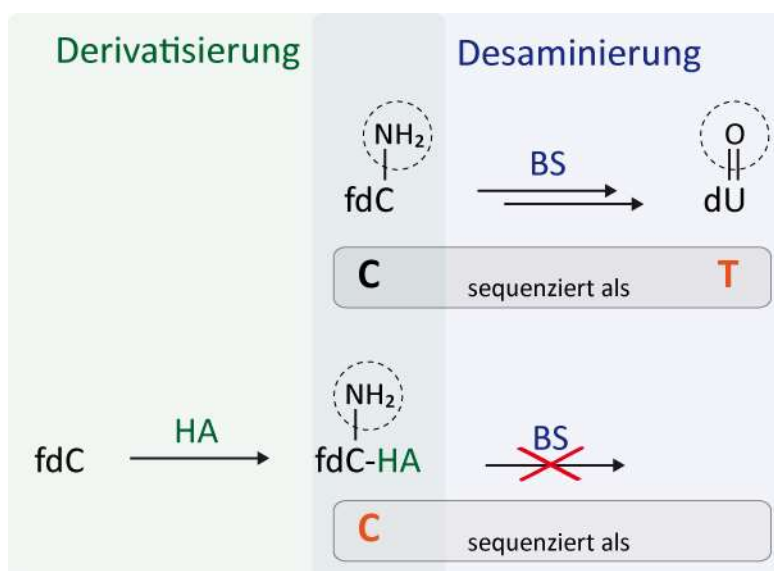


Abb. 40 Basis der vergleichenden fdC-Sequenzierung durch Hydroxylamin-Derivatisierung: Schematische Darstellung der fdC-Derivatisierung mit einem Hydroxylamin-Reagenz sowie der BS-vermittelten Desaminierung von fdC zu dU^[191,249]. HA: Hydroxylamin-Reagenz, BS: Bisulfit.

Der Schlüsselschritt der *fdC-HA-Seq* besteht in der Umsetzung mit unserem Hydroxylamin-Reagenz **81**. Der Prozess gliedert sich in drei Einheiten, siehe Abb. 41: Schritte der konventionellen BS-Seq, HP-Ligation (orange) sowie Hydroxylamin-Behandlung (grün). Abb. 41 zeigt den detaillierten Ablauf der *fdC-HA-Seq*: Zunächst wird die DNA-Probe mittels Restriktionsenzym-Verdau fragmentiert. Im zweiten Schritt erfolgt die optionale Ligation der in dieser Arbeit entwickelten Haarnadel-Oligonukleotide (HPs), vgl. Abschnitt 6.2.2.4. Da es sich bei der *fdC-HA-Seq* um eine vergleichende Sequenzierung handelt, erfolgt im nächsten Schritt die Aufteilung der Probe. Reaktion A) wird im Schlüsselschritt der *fdC-HA-Seq* mit unserem Hydroxylamin-Reagenz **81** behandelt. Reaktion B) bleibt als Referenz unbehandelt. Die Derivatisierung mit dem Hydroxylamin-Reagenz **81** findet bei milden 37 °C in H₂O über 16 h statt. In allen darauffolgenden Schritten werden beide Reaktionen identisch, aber getrennt voneinander, behandelt. Es folgt der reguläre Ablauf der konventionellen BS-Seq: BS-Behandlung, PCR, Sequenzierung und bioinformatische Auswertung.

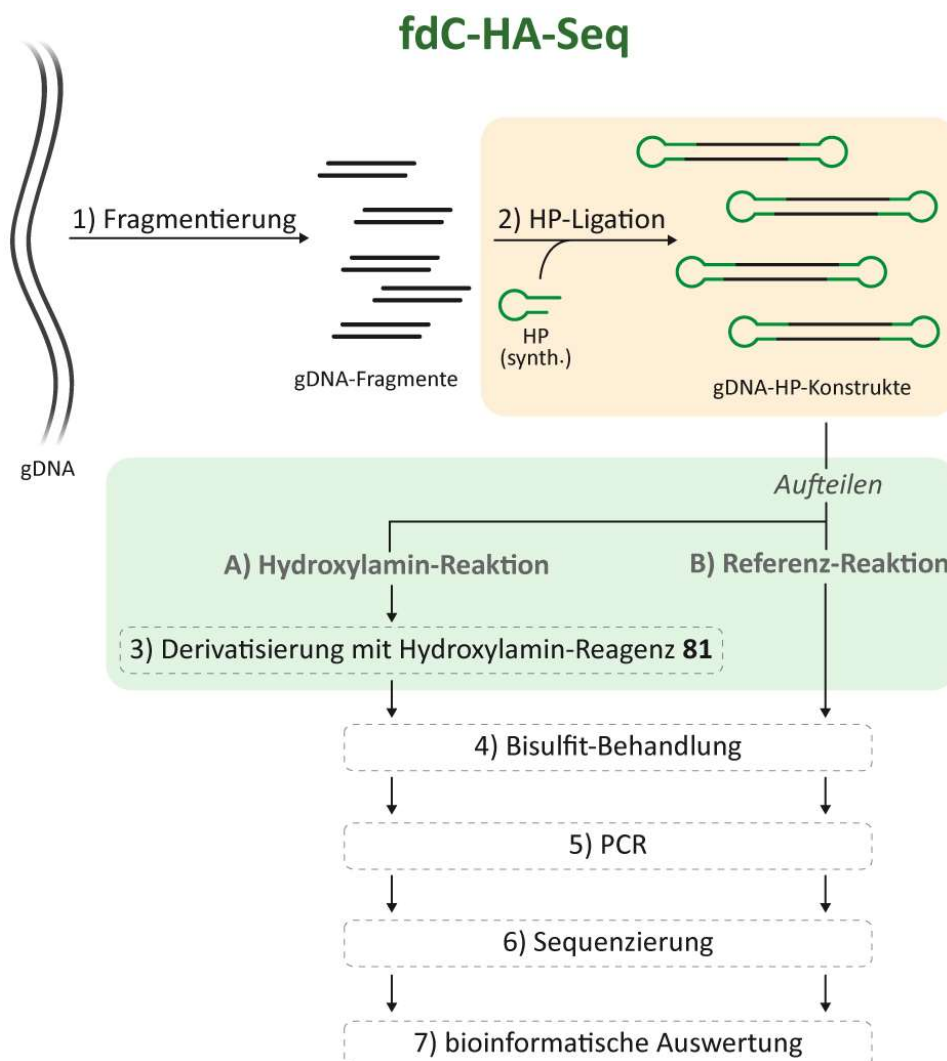


Abb. 41 Ablauf der *fdC-HA-Seq*. HP-Ligation: orange hinterlegt. HA-Derivatisierung: grün hinterlegt. Alle weiteren Schritte entsprechen der regulären BS-Sequenzierung. HP: Haarnadel-Oligonukleotid.

Die optionale Anwendung unserer HPs ermöglicht höchste Genauigkeit und Einzelstrang-spezifische Sequenzierung. Die Einführung des HPs ermöglicht drei zentrale Funktionen: **a)** Die Referenznukleoside ermöglichen die exakte Bestimmung der Desaminierungsrate von fdC sowie zusätzlich von dC, mdC, hmdC und cadC. **b)** In Kombination mit der individuellen Markierung jedes DNA-Fragments, kann die Sequenzierung von fdC in höchster Genauigkeit erfolgen. **c)** Darüber hinaus ermöglicht die kovalente Verknüpfung der beiden DNA-Einzelstränge durch das HP zudem die Einzelstrang-spezifische fdC-Sequenzierung. Somit kann zwischen hemi-, hybrid- und vollständig modifizierten Positionen unterschieden werden kann. Erst durch die Kombination der HP-Funktionen mit unserem Hydroxylamin-Reagenz **81** scheint eine verlässliche fdC-Sequenzierung möglich. Nichtsdestotrotz stellt die HP-Ligation einen optionalen Teil der fdC-HA-Seq dar. Sollte ein Labor keine Möglichkeit zum Bezug des entsprechenden HPs haben, ist eine vereinfachte Version der fdC-HA-Seq ohne Ligation des HPs möglich. So bleibt die breite Anwendbarkeit unserer Methode gewährleistet.

Die bioinformatische Auswertung kann für verschiedene Genauigkeitsstufen entwickelt werden. Die Auswertung der vergleichenden fdC-HA-Seq kann durch simple Subtraktion beider Datensätze erreicht werden, also den Sequenzierungs-Daten der Hydroxylamin-behandelten Reaktion und der Referenz-Reaktion. Für eine höhere Genauigkeit sollte die Auswertung mit einem *Hidden-Markov-Modell* erfolgen, wie dies bereits im AK WALTER am Beispiel der Sequenzierung von hmdC entwickelt wurde.^[250] Eine zusätzliche Steigerung der Genauigkeit sollte durch die Integration weiterer bioinformatischer Methoden^[251,252] erreicht werden können. Ein Prozess zur bioinformatischen Auswertung genomischer fdC-HA-Sequenzierungen wird aktuell von PASCAL GIEHR entwickelt.

6.2.3.3 Validierung des Hydroxylamin-Reagenzes **81** für die fdC-HA-Seq

*Unser Hydroxylamin-Reagenz **81** setzt fdC in DNA-Oligonukleotiden vollständig um.* Die Reaktivität des Hydroxylamin-Reagenzes **81** gegenüber fdC in DNA wurde zunächst von MATTHIAS KURZ anhand eines DNA-Oligonukleotids untersucht, siehe Abb. 42 A).^[246] Die Umsetzung erfolgte bei 37 °C in wässriger Lösung, sowohl in Anwesenheit als auch in Abwesenheit des Katalysators 4-Methoxyanilin. Zeitaufgelöste Reaktionskontrolle mittels RP-HPLC zeigte in Anwesenheit des Katalysators bereits nach 6 h Zielproduktbildung > 96 %, siehe Abb. 42 B). Nach 24 h wurde die quantitative Bildung des Zielprodukts sowohl mit als auch ohne Katalysator erreicht.

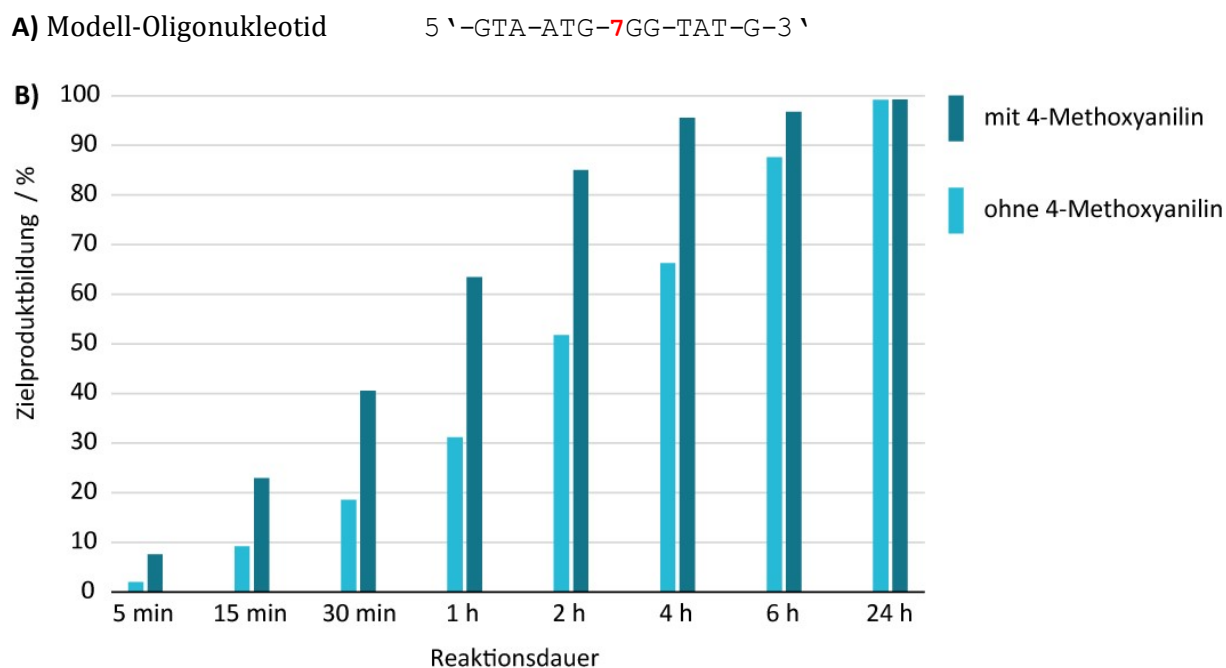


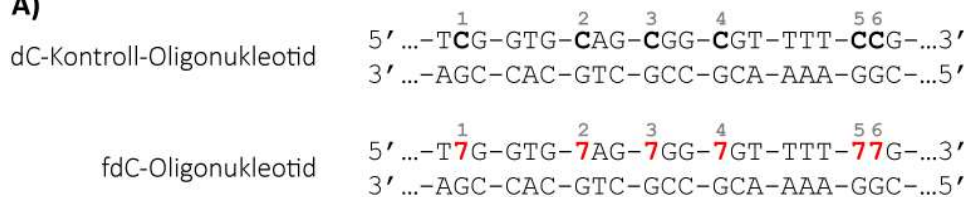
Abb. 42 Umsetzung des Hydroxylamin-Reagenzes **81** mit fdC in Modell-Oligonukleotid, in Anwesenheit sowie Abwesenheit des Katalysators 4-Methoxyanilin. **A)** Sequenz des Modell-Oligonukleotids. **7:** fdC. **B)** zeitaufgelöste Zielproduktbildung. Die Zielproduktbildung wurde mittels RP-HPLC bestimmt: Integration der Fläche unter dem Zielprodukt-Signal im Chromatogramm. Reaktionsbedingungen: Hydroxylamin-Reagenz **81** (1,5 mmol/L), Oligonukleotid (400 pmol), 4-Methoxyanilin (2,0 mmol/L), Na₂EDTA (0,1 mmol/L), HEPES (20 mmol/L, pH 7.4), H₂O, Reaktionsvolumen: 20 µL. Diese Untersuchung erfolgte durch MATTHIAS KURZ. Die experimentellen Details wurden durch ihn näher beschrieben.^[246] Abb. adaptiert aus^[246].

Die Reaktivität des Hydroxylamin-Reagenzes **81** wurde im nächsten Schritt an DNA-Oligonukleotiden mittels BS-Tiefensequenzierung untersucht. Um mögliche Nebeneffekte des Hydroxylamin-Reagenzes **81** während der BS-Seq zu untersuchen, erfolgte im nächsten Schritt die Sequenzierung von Oligonukleotiden die zuvor mit HA **81** derivatisiert wurden. Dazu erzeugte PASCAL GIEHR mittels PCR ein doppelsträngiges Kontroll-Oligonukleotid mit sechs dC-Positionen je Einzelstrang, siehe Abb. 43 A). Durch PCR in Anwesenheit von fdC- statt dC-Triphosphat, wurde das entsprechende fdC-Oligonukleotid mit sechs fdC-Positionen erzeugt. Anschließend wurde jedes Oligonukleotid gemäß dem Ablauf der fdC-HA-Seq behandelt und mittels Tiefensequenzierung analysiert. Die Fragmentierung mittels Restriktionsenzym, die Ligation des HPs sowie das Aufteilen der fdC-HA-Seq-Reaktion waren dabei nicht notwendig.

Unser Hydroxylamin-Reagenz **81** ermöglicht die verlässliche BS-Sequenzierung von fdC in DNA-Oligonukleotiden. Abb. 43 B) zeigt den Anteil desaminierter Referenznukleoside mit Hilfe eines Farbgradienten: je dunkler, desto höher der Anteil desaminierter Nukleoside. Dargestellt

sind die arithmetischen Mittelwerte von jeweils zwei Replikaten. Wie erwartet zeigt die HA-Behandlung des dC-Kontroll-Oligonukleotids keinen Effekt auf die BS-vermittelte Desaminierung der dC-Nukleotide. Der Anteil desaminierter fdC-Nukleoside liegt ohne HA-Behandlung zwischen 80 und 83 %. Dieses bekannte Phänomen^[184] verdeutlicht den einen Teil der Problematik der fdC-Sequenzierung: Erst durch Einsatz von Referenznukleosiden kann die fdC-Desaminierungsrate bestimmt werden. Ohne diesen Wert ist keine Korrektur der statistisch unvollständigen fdC-Desaminierung möglich. Ohne den Einsatz von Referenznukleosiden, wie in den Abschnitten 6.2.1 und 6.2.2 dieser Arbeit entwickelt, ist demnach keine verlässliche fdC-BS-Seq möglich. Der Anteil desaminierter fdC-Nukleoside nach Derivatisierung mit unserem Hydroxylamin-Reagenz **81** liegt hingegen zwischen 2 und 3 %. Diese nahezu quantitative Blockierung der BS-vermittelten Desaminierung adressiert den anderen Teil der Problematik der fdC-Sequenzierung: fCAB, das aktuell übliche Reagenz zur fdC-Sequenzierung, führt zu einer unvollständigen Derivatisierung von fdC, siehe folgender Abschnitt XX. Unser Reagenz ermöglicht hingegen eine nahezu quantitative Derivatisierung von fdC in DNA. Damit bietet unser Hydroxylamin-Reagenz **81** die notwendige Basis für eine verlässliche Sequenzierung genomischer fdC-Nukleoside.

A)



B)



Kontroll-Oligonukleotid	dC						Desaminierungsrate
	1	2	3	4	5	6	
BS	1,00	0,99	1,00	0,99	1,00	1,00	 0,00 0,25 0,50 0,75 1,00
HA 81+BS	0,99	0,99	0,99	0,99	0,99	0,99	
Nukleosid	1	2	3	4	5	6	
fdC-Oligonukleotid	fdC						Desaminierungsrate
	1	2	3	4	5	6	
BS	0,81	0,80	0,81	0,82	0,82	0,83	 0,00 0,25 0,50 0,75 1,00
HA 81+BS	0,02	0,03	0,03	0,03	0,02	0,02	
Nukleosid	1	2	3	4	5	6	

Abb. 43 Validierung des Hydroxylamin-Reagenzes **81** mittels lokaler BS-Tiefensequenzierung anhand zweier Oligonukleotide. **A)** Ausschnitte der verwendeten Oligonukleotid-Sequenzen und Nummerierung der Referenznukleoside. **7:** fdC **B)** Anteile der desaminierten Referenznukleoside (Desaminierungsraten),

nach alleiniger BS-Behandlung sowie nach Behandlung mit Hydroxylamin-Reagenz **81** sowie BS. Farbgradient: je dunkler, desto höher der Anteil desaminiertes Nucleoside. 1,00: Desaminierung aller Nucleoside. Jeder Wert repräsentiert den arithmetischen Mittelwert von zwei Replikaten. Alle BS-Behandlungen erfolgten mittels OPTI-Bisulfit.^[184] **Nucleosid** (unten) bezieht sich auf die Nummerierung in A).

6.2.3.4 Anwendung

In einer proof-of-concept-Studie wurde fdC-HA-Seq direkt mit fCAB verglichen, gleichzeitig wurde die Basis für die Sequenzierung von fdC im Genom gelegt. Die aktuell verbreitetste Methode zur BS-basierten Sequenzierung von fdC ist fCAB^[191]. Eine Methode, die ebenso wie das alternative redBS^[192] innerhalb der letzten 8 bzw. 7 Jahre nicht zur kommerziellen Reife entwickelt wurde. Um die Wirksamkeit unserer Sequenzierungs-Methode zu untersuchen, wurde fdC-HA-Seq in einer *proof-of-concept*-Studie direkt mit fCAB verglichen. Die in dieser Arbeit entwickelten HPs boten sich dafür als umfassendes und klar definiertes Modellsystem an. Da jedes HP sowohl fdC als auch dC, mdC, hmdC und cadC als Referenznucleoside enthält, konnten so die fdC-Desaminierungsrate exakt bestimmt werden. Gleichzeitig konnten Effekte der fdC-HA-Seq auf andere dC-Nucleotide untersucht werden. Zunächst wurde ein in dieser Arbeit synthetisiertes HP an die fragmentierte Ziel-DNA ligiert. Mittels spezifischer PCR-Primer wurden anschließend drei genomische Loci selektiert, für die vorhergehende Studien die Anwesenheit von fdC zeigten: Rik,^[158,192] Chr15^[158,192] und Tbx5^[253], vgl. Abschnitt 6.2.3.6. Anschließend erfolgte die Behandlung dieser HP+Locus-Konstrukte gemäß dem Ablauf der fdC-HA-Sequenzierung sowie nach dem fCAB-Protokoll.

fdC-HA-Seq schützt fdC verlässlich, fCAB nicht. Abb. 44 zeigt die Desaminierungsraten der Referenznucleoside der drei HP+Locus-Konstrukte nach Behandlung mit BS^[184], HA **81**+BS sowie fCAB+BS. Dargestellt ist jeweils der Median von 7-10 biologischen Replikaten. Um falsch positive fdC-Sequenzierungen auszuschließen, ist es relevant den Einfluss der Hydroxylamin-Behandlung auf die übrigen dC-Nucleoside (dC, mdC, hmdC und cadC) zu untersuchen. Abb. 44 zeigt, dass die Desaminierungsraten der übrigen Referenznucleoside weder durch fdC-HA-Seq noch durch fCAB in relevantem Maße beeinflusst wird. Die Behandlung des fdC-Referenznucleosids mit BS allein, zeigt die bekannte unvollständige Desaminierung.^[184] Hervorzuheben ist, dass fCAB zu einer ungleichmäßigen Schützung von fdC vor der BS-induzierten Desaminierung führt: In Konstrukt 2 (Chr15) wurde das fdC-Referenznucleosid des HPs nahezu vollständig vor der Desaminierung geschützt. Für die Konstrukte 1 (Rik) und 3 (Tbx5) wurden hingegen 37 bzw. 19 % der fdC-Referenznucleoside *nicht* vor der Desaminierung geschützt. Die Behandlung mit unserem

Hydroxylamin-Reagenz **81** bewirkt hingegen einen nahezu quantitativen Schutz von 95 %, 99 % sowie 98 %.

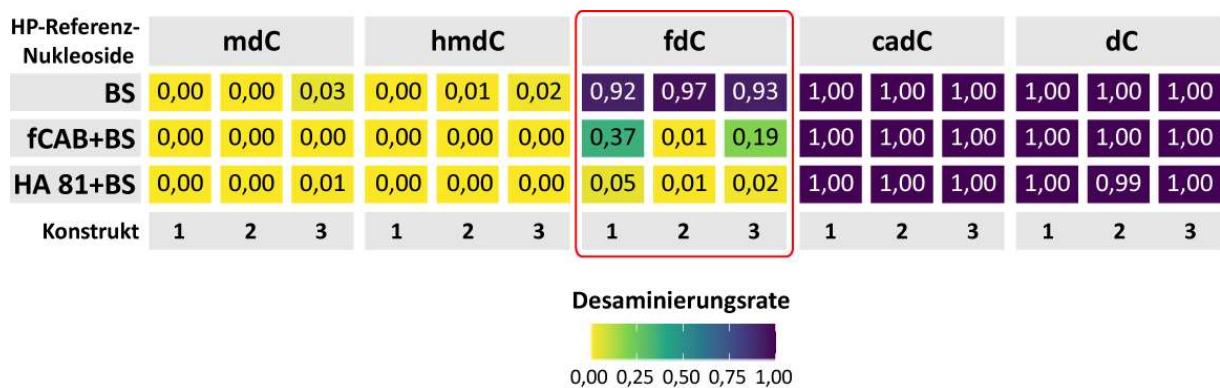


Abb. 44 Vergleich von fdC-HA-Seq mit fCAB. Angegeben sind die Desaminierungsraten der Referenznukleoside dreier HP+Locus-Konstrukte: **1** (*A-Tail*-HP+Rik), **2** (*A-Tail*-HP+Chr15), **3** (MspI-HP+Tbx5). Jeder Wert repräsentiert den Median von 7-10 biologischen Replikaten. Zum Vergleich erfolgte zusätzlich eine reine BS-Behandlung. Alle BS-Behandlungen erfolgten mittels OPTI-Bisulfit.^[184] Desaminierungsrate: Anteil desaminiertes Nucleoside. Farbgradient: je dunkler, desto höher der Anteil desaminiertes Nucleoside. 1,00: Desaminierung aller Nucleoside. Vgl. Abschnitt 6.2.3.6 für weitere Informationen zu den genomischen Loci.

6.2.3.5 Zusammenfassung, Einordnung und Ausblick

Zusammengenommen ermöglicht fdC-HA-Seq die robuste Sequenzierung von fdC mit dem im AK CARELL entwickelten Hydroxylamin-Reagenz **81**^[168]. Dabei erfolgte die fdC-Derivatisierung nahezu quantitativ, unter den bisher untersuchten Bedingungen. fCAB, die aktuell verbreitetste Methode zur fdC-BS-Seq, konnte in zwei der drei untersuchten Konstrukte hingegen 19 bzw. 37 % der fdC-Referenznucleoside nicht vor der BS-induzierten Desaminierung schützen. Durch Integration der in Abschnitt 6.2.2.4 entwickelten HPs ermöglicht fdC-HA-Seq zudem die Strang-spezifische fdC-Sequenzierung bei höchster Genauigkeit der Sequenzierung. Damit stellt fdC-HA-Seq eine überlegene Methode zur Sequenzierung von fdC dar.

Im nächsten Schritt sollte die Anwendung von fdC-HA-Seq auf biologische Fragestellungen erfolgen. Eine Auswahl relevanter genomischer Loci wurde dazu bereits etabliert. Darüber hinaus könnte eine Verkürzung der Behandlungsdauer mit dem Hydroxylamin-Reagenz **81** möglich sein: In einer Studie zur Umsetzung des HA **81** mit fdC in kurzen DNA-Oligonucleotiden zeigte der Katalysator 1,4-Diaminobenzol eine vollständige Umsetzung bereits nach 5 h.^[158] Eine quantitative Umsetzung mit dem hier als Katalysator eingesetztem 4-Methoxyanilin konnte

hingegen erst nach 24 h erreicht werden. In dieser Arbeit wurde fdC-HA-Seq für die lokale Tiefensequenzierung optimiert. Eine Anwendung auf andere BS-Seq-Techniken, wie *whole genome* (WGBS) oder *reduced representation* (RRBS) ist aber ebenso denkbar.

fdC-HA-Seq sollte ideal auf SMRT- und Oxford Nanopore-Sequenzierungen anwendbar sein. Langfristig betrachtet ist es wahrscheinlich, dass Sequenzierungs-Methoden der 3. Generation die BS-Seq ablösen. fdC-HA-Seq sollte leicht auf Sequenzierungs-Methoden der 3. Generation übertragbar sein. Sowohl Sequenzierungen mittels *Oxford Nanopore* als auch SMRT generieren Signale auf Basis der Struktur des sequenzierten Nucleosids, siehe Abb. 45. So basiert das Signal bei SMRT-Sequenzierungen auf der Dauer des Einbaus des komplementären Nucleotids. Ein Prozess der durch den sterischen Anspruch unseres Hydroxylamin-Reagenzes **81** beeinflusst werden sollte. Das Signal von *Oxford Nanopore*-Sequenzierungen basiert auf der Leitfähigkeit der Nanopore. Diese Leitfähigkeit wird durch die Nucleotide in der Pore moduliert. HA-derivatisiertes fdC sollte demnach klar von dC unterschieden werden können. Insbesondere aufgrund des permanent positiv geladenen quaternären Ammoniums unseres Hydroxylamin-Reagenzes **81**. Eine Anwendung von fdC-HA-Seq auf *Oxford Nanopore*-Sequenzierungen scheint demnach vielversprechend. Die Anwendung von fdC-HA-Seq auf SMRT-Sequenzierungen wird aktuell von PASCAL GIEHR untersucht.

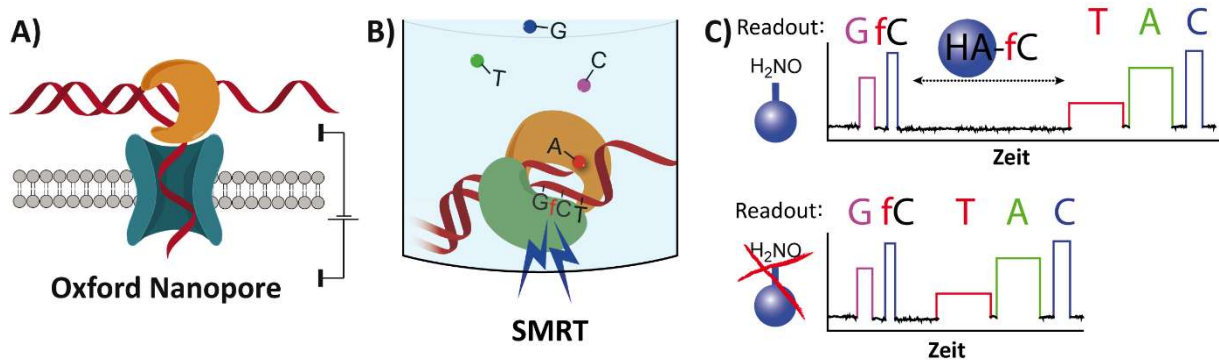


Abb. 45 Die Sequenzierungs-Methoden der 3. Generation SMRT sowie *Oxford Nanopore* generieren Signale auf Basis der chemischen Struktur des sequenzierten Nucleosids. **A)** Schematische Darstellung – *Oxford Nanopore*-Sequenzierungen. **B)** Schematische Darstellung – SMRT-Sequenzierungen. **C)** Veranschaulichung erwarteter Signalmuster der SMRT-Sequenzierung. oben: Sequenzierung nach Derivatisierung mit unserem Hydroxylamin-Reagenz **81**. unten: Sequenzierung ohne Derivatisierung.

6.2.3.6 Experimentelle Details

Die Anwendung der fdC-HA-Seq in Abschnitt 6.2.3.4 erfolgte auf Konstrukten aus genomischen Loci Haarnadel-Oligonukleotiden (HPs). Dazu wurde zunächst ein in dieser Arbeit synthetisiertes HP an die fragmentierte Ziel-DNA ligiert. Mittels spezifischer PCR-Primer wurden anschließend drei genomische Loci selektiert, für die vorhergehende Studien die Anwesenheit von fdC zeigten: Rik,^[158,192] Chr15^[158,192] und Tbx5^[253]. Im Folgenden die Beschreibung der Positionen sowie die für die bioinformatische Auswertung verwendeten Nukleotid-Sequenzen der Konstrukte. Die HP-Sequenz ist fett hervorgehoben.

Rik bezieht sich auf Position 30 020 539 des Chromosoms 16 *Mus musculus* (MM9), auf Exon 3 von 632428C04Rik.^[158,192]

A-Tail-HP-Rik (Konstrukt 1):

5'-TGTTGTTGGTGCCTGAGACTCATGGACCGTGGCATAAACACACCTCTGTCCTCAGATGCCGGCAGAGTCTCCAAGGATACTTGC GGTTTAGTTTGGAGATGACCGATAAAATCGCAGGGCTGGTTTTGAGGAGG**GNGNGATDDDDDDDDATTGNGNTTTCCTCAAACCAGCCCTGCGATTTTATCGGTTCATCTCCAAAC**TAAACCGCAAGTATCCTTGGAGACTCTGCCGGCATCTGAGGACAGAGGTGTGTTTATGCCACGGTCCATGAGTCTCAGGGACCAACAACACCGCCCCTCAGATGCTCCTTCAACTCTGGAGACTGGGGTACCCTCC TGAGCAAGCACACAGGGGCTGCTGTTACG-3'

Chr15 bezieht sich auf Position 8 846 677 des Chromosoms 15 *Mus musculus*, welche sich in nichtkodierender DNA befindet.^[158,192]

A-Tail-HP-Chr15 (Konstrukt 2):

5'-CGTCTGCGGCGGACGAAGTCCTAACACACTTCTTGCCCCGCGGAAGCATTCTGCAAAAATATCCACGAGGCTCAGTGCAGCTCCAGCTGCTGAAAAGCCCATCTGGGAAACAGCTCCGGAACCCCTTCAGACGCAAGCGAAGCAGGGCAAATGGCGATCTCTGAACCTCTGCCAGCTCATTTGCGGAGGG**GNGNGATDDDDDD****DDATTGNGNTTTCGCAAATGAGCTGGCAGAGGTTTCAGAGATCGCCATTTGCCCTGCTTCGCTTGCCT**CTGAAGGGGTTCCGGAGCTGTTTCCCAGATGGGCTTTTCAGCAGCTGGAGCTGCACTGAGCCTCGTGGA TATTTTTGCAGAATGCTTCCGCGGGGCAAGAAGTGTGGTTAGGACTTCGTCCGCCGCGAGACG-3'

Tbx5 bezieht sich auf folgenden Abschnitt des Chromosoms 5 *Mus Musculus*: 120 285 942-120 288 285.^[253]

MspI-HP-Tbx5 (Konstrukt 3):

5'-TGGAGTGAAGGGTCAAACAGCTTGTCCCCTATGTCGCTAGACACTCTCCAACCTCCCTCTCTTGTCTT
CAGAATAGAACCTCGCGCGGGCACAGGAACCCCTGCGCACCATGGCCGATACAGATGAGGGCTTTGGCC
TGGCGCGCACGCCTCTGGAGCCTGATTCCAAAGACAGGTCTTGCGATTTGTGGNGNGATDDDDDD
DATTGNGNCGCGAATCGCAAGACCTGTCTTTGGAATCAGGCTCCAGAGGCGTGCGCGCCAGGCCAAAG
CCCTCATCTGTATCGGCCATGGTGCGCAGGGTTCTGTGCCCCGCGGAGGTTCTATTCTGAGGACAAG
AGAGGGAGTTGGAGAGTGTCTAGCGACATAGGGGACAAGCTGTTTGACC-3'

C allgemeine Methoden

7 Methoden der chemischen Synthese

Bei wasserempfindlichen Reaktionen wurden trockene Lösungsmittel mit einem Wassergehalt $\varphi(\text{H}_2\text{O}) < 50$ ppm verwendet. Für Extraktionen und Säulenchromatographie wurden Lösungsmittel technischer Güte verwendet. Lösungsmittel technischer Güte wurden vor der Verwendung nicht destilliert. Chemikalien wurden in den kommerziell erhältlichen Qualitäten puriss., p.a. oder purum eingesetzt. Mischungsverhältnisse wurden als Volumenverhältnis V/V angegeben. Zur Feintrocknung von Substanzen und zur Erzeugung eines Hochvakuums wurde eine RZ-2.5-Vakuumpumpe der Firma *Vacuubrand* verwendet. Kleinstmengen der Lösungsmittel H_2O und/oder MeCN wurden mit Vakuumzentrifugen (*SpeedVac Plus SC110A* oder *SpeedVac SPD111V*) der Firma *Savant* entfernt. Zur Lyophilisierung wässriger Lösungen wurde die Probe mit $\text{N}_2(\text{l})$ eingefroren und unter Verwendung des Lyophilisators *ALPHA 2-4 LD plus* der Firma *Christ* getrocknet.

SCHLENK-Technik folgendermaßen angewandt: Alle Flüssigkeiten wurden unter Verwendung von Septen und mit Schutzgas-gespülten Einwegspritzen zugeführt. Feststoffe wurden unter Schutzgas-Gegenstrom zugegeben. Zur Entgasung von Reaktionsgemischen wurde das *freeze-pump-thaw*-Verfahren angewendet. Dazu wurde das Gemisch unter Schutzgas-Atmosphäre eingefroren und im Hochvakuum wieder aufgetaut. Dieser Vorgang wurde zwei Mal wiederholt.

8 DNA-Oligonukleotid-Herstellung

Die Synthese der Oligonukleotide erfolgte unter Anwendung der Phosphoramidit-Festphasen-Methode. Dazu wurde das Oligonukleotid mittels automatisierter Festphasen-Synthese (FPS) hergestellt, bevor es manuell von der Festphase abgespalten und entschützt wurde. Anschließend erfolgte dessen Aufreinigung mittels semi-präparativer oder analytischer RP-HPLC. Die Charakterisierung erfolgte durch MALDI-TOF-MS und analytischer RP-HPLC. Alle weiteren Sequenzen der im Rahmen dieser Arbeit synthetisierten Oligonukleotide finden sich in den ergänzenden Informationen zu den Publikationen in den Abschnitten 5 und 6.1.

Tab. 9 gibt einen Überblick der DNA-Oligonukleotid-Sequenzen die im Rahmen der Projekte in Abschnitt 6.2 synthetisiert wurden. Alle weiteren Sequenzen der im Rahmen dieser Arbeit synthetisierten Oligonukleotide finden sich in den ergänzenden Informationen zu den Publikationen in den Abschnitten 5 und 6.1.

Tab. 9 Übersicht der im Rahmen der Projekte in Abschnitt 6.2 synthetisierten DNA-Oligonukleotid-Sequenzen. N = A/T/G/C, D = A/T/G, W = A oder T. **5:** mdC, **6:** hmdC, **7:** fdC, **8:** cadC. HP: Haarnadel-Oligonukleotid. Alle weiteren Sequenzen der im Rahmen dieser Arbeit synthetisierten Oligonukleotide finden sich in den ergänzenden Informationen zu den Publikationen in den Abschnitten 5 und 6.1.

Bezeichnung	Sequenz (5' > 3')
<i>spike-in</i> -Oligo. A1	GTA- 6 GA-ACT-GCG-ATA-GTA- 5 TC-GTG-AG 7 -AGT
<i>spike-in</i> -Oligo. A2	GAA-T 7 T-CAG-CAT-GCT-ATG-TA 5 -G 8 G-TAA
<i>spike-in</i> -Oligo. A3	GTA- 6 GA-AGT-GGG-ATA-GTA- 5 TC-GTG-ATG-AGT
<i>spike-in</i> -Oligo. A4	GAA-T 7 T-GAG-CAT-GGT-ATG-TAG-G 8 G-TAA
<i>spike-in</i> -Oligo. B1	TTA-CG 6 -GTA- 8 AT-AGC-ATG-CTG-AGA-TT 5 -ACT
<i>spike-in</i> -Oligo. B2	GCT-AA 8 -GAG-TA 6 -TAT-CGC-AGT-T 7 G-TAC
<i>spike-in</i> -Oligo. B3	TTA- 7 G 6 -CTA-CAT-ACC-ATG-CT 8 -AGA-TTC-ACT
<i>spike-in</i> -Oligo. B4	CCT-AAC-GAG-TAC-TAT-CCC-ACT-TCG-TA 5
<i>spike-in</i> -Oligo. B5	TTA-CG 8 -GTA- 5 AT-AGC-ATG-CTG-AGA-TT 6 -ACT
HP MspI	C-GCG-G 5 G- 6 GA-TDD-DDD-DDD-ATC-G 7 G- 8 CG
HP DdeI	TN-AGC-G 5 G- 6 GA-TDD-DDD-DDD-ATC-G 7 G- 8 GC
HP BsaWI	CCG-GCG-G 5 G- 6 GA-TDD-DDD-DDD-ATC-G 7 G- 8 CG
HP Eco47I	GW-CGG-G 5 G- 6 GA-TDD-DDD-DDD-ATC-G 7 G- 8 CC
HP A-Tail	GG-G 5 G- 6 GA-TDD-DDD-DDD-ATC-G 7 G- 8 CCT

8.1 Synthese

Die DNA-Oligonukleotid-Synthesen erfolgten an *ABI 392*- sowie *ABI 394*-Geräten zur automatisierten Festphasensynthese der Firma *Applied Biosystems*. Die kommerziell erhältlichen Phosphoramidite Bz-dA-CE, Ac-dC-CE, iBu-dG-CE, dT-CE und Ac-mdC-CE wurden von den Firmen *Sigma-Aldrich* und *Link Technologies* bezogen. Als Festphase wurden *SynBase CPG-Träger 1000/110* (1 μmol) der Firma *Link Technologies* eingesetzt. Die zur Synthese benötigten Lösungen und deren Zusammensetzung sind in Tab. 10 aufgeführt. Die Kupplung von Phosphoramiditen kanonischer Nucleoside erfolgte über eine Kupplungszeit von 45 s. Phosphoramidite nicht-kanonischer Nucleoside wurden insgesamt für 3 min gekuppelt. Zusätzlich wurden die Reaktionsgemische aller Phosphoramidite nach 150 s acht Mal für je 0,1 s bewegt. Die Reaktionszeit zwischen diesen Durchmischungs-Schritten betrug jeweils 4 s. Die Synthese aller *spike-in*-Oligonukleotide erfolgte im *DMT-on*-Verfahren. Die Synthese aller Haarnadel-Oligonukleotide erfolgte im *DMT-off*-Verfahren, das heißt Abspaltung der

5'-terminalen DMT-Schutzgruppe im letzten Schritt der Oligonukleotid-Synthese. Die Synthese von Oligonukleotiden mit nicht kanonischen Nukleotiden in 3'-terminaler Position erfolgte unter Verwendung von *universal Q*-Festphasenträgern der Firma *LinkTech*. Zur Detritylierung der universellen Festphasenträger wurde die Dauer der ersten Detritylierung von 73 s auf 116 s erhöht. Im Anschluss an die Synthese wurden die Festphasen-Träger 15 min im Hochvakuum getrocknet und gegebenenfalls unter N₂-Atmosphäre bei -20 °C bis zur weiteren Verwendung gelagert.

Tab. 10 Lösungen für die DNA-Festphasensynthese. Bei selbst hergestellten Lösungen ist die Zusammensetzung angegeben. Bei kommerziell erworbenen Lösungen ist Produktname und Hersteller angegeben.

Lösung	Zusammensetzung bzw. Hersteller
Lösungen kanonischer Phosphoramidite	100 mmol/L in MeCN
Lösungen nicht-kanonischer Phosphoramidite	125-150 mmol/L in MeCN
Aktivator-Lösung	<i>Aktivator 42</i> (0,25 mol/L), <i>Sigma-Aldrich</i>
Detritylierungs-Lösung	Dichloressigsäure in CH ₂ Cl ₂ (3 %, V/V), <i>emp Biotech</i>
Oxidierungs-Lösung	<i>Oxidizing (ABI), for DNA/RNA synthesis, Baker Analyzed, 9488, J.T. Baker</i>
Capping-Lösung A	<i>Capping A (ABI), for DNA/RNA synthesis, Baker Analyzed, 9497, J.T. Baker</i>
Capping-Lösung B	<i>Capping B (ABI), for DNA/RNA synthesis, Baker Analyzed, 9485, J.T. Baker</i>
MeCN	<i>Acetonitrile, Low Water, Baker Bio-Analyzed Reagent, 9018-03 J.T. Baker</i> oder <i>ROTIDRY</i> ($\varphi(\text{H}_2\text{O}) \leq 10$ ppm), <i>Carl Roth</i>
CH ₂ Cl ₂	<i>HiPerSolv CHROMANORM, VWR</i>

spike-ins:

.Um einer geringen Ausbeute der Oligonukleotid-Synthese vorzubeugen, wurden alle modifizierten Phosphoramidite in einer um 25-50 % erhöhten Konzentration eingesetzt. Zudem wurden die Reaktionszeiten aller modifizierten Phosphoramidite auf 3 min vervierfacht. Eine zusätzliche Durchmischung bei der Phosphoramidit-Kupplung wurde erreicht, indem die Reaktionsgemische aller Phosphoramidite abschließend acht Mal für je 0,1 s bewegt wurde. Die Reaktionszeit zwischen diesen Durchmischung-Schritten betrug jeweils 4 s. Alle

Phosphoramidite wurden anschließend für 35 s detrityliert. Um eine vollständige Entschützung aller modifizierten Phosphoramidite sicherzustellen, wurden diese für insgesamt 80 s detrityliert. Die Synthese der *spike-in*-Oligonukleotide erfolgte an Polystyrol-Festphasenträgern im 400 µmol-Maßstab, siehe Tab. 11. Dazu wurden je *spike-in*-Oligonukleotid zwei 200 µmol-Festphasenträger eingesetzt, lediglich die *spike-in*-Oligonukleotide A3 und B5 wurden im 200 µmol-Maßstab synthetisiert. Alle Oligonukleotide außer *spike-in*-Oligonukleotid B4 wurden an LV200-Festphasenträgern synthetisiert.

Tab. 11 Überblick zu Ausbeute Synthese-Maßstab, Festphasenträger und Ausbeute aller *spike-in*-Oligonukleotide. *universal Q*: universeller Festphasenträger der Firma *LinkTech*. *LV200*: kanonische Festphasenträger der Firma *Applied Biosystems*. Alle Synthesen erfolgten an Polystyrol-Festphasenträgern.

<i>spike-in</i> - Oligonukleotid	Ausbeute		Synthese- Maßstab	Festphasenträger
	nmol	%		
A1	25,3	6,3	2x200 µmol	<i>LV200</i>
A2	10,0	2,5	2x200 µmol	<i>LV200</i>
A3	17,5	8,8	200 µmol	<i>LV200</i>
A4	17,0	4,3	2x200 µmol	<i>LV200</i>
B1	27,7	6,9	2x200 µmol	<i>LV200</i>
B2	28,08	7,0	2x200 µmol	<i>LV200</i>
B3	23,8	5,6	2x200 µmol	<i>LV200</i>
B4	93,6	23	2x200 µmol	<i>universal Q</i>
B5	17,53	8,8	200 µmol	<i>LV200</i>

8.2 Entschützung und Abspaltung

Zur basischen Entschützung und Abspaltung vom Festphasen-Träger wurde das Trägermaterial aus der Kartusche entfernt und mit einer geeigneten Reagenz inkubiert, siehe Tab. 12. Anschließend wurde die Suspension zentrifugiert (4 min, 21 130 g), der Überstand abgenommen und der Träger mit H₂O (2 × 450 µL) gewaschen. Die vereinigten Lösungen wurden mittels Vakuum-Zentrifugation auf ca. 500 µL eingengt.

Enthielt das Oligonukleotid fdC, so folgte anschließend dessen saure Entschützung. Dazu wurde die eingengte Lösung der basischen Entschützung mit HOAc_(aq) (80 %, 1 mL) versetzt und entschützt, siehe Tab. 12. Die Dauer der sauren Entschützung wurde mittels MALDI-TOF-MS bestimmt: Entschützung wurde beendet, wenn die Entschützung mittels MALDI-TOF-MS vollständig erschien oder wenn Depurinierungen detektiert werden konnten. Beendet wurden

Entschützungsreaktionen durch Zugabe von HNET_3OAc (1 mol/L, 1 mL). Die saure Entschätzung aller HPs erfolgte hingegen für 6 h, da deren randomisierte Sequenzen keine Detektion von Depurinierungen mittels MALDI-TOF-MS erlaubten.

Tab. 12 Bedingungen zur Abspaltung und Entschätzung der DNA-Oligonukleotide. HPs: Haarnadel-Oligonukleotide.

Oligonukleotid-Konfiguration	Entschätzungs-Bedingungen
<i>spike-ins</i> ohne hmdC und cadC, auch <i>spike-ins</i> mit 3'-terminalem mdC (universeller Festphasenträger)	A) $\text{NH}_4\text{OH}_{(\text{aq})}$ (30-32 %, 1 mL), 17 h, 25 °C, 700 rpm
HPs & <i>spike-ins</i> mit hmdC	B) NaOH (MeOH/H ₂ O, 4/1, 1 mL, 0,4 mol/L), 26 h, 25 °C, 700 rpm
<i>spike-ins</i> mit fdC & HPs	C) $\text{HOAc}_{(\text{aq})}$ (80 %, 1 mL), 20 °C, 700 rpm

Im letzten Schritt vor der Aufreinigung erfolgte die **Fällung** der Oligonukleotide, siehe Tab. 13. **HPs** wurden zusätzlich bereits vor der sauren Entschätzung gefällt. Außerdem erfolgte deren Fällung im Gegensatz zur Fällung der *spike-ins* bei -20 °C über Nacht.

Tab. 13 Bedingungen zur Fällung der DNA-Oligonukleotide. HPs: Haarnadel-Oligonukleotide.

Oligonukleotid-Typ	Fällungs-Prozess
Zeitpunkt der Fällung	
<i>spike-ins</i>: vor Aufreinigung	<ol style="list-style-type: none"> 1. Zugabe von $\text{NaOAc}_{(\text{aq})}$ (3 mol/L), finale Konz.: 0,3 mol/L 2. Zugabe des vierfachen Volumens an EtOH (abs.) 3. gevortext 4. Inkubation bei 0-4 °C, mind. 1 h 5. Zentrifugation (60 min, 21 130 g, 4 °C) 6. Abnahme des Überstands 7. Pellet lyophilisiert
HPs: vor saurer Entschätzung sowie vor Aufreinigung	<ol style="list-style-type: none"> 1. Zugabe von $\text{NaOAc}_{(\text{aq})}$ (3 mol/L), finale Konz.: 0,3 mol/L 2. Zugabe des vierfachen Volumens an EtOH (abs.) 3. gevortext 4. Inkubation bei -20 °C, über Nacht 5. Zentrifugation (60 min, 21 130 g, 4 °C) 6. Abnahme des Überstands 7. Pellet lyophilisiert

8.3 Aufreinigung und Charakterisierung

Zur Aufreinigung wurde das Pellet in $\text{HNEt}_3\text{OAc}_{(\text{aq})}$ (1 mol/L, pH 7) gelöst. Anschließend wurde filtriert und die resultierende Lösung mittels analytischer HPLC und MALDI-TOF-MS untersucht. Die Aufreinigung erfolgte mittels semi-präparativer oder analytischer HPLC. Die einzelnen Fraktionen wurden mittels analytischer HPLC sowie MALDI-TOF-MS auf Anwesenheit und Anteil des Produkts untersucht. Die Produktfraktionen wurden vereinigt, mittels Vakuum-Zentrifugation eingengt und lyophilisiert.

Zur Reduktion freier Ionen erfolgte anschließend eine „Entsalzung“ der Oligonukleotide mittels *Sep-Pak C18 Classic Cartridges* der Firma *Waters*. Dazu wurde die *Sep-Pak*-Kartusche mit MeCN (10 mL) gewaschen und mit $\text{HNEt}_3\text{OAc}_{(\text{aq})}$ (10 mL, 0.1 mol/L, pH 7) äquilibriert. Das Oligonukleotid wurde in $\text{HNEt}_3\text{OAc}_{(\text{aq})}$ (2 mL, 0.1 mol/L, pH 7) gelöst und auf die Kartusche aufgetragen (1 mL/min). Anschließend wurde mit H_2O (10 mL, 1 mL/min) gewaschen, bevor das Oligonukleotid mit MeCN/ H_2O (80/20, 6 mL, 1 mL/min) eluiert wurde.

Abschließend wurden Reinheit und Identität der Oligonukleotide mittels analytischer HPLC und MALDI-TOF-MS überprüft. Dazu wurden die Oligonukleotide in H_2O gelöst. Zudem erfolgte die Ermittlung der Ausbeute via Spektralphotometrie bei 260 nm (*NanoDrop ND-1000, PreQlab*). Die Berechnung der Konzentration erfolgte unter Verwendung der Extinktions-Koeffizienten aller Nukleotide des jeweiligen Oligonukleotids. Für randomisierte Positionen wurde der Mittelwert der Extinktions-Koeffizienten berechnet. Anschließend wurden die Oligonukleotide lyophilisiert und bei $-20\text{ }^\circ\text{C}$ gelagert.

9 Massenspektrometrie

9.1 HR-HESI-LC-MS

Untersuchungen mittels HR-HESI-LC-MS erfolgten selbstständig mit einem *LTQ Orbitrap XL*-Massenspektrometer der Firma *Thermo Finnigan*. Die chromatographische Auftrennung erfolgte durch ein *Dionex Ultimate 3000* HPLC-System der Firma *Thermo Scientific*. Vor der Injektion wurden alle Proben auf eine Zielsubstanz-Konzentration von etwa $c = 1\text{ mmol/L}$ verdünnt und durch einen Membranfilter ($0,2\text{ }\mu\text{m}$) filtriert, bevor $5\text{ }\mu\text{L}$ injiziert wurden. Die chromatographische Auftrennung erfolgte mittels *Interchim Utisphere120 3HDO C18*-Säule mit einer Flussrate von $Q = 150\text{ }\mu\text{L/min}$ und bei einer Säulentemperatur von $30\text{ }^\circ\text{C}$. Zur Auftrennung wurde ein System aus HR-HESI-LC-MS-Puffer A und HR-HESI-LC-MS-Puffer B einem Gradienten

von 0 % → 20 % Puffer B in 25 min eingesetzt. Die Elution wurde bei 243 nm und 260 nm durch einen *Dionex Ultimate 3000 Diode Array Detektor* der Firma *Thermo Scientific* überwacht. Anschließend wurde das Eluent ohne vorherige Spaltung direkt in die Ionenquelle injiziert. Die massenspektrometrische Analyse erfolgte im Positiv-Ionenmodus mit einer Auflösung von $R = 30000$ über einen Massenbereich von m/z 80 - 500. Die Datenauswertung erfolgte mit Hilfe des Programms *Xcalibur 2.1 SP1* der Firma *Thermo Fisher Scientific*.

Tab. 14 In dieser Arbeit verwendete HR-HESI-LC-MS-Puffer.

Puffer	Zusammensetzung
A	H ₂ O
B	MeCN
C	2 mmol/L NH ₄ COOH in H ₂ O
D	2 mmol/L NH ₄ COOH in 80% MeCN in H ₂ O

9.2 HR-ESI-MS

Massenspektrometrische Untersuchungen mittels HR-ESI-MS wurden durch die zentrale Analytik, Department Chemie und Pharmazie der Ludwig-Maximilians-Universität München durchgeführt. Elektrospray-Ionisation (ESI) Messungen wurden an einem *Thermo Finnigan LTQ FT Ultra* Fourier Transform Ionen Cyclotron Resonanz-Massenspektrometer durchgeführt. Die Auflösung war auf 100000 bei m/z 400 eingestellt. Methodenabhängig wurden Massenbereiche von 50 bis 2000 u gemessen. Die Spray-Kapillarenspannung am *IonMax*-ESI-Kopf betrug 4 kV, die Heizer-Kapillarentemperatur 250 °C, der Stickstoff *Sheath*-Gasfluss 20 und der *Sweep*-Gasfluss 5 Einheiten.

9.3 HR-EI-MS

Massenspektrometrische Untersuchungen mittels HR-EI-MS wurden durch die zentrale Analytik, Department Chemie und Pharmazie der Ludwig-Maximilians-Universität München durchgeführt. EI-Messungen erfolgten an einem *Thermo Q Exactive GC Orbitrap*- oder *Finnigan MAT95*-Sektorfeld-Massenspektrometer durchgeführt, bei einer Auflösung von etwa 5000 (*MAT95*) oder 50000 (bei m/z 200, *Q Exactive GC*) eingestellt. Methodenabhängig wurden Massenbereiche von 40 bis 1040 u gemessen. Elektronenstoß-Ionisation (EI) wurde bei einer Quellentemperatur von 250 °C und einer Elektronenenergie von 70 eV durchgeführt. Für EI-Spektrenvergleiche diente die

NIST Mass Spectral Library. Direktverdampfungs-EI-Proben (DEP/EI) wurden auf einem Platinfaden von 20 bis 1600 °C mit einer Rate von 120 °C/min geheizt. Direktinsertions-EI-Proben (DIP/EI) wurden in einem Aluminiumtiegel von 20 bis 400 °C mit einer Rate von 60 °C/min verdampft.

Gaschromatographie (GC/EI) wurde an einem *Thermo Trace 1300* Gaschromatographen mit temperaturprogrammierbarem Injektor (*Q Exactive GC*) oder einem *Varian 3400* mit split-splitlos-Injektor (*MAT95*) und *Autosampler* durchgeführt. Für die Trennung wurden *Thermo* oder *MacheryNagel* 5%-Phenyl-95%-Methyl-Silphenylenphasen Kapillar-Trennsäulen von 30 m Länge, 0,25 mm Innendurchmesser und 0,25 µm Filmdicke eingesetzt. Das Trägergas war Helium mit einer Flussrate von 1 mL/min. Die Temperatur der Kopplung zum Massenspektrometer betrug 250 °C, der Injektor wurde von 50 °C (1 min isotherm) mit einer Rate von 5 °C/s bis 250 °C (3 min isotherm) hochgeheizt (*Q Exactive GC*) oder hatte 250 °C (*MAT95*). Der Säulenofen wurde von 50 °C (1 min isotherm) bis 300 °C (4 min isotherm) mit einer Rate von 25 °C/min oder 10 °C/min geheizt. Jeweils 0,1-1 µl Probenlösung wurde mit einem Split von 1/10 oder 1/100 injiziert.

9.4 LR-(LC-)ESI-MS

Untersuchungen mittels LR-(LC-)ESI-MS erfolgten selbstständig an einem *MSQ Plus*-System (ESI) der Firma *Thermo Fisher Scientific*. Bei Bedarf erfolgte zunächst die chromatographische Auftrennung der Probe mittels *Ultimate 3000*-HPLC-System der *Thermo Fisher Scientific*. Die Auftrennung erfolgte mittels *Hypersil Gold*-Säule (3 µm, 100 x 2.1 mm) und war direkt an das Massenspektrometer gekoppelt. Als mobile Phasen dienten H₂O und MeCN mit einem Anteil von jeweils 0,05% Ameisensäure.

9.5 MALDI-TOF-MS

Massenspektrometrische Untersuchungen mittels MALDI-TOF-MS wurden selbstständig an einem *Autoflex II*-System der Firma *Bruker* durchgeführt. Zur Desorption wurde ein Stickstoff-Kartuschen-Laser mit einer Wellenlänge von 337 nm eingesetzt. Vor der Untersuchung von Oligonukleotiden wurden diese mittels Mikrodialyse (10 mL H₂O, min. 30 min, RT) über eine *MF-Millipore Membran VSWP02500* der Firma *Merck Millipore* aufgereinigt. Zur Messung von MALDI-Spektren wurde die Probensubstanz je nach Menge in 0,5 µL bzw. 2 × 0,5 µL HPA-Matrix

eingebettet (Zusammensetzung siehe Tabelle 9-a). Die Auswertung erfolgte durch Angabe der berechneten Masse sowie der für das Molekülionensignal gefundenen Masse.

Tabelle 9-a Zusammensetzung der in dieser Arbeit verwendeten HPA-Matrix für die MALDI-TOF-MS-Analyse.

Komponente	Anteil
HPA	gesättigt
Diammoniumhydrogencitrat	10 mg/mL
TFA	1 μ L/mL (0,1 %)
H ₂ O _{LC-MS-rein}	50 %
MeCN _{LC-MS-rein}	50 %

10 Spektroskopie

10.1 NMR-Spektroskopie

NMR-Spektren wurden an einem *Bruker Ascend 400*-Spektrometer aufgenommen. Alle ¹H-NMR-Spektren wurden auf die Resonanzlinie des undeuterten Lösungsmittelanteils kalibriert.^[254] ¹H-NMR-Daten wurden folgendermaßen angegeben: chemische Verschiebung δ in ppm relativ zu Tetramethylsilan, Multiplizität, Kopplungskonstanten ${}^nJ_{X/Y}$, Integral, Zuordnung. Zur Unterscheidung von diastereotopen Protonen einer Methylengruppe wurde dem Proton mit größerer chemischer Verschiebung der Zusatz „A“ und dem Proton mit geringerer chemischer Verschiebung der Zusatz „B“ hinzugefügt. Kopplungskonstanten ${}^nJ_{X/Y}$, zwischen dem Proton an Kohlenstoff X und dem Proton an Kohlenstoff Y über n Bindungen, sind in Hz angegeben. Aufgeführte Kopplungskonstanten wurden über alle auftretenden Werte im Signal gemittelt. Die Zuordnung von ¹H-Signalen erfolgte gemäß der in der Synthesvorschrift vermerkten Nummerierung der Verbindung. Die folgenden Abkürzungen wurden zur Beschreibung der Spinmultiplizität verwendet: s (Singulett), d (Dublett), t (Triplet), m (Multipllett) und m_c (zentrosymmetrisches Multipllett), wobei die Bezeichnungen auch entsprechend kombiniert wurden. Bei zentrosymmetrischen Multipletts wurde der Signalschwerpunkt angegeben, bei allen anderen Multipletts der gesamte Signalbereich. Alle NMR-Spektren wurden mit Hilfe der Software *MestReNova* der Firma *Mestrelab Research* ausgewertet.

10.2 Infrarot-Spektroskopie

Die Infrarot (IR)-Spektren wurden mit einem *Spectrum BX II* FT-IR-System der Firma *PerkinElmer* gemessen. Dabei wurde ein *DuraSampl/R II Diamond ATR*-Detektor der Firma *Smiths Detection* eingesetzt. Vor Messung der Probe wurde der Hintergrund aufgenommen. Das Hintergrundspektrum wurde anschließend vom Spektrum der Probe subtrahiert. Die Substanz wurde zunächst in CH_2Cl_2 gelöst und direkt auf die ATR-Einheit aufgetragen. Die Wellenzahl $\tilde{\nu}$ ist in cm^{-1} angegeben. Die Bandenintensitäten werden durch die folgenden Abkürzungen beschrieben: w (schwach), m (mittel), s (stark). Breite Bänder werden zusätzlich mit v (broad) bezeichnet.

11 Chromatographie

11.1 Dünnschichtchromatographie

Der Reaktionsverlauf sowie der Verlauf der *Flash*-Säulenchromatographie wurden durch qualitative Dünnschichtchromatographie (DC) überwacht. Es wurden Aluminium Fertigplatten mit Kieselgel *60 F₂₅₄* der Firma *Merck* verwendet. Die Substanzen wurden unter UV-Licht ($\lambda_1 = 254 \text{ nm}$, $\lambda_2 = 366 \text{ nm}$) und/oder durch Eintauchen in ein Nachweisreagenz und anschließendes Trocknen im Heißluftstrom visualisiert.

11.2 Säulenchromatographie

Flash-Säulenchromatographie (FSC) wurde an Kieselgel 60 (Korngröße: 63-200 μm) der Firma *Merck* durchgeführt. Alle produkthaltigen Fraktionen wurden vereinigt und die Lösungsmittel *in vacuo* entfernt. Dazu wurden Rotationsverdampfer der Firma *Heidolph Instruments (Laborota 4000-efficient)* unter Membranpumpenvakuum (*Vacuubrand Vario PC 2001*) verwendet. Anschließend wurde die Substanz im Hochvakuum getrocknet.

11.3 HPLC

Auftrennung von Gemischen mittels HPLC (Hochleistungsflüssigkeitschromatographie) erfolgte nach dem Umkehrphasen-Prinzip (RP-HPLC) an Anlagen der Firma *Waters*. Analytische Auftrennungen erfolgten an einer *Alliance 2695*-Anlage mit *Photo Diode Array Detector 2996* mit einer *EC 250/4 Nucleodur 100-3 C₁₈ ec*-Säule der Firma *Macherey & Nagel*. Präparative Auftrennungen erfolgten an einer *Alliance 1525*-Anlage mit *Dual Wavelength UV-Detector 2487*

mit einer *VP 250/10 Nucleodur 100-5 C₁₈ ec*-Säule der Firma *Macherey & Nagel*. Wenn nicht anders vermerkt erfolgten alle Auftrennungen unter Anwendung von Gradientenmethoden mit dem folgenden Puffersystem: HPLC-Puffer A (Triethylammonium-acetat (100 mmol/L) in H₂O), HPLC-Puffer B (Triethylammoniumacetat (100 mmol/L) in H₂O/MeCN (20/80). Gradienten wurden in Klammern als prozentuale Volumenanteile angegeben. MeCN-haltige Lösungen wurden mit MeCN (*HiPerSolv CHROMANORM*, $\varphi(\text{H}_2\text{O}) \leq 30$ ppm) der Firma *VWR* hergestellt. Zum Entfernen von Partikeln wurden die Proben vor der HPL-Chromatographie 10 min bei 21130 *g* zentrifugiert (analytische HPLC) oder durch einen Membranfilter (0,2 μm) filtriert (präparative HPLC).

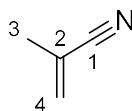
12 Nomenklatur

Die Benennung der Verbindungen erfolgte gemäß *IUPAC*-Empfehlungen. Ausgenommen davon sind Strukturen mit etablierten Trivialnamen. Die Nummerierung einer Verbindung zur Zuordnung von NMR-Signalen, nicht aber zur Benennung, erfolgte willkürlich und ohne Berücksichtigung der *IUPAC*-Empfehlungen. Homotope Kohlenstoffatome sind je Verbindung nur einmal gekennzeichnet.

13 Synthesevorschriften

13.1 Synthesen bezüglich m⁵C (2) und m⁵U (3)

13.1.1 Methacrylnitril (32)



herkömmliche Synthese

Methacrylformamid (**38**, 20,0 g, 235 mmol, 1,00 Äq.) wurde mit P₄O₁₀ (50,0 g, 353 mmol, 1,00 Äq.) unter Ar-Atmosphäre gemörsert. Das Reaktionsgemisch wurde 15 min bei 80 °C erhitzt, bevor die Temperatur auf 100 °C für weitere 75 min erhöht wurde. Dabei wurde das Produkt unter vermindertem Druck destilliert (380 mbar, $T_{\text{Kopf}} = 57$ °C) und in einem N_{2(l)}-gekühlten Auffangkolben gesammelt. Die Destillation ergab Methacrylnitril (**32**, 5,64 g, 84,1 mmol, 36 %) als farblose Flüssigkeit. Die Synthese erfolgte in Anlehnung an YANG *et al.*^[213]

¹H-NMR (400 MHz, CDCl₃): $\delta = 5,81$ (dq, $^4J_{4A/3} = 1,1$ Hz, $^2J_{4A/4B} = 1,1$ Hz, 1H, 4-H_A), 5,71 (qd, $^4J_{4B/3} = 1,7$ Hz, $^2J_{4B/4A} = 0,7$ Hz, 1H, 4-H_B), 1,98 (dd, $^4J_{3/4B} = 1,7$ Hz, $^2J_{3/4A} = 1,2$ Hz, 3H, 3-H) ppm.

¹³C{¹H}-NMR (101 MHz, CDCl₃): $\delta = 131,29$ (C-1), 119,22 (C-2), 118,10 (C-4), 20,76 (C-3) ppm.

HR-EI-MS von C₄H₆N⁺ [(M+H)⁺]:
gefunden 68,0495
berechnet 68,0495.

IR (ATR): $\tilde{\nu} = 2964$ (w), 2930 (w), 2228 (w), 1881 (w), 1625 (w), 1451 (m), 1380 (w), 1266 (w), 1018 (w), 936 (s), 748 (w) cm⁻¹.

13.1.2 2-Methyloxiran-2-carbonitril (**33**)



herkömmliche Synthese

Zu einer Lösung von Methacrylnitril (**32**, 1,40 g, 20,9 mmol, 1,00 Äq.) in CHCl_3 (20,0 mL) wurde mCPBA ($w_{\text{mCPBA}} = 77\%$, 7,81 g, 34,9 mmol, 1,67 Äq.) gegeben und das Reaktionsgemisch 20 h refluxiert. Der entstandene Niederschlag wurde abfiltriert und mit kaltem CHCl_3 gewaschen. Das Lösungsmittel wurde *in vacuo* entfernt und die Filtration noch einmal wiederholt. In einer ersten Vakuumdestillation wurden die flüchtigen Stoffe vom festen Rückstand getrennt (16 mbar, $T_{\text{Ölbad}} = 60\text{ °C}$). In einer zweiten Destillation (150 mbar, $T_{\text{Ölbad}} = 60\text{ °C}$) wurde restliches Methacrylnitril (**32**) aus dem zuvor erhaltenen Destillat entfernt. In einer dritten Destillation (75 mbar, $T_{\text{Kopf}} = 60\text{ °C}$) wurden die flüssigen Verunreinigungen aus dem zuvor erhaltenen zweiten Destillat entfernt. Der Rückstand dieser dritten Destillation enthielt das Zielprodukt **33** sowie einen geringen Anteil an Verunreinigungen. Diese wurden bei einer letzten Destillation (35 mbar, $T_{\text{Kopf}} = 48\text{--}51\text{ °C}$) zusammen mit dem Zielprodukt **33** abgetrennt. Der Rückstand dieser letzten Destillation enthielt das Zielprodukt 2-Methyloxiran-2-carbonitril (**33**, 305 mg, 3,67 mmol, 18 %) als farblose Flüssigkeit in $^1\text{H-NMR}$ -spektroskopisch reiner Form. Diese Synthese wurde entwickelt um die Wolframat-Komplex-vermittelte Synthese von 2-Methyloxiran-2-carbonitril (**33**)^[215] zu umgehen.

$^1\text{H-NMR}$ (400 MHz, CDCl_3): $\delta = 3,22$ (d, $^2J_{4\text{A}/4\text{B}} = 5,3$ Hz, 1H, 4- H_A), 2,84 (d, $^2J_{4\text{B}/4\text{A}} = 5,3$ Hz, 1H, 4- H_B), 1,66 (s, 3H, 3-H) ppm.

$^{13}\text{C}\{^1\text{H}\}$ -NMR (101 MHz, CDCl_3): $\delta = 118,72$ (C-1), 53,27 (C-4), 44,59 (C-2), 20,31 (C-3) ppm.

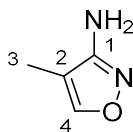
IR (ATR): $\tilde{\nu} = 3003$ (w), 2358 (w), 2246 (w), 1448 (w), 1386 (w), 1336 (w), 1244 (m), 1153 (w), 1111 (w), 1077 (m), 1009 (w), 917 (w), 871 (s), 802 (m), 756 (s), 668 (w), 655 (m) cm^{-1} .

Weder HR-EI- noch HR-LC-ESI-Massenspektrometrie führte zu aussagekräftigen Ergebnissen. Auch die zu Epoxid **33** publizierten EI-MS-Daten^[215] enthalten kein Molekül-Ion-Signal. Die angegebenen Fragment-Massen deuten auf den üblichen Hintergrund eines Massenspektrometers hin^[216] und ließen sich nicht reproduzieren.

präbiotische Modellreaktion

Methacrylnitril (**32**, 3,35 mg, 49,9 μmol , 1,00 Äq.) wurde in $\text{D}_2\text{O}/\text{H}_2\text{O}$ (4/1, 0,50 mL) gelöst. Zur Lösung wurde $\text{H}_2\text{O}_{2(\text{aq})}$ ($w_{\text{H}_2\text{O}_2} = 35\%$, 12,9 μL , 300 μmol , 6,00 Äq.) gegeben, das Reaktionsgemisch in ein NMR-Röhrchen überführt und bei RT geschwenkt. ^1H -NMR-spektroskopische Untersuchung des Reaktionsgemisches zeigte die Bildung von 2-Methyloxiran-2-carbonitril (**33**) nach 16,5 h.

13.1.3 3-Amino-4-methylisoxazol (34)



herkömmliche Synthese

Methacrylnitril (**32**, 188 μL , 2,24 mmol, 1,00 Äq.) wurde auf 0 °C gekühlt, bevor Br_2 (97,3 μL , 1,90 mmol, 0,85 Äq.) über 5 min zugetropft wurde. Das Reaktionsgemisch wurde auf 35 °C erwärmt und für 4,5 h gerührt. Anschließend wurde das Reaktionsgemisch auf 0 °C gekühlt, bevor Hydroxyharnstoff (221 mg, 2,91 mmol, 1,30 Äq.) sowie $\text{NaOH}_{(\text{aq})}$ (20 mol/L, 300 μL , 6,0 mmol, 2,70 Äq.) zugegeben wurden. Die Reaktion wurde auf RT erwärmt und für 13 h gerührt. Anschließend wurde das Reaktionsgemisch mit CH_2Cl_2 (3 x) extrahiert. Die vereinigten organischen Phasen wurden über MgSO_4 getrocknet und das Lösungsmittel *in vacuo* entfernt. wurde als farbloser Feststoff erhalten (13 mg, 133 μmol , 5,9 %). Die Synthese erfolgte in Anlehnung an KÖHN *et al.*[219]

$^1\text{H-NMR}$ (400 MHz, $\text{DMSO-}d_6$): δ = 8,09 (d, J = 0,5 Hz, 1H, 4-H), 5,45 (s, 2H, NH_2), 1,81 (d, J = 1.1 Hz, 3H, 3-H) ppm.

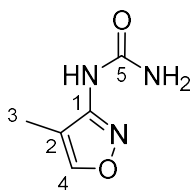
HR-HESI-MS von $\text{C}_4\text{H}_7\text{N}_2\text{O}^+$ [(M+H) $^+$]:
gefunden 90,0552
berechnet 99,0553.

präbiotische Modellreaktion

2-Methyloxiran-2-carbonitril (**33**, 7,34 μL , 96,3 μmol , 1,00 Äq.) wurde in $\text{NaHCO}_3_{(\text{aq})}$ (D_2O , pH 8, 1,00 mL) gelöst. Anschließend wurde Hydroxyharnstoff $_{(\text{aq})}$ (2 mol/L, 50,6 μL , 101 μmol , 1,05 Äq.) hinzugegeben und bei 40 °C reagiert (800 rpm). Nach 21 h wurde die Reaktion zunächst für 4,5 h bei RT und anschließend bei -20 °C gelagert. Die Reaktion wurde anschließend bei 70 °C für 78 h fortgesetzt. Der Reaktionsverlauf wurde mittels HR-LC-ESI-MS verfolgt. Bildung des Zielprodukts 3-Amino-4-methylisoxazol (**34**) konnte ab der ersten Messung (1 h bei 70 °C) detektiert werden. Die HR-LC-ESI-MS-Koinjektion der Referenzverbindung steht noch aus. Reaktionszeiten über 78 h führten zu keinem relevanten Zuwachs an Zielproduktbildung.

HR-HESI-MS von $\text{C}_4\text{H}_7\text{N}_2\text{O}^+$ [(M+H) $^+$]:
gefunden 99,0558
berechnet 99,0553.

13.1.4 N-(4-Methylisoxazol-3-yl)harnstoff (36)



herkömmliche Synthese

3-Amino-4-methylisoxazol (**34**, 50,0 mg, 510 μmol , 1,00 Äq.) wurde in trockenem THF (750 μL) bei RT gelöst. Nach Abkühlen auf 0 °C wurde 2,2,2-Trichloracetylisocyanat (60,7 μL , 510 μmol , 1,00 Äq.) über 30 min unter Ar-Atmosphäre hinzugegeben. Im Anschluss wurde auf 25 °C erwärmt und für 3 h reagiert (800 rpm). Die Reaktion wurde durch Zugabe von MeOH gestoppt und das Reaktionsgemisch in EtOH aufgenommen. Die Lösungsmittel wurden *in vacuo* entfernt und methanolisches NH_3 (7 mol/L, 700 μL) bei RT hinzugegeben. Die Lösungsmittel und überschüssiges NH_3 wurden direkt im Anschluss *in vacuo* entfernt. Aufreinigung mittels FSC ($\text{CH}_2\text{Cl}_2/\text{MeOH}$, 19/1) ergab 1-(4-Methylisoxazol-3-yl)harnstoff (**36**, 31,0 mg, 220 μmol , 43 %) als weißen Feststoff. Die Synthese erfolgte in Anlehnung an Abschnitt 4.1.

$^1\text{H-NMR}$ (400 MHz, $\text{DMSO-}d_6$): δ = 8,90 (s, 1H, NH), 8,44 (d, $^4J_{4/3}$ = 1,2 Hz, 1H, 4-H), 6,67 (s, 2H, NH_2), 1,89 (d, $^4J_{3/4}$ = 1,2 Hz, 3H, 3-H) ppm.

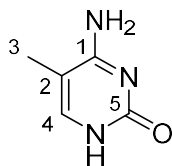
$^{13}\text{C}\{^1\text{H}\}\text{-NMR}$ (101 MHz, $\text{DMSO-}d_6$): δ = 158,37 (C-1), 156,35 (C-4), 154,75 (C-5), 107,86 (C-2), 5,96 (C-3) ppm.

HR-EI-MS von $\text{C}_5\text{H}_8\text{N}_3\text{O}_2^+$ [(M+H) $^+$]:
gefunden 142,0618
berechnet 142,0611.

präbiotische Modellreaktion

In einem Mörser wurde 3-Amino-4-methylisoxazol (**34**, 50,0 mg, 510 μmol , 1,00 Äq.) mit Harnstoff (91,8 mg, 1,53 mmol, 3,00 Äq.) und ZnCl_2 (69,5 mg, 510 μmol , 1,00 Äq.) vermischt. Das Reaktionsgemisch wurde im Anschluss bei 95 °C (800 rpm) zur Reaktion gebracht. Nach einer Reaktionszeit von 42,5 h konnte die Bildung von 4-Methylisoxazol-3-ylharnstoff (**36**) mittels HR-ESI-LC-MS (Verdünnung in H_2O : 1/1000) nachgewiesen werden. Die HR-LC-ESI-MS-Koinjektion der Referenzverbindung steht noch aus. Die Reaktion erfolgte in Anlehnung an Abschnitt 4.1.

13.1.5 5-Methylcytosin (**52**)



herkömmliche Synthese

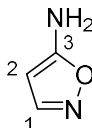
Eine Lösung von 4-Methylisoxazol-3-ylharnstoff (**36**, 20,0 mg, 142 μmol , 1,00 Äq.) in trockenem MeOH (4,00 mL) wurde mit einer katalytischen Mengen NaCO_3 versetzt. Das Reaktionsgemisch und die Atmosphäre darüber wurden mit H_2 gesättigt. Im Anschluss wurde eine kleine Spatelspitze Pd auf Aktivkohle ($w_{\text{Pd}} = 10\%$) zugegeben. Nach 50 min bei RT wurde das Reaktionsgemisch über Kieselgur filtriert. Entfernung des Lösungsmittels *in vacuo* ergab 5-Methylcytosin (**52**, quant.) als farblosen Feststoff.

$^{13}\text{C}\{^1\text{H}\}$ -NMR (101 MHz, MeOD): $\delta = 168,35$ (C-1), 160,57 (C-5), 141,69 (C-4), 103,35 (C-2), 13,01 (C-3) ppm.

HR-EI-MS von $\text{C}_5\text{H}_8\text{N}_3\text{O}^+$ [(M+H) $^+$]:
gefunden 126,0662
berechnet 126,0662.

13.2 Synthesen bezüglich s²U (1)

13.2.1 5-Aminoisoxazol (55)



herkömmliche Synthese

Hydroxylamin ($\omega = 50\%$ in H_2O , 19,0 mL, 348 mmol, 1,10 Äq.) wurde zu 3-Ethoxyacrylonitril (28,0 mL, 272 mmol, 1,00 Äq.) gegeben und das Reaktionsgemisch für 2 h bei 60 °C gerührt. Nach Aufreinigung mittels FSC (Hex/EtOAc, 2/1 \rightarrow 1/1) wurde 5-Aminoisoxazol (**55**, 11,1 g, 132 mmol, 48 %) in Form weißer Kristalle erhalten. Die Synthese erfolgte nach WIEDEMANN *et al.*^[223]

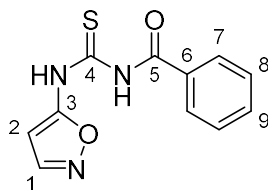
¹H-NMR (400 MHz, $\text{DMSO-}d_6$): $\delta = 8,01$ (d, $^3J_{1/2} = 1,9$ Hz, 1H, 1-H); 6,66 (s, 2H, 3-N-H); 4,88 (d, $^3J_{2/1} = 2,0$ Hz, 1H, 2-H) ppm.

¹³C{¹H}-NMR (101 MHz, $\text{DMSO-}d_6$): $\delta = 170,02$; 151,94; 76,42 ppm.

HR-HESI-LC-MS von $\text{C}_3\text{H}_5\text{N}_2\text{O}^+$ [(M+H)⁺]:
berechnet 85,0396
gefunden 85,0398.

IR (ATR): $\tilde{\nu} = 3309$ (w), 3250 (w), 3195 (w), 3164 (w), 2982 (vw), 2216 (vw), 2120 (vw), 2058 (vw), 2001 (vw), 1629 (vs), 1588 (s), 1494 (vs), 1374 (w), 1312 (w), 1199 (s), 1107 (s), 1033 (w), 990 (s), 910 (vs), 852 (s), 752 (vs), 689 (w) cm^{-1} .

13.2.2 *N*¹-Benzoyl-*N*³-(isoxazol-5-yl)thioharnstoff (60)



herkömmliche Synthese

Die gesamte Reaktion wurde unter Anwendung von SCHLENK-Technik durchgeführt. 5-Aminoisoxazol (**55**, 4,00 g, 47,6 mmol, 1,00 Äq.) wurde in trockenem THF (24 mL) gelöst und Benzoylisothiocyanat (**59**, 12,8 mL, 95,2 mmol, 2,00 Äq.) mittels Spritzenpumpe (2,00 mL/h) bei RT zugegeben. Das Reaktionsgemisch wurde für 21 h bei RT gerührt. Der entstandene weiße Niederschlag wurde abfiltriert, mit trockenem, kaltem THF gewaschen und *in vacuo* getrocknet. *N*¹-Benzoyl-*N*³-(isoxazol-5-yl)thioharnstoff (**60**, 8,66 g, 35,0 mmol, 74 %) wurde als weißes Pulver erhalten.

¹H-NMR (400 MHz, DMSO-*d*₆): δ = 13,61 (s, 1H, 3-N-H/5-N-H); 12,16 (s, 1H, 3-N-H/5-N-H); 8,60 (d, ³*J*_{1/2} = 2,0 Hz, 1H, 1-H); 8,01 – 7,97 (m, 2H, 7-H); 7,71 – 7,66 (m, 1H, 9-H); 7,55 (t, ³*J*_{8/9} = 7,8 Hz, 2H, 8-H); 6,93 (d, ³*J*_{2/1} = 2,0 Hz, 1H, 2-H) ppm.

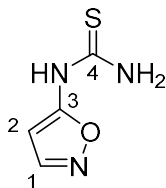
¹³C{¹H}-NMR (101 MHz, DMSO-*d*₆): δ = 177,14; 168,76; 160,47; 152,32; 133,55; 131,70; 128,97; 128,54; 91,30 ppm.

HR-EI-MS von C₁₁H₁₀N₃O₂S⁺ [M⁺]:
berechnet 247,0410
gefunden 247,0410.

IR (ATR): $\tilde{\nu}$ = 3166 (vw), 3001 (vw), 2184 (vw), 2054 (vw), 2009 (vw), 1981 (vw), 1926 (vw), 1669 (w), 1604 (s), 1551 (s), 1518 (s), 1490 (s), 1448 (s), 1336 (w), 1324 (vw), 1305 (w), 1260 (s), 1244 (s), 1190 (w), 1147 (s), 1086 (w), 1072 (w), 1018 (w), 999 (w), 942 (vw), 928 (w), 911 (s), 828 (w), 784 (w), 752 (s), 711 (vs), 671 (vs) cm⁻¹.

Schmelzbereich: 135,3 - 158,8 °C.

13.2.3 *N*-(Isoxazol-5-yl)thioharnstoff (57)



herkömmliche Synthese

Die gesamte Reaktion wurde unter Anwendung von SCHLENK-Technik durchgeführt. Zu einer Lösung von NaOMe (251 mg, 4,65 mmol, 1,15 Äq.) in trockenem MeOH (20 mL) wurde *N*¹-Benzoyl-*N*³-(isoxazol-5-yl)thioharnstoff (**60**, 1,00 g, 4,04 mmol, 1,00 Äq.) gegeben. Die Reaktion wurde bei RT gerührt und nach 18 h durch Zugabe von H₂O (1 mL) beendet. Anschließend wurden die Lösungsmittel *in vacuo* entfernt. Nach Zugabe von H₂O (15 mL) wurde mit EtOAc (3 x 15 mL) extrahiert. Die vereinigten organischen Phasen wurden über MgSO₄ getrocknet und das Lösungsmittel *in vacuo* entfernt. Nach Aufreinigung mittels FSC (CH₂Cl₂/MeOH, 100/1) wurde *N*-(Isoxazol-5-yl)thioharnstoff (**55**, 103 mg, 719 μmol, 18 %) als gelber Feststoff erhalten.

¹H-NMR (400 MHz, DMSO-*d*₆): δ = 11,06 (s, 1H, 3-N-H); 8,55 (s, 1H, 4-N-H_A); 8,41 (d, ³J_{1/2} = 2,0 Hz, 1H, 1-H); 7,40 (s, 1H, 4-N-H_B); 6,50 (d, ³J_{2/1} = 2,0 Hz, 1H, 2-H) ppm.

¹³C{¹H}-NMR (101 MHz, DMSO-*d*₆): δ = 179,58; 162,21; 151,80; 87,74 ppm.

HR-HESI-LC-MS von C₄H₆N₃OS⁺ [(M+H)⁺]:
berechnet 144,0226
gefunden 144,0228.

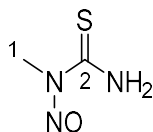
IR (ATR): $\tilde{\nu}$ = 3335 (w), 3281 (w), 3135 (w), 3032 (s), 2923 (s), 2852 (w), 2789 (w), 2011 (vw), 1632 (vs), 1606 (s), 1579 (vs), 1480 (s), 1352 (s), 1321 (vs), 1190 (s), 1087 (s), 1035 (s), 1000 (s), 912 (vs), 896 (s), 828 (w), 776 (vs), 687 (vs), 665 (vs) cm⁻¹.

Schmelzbereich: 128,0 – 130,2 °C.

präbiotische Modellreaktion

Die Reaktion erfolgte in einem geschlossenen System (*pressure tube*) unter Ar-Atmosphäre. P₄S₁₀ (38,5 mg, 86,5 μmol, 0,22 Äq.) wurde unter Ar abgewogen und zu *N*-(Isoxazol-

13.2.4 *N*¹-Methyl-*N*¹-(nitroso)thioharnstoff (**66**)



präbiotische Modellreaktion

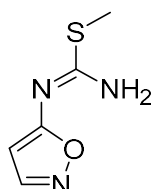
N-(Methyl)thioharnstoff (**65**, 3,00 g, 33,3 mmol, 1,00 Äq.) wurde in H₂O (20 mL) gelöst und auf -10 °C gekühlt. Natriumnitrit (2,53 g, 36,6 mmol, 1,10 Äq.) wurde zugegeben, bevor HCl_(aq) (10 mol/L, 4,66 mL) über einen Zeitraum von 20 min zugetropft wurde. Der entstandene Niederschlag wurde abfiltriert, mit kaltem H₂O gewaschen und *in vacuo* getrocknet. Das Zielprodukt *N*¹-Methyl-*N*¹-(nitroso)thioharnstoff (**66**, 1,19 g, 10,0 mmol, 30 %) wurde als gelber Feststoff erhalten. Die Reaktion erfolgte in Anlehnung an SCHNEIDER *et al.*^[231]

¹H-NMR (400 MHz, DMSO-*d*₆): δ = 9,88 (s, 1H, 2-N-H_A); 9,73 (s, 1H, 2-N-H_B); 3,39 (s, 3H, 1-H) ppm.

¹³C{¹H}-NMR (101 MHz, DMSO-*d*₆): δ = 182,16; 30,57 ppm.

Im Folgenden sind Synthesen aufgeführt, die weiterer Validierung bedürfen.

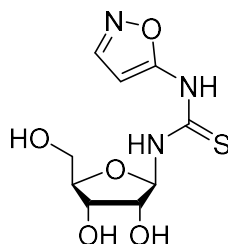
13.2.5 *S*-Methyl-*N*-(isoxazol-5-yl)thioharnstoff (**67**)



5-Aminoisoxazol (**55**, 2,0 mg, 24 μmol , 1,0 \AA q.) wurde in H_2O (1,00 mL) gelöst und *N*¹-Methyl-*N*¹-(nitroso)thioharnstoff (**66**, 28,3 mg, 238 μmol , 10,0 \AA q.) sowie NaHCO_3 (20,0 mg, 238 μmol , 10,0 \AA q.) wurden zugegeben. Anschließend wurde das Reaktionsgemisch für 1,5 h bei 55 °C geschüttelt (800 rpm). HR-LC-ESI-MS zeigte die Bildung des Zielprodukts *S*-Methyl-*N*-(isoxazol-5-yl)thioharnstoff (**67**). Die HR-LC-ESI-MS-Koinjektion der Referenzverbindung steht noch aus. Die Bildung des Zielprodukts **67** konnte statt mit NaHCO_3 auch in Anwesenheit einer der folgenden Katalysatoren detektiert werden: Borax (1,0 \AA q.), $(\text{NH}_4)_2\text{S}$ (3,0 \AA q.), Na_2HPO_4 (10,0 \AA q.) und CaCO_3 (5,00 \AA q.). Zudem konnte das Zielprodukt **67** mit jedem dieser Katalysatoren statt bei 55 °C auch bei Anwendung eines Temperaturgradienten (RT \rightarrow 70 °C) detektiert werden.

HR-HESI-LC-MS von $\text{C}_5\text{H}_8\text{N}_3\text{OS}^+$ [(M+H) ⁺]:	berechnet	158,0383
	gefunden	158,0388.

13.2.6 *N*¹-(β -D-Ribosyl)-*N*³-(isoxazol-5-yl)thioharnstoff (β -58a)



herkömmliche Synthese

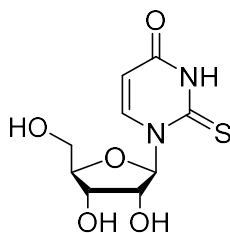
Die gesamte Reaktion wurde unter Anwendung von SCHLENK-Technik durchgeführt. 1'-*O*-Acetyl-2',3',5'-tri-*O*-benzoyl- β -D-ribose (**73**, 8,00 g, 15,9 mmol, 1,00 Äq.) wurde in trockenen CH₂Cl₂ (140 mL) gelöst. TiCl₄ (2,09 mL, 19,1 mmol, 1,20 Äq.) wurde zugegeben und das Reaktionsgemisch für 2 h bei RT gerührt. Anschließend wurde H₂O (120 mL) zugegeben und über Kieselgur filtert. Die organische Phase wurde separiert und die wässrige Phase mit CH₂Cl₂ (3 x 100 mL) extrahiert. Die vereinten organischen Phasen wurden mit NaCl_(aq) (ges., 200 mL) gewaschen, über MgSO₄ getrocknet, filtriert und das Lösungsmittel *in vacuo* entfernt. Das Rohprodukt wurde in trockenem Aceton (78,6 mL) aufgenommen und Kaliumthiocyanat (4,58 g, 47,2 mmol, 3,00 Äq.) zugegeben. Das Reaktionsgemisch wurde für 25 h bei RT gerührt und das Lösungsmittel *in vacuo* entfernt. Das entstandene Rohprodukt wurde in trockenem CH₂Cl₂ (149 mL) aufgenommen und 5-Aminoisoxazol (**55**, 1,51 g, 17,9 mmol, 1,20 Äq.) zugegeben. Das Reaktionsgemisch wurde vier Tage bei RT und für weitere zwei Tage bei 40 °C gerührt. Anschließend wurde das Lösungsmittel *in vacuo* entfernt. Nach Aufreinigung mittels FSC (Hex./EtOAc, 9/1 → 7/3) wurde *N*¹-(2',3',5'-tri-*O*-benzoyl- β -D-ribofuranosyl)-*N*²-(isoxazol-5-yl)thioharnstoff (**β -76a**, 125 mg, 210 μ mol, 1,3 %) als gelblicher Feststoff erhalten. Das Benzoyl-geschützte Derivat **β -76a** wurde mit methanolischer Ammoniak-Lösung (7 mol/L, 2 mL) versetzt und für 18 h bei RT gerührt, bevor das Lösungsmittel *in vacuo* entfernt wurde. Durch Isomerisierung während der Aufreinigung mittels RP-HPLC (H₂O/MeCN, 96/4 → 92/8), wurde ein Gemisch der Furanosid-Isomere (**β -58a**, **α -58a**) sowie Pyranosid-Isomere (**β -58b**, **α -58b**) des Zielprodukts erhalten. Somit konnte keine Analytik für das Zielprodukt **β -58a** durchgeführt werden. In Zukunft sollte diese Aufreinigung bei pH 7 stattfinden, z. B. in NEt₃NH₄OAc-Puffer (TEAA). Die Synthese erfolgte in Anlehnung an die in Abschnitt 4.1 etablierte Synthese der analogen Harnstoff-Verbindungen.

präbiotische Modellreaktion

Zu *N*-(Isoxazol-5-yl)thioharnstoff (**55**, 7,16 mg, 50,0 μmol , 1,00 Äq.) wurden wässrige Lösungen von *D*-Ribose (3 mol/L, 37,5 mg, 250 μmol , 5,00 Äq.) und $\text{ZnCl}_{2(\text{aq})}$ (100 mmol/L, 170 μg , 1,25 μmol , 0,025 Äq.) gegeben. Das Reaktionsgemisch wurde für 24 h bei 70 °C und 400 rpm geschüttelt. Der Rückstand wurde in H_2O (800 μL) aufgenommen und filtriert. Mittels HR-HESI-LC-MS konnte *N*¹-Ribosyl-*N*³-(isoxazol-5-yl)thioharnstoff (**58**) nachgewiesen werden. Die HR-LC-ESI-MS-Koinjektion der Referenzverbindungen steht noch aus.

HR-HESI-LC-MS von $\text{C}_3\text{H}_5\text{N}_2\text{O}^+$ [(M+H) ⁺]:	berechnet	276,0663
	gefunden	276,0658.

13.2.7 2-Thiouridin (β -1a)



Die gesamte Reaktion wurde unter Anwendung von SCHLENK-Technik durchgeführt. Zunächst wurde $\text{Na}_2\text{CO}_{3(\text{aq})}$ (75 mmol/L) für 1,5 h bei RT mit $\text{N}_{2(\text{g})}$ durchströmt. Anschließend wurde $\text{Na}_2\text{CO}_{3(\text{aq})}$ (75 mmol/L, 1,00 mL) zu einem Gemisch von N^1 -(2',3',5'-tri-*O*-benzoyl- β -D-ribofuranosyl)- N^2 -(isoxazol-5-yl)thioharnstoff (**β -76a**, 13,3 mg, 22,7 μmol , 1,00 Äq.), Butan-2,3-diol-1,4-dithiol (5,3 mg, 34 μmol , 1,5 Äq.) und Fe^{2+} (1,5 mg, 12 μmol , 0,55 Äq.) gegeben. Das schwach rote Reaktionsgemisch wurde auf 99 °C erhitzt und bei 1000 rpm für 8 h geschüttelt. Reaktionskontrolle mittels ^1H -NMR-Spektroskopie, LR-(LC-)ESI-MS sowie Dünnschichtchromatographie zeigten keine Bildung des Zielprodukts **β -1a**. Nach Aufbewahrung bei -20 °C über Nacht, wurde das Reaktionsgemisch durch Zugabe von $\text{HCl}_{(\text{aq})}$ (0,5 mol/L) auf pH 7 eingestellt. Nach Extraktion mit CH_2Cl_2 (1 mL) wurde die wässrige Phase mit NaCl gesättigt und mit EtOAc (1 x 1 mL, 2 x 600 μL) extrahiert. Anschließend wurden die Lösungsmittel der vereinigten organischen Phasen *in vacuo* entfernt. Zum Aufnehmen des Rohprodukts wurde H_2O (3,4 mL) zugegeben und für 20 min bei 40 °C geschwenkt. Eine HR-LC-ESI-MS-Koinjektion der Referenzverbindung s^2U (**β -1a**) bestätigte die Bildung des Zielprodukts s^2U (**β -1a**) unter präbiotisch plausiblen Bedingungen. Die Reaktion erfolgte in Anlehnung an die in Abschnitt 4.1 etablierte Bedingungen zu Reduktion und Ringschluss des 3-Aminoisoxazol-Analogs.

HR-HESI-LC-MS von $\text{C}_9\text{H}_{13}\text{N}_2\text{O}_5\text{S}^+$ [(M+H) $^+$]:	berechnet	261,0540
	gefunden	261,0551.

Alle weiteren Synthesevorschriften finden sich in den ergänzenden Informationen der dieser Arbeit beigefügten Publikationen, siehe Abschnitte 4.1, 5 und 6.1.

D Anhang

14 Abkürzungsverzeichnis

*	diese Autoren haben zu gleichen Teilen beigetragen
Δ	Differenz/in Reaktionsgleichungen: erhöhte Temperatur
$^1\text{H-NMR}$	$^1\text{H-nuclear magnetic resonance}/^1\text{H-Kernspinresonanz-Spektroskopie}$
$^{13}\text{C}\{^1\text{H}\}\text{-NMR}$	$^1\text{H-entkoppelte-}^{13}\text{C-Kernspinresonanz-Spektroskopie}$
A	Adenosin/im Kontext einer Nukleotidsequenz: Adenin
$A_{260\text{ nm}}$	spektrale Absorption bei $\lambda = 260\text{ nm}$
Abb.	Abbildung
abs.	absolut
Ac	acetyl-
Ac-dC-CE	5'-DMT- <i>N</i> ⁴ -acetyl-cytidin- 3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidit
Ac-mdC-CE	5'-DMT- <i>N</i> ⁴ -acetyl- <i>N</i> ⁵ -methyl-cytidin- 3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidit
ALKBH5	AlkB-Homolog 5, RNA-Demethylase
Am	2'- <i>O</i> -Methyladenosin
AMA	Ammoniumhydroxid-Methylamin
aq	in wässriger Lösung
Äq.	Äquivalente
arb. U.	<i>arbitrary units</i> /willkürliche Einheiten
ATR	abgeschwächte Totalreflexion
BER	Basen-Exzisions-Reperatur
bp	Basenpaare
BS	Bisulfit
Bz	benzoyl-
Bz-dA-CE	5'-DMT- <i>N</i> ⁶ -benzoyl-adenosin 3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidit

C	Cytidin/im Kontext einer Nukleotidsequenz: Cytosin
c_0	Ausgangs-Stoffmengenkonzentration
c'	Ziel-Stoffmengenkonzentration
ca.	circa
CAB	chemisch assistierte Bisulfit-Sequenzierung
caCAB	carboxyl-chemisch assistierte Bisulfit-Sequenzierung
cadC	2'-Deoxy-5-carboxycytidin
CAM	Cerammoniummolybdat
CE	cyanoethyl-
CLEVER-Seq	<i>chemical-labelling-enabled C-to-T conversion sequencing</i>
Cm	2'-O-Methylcytidin
CpG(s)	CpG-Dinukleotid(e)
CV	<i>column volume</i> /Säulenvolumen; CV: In der Chromatographie das Volumen der mobilen Phase, welches nötig ist um die stationäre Phase vollständig auszufüllen.
d	Duplett
D	Adenin oder Guanin oder Thymin/Uracil
dA	2'-Deoxyadenosin
DC	Dünnschichtchromatographie
dC	2'-Deoxycytidin
ddPCR	<i>droplet digital PCR</i> /Tröpfchen-basierte digitale PCR
Dest.	Destillation
dG	2'-Deoxyguanosin
d. h.	das heißt
DMF	<i>N,N</i> -Dimethylformamid
DMSO	Dimethylsulfoxid
DMT	4,4'-Dimethoxytrityl
DNA	deoxyribonucleic acid/Desoxyribonukleinsäure
DNMT	Proteinfamilie der DNA-Methyltransferasen
dNTPs	Deoxyribonukleosidtriphosphate
dr	<i>diastereomeric ratio</i> /Diastereomerenverhältnis
DRACH	Sequenzmotiv bestehend aus: A/G/U – A/G – A – C – A/U/C
ds	<i>double-stranded</i> /doppelsträngig
dT	2'-Deoxythymidin
dT-CE	5'-DMT-Thymidin-3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidit

DTT	Dithiothreitol (Butan-2,3-diol-1,4-dithiol)
dU	2'-Deoxyuridin
e⁻	Elektron
EI	Elektronenstoßionisation
ESI	Elektrosprayionisation
<i>et al.</i>	<i>et alii</i> /und andere
EtOH	Ethanol
Et	ethyl-
etc.	<i>et cetera</i> /und so weiter
FA	<i>formic acid</i> /Ameisensäure
FaPy(s)	Formamidopyrimidin(e)
fdC	2'-Deoxy-5-formylcytidin
fCAB	formyl-chemisch assistierte Bisulfit-Sequenzierung
FMR1	fragile X mental retardation 1 / fragiles X mentale Retardierung 1
FRET	Förster-Resonanzenergietransfer
FSC	<i>Flash</i> -Säulenchromatographie
FTO	<i>Fat mass and obesity-associated</i> -Protein
g	gasförmig
G	Guanosin/im Kontext einer Nukleotidsequenz: Guanin
G3BP1	Ras GTPase-activating protein-binding protein 1/ Ras GTPase-aktivierendes Protein-bindendes Protein 1
gDNA	genomische DNA
gem.	gemessen
ges.	gesättigt
Gm	2'- <i>O</i> -Methylguanosin
Gya	Milliarden Jahre vor unserer Zeit
H	Adenin oder Cytosin oder Thymin/Uracil
HA	Hydroxylamin-Reagenz 81
hmdC	2'-Deoxy-5-hydroxycytidin
hnRNP	<i>heterogeneous nuclear ribonucleoproteins</i> / heterogene Kern-Ribonukleoproteine
HOAc	Essigsäure
HP(s)	Haarnadel-Oligonukleotid(e)

HPBS	<i>hairpin</i> BS-Seq/BS-Seq unter Verwendung von HPs
HPA	3-hydroxypyridine-2-carboxylic acid/3-Hydroxypyridin-2-carbonsäure
HPLC	<i>high performance liquid chromatography</i> / Hochleistungsflüssigkeitschromatographie
HPoxBS	oxBS unter Verwendung von HPs
HR-MS	<i>high resolution mass spectrometry</i> /hochauflösende Massenspektrometrie
I	Inosin
iBu	isobutyryl-
iBu-dG-CE	5'-(4,4'-DMT)- <i>N</i> ² -isobutyryl-guanosin- 3'-[[2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidit
Im	<i>O</i> ² -Methylinosin
IR	Infrarotspektroskopie –
Konz.	Konzentration
konz.	konzentriert
l	<i>liquidus</i> /flüssig
LC	<i>liquid chromatography</i> /Flüssigkeitschromatographie
LR-MS	<i>low resolution mass spectrometry</i> / niedrigauflösende Massenspektrometrie
LUCA	<i>the last universal common ancestor</i> /der letzte gemeinsame Urvorfahr
m	<i>medium</i> /mittelstarkes Signal (Infrarotspektroskopie)
M	Molekül
m ¹ G	<i>N</i> ¹ -Methylguanosin
m ² ₂ G	<i>N</i> ² , <i>N</i> ² -Dimethylguanosin
m ⁵ C	5-Methylcytosin
m ⁵ U	5-Methyluridin = Ribothymidin (T)
m ⁶ A	<i>N</i> ⁶ -Methyladenosin
m ⁶ ₂ A	<i>N</i> ⁶ , <i>N</i> ⁶ -Dimethyladenosin
MALDI	Matrix-unterstützte Laser-Desorption/Ionisation
max.	maximal
mCPBA	<i>meta</i> -Chlorperbenzoesäure
mdC	2'-Deoxy-5-methylcytidin
Me	methyl-

MeCN	Acetonitril
MeOD	Methanol-d1
MeOH	Methanol
mESC	<i>murine embryonic stem cells</i> /embryonale Stammzellen der Maus.
METTL	Familie der <i>methyltransferase like</i> -Proteine
Mio.	Million(en)
mod.	Modifiziert
Mrd.	Milliarde(n)
mRNA	<i>messenger</i> RNA/Boten-RNA
MS	Massenspektrometrie
MS/MS	Tandem-Massenspektrometrie
ms ² A	2-Thiomethyladenosin
n	
<i>n</i>	Anzahl
N	Adenin oder Thymin oder Guanin oder Cytosin
NaOMe	Natriummethanolat
NEt ₃	Triethylamin
NGS	<i>next generation sequencing</i> /Sequenzierung der nächsten Generation
NOMe	<i>Nucleosome Occupancy and Methylome sequencing</i>
Nt.	Nukleotid(e)
od.	
od.	oder
Oligo.	Oligonukleotid
oxBS	oxidative Bisulfit-Sequenzierung
ox-mdC(s)	hmdC, fdC, cadC
PCR	
PCR	<i>polymerase chain reaction</i> /Polymerase-Kettenreaktion
pH	<i>potentia Hydrogenii</i> /Potenz des Wasserstoffs, pH $-\log_{10} a(\text{H}^+)$
ppm	<i>parts per million</i> /Millionstel, 1 ppm = $1 \cdot 10^{-6}$
q	
quant.	quantitative Ausbeute
R	
R	Adenin oder Guanin
R _F	<i>retardation factor</i> /Retardierungsfaktor (Dünnschicht-Chromatographie)
redBS	reduzierte Bisulfit-Sequenzierung
Redox-	Reduktions-Oxidations-

Rkt.	Reaktion
RNA	ribonucleic acid/Ribonukleinsäure
RP-HPLC	<i>reversed phase high performance liquid chromatography</i> / Umkehrphasen-Hochleistungsflüssigkeitschromatographie
rpm	<i>revolutions per minute</i> /Umdrehungen pro minute
RRBS	<i>reduced representation bisulfite sequencing</i>
rU*	ddPCR-basierte Methode fdC-Sequenzierung ohne Bisulfit ^[158]
s	<i>strong</i> /starkes Signal (Infrarotspektroskopie)
s ² C	2-Thiocytidin
s ² U	2-Thiouridin
SILAC	<i>stable isotope labeling by/with amino acids in cell culture</i> / Markierung von/mit Aminosäuren durch stabile Isotopen in Zellkultur
SNuPE	<i>single nucleotide primer extension</i> /Einzelnukleotid-Primer-Extension
ss	<i>single-stranded</i> /einzelsträngig
synth.	synthetisch (Herstellung mittels Oligonukleotid-Festphasensynthese)
t	Triplett
T	Ribothymidin = 5-Methyluridin (m ⁵ U)/ /im Kontext einer Nukleotidsequenz: Thymin
T _{Kopf}	Kopf-Temperatur (Destillation)
Tab.	Tabelle
TBAB	Tetrabutylammoniumbromid
TBDMS	<i>tert</i> -Butyldimethylsilyl
TDG	Thymin-DNA-Glykosylase
TEAA	Triethylammoniumacetat-Puffer
TET	Proteinfamilie der <i>ten-eleven translocation methylcytosine dioxygenases</i>
TF(s)	Transkriptionsfaktor(en)
THF	Tetrahydrofuran
TOF-MS	<i>time-of-flight mass spectrometry</i> /Flugzeit-Massenspektrometrie
tRNA	Transfer-RNA
U	Uridin/im Kontext einer Nukleotidsequenz: Uracil
u. a.	unter anderem/anderen
Um	2'-O-Methyluridin
UMI	<i>unique molecular identifier</i> /Identifikationssequenz
unmod.	unmodifiziert

usw.	und so weiter
UTR	untranslatierte Region
UV-	Ultraviolett(strahlung)
vgl.	vergleiche
vs	<i>very strong</i> /sehr starkes Signal (Infrarotspektroskopie)
vw	<i>very weak</i> /sehr schwaches Signal (Infrarotspektroskopie)
w	<i>weak</i> /schwaches Signal (Infrarotspektroskopie)
W	Adenin oder Thymin
WGBS	<i>whole genome bisulfite sequencing</i> /genomweite Bisulfitsequenzierung
WT	Wildtyp
YTH	Proteindomäne YT521-B-Homologie
z	Ladungszahl eines Ions
z. B.	zum Beispiel

15 Literaturverzeichnis

- [1] National Research Council 2007, *The Limits of Organic Life in Planetary Systems*, The National Academies Press, Washington, DC, **2007**.
- [2] A. Rich, in *Horizons Biochem.* (Hrsg.: M. Kasha, B. Pullman), Academic Press, New York, **1962**, S. 103–126.
- [3] K. Kruger, P. J. Grabowski, A. J. Zaugg, J. Sands, D. E. Gottschling, T. R. Cech, *Cell* **1982**, *31*, 147–157.
- [4] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, *Cell* **1983**, *35*, 849–857.
- [5] W. Gilbert, *Nature* **1986**, *319*, 618.
- [6] S. A. Benner, H.-J. Kim, E. Biondi, in *Prebiotic Chemistry and Chemical Evolution of Nucleic Acids* (Hrsg.: C. Menor-Salván), Springer International Publishing, Cham, **2018**, S. 31–84.
- [7] T. R. Cech, *Cold Spring Harb. Perspect. Biol.* **2012**, *4*, 1–5.
- [8] A. W. Schwartz, *Mol. Orig. Life* **2010**, 237–254.
- [9] N. V. Hud, F. A. L. Anet, *J. Theor. Biol.* **2000**, *205*, 543–562.
- [10] N. Lahav, S. Nir, *Orig. Life Evol. Biosph.* **1997**, *27*, 377–395.
- [11] B. J. Cafferty, D. M. Fialho, N. V. Hud, in *Prebiotic Chemistry and Chemical Evolution of Nucleic Acids* (Hrsg.: C. Menor-Salván), Springer International Publishing, Cham, **2018**, S. 143–174.
- [12] M. C. Weiss, F. L. Sousa, N. Mrnjavac, S. Neukirchen, M. Roettger, S. Nelson-Sathi, W. F. Martin, *Nat. Microbiol.* **2016**, *1*, 1–8.
- [13] S. J. Mojzsis, G. Arrhenius, K. D. McKeegan, T. M. Harrison, A. P. Nutman, C. R. L. Friend, *Nature* **1996**, *384*, 55–59.
- [14] E. A. Bell, P. Boehnke, T. M. Harrison, W. L. Mao, *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 14518–14521.
- [15] E. J. Javaux, *Nature* **2019**, *572*, 451–460.
- [16] B. K. D. Pearce, R. E. Pudritz, *Astrophys. J.* **2015**, *807*, 85.
- [17] M. P. Callahan, K. E. Smith, H. J. Cleaves, J. Ruzicka, J. C. Stern, D. P. Glavin, C. H. House, J. P. Dworkin, *Proc. Natl. Acad. Sci.* **2011**, *108*, 13995–13998.
- [18] D. Deamer, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 11264–11265.
- [19] B. J. Wood, M. J. Walter, J. Wade, *Nature* **2006**, *441*, 825–833.
- [20] S. Kadoya, J. Krissansen-Totton, D. C. Catling, *Geochemistry, Geophys. Geosystems* **2020**, *21*, 1–18.
- [21] C. Sagan, C. Chyba, *Science* **1997**, *276*, 1217–1221.
- [22] N. T. Arndt, E. G. Nisbet, *Annu. Rev. Earth Planet. Sci.* **2012**, *40*, 521–549.
- [23] S. A. Benner, E. A. Bell, E. Biondi, R. Brassler, T. Carell, H. Kim, S. J. Mojzsis, A. Omran, M. A. Pasek, D. Trail, *ChemSystemsChem* **2019**, *1*, 1–21.
- [24] O. Leslie E., *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99–123.

- [25] „Ammonia“, can be found under <https://en.wikipedia.org/wiki/Ammonia#Structure>, **2020**.
- [26] „pH and composition of carbonic acid solutions“, can be found under https://en.wikipedia.org/wiki/Carbonic_acid#pH_and_composition_of_carbonic_acid_solutions, **2020**.
- [27] F. P. Terraglio, R. M. Manganelli, *J. Air Pollut. Control Assoc.* **1967**, *17*, 403–406.
- [28] P. Boehnke, E. A. Bell, T. Stephan, R. Trappitsch, C. Brenhin Keller, O. S. Pardo, A. M. Davis, T. Mark Harrison, M. J. Pellin, *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 6353–6356.
- [29] M. J. Russell, A. J. Hall, W. Martin, *Geobiology* **2010**, *8*, 355–371.
- [30] J. L. Bada, J. H. Chalmers, H. J. Cleaves, *Phys. Chem. Chem. Phys.* **2016**, *18*, 20085–20090.
- [31] B. Burcar, M. Pasek, M. Gull, B. J. Cafferty, F. Velasco, N. V. Hud, C. Menor-Salván, *Angew. Chemie* **2016**, *128*, 13443–13447.
- [32] J. L. Bada, J. Korenaga, *Life* **2018**, *8*, 55–70.
- [33] M. A. A. Schoonen, Y. Xu, *Astrobiology* **2001**, *1*, 133–142.
- [34] B. T. Burcar, L. M. Barge, D. Trail, E. B. Watson, M. J. Russell, L. B. McGown, *Astrobiology* **2015**, *15*, 509–522.
- [35] H. Follmann, C. Brownson, *Naturwissenschaften* **2009**, *96*, 1265–1292.
- [36] R. M. Hazen, *Am. J. Sci.* **2013**, *313*, 807–843.
- [37] S. M. Morrison, S. E. Runyon, R. M. Hazen, *Life* **2018**, *8*, 1–20.
- [38] H. Becker, M. F. Horan, R. J. Walker, S. Gao, J. P. Lorand, R. L. Rudnick, *Geochim. Cosmochim. Acta* **2006**, *70*, 4528–4550.
- [39] S. J. Mojzsis, *Dev. Precambrian Geol.* **2007**, *15*, 923–970.
- [40] J. F. Kasting, in *Geol. Soc. Am. Spec. Pap.*, Geological Society Of America, **2014**, S. 19–28.
- [41] F. Tian, J. F. Kasting, K. Zahnle, *Earth Planet. Sci. Lett.* **2011**, *308*, 417–423.
- [42] K. J. Zahnle, *J. Geophys. Res.* **1986**, *91*, 2819–2834.
- [43] S. L. Miller, G. Schlesinger, *Adv. Sp. Res.* **1983**, *3*, 47–53.
- [44] H. Genda, T. Iizuka, T. Sasaki, Y. Ueno, M. Ikoma, *Earth Planet. Sci. Lett.* **2017**, *470*, 87–95.
- [45] H. Genda, R. Brassler, S. J. Mojzsis, *Earth Planet. Sci. Lett.* **2017**, *480*, 25–32.
- [46] R. Brassler, S. J. Mojzsis, S. C. Werner, S. Matsumura, S. Ida, *Earth Planet. Sci. Lett.* **2016**, *455*, 85–93.
- [47] W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1972**, *67*, 25–33.
- [48] W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Evol.* **1972**, *1*, 249–257.
- [49] H. J. Kim, S. A. Benner, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 11315–11320.
- [50] H.-J. Kim, J. Kim, *Astrobiology* **2019**, *19*, 669–674.
- [51] I. Nam, H. G. Nam, R. N. Zare, *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 36–40.
- [52] I. Suárez-Marina, Y. M. Abul-Haija, R. Turk-MacLeod, P. S. Gromski, G. J. T. Cooper, A. O. Olivé, S. Colón-Santos, L. Cronin, *Commun. Chem.* **2019**, *2*, 1–8.

- [53] S. Becker, I. Thoma, A. Deutsch, T. Gehrke, P. Mayer, H. Zipse, T. Carell, *Science* **2016**, *352*, 833–836.
- [54] S. Becker, C. Schneider, H. Okamura, A. Crisp, T. Amatov, M. Dejmek, T. Carell, *Nat. Commun.* **2018**, *9*, 1–9.
- [55] M. Yadav, R. Kumar, R. Krishnamurthy, *Chem. Rev.* **2020**, *120*, 4766–4805.
- [56] R. Saladino, E. Carota, G. Botta, M. Kapralov, G. N. Timoshenko, A. Y. Rozanov, E. Krasavin, E. Di Mauro, *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E2746–E2755.
- [57] J. Xu, M. Tsanakopoulou, C. J. Magnani, R. Szabla, J. E. Šponer, J. Šponer, R. W. Góra, J. D. Sutherland, *Nat. Chem.* **2017**, *9*, 303–309.
- [58] M. W. Powner, B. Gerland, J. D. Sutherland, *Nature* **2009**, *459*, 239–242.
- [59] D. M. Fialho, T. P. Roche, N. V. Hud, *Chem. Rev.* **2020**, *120*, 4806–4830.
- [60] S. J. Mojzsis, *Dev. Precambrian Geol.* **2007**, *15*, 923–970.
- [61] R. K. Kumar, D. R. Davis, *Nucleic Acids Res.* **1997**, *25*, 1272–1280.
- [62] S. M. Testa, M. D. Disney, D. H. Turner, R. Kierzek, *Biochemistry* **1999**, *38*, 16655–16662.
- [63] W. Saenger, **1984**.
- [64] G. Jäger, R. Leipuviene, M. G. Pollard, Q. Qian, G. R. Björk, *J. Bacteriol.* **2004**, *186*, 750–757.
- [65] D. E. Condon, S. D. Kennedy, B. C. Mort, R. Kierzek, I. Yildirim, D. H. Turner, *J. Chem. Theory Comput.* **2015**, *11*, 2729–2742.
- [66] P. Davanloo, M. Sprinzl, K. Watanabe, M. Albani, H. Kersten, *Nucleic Acids Res.* **1979**, *6*, 1571–1581.
- [67] K. Watanabe, S. Yokoyama, F. Hansske, H. Kasai, T. Miyazawa, *Biochem. Biophys. Res. Commun.* **1979**, *91*, 671–677.
- [68] A. T. Larsen, A. C. Fahrenbach, J. Sheng, J. Pian, J. W. Szostak, *Nucleic Acids Res.* **2015**, *43*, 7675–7687.
- [69] S. Yokoyama, T. Watanabe, K. Murao, H. Ishikura, Z. Yamaizumi, S. Nishimura, T. Miyazawa, *Proc. Natl. Acad. Sci.* **1985**, *82*, 4905–4909.
- [70] B. D. Heuberger, A. Pal, F. Del Frate, V. V. Topkar, J. W. Szostak, *J. Am. Chem. Soc.* **2015**, *137*, 2769–2775.
- [71] W. A. Cantara, Y. Bilbille, J. Kim, R. Kaiser, G. Leszczyńska, A. Malkiewicz, P. F. Agris, *J. Mol. Biol.* **2012**, *416*, 579–597.
- [72] V. I. Lim, J. F. Curran, *RNA* **2001**, *7*, 942–957.
- [73] 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.
- [74] L. Frick, J. P. Mac Neela, R. Wolfenden, *Bioorg. Chem.* **1987**, *15*, 100–108.
- [75] J. Xu, N. J. Green, C. Gibard, R. Krishnamurthy, J. D. Sutherland, *Nat. Chem.* **2019**, *11*, 457–462.

- [76] G. R. Björk, K. Jacobsson, K. Nilsson, M. J. O. Johansson, A. S. Byström, O. P. Persson, *EMBO J.* **2001**, *20*, 231–239.
- [77] C. Schneider, S. Becker, H. Okamura, A. Crisp, T. Amatov, M. Stadlmeier, T. Carell, *Angew. Chem. Int. Ed.* **2018**, *57*, 5943–5946.
- [78] T. P. Hoernes, N. Clementi, K. Faserl, H. Glasner, K. Breuker, H. Lindner, A. Hüttenhofer, M. D. Erlacher, *Nucleic Acids Res.* **2016**, *44*, 852–862.
- [79] J. J. Xuan, W. J. Sun, P. H. Lin, K. R. Zhou, S. Liu, L. L. Zheng, L. H. Qu, J. H. Yang, *Nucleic Acids Res.* **2018**, *46*, D327–D334.
- [80] P. Boccaletto, M. A. Machnicka, E. Purta, P. Piątkowski, B. Bagiński, T. K. Wirecki, V. de Crécy-Lagard, R. Ross, P. A. Limbach, A. Kotter, u. a., *Nucleic Acids Res.* **2018**, *46*, D303–D307.
- [81] W. E. Cohn, E. Volkin, *Nature* **1951**, *167*, 483–484.
- [82] R. Desrosiers, K. Friderici, F. Rottman, *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 3971–3975.
- [83] J. M. Adams, S. Cory, *Nature* **1975**, *255*, 28–33.
- [84] Y. Aloni, R. Dhar, G. Khoury, *J. Virol.* **1979**, *32*, 52 LP-60.
- [85] K. Beemon, J. Keith, *J. Mol. Biol.* **1977**, *113*, 165–179.
- [86] G. Cao, H.-B. Li, Z. Yin, R. A. Flavell, *Open Biol.* **2016**, *6*, 160003.
- [87] K. D. Meyer, D. P. Patil, J. Zhou, A. Zinoviev, M. A. Skabkin, O. Elemento, T. V. Pestova, S. B. Qian, S. R. Jaffrey, *Cell* **2015**, *163*, 999–1010.
- [88] J. Zhou, J. Wan, X. Gao, X. Zhang, S. R. Jaffrey, S.-B. Qian, *Nature* **2015**, *526*, 591.
- [89] X. Wang, B. S. Zhao, I. A. Roundtree, Z. Lu, D. Han, H. Ma, X. Weng, K. Chen, H. Shi, C. He, *Cell* **2015**, *161*, 1388–1399.
- [90] W. Xiao, S. Adhikari, U. Dahal, Y. S. Chen, Y. J. Hao, B. F. Sun, H. Y. Sun, A. Li, X. L. Ping, W. Y. Lai, u. a., *Mol. Cell* **2016**, *61*, 507–519.
- [91] X. Zhao, Y. Yang, B. F. Sun, Y. Shi, X. Yang, W. Xiao, Y. J. Hao, X. L. Ping, Y. S. Chen, W. J. Wang, u. a., *Cell Res.* **2014**, *24*, 1403–1419.
- [92] J.-M. Fustin, M. Doi, Y. Yamaguchi, H. Hida, S. Nishimura, M. Yoshida, T. Isagawa, M. S. Morioka, H. Kakeya, I. Manabe, u. a., *Cell* **2013**, *155*, 793–806.
- [93] H. J. Kim, N. C. Kim, Y. D. Wang, E. A. Scarborough, J. Moore, Z. Diaz, K. S. MacLea, B. Freibaum, S. Li, A. Molliex, u. a., *Nature* **2013**, *495*, 467–473.
- [94] M. E. Hess, S. Hess, K. D. Meyer, L. A. W. Verhagen, L. Koch, H. S. Brönneke, M. O. Dietrich, S. D. Jordan, Y. Saletore, O. Elemento, u. a., *Nat. Neurosci.* **2013**, *16*, 1042–1048.
- [95] H. Li, T. O. Kilpeläinen, C. Liu, J. Zhu, Y. Liu, C. Hu, Z. Yang, W. Zhang, W. Bao, S. Cha, u. a., *Diabetologia* **2012**, *55*, 981–995.
- [96] I. Kalnina, L. Zaharenko, I. Vaivade, V. Rovite, L. Nikitina-Zake, R. Peculis, D. Fridmanis, K. Geldnere, J. A. Jacobsson, M. S. Almen, u. a., *Gene* **2013**, *527*, 462–468.
- [97] T. Sun, R. Wu, L. Ming, *Biomed. Pharmacother.* **2019**, *112*, 108613.

- [98] G. Jia, Y. Fu, X. Zhao, Q. Dai, G. Zheng, Y. Yang, C. Yi, T. Lindahl, T. Pan, Y.-G. Yang, u. a., *Nat. Chem. Biol.* **2011**, *7*, 885–887.
- [99] G. Zheng, J. A. Dahl, Y. Niu, P. Fedorcsak, C. M. Huang, C. J. Li, C. B. Vågbo, Y. Shi, W. L. Wang, S. H. Song, u. a., *Mol. Cell* **2013**, *49*, 18–29.
- [100] M. J. Clancy, M. E. Shambaugh, C. S. Timpte, J. A. Bokar, *Nucleic Acids Res.* **2002**, *30*, 4509–4518.
- [101] S. Zhong, H. Li, Z. Bodi, J. Button, L. Vespa, M. Herzog, R. G. Fray, *Plant Cell* **2008**, *20*, 1278–1288.
- [102] Y. Saletore, K. D. Meyer, J. Korlach, I. D. Vilfan, S. Jaffrey, C. E. Mason, *Genome Biol.* **2012**, *13*, 175.
- [103] D. Dominissini, S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, S. Osenberg, K. Cesarkas, J. Jacob-Hirsch, N. Amariglio, M. Kupiec, u. a., *Nature* **2012**, *485*, 201–206.
- [104] K. D. Meyer, Y. Saletore, P. Zumbo, O. Elemento, C. E. Mason, S. R. Jaffrey, *Cell* **2012**, *149*, 1635–1646.
- [105] B. Linder, A. V. Grozhik, A. O. Olarerin-George, C. Meydan, C. E. Mason, S. R. Jaffrey, *Nat. Methods* **2015**, *12*, 767–772.
- [106] S. Schwartz, M. R. Mumbach, M. Jovanovic, T. Wang, K. Maciag, G. G. Bushkin, P. Mertins, D. Ter-Ovanesyan, N. Habib, D. Cacchiarelli, u. a., *Cell Rep.* **2014**, *8*, 284–296.
- [107] S. Zaccara, R. J. Ries, S. R. Jaffrey, *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 608–624.
- [108] H. Shi, J. Wei, C. He, *Mol. Cell* **2019**, *74*, 640–650.
- [109] J. A. Bokar, M. E. Shambaugh, D. Polayes, A. G. Matera, F. M. Rottman, *RNA* **1997**, *3*, 1233–1247.
- [110] J. Liu, Y. Yue, D. Han, X. Wang, Y. Fu, L. Zhang, G. Jia, M. Yu, Z. Lu, X. Deng, u. a., *Nat. Chem. Biol.* **2014**, *10*, 93–95.
- [111] X. Wang, Z. Lu, A. Gomez, G. C. Hon, Y. Yue, D. Han, Y. Fu, M. Parisien, Q. Dai, G. Jia, *Nature* **2014**, *505*, 117.
- [112] X. Wang, B. S. Zhao, I. A. Roundtree, Z. Lu, D. Han, H. Ma, X. Weng, K. Chen, H. Shi, C. He, *Cell* **2015**, *161*, 1388–1399.
- [113] H. Shi, X. Wang, Z. Lu, B. S. Zhao, H. Ma, P. J. Hsu, C. Liu, C. He, *Cell Res.* **2017**, *27*, 315–328.
- [114] C. R. Alarcón, H. Goodarzi, H. Lee, X. Liu, S. Tavazoie, S. F. Tavazoie, *Cell* **2015**, *162*, 1299–1308.
- [115] N. Liu, Q. Dai, G. Zheng, C. He, M. Parisien, T. Pan, *Nature* **2015**, *518*, 560–564.
- [116] N. Liu, K. I. Zhou, M. Parisien, Q. Dai, L. Diatchenko, T. Pan, *Nucleic Acids Res.* **2017**, *45*, 6051–6063.
- [117] R. R. Edupuganti, S. Geiger, R. G. H. Lindeboom, H. Shi, P. J. Hsu, Z. Lu, S. Y. Wang, M. P. A. Baltissen, P. W. T. C. Jansen, M. Rossa, u. a., *Nat. Struct. Mol. Biol.* **2017**, *24*, 870–878.

- [118] C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. T. C. Jansen, C. Bauer, M. Münzel, M. Wagner, M. Müller, F. Khan, u. a., *Cell* **2013**, *152*, 1146–1159.
- [119] „Wiktionary, Eintrag „έπί“, can be found under <https://de.wiktionary.org/wiki/έπί>, **2020**.
- [120] R. Chahwan, S. N. Wontakal, S. Roa, *Trends Genet.* **2010**, *26*, 443–448.
- [121] M. Raff, B. Alberts, J. Lewis, A. Johnson, K. Roberts, *Molecular Biology of the Cell*, Garland Science, New York, **2002**.
- [122] O. J. Rando, K. J. Verstrepen, *Cell* **2007**, *128*, 655–668.
- [123] A. K. Lancaster, J. Masel, *Evolution (N. Y.)* **2009**, *63*, 2350–2362.
- [124] G. C. Burdge, S. P. Hoile, T. Uller, N. A. Thomas, P. D. Gluckman, M. A. Hanson, K. A. Lillycrop, *PLoS One* **2011**, *6*, e28282.
- [125] A. Caspi, K. Sugden, T. E. Moffitt, A. Taylor, I. W. Craig, H. Harrington, J. McClay, J. Mill, J. Martin, A. Braithwaite, *Science* **2003**, *301*, 386–389.
- [126] C. F. Rider, C. Carlsten, *Clin. Epigenetics* **2019**, *11*, 131.
- [127] S. B. Baylin, P. A. Jones, *Nat. Rev. Cancer* **2011**, *11*, 726–734.
- [128] A. K. Short, K. A. Fennell, V. M. Perreau, A. Fox, M. K. O’Byrne, J. H. Kim, T. W. Bredy, T. Y. Pang, A. J. Hannan, *Transl. Psychiatry* **2016**, *6*, e837–e837.
- [129] T.-F. Yuan, A. Li, X. Sun, H. Ouyang, C. Campos, N. B. F. Rocha, O. Arias-Carrión, S. Machado, G. Hou, K. F. So, *Mol. Neurobiol.* **2016**, *53*, 6367–6376.
- [130] F. M. Vassoler, G. Sadri-Vakili, *Neuroscience* **2014**, *264*, 198–206.
- [131] R. J. Schmitz, Z. A. Lewis, M. G. Goll, *Trends Genet.* **2019**, *35*, 818–827.
- [132] M. Ehrlich, M. A. Gama-Sosa, L.-H. Huang, R. M. Midgett, K. C. Kuo, R. A. McCune, C. Gehrke, *Nucleic Acids Res.* **1982**, *10*, 2709–2721.
- [133] M. G. Goll, T. H. Bestor, *Annu. Rev. Biochem.* **2005**, *74*, 481–514.
- [134] J. U. Guo, Y. Su, J. H. Shin, J. Shin, H. Li, B. Xie, C. Zhong, S. Hu, T. Le, G. Fan, u. a., *Nat. Neurosci.* **2014**, *17*, 215–222.
- [135] 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, u. a., *Nature* **2009**, *462*, 315–322.
- [136] A. Fusco, *Epigenomes* **2018**, *2*, 22.
- [137] V. Patil, R. L. Ward, L. B. Hesson, *Epigenetics* **2014**, *9*, 823–828.
- [138] M. G. P. Van Der Wijst, A. Y. Van Tilburg, M. H. J. Ruiters, M. G. Rots, *Sci. Rep.* **2017**, *7*, 1–15.
- [139] V. Patil, C. Cuenin, F. Chung, J. R. R. Aguilera, N. Fernandez-Jimenez, I. Romero-Garmendia, J. R. Bilbao, V. Cahais, J. Rothwell, Z. Herceg, *Nucleic Acids Res.* **2019**, *47*, 10072–10085.
- [140] Z. D. Smith, A. Meissner, *Nat. Rev. Genet.* **2013**, *14*, 204–220.
- [141] A. M. Deaton, A. Bird, *Genes Dev.* **2011**, *25*, 1010–1022.
- [142] S. Saxonov, P. Berg, D. L. Brutlag, *Proc. Natl. Acad. Sci.* **2006**, *103*, 1412 LP-1417.
- [143] L. D. Moore, T. Le, G. Fan, *Neuropsychopharmacology* **2013**, *38*, 23–38.
- [144] L. Li, Y. Gao, Q. Wu, A. S. L. Cheng, K. Y. Yip, *Genome Res.* **2019**, *29*, 543–553.

- [145] É. Héberlé, A. F. Bardet, *Essays Biochem.* **2019**, *63*, 727–741.
- [146] M. B. Stadler, R. Murr, L. Burger, R. Ivanek, F. Lienert, A. Schöler, E. van Nimwegen, C. Wirbelauer, E. J. Oakeley, D. Gaidatzis, u. a., *Nature* **2011**, *480*, 490–495.
- [147] M. J. Ziller, H. Gu, F. Müller, J. Donaghey, L. T.-Y. Tsai, O. Kohlbacher, P. L. De Jager, E. D. Rosen, D. A. Bennett, B. E. Bernstein, *Nature* **2013**, *500*, 477.
- [148] E. Hodges, A. Molaro, C. O. Dos Santos, P. Thekkat, Q. Song, P. J. Uren, J. Park, J. Butler, S. Rafii, W. R. McCombie, *Mol. Cell* **2011**, *44*, 17–28.
- [149] S. M. Iguchi-Arigo, W. Schaffner, *Genes Dev.* **1989**, *3*, 612–619.
- [150] F. Watt, P. L. Molloy, *Genes Dev.* **1988**, *2*, 1136–1143.
- [151] J. Ben-Hattar, J. Jiricny, *Gene* **1988**, *65*, 219–227.
- [152] D. Schübeler, *Nature* **2015**, *517*, 321–326.
- [153] A. P. Bird, A. P. Wolffe, *Cell* **1999**, *99*, 451–454.
- [154] J. S. Hardwick, A. N. Lane, T. Brown, *BioEssays* **2018**, *40*, 1–13.
- [155] H. Stroud, S. Feng, S. Morey Kinney, S. Pradhan, S. E. Jacobsen, *Genome Biol.* **2011**, *12*, DOI 10.1186/gb-2011-12-6-r54.
- [156] T. Carell, M. Q. Kurz, M. Müller, M. Rossa, F. Spada, *Angew. Chem. Int. Ed.* **2018**, *57*, 4296–4312.
- [157] M. Bachman, S. Uribe-Lewis, X. Yang, H. E. Burgess, M. Iurlaro, W. Reik, A. Murrell, S. Balasubramanian, *Nat. Chem. Biol.* **2015**, *11*, 555.
- [158] M. Su, A. Kirchner, S. Stazzoni, M. Müller, M. Wagner, A. Schröder, T. Carell, *Angew. Chem. Int. Ed.* **2016**, *55*, 11797–11800.
- [159] H. Kamiya, H. Tsuchiya, N. Karino, Y. Ueno, A. Matsuda, H. Harashima, *J. Biochem.* **2002**, *132*, 551–555.
- [160] T. Shibutani, S. Ito, M. Toda, R. Kanao, L. B. Collins, M. Shibata, M. Urabe, H. Koseki, Y. Masuda, J. A. Swenberg, u. a., *Sci. Rep.* **2015**, *4*, 5220.
- [161] L. Shen, H. Wu, D. Diep, S. Yamaguchi, A. C. D'Alessio, H. L. Fung, K. Zhang, Y. Zhang, *Cell* **2013**, *153*, 692–706.
- [162] F. Neri, D. Incarnato, A. Krepelova, S. Rapelli, F. Anselmi, C. Parlato, C. Medana, F. DalBello, S. Oliviero, *Cell Rep.* **2015**, *10*, 674–683.
- [163] X. Lu, D. Han, B. S. Zhao, C. X. Song, L. S. Zhang, L. C. Doré, C. He, *Cell Res.* **2015**, *25*, 386–389.
- [164] M. Schneider, C. Trummer, A. Stengl, P. Zhang, A. Szwagierczak, M. Cristina Cardoso, H. Leonhardt, C. Bauer, I. Antes, *PLoS One* **2020**, *15*, 1–23.
- [165] A. Sadakierska-Chudy, R. M. Kostrzewa, M. Filip, *Neurotox. Res.* **2015**, *27*, 84–97.
- [166] A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, *286*, 35334–35338.
- [167] F. Guo, X. Li, D. Liang, T. Li, P. Zhu, H. Guo, X. Wu, L. Wen, T. P. Gu, B. Hu, u. a., *Cell Stem Cell* **2014**, *15*, 447–459.

- [168] R. Rahimoff, O. Kosmatchev, A. Kirchner, T. Pfaffeneder, F. Spada, V. Brantl, M. Müller, T. Carell, *J. Am. Chem. Soc.* **2017**, *139*, 10359–10364.
- [169] A. Schön, E. Kaminska, F. Schelter, E. Ponkkonen, E. Korytiaková, S. Schiffers, T. Carell, *Angew. Chem. Int. Ed.* **2020**, *59*, 5591–5594.
- [170] K. Iwan, R. Rahimoff, A. Kirchner, F. Spada, A. S. Schröder, O. Kosmatchev, S. Ferizaj, J. Steinbacher, E. Parsa, M. Müller, u. a., *Nat. Chem. Biol.* **2017**, *14*, 72.
- [171] C. Xu, V. G. Corces, *Science* **2018**, *359*, 1166–1170.
- [172] C. Shao, M. Lacey, L. Dubeau, M. Ehrlich, *Epigenetics* **2009**, *4*, 165–175.
- [173] R. Nishiyama, L. Qi, M. Lacey, M. Ehrlich, *Mol. Cancer Res.* **2005**, *3*, 617–626.
- [174] S. Sun, Y. R. Lee, B. Enfield, *Cancer Inform.* **2019**, *18*, DOI 10.1177/1176935119872959.
- [175] J. Sharif, T. A. Endo, M. Nakayama, M. M. Karimi, M. Shimada, K. Katsuyama, P. Goyal, J. Brind'Amour, M. A. Sun, Z. Sun, u. a., *Cell Stem Cell* **2016**, *19*, 81–94.
- [176] J. Sharif, H. Koseki, *Science* **2018**, *359*, 1102–1103.
- [177] C.-X. Song, J. Diao, A. T. Brunger, S. R. Quake, *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 4338–4343.
- [178] Q. Zhu, R. Stöger, R. Alberio, *Front. Cell Dev. Biol.* **2018**, *6*, DOI 10.3389/fcell.2018.00024.
- [179] C. D. Laird, N. D. Pleasant, A. D. Clark, J. L. Sneed, K. M. A. A. Hassan, N. C. Manley, J. C. Vary, T. Morgan, R. S. Hansen, R. Stöger, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 204–209.
- [180] P. Giehr, C. Kyriakopoulos, K. Lepikhov, S. Wallner, V. Wolf, *Nucleic Acids Res.* **2018**, 1–10.
- [181] E. L. van Dijk, Y. Jaszczyszyn, D. Naquin, C. Thermes, *Trends Genet.* **2018**, *34*, 666–681.
- [182] C. Xu, V. G. Corces, *Nat. Protoc.* **2019**, *14*, 202–216.
- [183] M. Frommer, L. E. McDonald, D. S. Millar, C. M. Collis, F. Watt, G. W. Grigg, P. L. Molloy, C. L. Paul, *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 1827–1831.
- [184] S. Tierling, B. Schmitt, J. Walter, *Genet. Epigenet.* **2018**, *10*, 1–7.
- [185] K. Tanaka, A. Okamoto, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1912–1915.
- [186] D. Shortle, D. Nathans, *Proc. Natl. Acad. Sci.* **1978**, *75*, 2170 LP-2174.
- [187] H. Hayatsu, in (Hrsg.: W.E.B.T.-P. in N.A.R. and M.B. Cohn), Academic Press, **1976**, S. 75–124.
- [188] R. Feil, J. Charlton, A. P. Bird, J. Walter, W. Reik, *Nucleic Acids Res.* **1994**, *22*, 695–696.
- [189] M. Wan, D. A. Bell, *Analysis of genome-wide methylation using reduced representation bisulfite sequencing (RRBS) technology*, Elsevier Inc., **2020**.
- [190] P. P. Giehr, The role of Dnmts and Tets in shaping the DNA methylation landscape of mouse embryonic stem cells, Universität des Saarlandes, **2019**.
- [191] C. X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S. Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, u. a., *Cell* **2013**, *153*, 678–691.
- [192] M. J. Booth, G. Marsico, M. Bachman, D. Beraldi, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 435–440.

- [193] C. Zhu, Y. Gao, H. Guo, B. Xia, J. Song, X. Wu, H. Zeng, K. Kee, F. Tang, C. Yi, *Cell Stem Cell* **2017**, *20*, 720–731.e5.
- [194] M. Rossa, *persönliche Kommunikation mit Dr. Pascal Giehr*, **2020**.
- [195] M. Rossa, *persönliche Kommunikation mit Dr. Sascha Tierling*, **2020**.
- [196] B. E. Miner, R. J. Stöger, A. F. Burden, C. D. Laird, R. S. Hansen, *Nucleic Acids Res.* **2004**, *32*, 30–33.
- [197] P. Giehr, J. Walter, in *DNA Methylation Protoc.*, **2017**, S. 573–586.
- [198] R. E. Notari, *J. Pharm. Sci.* **1967**, *56*, 804–809.
- [199] R. Shapiro, R. S. Klein, *Biochemistry* **1967**, *6*, 3576–3582.
- [200] R. Shapiro, R. S. Klein, *Biochemistry* **1966**, *5*, 2358–2362.
- [201] R. Shapiro, R. E. Servis, M. Welcher, *J. Am. Chem. Soc.* **1970**, *92*, 422–424.
- [202] F. Raulin, *Space Sci. Rev.* **2005**, *116*, 471–487.
- [203] G. J. Molina-cuberos, K. Schwingenschuh, J. Lo, R. Rodrigo, L. M. Lara, V. Anicich, *J. Geophys. Res.* **2002**, *107*, 5099.
- [204] W. Reid Thompson, T. J. Henry, J. M. Schwartz, B. N. Khare, C. Sagan, *Icarus* **1991**, *90*, 57–73.
- [205] M. Lattelais, Y. Ellinger, A. Matrane, J.-C. Guillemin, *Phys. Chem. Chem. Phys.* **2010**, *12*, 4165–4171.
- [206] K. Hasan, N. Brown, C. M. Kozak, *Green Chem.* **2011**, *13*, 1230–1237.
- [207] G. B. Payne, P. H. Deming, P. H. Williams, *J. Org. Chem.* **1961**, *26*, 659–663.
- [208] T. Encrenaz, T. Greathouse, F. Lefèvre, S. Atreya, *Planet. Space Sci.* **2012**, *68*, 3–17.
- [209] R. W. Carlson, *Science* **1999**, *283*, 2062–2064.
- [210] M. J. Borda, A. R. Elsetinow, M. A. Schoonen, D. R. Strongin, *Astrobiology* **2001**, *1*, 283–288.
- [211] J. Haqq-Misra, J. F. Kasting, S. Lee, *Astrobiology* **2011**, *11*, 293–302.
- [212] A. F. Holleman, E. Wiberg, N. Wiberg, *Lehrbuch der Anorganischen Chemie*, De Gruyter, Berlin, **2007**.
- [213] Z. Yang, D. Liu, Y. Liu, M. Sugiya, T. Imamoto, W. Zhang, *Organometallics* **2015**, *34*, 1228–1237.
- [214] M. Y. Hyun, S. H. Kim, Y. J. Song, H. G. Lee, Y. D. Jo, J. H. Kim, I. H. Hwang, J. Y. Noh, J. Kang, C. Kim, *J. Org. Chem.* **2012**, *77*, 7307–7312.
- [215] K. Kamata, K. Sugahara, K. Yonehara, R. Ishimoto, N. Mizuno, *Chem. Eur. J.* **2011**, *17*, 7549–7559.
- [216] J. Tietze, *persönliche Kommunikation mit Dr. Werner Spahl*, **2018**.
- [217] S. Islam, D.-K. Bučar, M. W. Powner, *Nat. Chem.* **2017**, *9*, 584–589.
- [218] K. Kamata, K. Sugahara, K. Yonehara, R. Ishimoto, N. Mizuno, *Chem. – A Eur. J.* **2011**, *17*, 7549–7559.

- [219] A. Köhn, H. Ahrens, R. Braun, I. Heinemann, J. Tiebes, C. Waldruff, H. Dietrich, E. Gatzweiler, C. H. Rosinger, D. Schmutzler, *Preparation of N-(isoxazol-3-yl)-aryl-carboxylic acid amides and use thereof as herbicides*, **2014**.
- [220] S. Becker, J. Feldmann, S. Wiedemann, H. Okamura, C. Schneider, K. Iwan, A. Crisp, M. Rossa, T. Amatov, T. Carell, *Science* **2019**, *366*, 76–82.
- [221] A. J. Boulton, A. R. Katritzky, *Proc. Chem. Soc.* **1962**, 257.
- [222] J.-J. Li, E. J. Corey, *Name Reactions in Heterocyclic Chemistry II*, John Wiley & Sons, Ltd, **2011**.
- [223] S. Wiedemann, T. Carell, unveröffentlichte Arbeiten, Ludwig-Maximilians-Universität München, **2018**.
- [224] A. J. Tölke, Untersuchungen zur Bildung von s2U und s2C unter präbiotisch plausiblen Bedingungen, Ludwig-Maximilians-Universität München, **2020**.
- [225] V. V. Illarionov, T. I. Sokolova, S. I. Vol'fkovich, *Izv. Akad. Nauk SSSR* **1945**, 94–130.
- [226] P. Dutoit, *Process for the preparation of pentasulphide of phosphorus*, **1932**, US1852959A.
- [227] G. Brauer, Hrsg., *Handbook of preparative inorganic chemistry*, Academic Press, New York, **2012**.
- [228] S. N. Britvin, M. N. Murashko, E. Vapnik, Y. S. Polekhovskiy, S. V. Krivovichev, *Geol. Ore Depos.* **2017**, *59*, 619–625.
- [229] G. Wächtershäuser, *Prog. Biophys. Mol. Biol.* **1992**, *58*, 85–201.
- [230] B. Herschy, S. J. Chang, R. Blake, A. Lepland, H. Abbott-Lyon, J. Sampson, Z. Atlas, T. P. Kee, M. A. Pasek, *Nat. Commun.* **2018**, *9*, 1–7.
- [231] C. Schneider, S. Becker, H. Okamura, A. Crisp, T. Amatov, M. Stadlmeier, T. Carell, *Angew. Chem. Int. Ed.* **2018**, *57*, 5943–5946.
- [232] Y. L. Yung, M. B. McElroy, *Science* **1979**, *203*, 1002–1004.
- [233] K. Golankiewicz, M. Szajda, E. Wyrzykiewicz, *Pol. J. Chem.* **1978**, *52*, 1085–1087.
- [234] E. Golankiewicz, Krzysztof; Szajda, Maria; Wyrzykiewicz, *Pol. J. Chem.* **1980**, *54*, 363–365.
- [235] C. Landreau, D. Deniaud, A. Reliquet, F. Reliquet, J. C. Meslin, *J. Heterocycl. Chem.* **2001**, *38*, 93–98.
- [236] A. Robin, K. Julienne, S. Raimbault, J.-C. Meslin, D. Deniaud, *Synlett* **2005**, *2005*, 2805–2807.
- [237] P. F. Crain, J. D. Alfonzo, J. Rozenski, S. T. Kapushoc, J. A. McCloskey, L. Simpson, *RNA* **2002**, *8*, 752–761.
- [238] V. F. Reichle, D. P. Petrov, V. Weber, K. Jung, S. Kellner, *Nat. Commun.* **2019**, *10*.
- [239] F. B. Dains, R. Q. Brewster, C. P. Olander, C. S. Marvel, L. T. Sandborn, *Org. Synth.* **1926**, *6*, 72–73.
- [240] B. Bestgen, I. Kufareva, W. Seetoh, C. Abell, R. W. Hartmann, R. Abagyan, M. Le Borgne, O. Filhol, C. Cochet, T. Lomberget, u. a., *J. Med. Chem.* **2019**, *62*, 1817–1836.
- [241] I. R. Taylor, B. M. Duniyakh, T. Komiyama, H. Shao, X. Ran, V. A. Assimon, C. Kalyanaraman, J. N. Rauch, M. P. Jacobson, E. R. P. Zuiderweg, u. a., *J. Biol. Chem.* **2018**, *293*, 4014–4025.

- [242] X. Yang, S. Cai, X. Liu, P. Chen, J. Zhou, H. Zhang, *Bioorganic Med. Chem.* **2019**, *27*, 1605–1618.
- [243] J. F. Wolfe, B. H. Loo, F. E. Arnold, *Macromolecules* **1981**, *14*, 915–920.
- [244] H. Zhou, S. Wu, S. Zhai, A. Liu, Y. Sun, R. Li, Y. Zhang, S. Ekins, P. W. Swaan, B. Fang, u. a., *J. Med. Chem.* **2008**, *51*, 1242–1251.
- [245] H. Matsuki, M. Takahashi, M. Higuchi, G. N. Makokha, M. Oie, M. Fujii, *Genes to cells* **2013**, *18*, 135–146.
- [246] M. Q. Kurz, Synthese und Eigenschaften modifizierter Nukleoside in DNA und RNA (fdC, hmdC, Gm, m6A, m62A und ms2A), Ludwig-Maximilians-Universität München, **2020**.
- [247] A. S. Schröder, J. Steinbacher, B. Steigenberger, F. A. Gnerlich, S. Schiesser, T. Pfaffeneder, T. Carell, A. S. Schröder, J. Steinbacher, B. Steigenberger, u. a., *Angew. Chem. Int. Ed.* **2014**, *53*, 315–318.
- [248] H. Han, C. C. Cortez, X. Yang, P. W. Nichols, P. A. Jones, G. Liang, *Hum. Mol. Genet.* **2011**, *20*, 4299–4310.
- [249] M. J. Booth, E. A. Raiber, S. Balasubramanian, *Chem. Rev.* **2015**, *115*, 2240–2254.
- [250] P. Giehr, C. Kyriakopoulos, G. Ficzb, V. Wolf, J. Walter, *PLoS Comp Biol* **2016**, *12*, 1–16.
- [251] J. Qu, M. Zhou, Q. Song, E. E. Hong, A. D. Smith, *Bioinformatics* **2013**, *29*, 2645–2646.
- [252] T. Äijö, X. Yue, A. Rao, H. Lähdesmäki, *Bioinformatics* **2016**, *32*, i511–i519.
- [253] H. Wu, X. Wu, L. Shen, Y. Zhang, *Nat. Biotechnol.* **2014**, *32*, 1231–1240.
- [254] H. E. Gottlieb, V. Kotlyar, A. Nudelman, *J. Org. Chem.* **1997**, *62*, 7512–7515.