

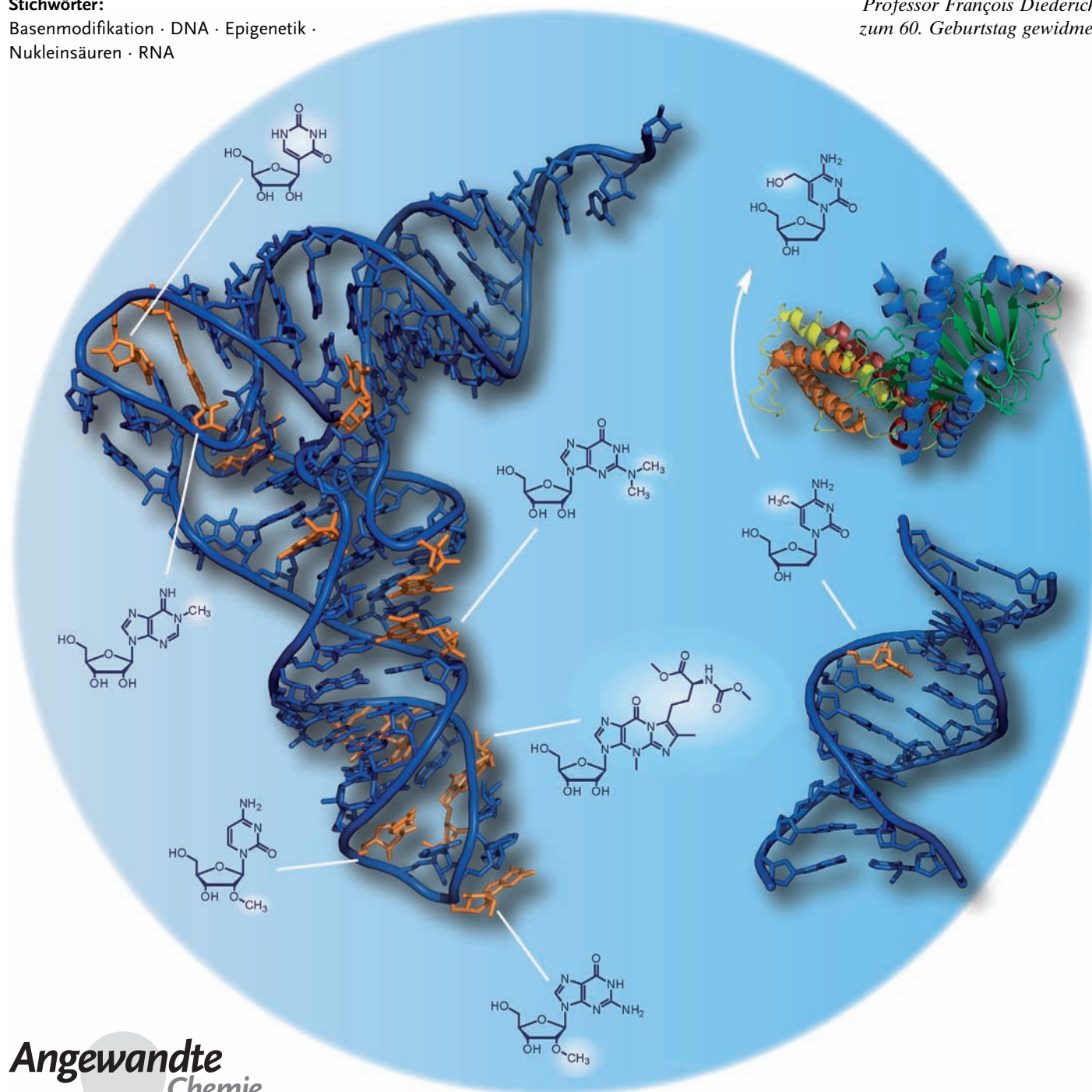
Struktur und Funktion nicht-kanonischer Nukleobasen

Thomas Carell,* Caterina Brandmayr, Antje Hienzsch, Markus Müller,
David Pearson, Veronika Reiter, Ines Thoma, Peter Thumbs und Mirko Wagner

Stichwörter:

Basenmodifikation · DNA · Epigenetik ·
Nukleinsäuren · RNA

Professor François Diederich
zum 60. Geburtstag gewidmet



DNA und RNA enthalten neben den vier kanonischen Nukleobasen eine Vielzahl modifizierter Nukleoside, die ihren chemischen Informationsgehalt vergrößern. RNA ist dabei besonders reich an Modifikationen, was offensichtlich das Resultat der Anpassung an ihre komplexen und vielfältigen Aufgaben ist. Die modifizierten Nukleoside und deren chemische Struktur etablieren hierbei eine zweite Informationsebene, welche für die Funktion der RNA-Verbindungen von zentraler Bedeutung ist. Auch die chemische Diversität der DNA ist größer als zunächst erwartet. Neben den vier kanonischen Basen enthält die DNA höherer Organismen insgesamt vier epigenetische Basen: m^5dC , hm^5dC , f^5dC und ca^5dC . Alle Zellen eines Organismus enthalten das gleiche genetische Material. Daher erfordert die Erfüllung der zahlreichen unterschiedlichen Aufgaben und Funktionen innerhalb dieses Organismus die kontrollierte Aktivierung und Inaktivierung Zelltyp-spezifischer Gene. Die Regulation der zugrundeliegenden Unterdrückungs- und Aktivierungsprozesse erfordert eine zweite, epigenetische Informationsebene, die augenscheinlich direkt mit erhöhter chemischer Diversität verbunden ist. Diese Diversität wird von den modifizierten, nicht-kanonischen Nukleosiden sowohl in DNA als auch in RNA bereitgestellt.

1. Einleitung

Desoxyribonukleinsäure (DNA) und Ribonukleinsäure (RNA) sind die Grundbausteine allen Lebens. Diese Moleküle kodieren die genetische Information und sind die Blaupause für jeden lebenden Organismus. Alle Organismen benutzen im Wesentlichen die gleiche genetische Sprache, die von den vier kanonischen Nukleobasen Desoxyadenosin (dA), Desoxycytosin (dC), Desoxyguanosin (dG) und Desoxythymidin (dT) gebildet wird. In der RNA finden sich die entsprechenden kanonischen Ribonukleoside A, C, G und U (Abbildung 1). Diese acht organischen Moleküle sind die zentralen Bausteine des genetischen Systems. Das DNA-

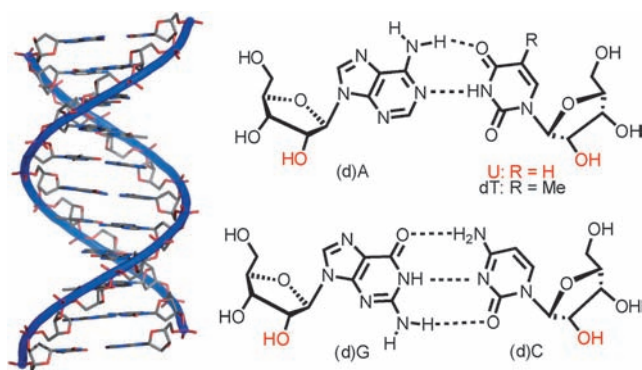


Abbildung 1. Darstellung der Watson-Crick-B-DNA-Doppelhelix und der Watson-Crick-Basenpaare. Diese Sequenz (dA, dC, dG, und dT) bildet den Sequenzcode, der die erste Ebene der genetischen Information darstellt. Schwarz: allgemeine Komponenten; rot: Bestandteile spezifisch für RNA.

Aus dem Inhalt

1. Einleitung	7221
2. RNA-Modifikationen	7223
3. Modifizierte Basen in der DNA	7235
4. Zusammenfassung	7238

Molekül selber wurde entdeckt von Miescher im Jahre 1869,^[1] Chargaff formulierte später die Chargaff-Regeln,^[2] die unter anderem die relative Häufigkeit der vier kanonischen DNA-Basen beschreiben. Diese Basen sind mit dem Zucker Desoxyribose verknüpft, was die vier Nukleoside ergibt, welche über Phosphordiester-

brücken verknüpft sind.^[3] Zentrale Arbeiten 1944 von Avery und 1952 von Hershey zeigten, dass DNA der Träger der vererbten genetischen Information ist.^[4] Kurz darauf wurde mithilfe bedeutender wissenschaftlicher Beiträge von Franklin, Goslin und Wilkins^[5] die doppelhelikale Struktur, der DNA-Doppelstrang (Abbildung 1), durch Watson und Crick gelöst^[6] – ein wissenschaftlicher Meilenstein des letzten Jahrhunderts. Die zentralen Bausteine von DNA und RNA sind die Nukleobasen selbst, die mit der (Desoxy)ribose über den N1-Stickstoff im Fall der Pyrimidine (T, C und U) und über den N9-Stickstoff im Fall der Purinbasen (A und G) verknüpft sind (Abbildung 1). Wir begreifen heute, dass die Basis des genetischen Systems spezifische Wasserstoffbrücken sind, die sich zwischen den Pyrimidinbasen (d)A/dT, U und den Purinbasen (d)G/(d)C bilden können. Diese spezifischen Basenpaarungseigenschaften etablieren selektive, nichtkovalente Wechselwirkungen, welche die Basis für die Replikationsfähigkeit von DNA und die Fähigkeit der DNA, in RNA transkribiert zu werden, sind. Während DNA die vererbte genetische Information in den Zellen (in den

[*] Prof. Dr. T. Carell, MChem. C. Brandmayr, Dipl.-Chem. A. Hienzsch, Dr. M. Müller, Dr. D. Pearson, Dipl.-Chem. V. Reiter, M. Sc. I. Thoma, M. Sc. P. Thumbs, Dipl.-Chem. M. Wagner
Zentrum für Integrative Proteinforschung, Department für Chemie, Ludwig-Maximilians-Universität München
Butenandtstraße 5–13, 81377 München (Deutschland)
Prof. Dr. T. Carell, MChem. C. Brandmayr, Dipl.-Chem. A. Hienzsch, Dr. M. Müller, Dr. D. Pearson, Dipl.-Chem. V. Reiter, M. Sc. I. Thoma, M. Sc. P. Thumbs, Dipl.-Chem. M. Wagner
Department für Chemie, Ludwig-Maximilians Universität München
Butenandtstraße 5–13, 81377 München (Deutschland)
E-Mail: thomas.carell@lmu.de

Zellkernen) speichert, erfüllt der transkribierte RNA-Strang, als Kopie eines DNA-Segementes, eine Vielzahl von Funktionen. RNA-Stränge transportieren nicht nur die genetische Erbinformation in das Zytosol, sie können darüber hinaus auch die Genaktivität regulieren.^[7] RNA-Moleküle können darüber hinaus auch katalytische Eigenschaften haben, so z.B. im Ribosom, wo sie die Bildung der Peptidbindungen katalysieren.^[8] Neben der Kodierung im Fall der Boten-RNA (mRNA), der bekannten ribosomalen RNA (rRNA) und der Transfer-RNA (tRNA) gibt es eine ganze Reihe neuartiger, nichtkodierender RNA-Moleküle. Diese Spezies haben in den letzten zwei Dekaden unser Bild über die RNA in den Zellen drastisch verändert und machen noch heute RNA-Forschung zu einem der dynamischsten und am schnellsten wachsenden Feldern innerhalb der Lebenswissenschaften. Die neuen RNA-Spezies sind z. B. nukleare und nukleolare RNA (snRNA und snoRNA), mikroRNA (miRNA) und lange nichtkodierende RNA (lncRNA),^[9] welche wichtige zelluläre Funktionen einnehmen, die weit über einen Informationstransport hinausgehen.^[10] Ganz besonders interessant ist in diesem Zusammenhang die Entdeckung, dass lncRNAs und miRNAs an der Orchestrierung der Details zellulärer Differenzierungs- und Entwicklungsprogramme beteiligt sind (Tabelle 1).^[11] Wir wissen heute, dass nur ein kleiner Teil des humanen Genoms für Proteine kodiert, und dass ein wesentlich größerer Teil in eine Vielzahl dieser nicht-kodierenden RNAs überschrieben wird.^[12]

Wir beginnen gerade erst zu verstehen, wie die unterschiedlichen RNA-Moleküle innerhalb der Zelle ein komplexes, intensiv vernetztes Geflecht bilden und als solches gezielte regulatorische Funktionen übernimmt. Für all diejenigen, die sich für diese Gebiete interessieren, ist es wichtig zu wissen, dass viele RNA-Moleküle neben den vier kanonischen Basen eine große Anzahl von modifizierten Strukturen aufweisen. Es wird immer deutlicher, dass für die komplexe Funktion, die die RNA jenseits von Transport und Translation der DNA-Sequenzinformation übernimmt, die chemische Diversität, die von den vier kanonischen Nukleobasen aufgebaut werden kann, nicht ausreichend ist. Neben der Information, die in der Basenabfolge der RNA gespeichert ist, findet man in den verschiedenen RNA-Spezies an einer

Tabelle 1: Abkürzungen und Beschreibung der verschiedenen hier vorgestellten RNA-Spezies.

Abkürzung	Bezeichnung	Funktion
tRNA	Transfer-RNA	mRNA dekodierender Aminosäure-Überträger
mRNA	Boten-RNA	kodierend für Proteinsequenz
rRNA	ribosomale RNA	Nukleinsäure-Anteil des Ribosoms
sncRNA	small non-coding RNA	
miRNA	micro RNA	Regulation der Genexpression
piRNA	pivi-interacting RNA	Regulation der Genexpression
siRNA	short interfering RNA	Regulation der Genexpression
snRNA	small nuclear RNA	Beteiligung z.B. am Intron-Splicing
snoRNA	small nucleolar RNA	Beteiligung an der RNA-Modifikation
lncRNA	long non-coding RNA	nicht-Protein kodierendes Transkript mit regulatorischer oder unbekannter Funktion

Vielzahl von Stellen strukturell unterschiedliche posttranslationale Modifikationen. Die Position des modifizierten Nukleosids und dessen chemische Struktur etablieren hierbei eine zweite Informationsebene, die für die Funktion der RNA-Verbindungen ebenfalls von zentraler Bedeutung ist. Viele der modifizierten Nukleoside wurden vor langer Zeit in tRNA, rRNA und mRNA gefunden, doch beginnen Wissenschaftler erst jetzt gezielt nach solchen Modifikationen in vielen nichtkodierenden RNA-Spezies zu suchen. Die chemische Diversität der DNA hingegen ist deutlich weniger komplex und im Wesentlichen limitiert auf methylierte und hydroxymethylierte Basen wie N4- und C5-methylierte Cytosine, N6-methyliertes Adenin sowie C5-hydroxymethyliertes Cytosin und Uracil (Nukleobase J).^[13] Vor kurzem wurden allerdings neue DNA-Basen entdeckt, die uns zeigen, dass auch im Fall der DNA die Komplexität, die durch die vier kanonischen Basen erzeugt werden kann, nicht ausreichend ist. Über die modifizierten DNA-Basen wissen wir heute, dass sie epigenetische Informationen kodieren, die für die Orchestrierung zellulärer Entwicklungsprogramme von entscheidender Bedeutung sind. Die modifizierten Nukleoside in der RNA haben demgegenüber viele unterschiedliche Funktionen, die wir im Folgenden näher betrachten wollen.

Im ersten Teil des Aufsatzes geben wir eine Einführung in das Gebiet der RNA-Modifikationen und beschreiben eine Reihe neuartiger Entwicklungen von, aus Sicht der Autoren, besonderem Interesse. Aufgrund des großen Umfangs des Gebietes ist es nicht unsere Absicht, einen allumfassenden Aufsatz zu schreiben. Teilaspekte des Gebiets der RNA-Modifizierung sind in exzellenten Übersichtsartikeln bereits zusammengefasst worden. Das betrifft die Wobble-Basenpaarbildung,^[14] die Modifikationen neben dem Antikodon,^[15] die Frage, wie die Modifikationen die RNA-Stabilität beeinflussen^[16] sowie verschiedene andere Aspekte.^[17] Im zweiten, kürzeren Teil des Aufsatzes diskutieren wir modifizierte DNA-Basen und konzentrieren uns hier auf den Entdeckungsprozess. Wir schildern, wie die Entdeckung dieser Basen beginnt, unser Verständnis der epigenetischen Programmierung der Erbinformation zu revolutionieren.



Thomas Carell studierte Chemie an den Universitäten Münster und Heidelberg und promovierte 1993 am Max-Planck-Institut für Medizinische Forschung. Nach einem Postdoktorat bei Julius Rebek am MIT in Cambridge, USA, habilitierte er 1999 an der ETH Zürich. 2000 folgte er einem Ruf auf den Lehrstuhl für Organische Chemie an der Universität Marburg. Seit 2004 ist er Lehrstuhlinhaber für Organische Chemie an der Ludwig-Maximilians-Universität München. Er ist Mitglied der Leopoldina, Empfänger des Leibniz-Preises und Träger des Bundesverdienstkreuzes.

2. RNA-Modifikationen

Bis heute sind über 100 modifizierte RNA-Nukleoside entdeckt worden. Die chemischen Strukturveränderungen sind hierbei nicht auf spezielle Watson-Crick-Basen limitiert. Strukturvariationen werden an einer Vielzahl von Positionen in allen vier kanonischen Basen gefunden, wie es in Abbildung 2 dargestellt ist. Neben sehr einfachen Modifikationen wie der Methylierung und Acetylierung (N^1 -Methylguanosin, m^1G ; N^2,N^2 -Dimethylguanosin, m^2G ; N^4 -Acetylcytosin, ac^4C) kennen wir sogenannte hypermodifizierte Basen, in denen z.B. eine andere heterocyclische Ringstruktur, normalerweise das 7-Desazapurin (Abbildung 2), vorkommt. Auch wenn angenommen wird, dass alle RNA-Spezies in der Zelle modifizierte Nukleoside enthalten, so ist die größte Modifizierungsdichte assoziiert mit der tRNA. Ein Drittel der modifizierten Nukleoside sind in der ribosomalen RNA anwesend, und ungefähr 10% der bekannten modifizierten Nukleoside werden in mRNA gefunden. Bis heute wurden 14

modifizierte Nukleoside in anderen RNA-Spezies entdeckt, und hier vor allem in den kleinen, nichtkodierenden RNAs.^[18] Während einige Modifikationen wie das Pseudouridin Ψ in fast allen RNA-Typen vorkommen,^[19] sind andere RNA-typspezifisch. Fokussiert man auf tRNA, zeigt die Sequenzanalyse, dass hier insgesamt 10–15% der kanonischen Nukleotide chemisch modifiziert sind. Im Schnitt enthält damit jedes tRNA-Molekül acht modifizierte Nukleoside, und von diesen befinden sich viele in der Antikodonschleife.^[20] Die anderen modifizierten Nukleoside sind an verschiedenen Stellen über die gesamte tRNA verteilt. In den letzten Jahren konnten die chemischen Strukturen dieser Modifikationen, ihre Position sowie viele Enzyme, die an der Biosynthese der Modifikationen beteiligt sind, aufgeklärt werden. Für einige Modifikationen ist die biochemische und biophysikalische Funktion in der tRNA in etwas größerer Detailgenauigkeit bekannt.

Um die Vielzahl der Nukleosidmodifikationen beschreiben zu können, wurde eine spezielle Nomenklatur entwickelt.

Wie in Abbildung 3 dargestellt, werden die chemischen Modifikationen mit einem Ein- oder Zweibuchstabencode abgekürzt. Die Position der chemischen Modifikation an den Pyrimidin- oder Purinringen wird als hochgestellte Nummer nach dem Ein- oder Zweibuchstabencode angegeben. Die Anzahl der modifizierenden Gruppen, die sich an den Basen an dieser speziellen Position befinden, wird durch eine entsprechende tiefgestellte Nummer beschrieben. Die Nummerierung erfolgt hierbei in Übereinstimmung mit der Nummerierung für Pyrimidine und Purine. Speziell hypermodifizierte Nukleoside wie das Queuosin (Q) oder das Wybutosin (yW) werden durch eigene Abkürzungen, ähnlich wie bei den kanonischen Nukleobasen, beschrieben.^[18,21] Modifikationen die sich an der Zuckerstruktur befinden, werden nach dem Basenkürzel angegeben.

Eine besondere Modifikation ist das Pseudouridin (Ψ). Deren Base, die in Abbildung 2 dargestellt ist, ist mit dem Zuckerrückgrat nicht über das Pyrimidin N1, sondern über C5 verknüpft, was eine Isomerisierung darstellt.^[19] Hypermodifizierte RNA-Basen sind in ihrer Struktur drastisch verändert. Hierzu gehört z.B. das N^6 -Threonylcarbamoyladenysin (t^6A) und das Lysidin (k^2C), in denen sich Aminosäureseitenketten am Purin- bzw. Pyrimidingerüst befinden (Nomenklatur siehe Abbildung 3). Weitere Beispiele für noch komplexere hypermodifizierte Basen sind das Queuosin (Q) und Wybutosin (yW), in denen mithilfe einer komplexen biosynthetischen Maschine das gesamte heterocyclische Gerüst umgebaut wird. Abbildung 4 zeigt die Vielfalt und die Verteilung der bekannten modifizierten Nukleoside in einem schematischen 2D-tRNA-Diagramm. Die komplexen Strukturen und die Diversität der existierenden RNA-Modifikationen führten zur Entwicklung spezieller Datenbanken, die konstant erweitert werden. Zwei der wichtigsten Datenbanken

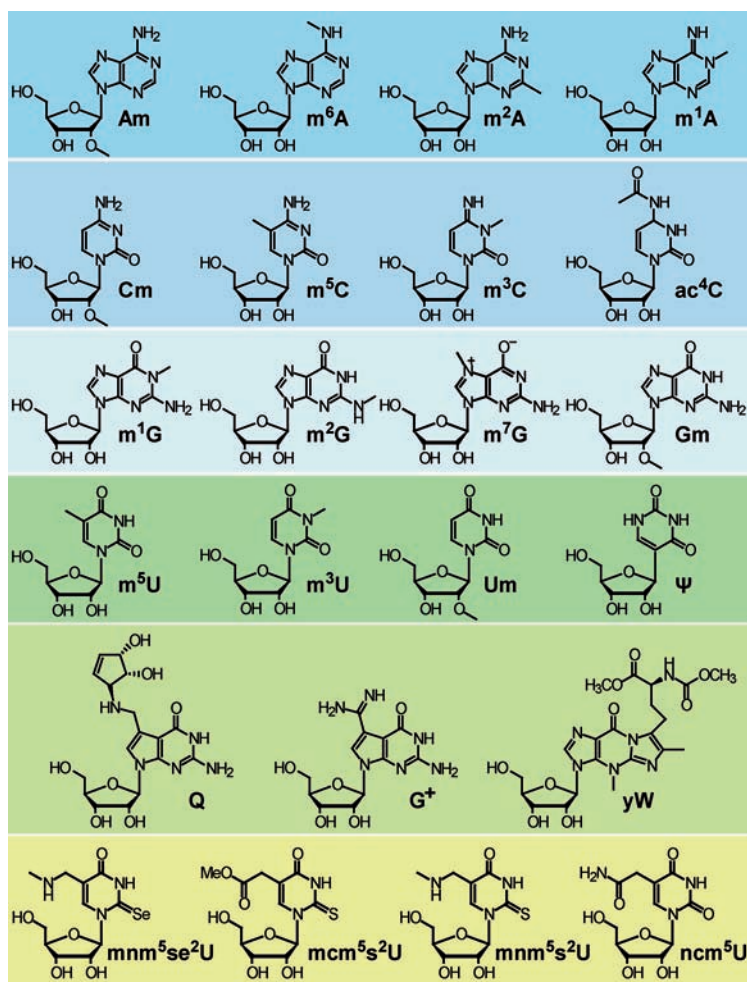


Abbildung 2. Ausgewählte RNA-Modifikationen. Von oben nach unten: A-abgeleitete Modifikationen (dunkelblau), C-abgeleitete Modifikationen (mittelblau), G-abgeleitete Modifikationen (hellblau), U-abgeleitete Modifikationen (dunkelgrün), hypermodifizierte Purinbasen (hellgrün), hypermodifizierte Pyrimidin(Uridin)-Basen (gelb). Für eine komplette Übersicht über die Modifikationen, die bisher gefunden wurden, siehe: <http://s59.cas.albany.edu/RNAmods/> und <http://modomics.genesilico.pl/>.

Abk.	Bezeichnung	Abk.	Bezeichnung
m	methyl	se	seleno
ac	acetyl	cmo	5-oxoessigsäure
t	threonylcarbamoyl	mcmo	5-oxoessigsäure-methylester
tm	taurinomethyl	mo	methoxy
i	isopentenyl	ho	hydroxy
io	cis-hydroxyisopentenyl	f	formyl
s	thio	gal	galactosyl
ms	methylthio	glu	glutamyl
inm	isopentenylaminomethyl	man	mannosyl
g	glycylcarbamoyl	r	2'-O-ribosyl
hn	hydroxynorvalylcarbamoyl	I	Inosin
hm	hydroxymethyl	Ψ	Pseudouridin
nm	aminomethyl	Q	Queuosin
acp	(3-amino-3-carboxypropyl)	yW	Wybutosin
cm	carboxymethyl	OHyW	Hydroxywybutosin
chm	carboxyhydroxymethyl	C ⁺	Agmatidin
cnm	carboxymethylaminomethyl	k ⁺ C	Lysidin
ncm	carbamoylmethyl	D	Dihydrouridin
nmn	methylaminomethyl	G ⁺	Archaeosin

Abbildung 3. Nomenklatur zur Beschreibung der häufigsten modifizierten RNA-Nukleoside. Die möglichen Modifikationsstellen an den Purinen und Pyrimidinen werden in Violett (oben) angegeben. In der Tabelle (unten) schreiben wir zunächst die Abkürzungen, die für die funktionellen Gruppen an den modifizierten Nukleosiden benutzt werden; im zweiten Teil der Tabelle schreiben wir einige Abkürzungen für spezielle modifizierte Nukleoside.

sind die „RNA Modification Database“ (<http://s59.cas.albany.edu/RNAMods/>) und „Modomics: A Database of RNA Modifications“ (<http://modomics.genesilico.pl/>). Diese Datenbanken enthalten Informationen über die chemische Struktur der verschiedenen RNA-Modifikationen, ihre Trivialnamen und ihre Abkürzungen. Zusätzlich wird Literatur über die chemische Synthese der Verbindungen und ihre biologischen Funktionen angegeben. Eine weitere wichtige Datenbank ist tRNAdb (<http://trnadb.bioinf.uni-leipzig.de/>), welche die modifizierten Nukleoside in einen Kontext mit den tRNA-Sequenzen stellt.

2.1. Kürzlich entdeckte RNA-Modifikationen

Mithilfe neuer Techniken der Flüssigchromatographiegekoppelten Massenspektrometrie (LC-MS) mit empfindlichen Massenspektrometern gelang es kürzlich, weitere RNA-Nukleoside zu entdecken (Abbildung 5). In den letzten zehn Jahren wurden ca. 14 neue Modifikationen in Datenbanken hinterlegt, unter anderem das Agmatidin (agm²C oder C⁺), 8-Methyladenosin (m⁸A), Glutamylqueuosin (gluQ) und 5-Taurinomethyluridin (tm⁵U).^[18] Die Agmatidin-Modifikation wurde 2010 durch die Gruppen von Suzuki und RajBhandary entdeckt. Diese Modifikation ist ein Cytidinderivat, das in der

ersten Position in der AUA-dekodierenden tRNA_{2^{Ile}} von Archäen gefunden wird.^[23] Agmatidin ist verwandt mit der bakteriellen Modifikation Lysidin (k²C), unterscheidet sich aber hinsichtlich der Aminosäureseitenkette, die mit dem Cytidin an C2-Position verknüpft ist. Beide Modifikationen haben die gleiche Aufgabe. Sie reprogrammieren das Kodierungspotential von C, da die modifizierten Basen k²C und C⁺ präferentiell mit A statt mit G paaren.^[23b,24]

Drei Jahre zuvor entdeckten Wissenschaftler das 8-Methyladenosin (m⁸A) als eine neue Base in der bakteriellen 23S-rRNA. Diese Base wird in der Position A2503 gefunden, wo man zuvor auch die Modifikation m²A gefunden hatte.^[25] Die Methylierung der Position A2503 zum m⁸A wird durch die Methyltransferase Cfr durchgeführt. Die Methylierung etabliert eine starke Resistenz der entsprechenden Ribosomen gegen fünf unterschiedliche Klassen von Antibiotika, die auf die Peptidyltransferaseaktivität zielen.^[26] Diese Inhibitoren können am modifizierten Ribosom nicht mehr binden. Eine chemische Synthese von m⁸A wurde bereits 1993 beschrieben, und es wurde entdeckt, dass das Nukleosid selbst ein selektiver Inhibitor des Vacciniavirus ist.^[27] Dieses Resultat ist interessant, weil es zeigt, dass auch die Basen selbst interessante biologische Funktionen haben können. Hierzu gehört auch die als Phytohormon bekannte Modifikation i⁶A, die antiproliferative Wirkung auf Tumorzellen zeigt.^[28] Zwei weitere neu entdeckte Modifikationen, das 5-Taurinomethyluridin (tm⁵U) und das 5-Taurinomethyl-2-thiouridin (tm⁵s²U) wurden in der mitochondrialen tRNA von Säugern gefunden. Diese Modifikationen wurden mit hochauflösender Massenspektrometrie in Kombination mit MSⁿ-Fragmentierungsanalysen identifiziert. Die Strukturen wurden durch Totalsynthese und nachfolgende NMR-Analyse bestätigt.^[29] Fehlen diese Modifikationen, so kommt es zur Bildung von Mitochondriopathien (siehe Abschnitt 2.5.1). Die Analyse von seltenen RNA-Spezies mithilfe von hochempfindlichen LC-MS-Methoden konnte kürzlich verschiedene Addukte mit kleinen Molekülen wie NAD, CoA und möglicherweise weiteren kleinen Molekülen aufdecken. Die Verknüpfungsstellen mit dem NAD wurde bis heute noch nicht identifiziert, auch wenn die NAD-Modifikation relativ häufig in *Escherichia coli* und in *Streptomyces venezuelae* vorkommt (ungefähr 3000 Kopien pro Zelle).^[30]

Einige der hypermodifizierten RNA-Basen wie das Q oder das yW besitzen ihre eigene chemische Diversität. Für Queuosin ist z. B. bekannt, dass diese Basen in verschiedenen glycosylierten Formen, unter anderem als Galactose- (galQ) und Mannose(manQ)-Derivat existiert.^[31] Die Funktion der zusätzlichen Zuckereinheiten ist nicht geklärt. Neben den zuckermodifizierten Formen von Q ist noch das Q-Epoxid (oQ) bekannt, bei dem es sich vermutlich um ein Biosyntheseintermediat handelt (Abbildung 5).^[32] Kürzlich wurde ein neues Q-Derivat, das Glutamylqueuosin (gluQ) entdeckt, das statt der Zuckereinheit eine Aminosäure am Aminocyclopentendiol-Ring enthält.^[33] Diese Modifikation wurde in *E.-coli*-tRNA mithilfe spezieller Isolierungsbedingungen, welche die schnelle Hydrolyse der Glutamyl-Queuosin-Bindung verhindert, entdeckt. Dieses Resultat ist sehr wichtig, denn es zeigt, dass viele andere Modifikationen möglicherweise aufgrund zu harscher Isolations- und Detektions-

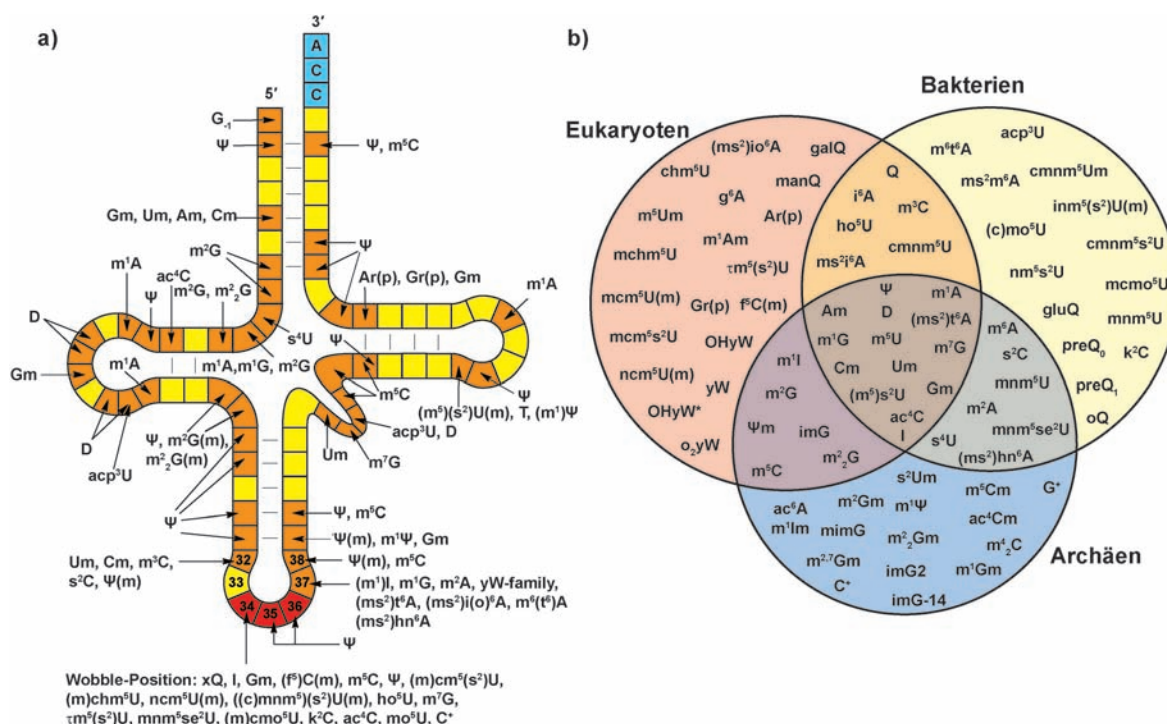


Abbildung 4. a) 2D-Darstellung von tRNA, welche (nicht vollständig) die Positionen wichtiger Modifikationen aufzeigt, die in Bakterien, Archäen und Eukaryoten an der jeweiligen Stelle gefunden werden. Der Antikodon ist in Rot hervorgehoben, häufig modifizierte Positionen sind in Orange dargestellt. Das universelle CCA-3'-Ende wird in Blau gezeigt. Position 34 ist die Wobble-Position und Position 37 gleich neben dem Antikodon ist eine semi-invariable Stelle, die immer von einer Purinbase eingenommen wird. Diese beiden Positionen enthalten oft die hypermodifizierten Basen. Daten wurden der Literatur^[17c] entnommen. b) Modifikationen gruppiert nach ihrem Vorkommen in den drei Domänen des Lebens. Abbildung wurde angepasst nach Lit. [22].

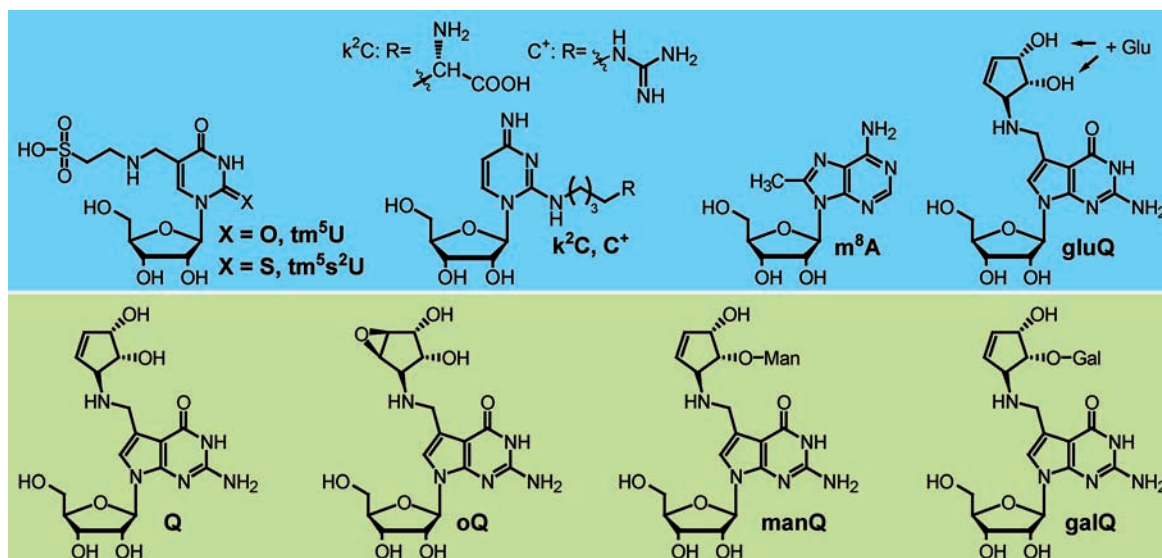


Abbildung 5. Neu entdeckte RNA-Basen: k²C (blau) und Derivate der hypermodifizierten RNA-Base Q (grün).

bedingungen bisher unentdeckt blieben. Bis heute ist weder die exakte Position der Glutamylgruppe noch deren Funktion an der Q-Base bekannt.^[33a] Zusätzlich fehlen bislang Synthesen der Zucker- und Aminosäure-modifizierten Q-Basen.

Man kann annehmen, dass sich der Entdeckungsprozess neuer Nukleoside mit der Entwicklung empfindlicherer Massenspektrometrie und milderer Extraktionsprotokolle fortsetzen wird.

2.1.1. Die Familie der Q-Basen

Die neu entdeckte Base gluQ bestätigt die Q-Base als eine der variabelsten Modifikationen.^[33] Zentrales Strukturelement der Q-Base ist das Desazaguanosin-Ringsystem, das ein Kohlenstoffatom statt eines Stickstoffs in Position 7 trägt und eine Cyclopentendiol-Einheit, die mit dieser C7-Position über eine Methylamingruppe verknüpft ist. Die weiteren Q-Spezies galQ und manQ, in denen eine Aldohehexose mit der nicht-allylischen Alkoholgruppe des Cyclopentens verknüpft ist, existieren alle als β -Anomere.^[31] Das Nukleosid Queuosin selber wurde in den letzten 40 Jahren in tRNA^{Asn}, tRNA^{Asp}, tRNA^{His} und tRNA^{Tyr} nachgewiesen. Q-Derivate sind sowohl in Bakterien als auch in Eukaryoten beschrieben worden, mit der Ausnahme von Bäckerhefe und Mykoplasma.^[34] Es ist aber bekannt, dass sich die höher modifizierten Q-Derivate nur in bestimmten Ästen des Stammbaums wiederfinden. gluQ wurde z. B. nur in Bakterien gefunden,^[33b] während sich galQ und manQ nur in Tieren nachweisen ließen.^[31,35] Alle Q-Derivate befinden sich in der Position 34 der jeweiligen tRNA, der Wobble-Position in der Antikodonschleife.^[36] In Archäen wurden bislang keine Q-Derivate gefunden. Hier findet sich stattdessen das strukturell ähnliche 7-Desazaguanosin-Nukleosid Archaeosin G⁺ (Abbildung 2). G⁺ befindet sich allerdings in tRNA an der Position 15 im sogenannten Dihydrouridin-Loop, und es wird angenommen, dass diese Base hauptsächlich für die Struktur und Stabilität der tRNA verantwortlich ist.^[37]

Die komplexe Biochemie, die gebraucht wird, um solche Basen in der RNA zu etablieren, wird am besten am Beispiel der Q-Base dargestellt. Prokaryoten und Eukaryoten bauen diese Base mithilfe unterschiedlicher Prozeduren ein. Eukaryoten verwenden das Queuin für die Biosynthese, bei dem es sich einfach um den Q-Heterocyclus handelt.^[38] Diese Base wird nicht de novo synthetisiert, sondern mit der Nahrung aufgenommen.^[39] Eukaryoten besitzen eine spezielle tRNA-Guanin-Transglycosylase (TGT), welche die glykosidische Bindung von G34 spaltet und eine abasische Stelle bildet. An dieser Stelle wird anschließend der Queuosin-Heterocyclus mit dem Zucker verknüpft. Die Biosynthese beinhaltet daher einen Heterocyclus-Austauschprozess, der direkt an der tRNA erfolgt.^[38,40] Bakterien sind in der Lage, die Q-Base de novo zu synthetisieren.^[41] Zu diesem Zweck besitzen sie eine komplexe Biosynthesemaschinerie mit mehreren Enzymen.^[34,42] Einer der interessantesten Schritte in der Biosynthese ist die Bildung des reaktiven Epoxids. Interessanterweise wird der Cyclopentenring direkt aus der Riboseeinheit eines SAM-Cofaktors gebildet.^[43] Der letzte Schritt der Biosynthese von Q wird durch ein von Vitamin B12 abhängiges Protein katalysiert, welches oQ in Q umwandelt.^[44] Dieses Enzym wurde von Bandarian und Mitarbeitern entdeckt.^[45] Es handelt sich dabei um eine Eisen-Schwefel-Reduktase, deren Struktur noch nicht bekannt ist. Auch ist die Funktion des Vitamin B12 noch nicht aufgeklärt.^[45]

Trotz umfangreicher Forschungen bezüglich der Q-Base ist deren Funktion noch nicht vollständig geklärt.^[46] Es wird aber angenommen, dass sie die molekulare Erkennung des mRNA-Kodons durch die tRNA am Ribosom maßgeblich beeinflusst. Dies spielt z. B. eine Rolle beim Überlesen von

Stop-Kodons bei bestimmten Viren.^[47] Die Q enthaltenden tRNAs erkennen Kodons mit der Sequenz NAC und NAU (die Sequenz wird, wenn nicht anders angegeben, immer in 5'-3'-Richtung gelesen), wobei das N für A, C, U oder G stehen kann. tRNAs, die statt der G-Base die Q-Base beinhalten, interagieren präferentiell mit Kodons, die mit einem C enden (zur Illustration siehe Abbildung 6B).^[48] Allerdings wird nur mit einer Q-Base an Position 34 der tRNA^{Tyr} das Stop-Kodon UAG in der RNA des Tabakmosaikvirus (TMV) nicht mehr fälschlicherweise erkannt. Die Translation bricht ab, was die Produktion viraler Proteine verhindert.^[49] Weitere Zusammenhänge zwischen der Q-Base und Krankheiten werden in Abschnitt 2.5 diskutiert.

2.2. Funktion von RNA-Modifikationen

Bis heute wurden vielzählige Informationen zur Funktion spezieller Nukleosidmodifikationen in speziellen Positionen gesammelt. Die Modifikationen in oder neben dem tRNA-Antikodon wurden hierbei am intensivsten untersucht. Eine Ausnahme ist die in mRNA vorkommende Base m⁷G, welche die 5'-cap-Struktur in Eukaryoten bildet, die zusammen mit dem poly-A-Schwanz verschiedene Funktionen übernimmt. m⁷G hilft unter anderem, den mRNA-Leserahmen zu definieren. Es beeinflusst den nuklearen mRNA-Export, es schützt die mRNA vor Abbau, und es reguliert die Spleißprozesse.^[50] Entscheidend ist, dass die cap-Struktur das Molekül als mRNA definiert. Interessanterweise wird das m⁷G mit der mRNA über eine relativ labile Triphosphatgruppe verknüpft.^[51] Eine weitere, gut untersuchte Modifikation in mRNA ist das Inosin, das durch Desaminierung von Adenosin gebildet wird. Aufgrund des dadurch veränderten Basenpaarungsmodus (Inosin paart mit C statt mit U) beeinflusst diese Modifikation direkt den genetischen Code. Die Desaminierung ist Teil des RNA-Editierens.^[52] In tRNA kann die große Zahl der Nukleosidmodifikationen bezüglich der Funktionalität grob in drei Gruppen eingeteilt werden.^[17c] Die erste Gruppe von Modifikationen befindet sich direkt in oder neben dem Antikodon (z. B. die Q-Base in Position 34). Diese Modifikationen sind an der Feinjustierung des Translationsprozesses beteiligt und bestimmen damit die translationale Fitness von Organismen.^[14a] Eine zweite Klasse von tRNA-Modifikationen außerhalb der Antikodonschleife kontrolliert die Stabilität und die Faltung der tRNA. Eine dritte Gruppe von tRNA-Modifikationen bestimmt die Identität der jeweiligen tRNA. Diese Modifikationen sind besonders wichtig für die Interaktion von tRNA mit der entsprechenden Synthetase, die eine spezifische Aminosäure auf die zugehörige tRNA-Spezies lädt.^[53] Die Modifikationen dieser Gruppe können sich in ganz unterschiedlichen Positionen der tRNA befinden, auch im Antikodon.

Viele der am stärksten modifizierten Nukleoside gehören zu der ersten Gruppe. Sie befinden sich in der Wobble-Position 34 oder in Position 37 (Abbildung 6) direkt 3' neben dem Antikodon. Position 37 ist die semi-invariable Position, die immer von einer, normalerweise modifizierten Purinbase besetzt wird. Die bekanntesten modifizierten Basen sind das N⁶-Threonylcarbanyladenoin (t⁶A) und das N⁶-Isopentenyl-

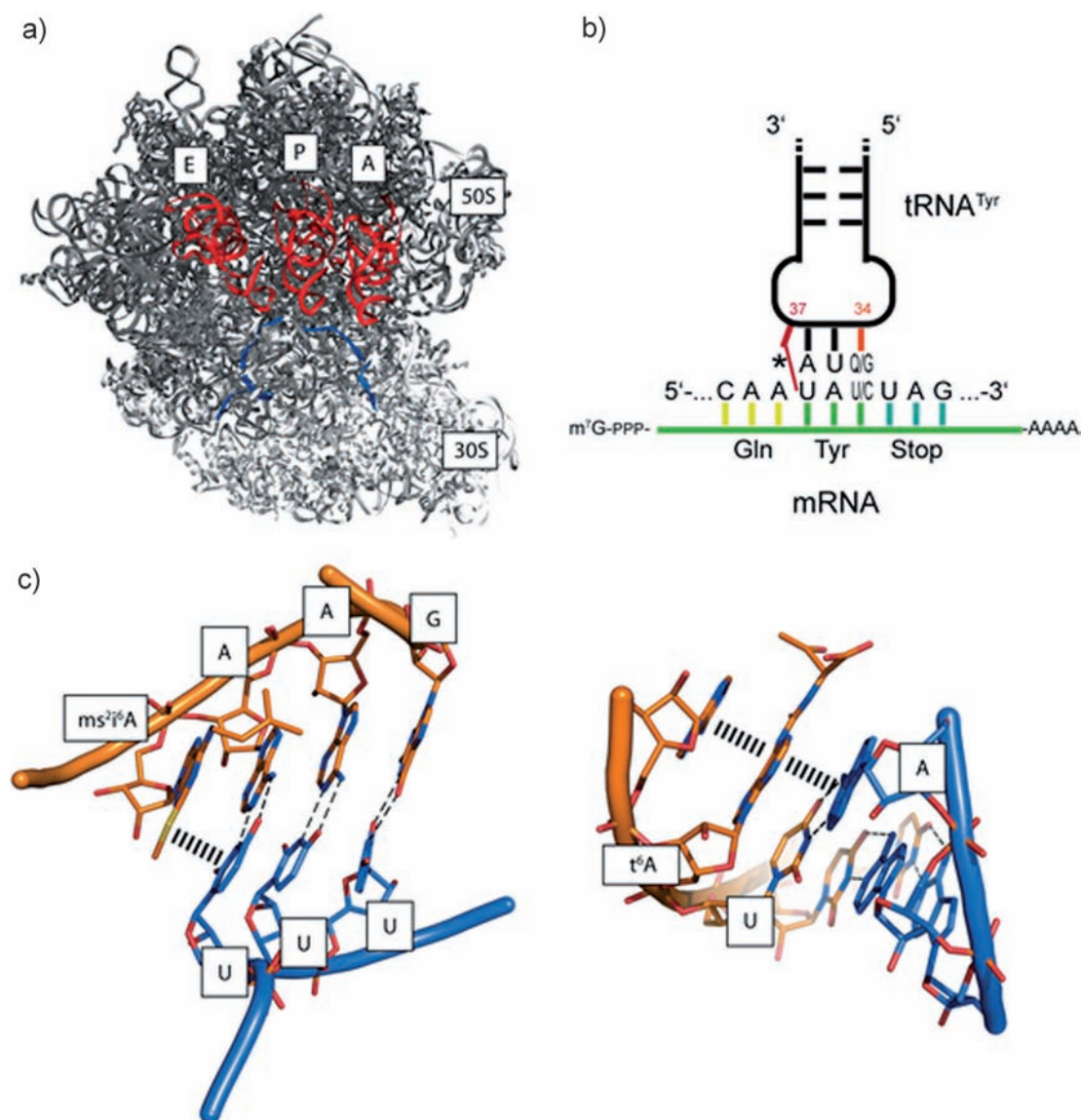


Abbildung 6. Kristallstrukturen, welche die Kodon:Antikodon-Bindung am Ribosom zeigen. a) Struktur des 70S-Ribosoms mit gebundener mRNA und tRNA in der A-, P- und E-Stelle. Die tRNA wird in Rot gezeigt, die mRNA in Blau; dunkelgrau ist das 50S und hellgrau die 30S-Untereinheit des Ribosoms. Teile der 30S-Untereinheit wurden aus Übersichtsgründen weggelassen. b) Illustration der Kodon/Antikodon-Interaktion und der Rolle der modifizierten Basen an den Positionen 34 und 37. * = ms^2i^6A , Q = Queuosin. c) Ausschnitte aus der tRNA/mRNA-Interaktion am Ribosom, welche die zusätzlichen stabilisierenden Wechselwirkungen zeigen, die durch die großen Reste der modifizierten Base an Position 37 neben der Anticodonschleife erzeugt werden.^[56] Die tRNA wird in Orange gezeigt, die mRNA in Violett. Die Stapelwechselwirkung zwischen der modifizierten Base und der tRNA/mRNA werden mit gestrichelten Linien angedeutet. Die Daten wurden den Kristallstrukturen 3I8G (links) und 1XMQ (rechts) entnommen.

adenosin (i^6A), sowie das guaninmodifizierte N^1 -Methylguanidin (m^1G) und das strukturell hochkomplexe Wybutosin (yW), das sich speziell in Phenylalanin kodierenden tRNAs befindet. Diese Modifikationen beeinflussen direkt den Dekodierungsprozess. Sie werden in der Regel nicht für die Interaktion der tRNA mit der entsprechenden Aminoacylsynthetase gebraucht. Diese Basen reduzieren die konformative Flexibilität der Antikodonschleife. Sie erzwingen eine rigide Konformation und optimieren dadurch die Interaktion des Antikodons mit dem Kodon der mRNA.^[15] Diese Stabilisierung, so wird angenommen, erhöht die Geschwindigkeit und die Genauigkeit der tRNA-Bindung zum

entsprechenden Kodon an der ribosomalen A-Seite. Darüber hinaus helfen die Modifikationen, die Positionierung in der ribosomalen P-Stelle zu optimieren, wodurch sie Leseramenverschiebungen verhindern.^[54] Kristallstrukturen^[55] zeigen, dass die großen Seitenketten dieser Basen an Position 37 (z.B. in t^6A , ms^2i^6A und yW) helfen, die schwache Kodon:Antikodon-Wechselwirkung durch zusätzliche Basenstapelung mit der Base 38 der tRNA mit der ersten Base des mRNA-Kodons zu stabilisieren (Abbildung 6).^[56] Ohne diese zusätzlich stabilisierenden Wechselwirkungen könnten U/A-reiche Kodon:Antikodon-Interaktionen (z.B. das AAA-Kodon) nicht effizient gebildet werden. Darüber hinaus

halten die großen Seitenketten die Antikodonschleife offen und verhindern so eine Basenpaarung zwischen den Positionen 32 und 33 mit der Position 38 in der Antikodonschleife (Abbildung 4). Solch eine Interloop-Interaktion würde die Größe der Schleife drastisch verändern und damit die Antikodonerkenennung deutlich behindern.^[57]

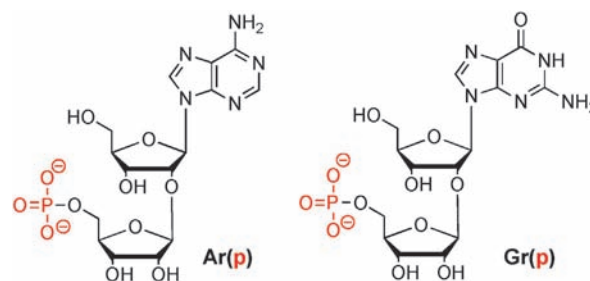
Modifikationen in der Wobble-Position 34 beinhalten das Q und eine große Zahl 5'-modifizierter Uridine (siehe auch Abbildung 2). Diese Basen beeinflussen ganz sicher die Natur der Wobble-Basenpaarung, neben ihrem Einfluss auf die Stabilität der Antikodonschleifenstruktur. Kürzlich erschienene Übersichten haben die molekularen Mechanismen, die der Wobble-Basenpaarung^[14a] zugrunde liegen, intensiv beleuchtet und auch den Einfluss der Wobble-Basenpaarung auf den Dekodierungsprozess des genetischen Materials beschrieben.^[14b] Die Mechanismen der modifizierten Wobble-Basenpaarung sind komplex und sehr spezifisch für bestimmte tRNAs, wie es z.B. durch kürzlich erschienene Röntgenstrukturen gezeigt wurde, welche die Modifikationen mnm⁵U34 und cmo⁵U34 beinhalten.^[56a,58] Diese Strukturen zeigen, wie die Modifikationen die modifizierte Base stabilisieren und korrekt anordnen, sodass optimale Basenstapel und Wasserstoffbrücken am Ribosom gebildet werden können.

Anhand der Lysidinmodifikation (k²C, Abbildung 5) kann eindrucksvoll gezeigt werden, wie modifizierte Basen sowohl den Dekodierungsprozess als auch die Identität der tRNA beeinflussen können. Diese Base wird in der Position 34 von tRNAs gefunden, die für die Aminosäure Isoleucin mithilfe des Antikodons CAU (tRNA^{Ile}_{CAU}) kodiert.^[24b] Das CAU-Antikodon mit einem unmodifizierten Cytidin an Position 34 erkennt typischerweise das Methionin-Kodon AUG in der mRNA. Wenn jedoch das Cytidin posttranskriptional zu einem Lysidin modifiziert wird, dann wird das Isoleucin-Kodon AUG erkannt. Die Modifikation Lysidin hat eine duale Funktion, weil sie darüber hinaus auch die Wechselwirkung mit der tRNA-Synthetase beeinflusst. Während die Isoleucin-Aminoacyl-tRNA-Synthetase (IleRS) die tRNA^{Ile}_{k²CAU} erkennt und ein Isoleucin auf diese tRNA lädt, wird die unmodifizierte tRNA (tRNA^{Ile}_{CAU}) sehr viel langsamer erkannt. Die an dieser Position unmodifizierte tRNA wird jedoch schnell durch die MetRS mit Methionin beladen.^[24a] Diese Modifikation gehört daher zu den beiden Klassen 1 und 3.

Modifikationen der zweiten Gruppe beeinflussen die korrekte Faltung des tRNA-Moleküls und stabilisieren die tRNA. Zusammenhänge, die kürzlich in einem hervorragenden Übersichtsartikel dargestellt wurden.^[16,22] In Übereinstimmung mit dieser Ansicht wurde kürzlich gezeigt, dass die Schmelztemperatur von komplett unmodifizierter und teilmodifizierter tRNA niedriger ist als die von normal modifizierter tRNA.^[59] Es wurde gezeigt, dass tRNA, welche mindestens die zwei Modifikationen m⁷G46 und m⁵C oder ac⁴C und Um44 nicht beinhalten, temperaturempfindlich sind und diese untermodifizierten tRNAs schnell abgebaut werden.^[20,60] Für die Modifikation Pseudouridin^[61] und 2'-O-Methyladenosin^[62] ist bekannt, dass sie die Doppelhelixstrukturen innerhalb eines tRNA-Moleküls stabilisieren. Die modifizierte Base m¹A ist wichtig für die korrekte Faltung der

humanen mitochondrialen tRNA^{Lys}.^[63] Diese Base scheint ein spezielles Faltungsintermediat zu stabilisieren.

Die dritte Klasse von Modifikationen bestimmt die Identität der entsprechenden tRNA, so wie das schon beschriebene Lysidin, welches als spezifisches Erkennungssignal für die IleRS dient.^[24a] Eine ganze Reihe anderer Modifikationen ist ebenfalls für die Erkennung der tRNA durch die entsprechende Aminoacyl-tRNA-Synthetase bekannt, so z.B. das s²U^[64] und das m¹G.^[65] Ein weiteres interessantes Beispiel sind die einzigartigen Phosphoribosylmodifikationen, die sich in Position 64 der Initiator tRNA_i^{Met} in Hefe, Pflanzen und Pilzen befinden. Diese tRNA wird benötigt, um die Peptidsynthese mit Methionin am AUG-Startkodon zu initiieren. Aber sie ist nicht an der weiteren Elongation der Peptidkette und am weiteren Einbau von Methionin beteiligt. Die Phosphoribosylmodifikation verhindert, dass die tRNA_i^{Met} mit dem Elongationsfaktor interagieren kann. Damit schränkt sie die Funktion dieser tRNA auf den Initiierungsprozess ein.^[66] Diese Modifikation ist bekannt als 2'-O-Ribosyladenosin Ar(p) oder 2'-O-Ribosylguanosin Gr(p). Die große Modifikation weist hierbei in den T-Arm des tRNA-Moleküls und



vergrößert dessen sterische Anforderung stark.^[67] Während in Hefe tRNA_i^{Met} normalerweise nicht in der Lage ist, mit dem Elongationsfaktor EF-1α zu interagieren, wurde gezeigt, dass eine durch chemische Modifikation nicht-phosphoribosylierte Variante von tRNA_i^{Met} an diesen Elongationsfaktor binden kann.^[68] Folglich konnte eine Initiator-tRNA ohne die Ar(p)-Modifikation Methionin in vitro auch an internen AUG-(Met)-Kodons einfügen, also eine Elongatorfunktion übernehmen. Auch wurde gezeigt, dass ein Stamm, dem das Gen für das phosphoribosylierende Enzym fehlt und der auch keine Elongator-tRNA^{Met} bilden kann, erfolgreich mit tRNA_i^{Met} als Elongator-tRNA^{Met} leben kann.^[69] Diese Experimente zeigen deutlich, dass das phosphoribosylierte Nukleosid die Rolle der tRNA als Initiator-tRNA definiert und verhindert, dass diese tRNA an der Elongation beteiligt wird. Es sei darauf hingewiesen, dass in Bakterien und höheren Eukaryoten andere Strategien zur Differenzierung zwischen normaler tRNA und initialer tRNA gefunden werden.

Zusammenfassend stellen wir fest: Auch wenn der genaue Mechanismus, wie eine bestimmte Nukleosidmodifikation den Organismus beeinflusst, häufig noch unbekannt ist, erhalten wir doch Einblick in ihre Funktionen, z.B. wie sie die Stabilität der Antikodonschleife^[56b] spezifisch verändern, um die Genauigkeit des Dekodierungsprozesses zu beeinflussen.^[70] Wir lernen, wie die Modifikationen spezifische Kodon- und Antikodonpräferenzen einstellen, vor allem um Lese-

rahmenverschiebungen zu verhindern.^[54] Diese Beobachtungen führten schon früh zu der Auffassung, dass die Dekodierung der mRNA-Sequenz durch tRNA über ein sogenanntes „Extended Codon“ abläuft, das neben der Information der drei Antikodonbasen auch die Informationen verwendet, die durch Basenmodifikationen eingeführt werden.^[14a,71]

2.3. Biosynthese der RNA-Modifikationen

Wie bereits in Abschnitt 2.1.1 am Beispiel des Queuosins dargestellt, ist die Chemie zur Einführung einiger Modifikationen oft komplex und erfordert dedizierte Biosynthesenzyme, mit häufig unbekannten Reaktionsmechanismen. Informationen über die enzymatische Maschinerie, die hinter dem Einbau von Modifikationen in RNA steht, findet sich gesammelt in der Modomics-Datenbank unter <http://modomics.genesilico.pl>.

Diese biosynthetischen Enzyme operieren häufig in Komplexen, in denen die vielfältigen Schritte zur Erzeugung einer RNA-Modifikation orchestriert ablaufen können.^[17b] Unter den kürzlich identifizierten Komplexen ist der KEOPS-Komplex, der für die Synthese der universellen Base t⁶A37 in tRNAs, die ANN-Kodons erkennen, zuständig ist,^[72] sowie das Säugerenzym ABH8 (ein AlkB-Homolog), das sowohl die Methylierung als auch die Hydroxylierung bei der Synthese von mcm⁵U34/mcm⁵U34 katalysiert.^[73] Das ABH8-Enzym ist besonders interessant, da es einen Einfluss beim Überleben nach DNA-Schädigung ausübt und somit eine Rolle von tRNA-Modifikationen bei der Antwort auf DNA-Schädigung impliziert. Darüber hinaus ist die hydroxylierende Untereinheit von ABH8 eine Fe^{II}/α-Ketoglutarat-abhängige Dioxygenase, ähnlich wie die später beschriebenen TET-Enzyme, die für die Hydroxylierung von 5-Methylcytosin in DNA zuständig sind (Abschnitt 3.2).

Obwohl bereits viel über die Enzyme und Stoffwechselwege, die an der RNA-Modifikation beteiligt sind, bekannt ist, fehlen Detailkenntnisse über die Erkennung der genauen Position, an der die Modifikationen eingeführt werden. Ein Teil der modifizierenden Enzyme ist sicherlich strukturspezifisch. Sie benötigen eine ausgebildete Struktur der RNA für den Start der Modifizierungsreaktion. Andere Enzyme nutzen Template aus kleiner nukleolarer RNA (snoRNA), die mit einer komplementären Ziel-RNA-Sequenz interagieren, in die die Modifikation eingeführt wird. Diese Templat-RNA-Stränge sind Teil von kleinen, ribonukleären Partikeln (snoRNPs). Die snoRNAs vom Box-C/D-Typ erkennen beispielsweise komplementäre Sequenzen im Ribosom. Nach erfolgter Erkennung wird das Protein Fibrillarin, ein Bestandteil des snoRNP, aktiviert und katalysiert eine 2'-O-Ribosylierung mittels des Cofaktors SAM.^[74] Box-H/ACA-snoRNAs hingegen rekrutieren das Enzym Dyskerin zur Modifikationsstelle, welches die Umwandlung von Uridin in Pseudouridin katalysiert.^[75] Kürzlich wurden weitere modifizierende Templat-RNAs in den Cajal-Körperchen der Zellen entdeckt (scaRNA), die eine Rolle bei der Modifikation von snRNA zu spielen scheinen.^[76]

tRNAs, auf die wir uns in diesem Teil des Aufsatzes aufgrund ihrer großen chemischen Diversität konzentrieren wollen, werden zunächst als Primärtranskripte (pri-tRNAs) im Zellkern hergestellt. Sie durchlaufen eine Prozessierung der 3'- und 5'-Enden und werden häufig vor der Einführung von Modifikationen gespleißt. In anderen Fällen, wie bei der Lysidin-tRNA-Modifikation, findet die Modifizierung bereits an der Vorläufer-tRNA statt, vermutlich um die Konsequenzen von unmodifizierter tRNA zu vermeiden.^[77] In jedem Fall muss die Zelle die Art und Position der Modifikationen streng kontrollieren. Lange Zeit nahm man an, dass vollständig modifizierte tRNA unwiderruflich aus dem Zellkern in das Zytoplasma exportiert wird, wo die Aminosäuren auf den CCA-Terminus der tRNA übertragen werden.^[78] Man hat jedoch in den letzten Jahren entdeckt, dass diese vereinfachte Betrachtung nicht korrekt ist. Die Prozesse der Aminoacylierung^[79] und des Spleißens^[80] können sowohl im Zellkern als auch im Zytoplasma stattfinden. Wir wissen heute, dass auch unvollständig prozessierte RNA vom Zellkern ins Zytoplasma exportiert wird und für die Vervollständigung des Modifikationsprozesses auch wieder in den Zellkern reimportiert werden kann.^[81] Wir folgern daraus, dass die RNA-Reifung ein höchst komplexer Prozess ist, der in verschiedenen Organismen unterschiedlich ablaufen kann.^[82] Um diesen Prozess beispielhaft zu verdeutlichen, werden wir im Folgenden die Prozessierung der tRNA^{Phe} aus *Saccharomyces cerevisiae* bis zur Einführung der hypermodifizierten Base Wybutosin näher beschreiben.

2.3.1. Verteilung des tRNA-Reifungsprozesses in der Zelle am Beispiel yW-modifizierter tRNA

Transkription^[83] und 3'-Prozessierung^[84] der pri-tRNA^{Phe} aus Hefe finden innerhalb des Nukleolus des Zellkerns statt (Abbildung 7). Die Schritte der 3'-Prozessierung und das Anhängen des CCA-3'-Endes, sowie die Einführung der ersten 10 Nukleosidmodifikationen in die tRNA^{Phe} finden anschließend im Kernplasma und an der inneren Kernmembran statt.^[85] Während dieses Vorgangs bleibt das 19 Nukleotide lange Intron dieser prä-tRNA^{Phe} vorhanden, was die Voraussetzung für die Bildung der m⁵C-Modifikation an Position 40 ist.^[86] Vor der Synthese des Wybutosins und dreier weiterer Modifikationen muss dieses Intron zwischen Position 37 und 38 jedoch aus der tRNA herausgeschnitten werden. Die für dieses Spleißen zuständigen Endonukleasen befinden sich an der äußeren Membran der Mitochondrien.^[80] Die Intron-enhaltende prä-tRNA^{Phe} wird daher aus dem Zellkern exportiert, um an der Mitochondrienmembran den Spleißvorgang zu durchlaufen. Anschließend werden von zytosolischen Methyltransferasen die Modifikationen m²G10 und Cm32/Gm34 eingeführt, bevor die tRNA^{Phe} wieder in den Zellkern reimportiert wird.^[85] Im Zellkern kann nun der erste Schritt der Wybutosin-Synthese stattfinden, die S-Adenosylmethionin(SAM)-abhängige Methylierung von Guanotin 37 in m¹G durch TRM5. Es folgt der erneute Export der tRNA in das Zytoplasma, wo alle weiteren Schritte der yW-Biosynthese stattfinden.^[85] Die letzte Aminoacylierung findet wahrscheinlich im Zytoplasma an der vollständig modifizierten tRNA^{Phe} mit Wybutosin in Position 37 statt. Es wurde

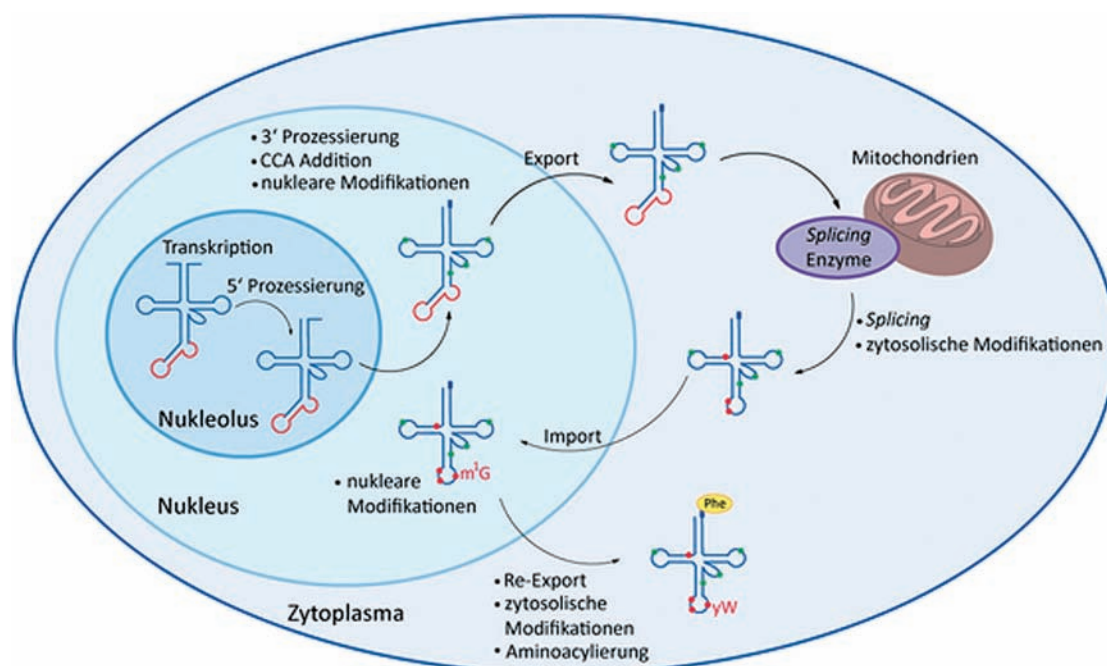


Abbildung 7. Darstellung der Biosynthese der yW-Base und der Orte innerhalb der Zelle, an denen die einzelnen Reifungsschritte stattfinden.

jedoch gezeigt, dass die Beladung mit Aminosäuren auch im Zellkern stattfinden kann.^[87]

Man nimmt heute an, dass der standardmäßige Transport von tRNA in und aus dem Zellkern regulatorische Funktion hat und einen Korrekturmechanismus für falsch modifizierte oder gefaltete tRNA darstellt.^[81b] So ist bekannt, dass das eukaryotische Transportprotein Exportin-t intakte aminoacylierte tRNA wesentlich besser erkennt als prä-tRNA oder tRNA mit strukturellen Defekten.^[79b,87] Hypomodifizierte oder fehlgefaltete tRNA wird so im Zellkern erkannt und dem Abbau zugeführt.^[88] Zudem ist bekannt, dass unter bestimmten Stressbedingungen tRNA im Zellkern akkumuliert und so wichtige Impulse für die Produktion von Aminosäuren liefert.^[81b,89] Werden diese Aminosäuren nachträglich in das Medium gegeben und auf die tRNA geladen, werden diese unverzüglich aus dem Zellkern exportiert.^[89] Diese Beobachtung zeigt, dass die Verteilung der tRNA zwischen Zellkern und Zytoplasma reversibel ist und auch regulatorische Funktion besitzen kann. Eine interessante Beobachtung ist auch, dass Im- und Export von tRNA auch mit der Erkennung von DNA-Schäden verknüpft ist. Werden zu viele DNA-Schäden im Genom der Hefe erkannt, wird das Exportprotein Los1 im Zytoplasma zurückgehalten. Dies führt zu einer Akkumulation unprozessierter tRNA im Zellkern, was wiederum zu einer reduzierten Translation und dem Verbleib in der G1-Phase des Zellzyklus führt.^[90]

2.3.2. Biosynthese der yW-Base

Die Biosynthese der hypermodifizierten Basen Wybutosin (yW) und ihrer hydroxylierten Form, Hydroxywybutosin (OHyW), wurde in den letzten Jahren vollständig aufgeklärt (Abbildung 8). Von den 6 an der Biosynthese von Wybutosin beteiligten Enzymen (TRM5, TYW1-5) wurden TYW1,

TYW2, TYW5 und TYW5 von Nureki und Mitarbeitern strukturell charakterisiert.^[91] Der erste Schritt, die Methylierung an der N1-Position des Guanosins durch TRM5, ist oben bereits erwähnt worden.^[92] Dem schließt sich die Bildung der für yW charakteristischen, tricyclischen Ringstruktur mittels Pyruvat an.^[93] Diese Reaktion wird von dem Enzym TYW1 katalysiert.^[91a,94] Im nächsten Schritt erfolgt die Verknüpfung der α -Amino- α -carboxypropyl(acp)-Gruppe von SAM mit dem tricyclischen Grundgerüst durch TYW2.^[91d,94b,95] Danach wird von TYW3 eine weitere Methylgruppe an Position 4 eingeführt.^[94b] Der letzte Schritt der Biosynthese von yW besteht aus zwei Reaktionen, die beide von TYW4 katalysiert werden.^[91b] In diesen Schritten wird zunächst die Carboxyfunktion der acp-Gruppe methyliert und dann die Aminofunktion methoxycarbonyliert. Dieser Prozess beinhaltet die Fixierung von CO₂. Die noch komplexere, in höheren Eukaryoten zu findende Modifikation Hydroxywybutosin, trägt zusätzlich eine Hydroxygruppe in β -Position zur acp-Seitenkette. Die Synthese dieser Modifikation benötigt eine ungewöhnliche C-H-Aktivierung. In wegweisenden Arbeiten haben die Gruppen um Nureki und Suzuki 2010 das humane Jumonji-C-Domänenprotein TYW5 isoliert, das diese ungewöhnliche Reaktion katalysiert.^[96] Wie auch das zuvor erwähnte ABH8 gehört TYW5 zur Familie der Fe^{II}/ α -KG abhängigen Dioxygenasen, auf die wir in Abschnitt 3.2 näher eingehen wollen. Die Kristallstruktur des humanen TYW5 (hTYW5) zeigt ein konserviertes His-X-(Asp/Glu)-X_n-His-Motiv, das den Eisen-Chelatkomplex bildet (Abbildung 8D).^[91c] Des Weiteren wird die acp-Seitenkette von zwei konservierten Argininresten komplexiert. Der vorgeschlagene Mechanismus der C-H-Aktivierung erfolgt nach dem Muster anderer Dioxygenasen.^[97] Das Enzym bindet zunächst Fe^{II}, molekularen Sauerstoff, α -Ketoglutarat und Wybutosin und bildet nun eine reaktive Oxoferryl-Spezies

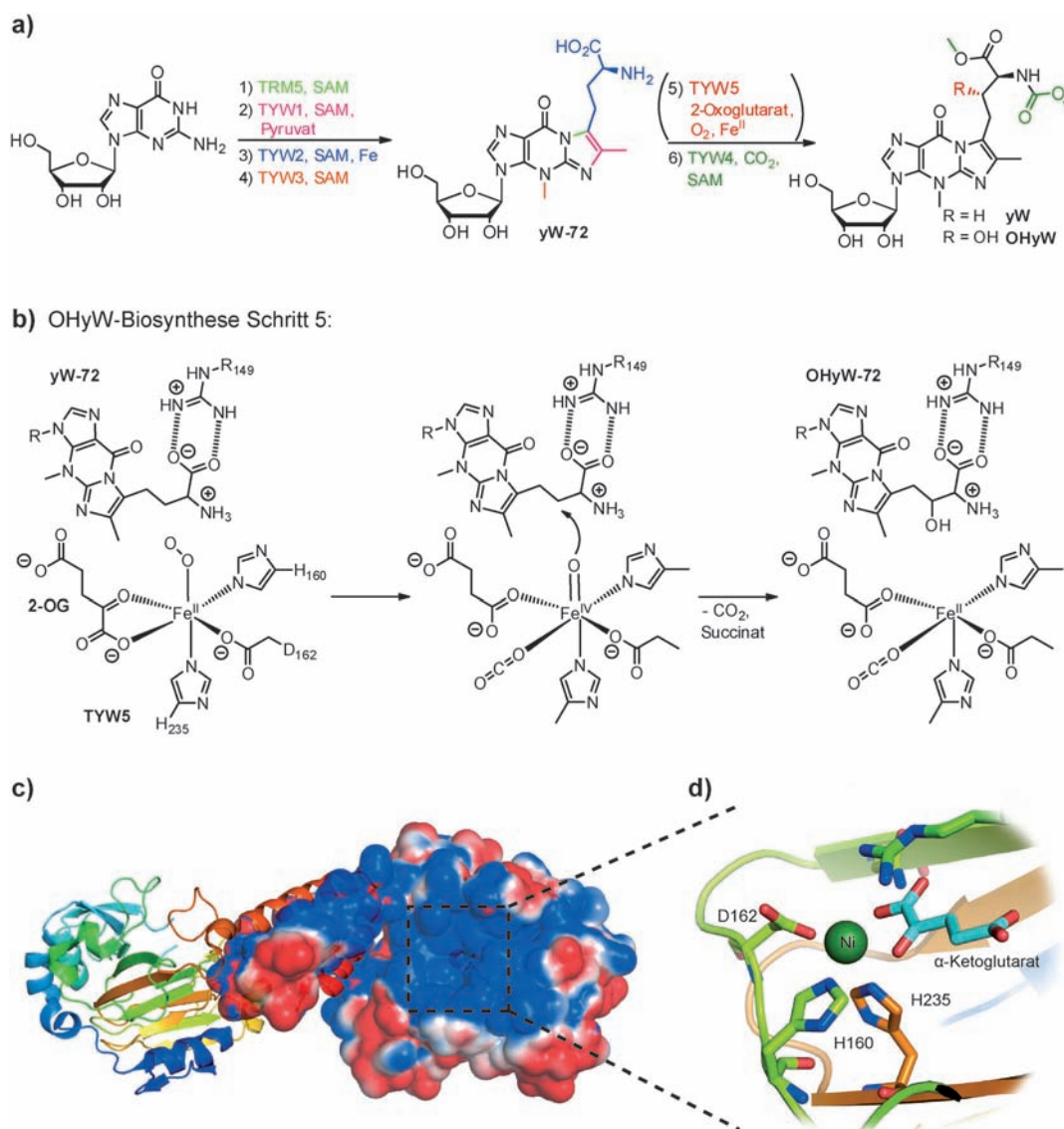


Abbildung 8. a) Biosynthese von Wybutosin und Hydroxywybutosin. b) Vorgeschlagerener Mechanismus für die Einführung der Hydroxygruppe in die β -Position der acp-Seitenketten. Abbildung verändert nach Lit. [91c]. c) Schematische Ansicht des TYW5-Dimers. Die linke Untereinheit verdeutlicht die Sekundärstruktur, während auf der rechten Untereinheit die Oberflächenladungen gezeigt sind. Blau steht für positive Ladung, rot für negative Ladung. Das aktive Zentrum ist von positiver Ladung umgeben, was auf die Schnittstelle zur RNA-Bindung hindeutet. d) Vergrößerung des aktiven Zentrums von TYW5, das mit Nickel anstatt Eisen und α -Ketoglutarat, aber ohne das RNA-Substrat kristallisiert wurde.

mit der Entstehung von CO_2 und Succinat. Diese Oxoferrylspezies greift nun die β -Position der acp-Seitenkette an und führt zur Bildung des Hydroxywybutosin-Derivats OHyW-72,^[91c] das anschließend von TYW4 zu OHyW überführt wird. Es ist faszinierend zu beobachten, wie die Natur komplexe Radikalchemie für die notwendigen Transformationen nutzt, und zwar auf einem komplexen DNA- bzw. RNA-Substrat, das normalerweise sehr leicht durch Radikale zerstört wird.^[98]

2.3.3. Modifikationen in nicht-kodierender RNA

Nicht-kodierende RNA (ncRNA) sind Transkripte, die nicht in Proteine übersetzt werden.^[99] Diese RNA-Familie beinhaltet tRNA, rRNA, snoRNA, miRNA, siRNA, piRNA (etwas längere miRNA, assoziiert mit piwi-Proteinen in

Zellen der Keimbahnlinie) und auch lange, nicht-kodierende RNA (lncRNA). Die kleinen, nicht-kodierenden RNAs sind dafür bekannt sowohl interne als auch externe Modifikationen zu tragen. 2005 zeigten Chen und Mitarbeiter, dass die 3'-Enden der miRNA in *Arabidopsis thaliana* 2'-O-methyliert sind.^[100] Sie zeigten, dass die Methyltransferase HEN1 Methylgruppen in miRNA und siRNA einführt, um diese Moleküle zu stabilisieren. In *hen1*-Mutanten war die Menge der miRNA verringert und ihre Länge unterschiedlich.^[101] Es wurde gezeigt, dass diese Methylierung die miRNA vor der Uridinylierung mit bis zu 5 Uridinen schützt, was andernfalls zum Abbau der miRNA führen würde.^[102] Während die 3'-Enden tierischer miRNA generell unmodifiziert sind, sind die 3'-Enden der piRNA in Maus,^[103] Zebrafisch^[104] und *Drosophila*^[105] 2'-O-methyliert. In *Drosophila* wird auch die siRNA

an ihrem 3'-Ende 2'-O-methyliert. Um diese Modifikation des 3'-Endes in *Drosophila*-siRNA zu bewerkstelligen, assoziiert das HEN1 ortholog (DmHEN1) mit dem Protein Argonaut 2 (Ago2), was impliziert, dass diese Modifizierung möglicherweise den letzten Schritt beim Aufbau des RISC-Komplexes darstellt.^[105] Interne 2'-O-Methylierungen sind beispielsweise für die kleine, nicht-kodierende BC1-RNA wichtig, die in Neuronen stark exprimiert und in Synapsen angereichert wird. Die 2'-O-Methylierung beeinflusst die RNA-Struktur und moduliert so die Interaktion mit wichtigen Proteinen. Wenn die Modifikation fehlt, bindet BC1-RNA an Proteine. Auf diese Weise wird über Modifikationen die Translation bestimmter mRNAs in Synapsen gesteuert.^[106] Ein kürzlich erschienener Übersichtsartikel von Kim et al. behandelt diese wichtigen Aspekte der ncRNA-Modifikation bei der dynamischen Regulation der Gen-Stummschaltung.^[107] Es gibt einige weitere interne Modifikationen in nicht-kodierenden RNAs. Adenosin-nach-Inosin-Editierung wurde zum Beispiel für einige pri-miRNA gezeigt. Eine A-nach-I modifizierte pri-miRNA verlor die Fähigkeit zu prä-miRNA bzw. miRNA prozessiert zu werden. Dies gab Anlass zur Spekulation, dass diese Art von Editierung strukturelle Veränderungen hervorruft, welche die Weiterprozessierung verhindern.^[108] Ein weiterer Editierungsprozess, die Isomerisierung von Uridin zum Pseudouridin (siehe Abschnitt 2.3) findet in verschiedenen nicht-kodierenden, strukturellen RNAs, wie rRNA und tRNA, sowie in snRNA^[109] und snoRNA statt. Auch die RNA-Komponente der humanen Telomerase enthält potentielle Pseudouridinylierungsstellen.^[110] Pseudouridinylierung kann die Regulation der Genexpression beeinflussen, was am Beispiel einer Coaktivator-RNA, der sogenannten Steroidrezeptoraktivator-RNA (SRA), gezeigt wurde. Diese SRA wird von der Pseudouridinylat-Synthase aus der Maus (mPUS1p) an verschiedenen Positionen pseudouridinyliert, und es wird angenommen, dass SRA dadurch eine unterschiedliche Struktur einnimmt, um mit weiteren Proteinen, die an der Regulation der Transkription bestimmter Zielgene beteiligt sind, zu interagieren.^[111]

Zusammenfassend können wir sagen, dass tRNA zwar die RNA-Spezies mit den meisten und chemisch komplexesten Modifikationen (z. B. Queuosin, Wybutosin) darstellt, Modifikationen aber von grundsätzlicher Bedeutung auch für andere RNA-Spezies sind. Modifizierte Basen werden auch in ribosomalen RNAs und kleinen nicht-kodierenden RNAs (sncRNA) wie miRNA und piRNA gefunden, aber auch – erst kürzlich – in langen, nicht-kodierenden RNAs (lncRNA), wo sie die Faltung, Stabilität und nicht zuletzt die Interaktion mit kritischen Proteinen beeinflussen. Nicht-kanonische Basen sind in hohem Maß für die diversen Funktionen dieser nicht-kodierenden RNAs zuständig.

2.4. Quantifizierung und genomweite Kartierung von RNA-Modifikationen

Die Detektion und strukturelle Charakterisierung von modifizierten RNA-Nukleosiden ist ein hochentwickeltes Gebiet, und wir verweisen an dieser Stelle auf exzellente Übersichtsartikel.^[112] Es ist nach wie vor schwierig, modifi-

zierte Nukleoside in unterschiedlichen RNA-Spezies und Geweben zu lokalisieren und zu quantifizieren. Die etablierten Sequenzieretechniken beruhen auf vorangehender Amplifikation revers transkribierter RNA durch die Polymerasekettenreaktion (PCR). Informationen über in der RNA befindliche Modifikationen, die sich nicht auf das Basenpaarungsverhalten auswirken, gehen bei diesen Schritten verloren. Zudem werden die für die reverse Transkription benötigten reversen Transkriptasen durch viele Modifikationen blockiert, und es entstehen unvollständige cDNA-Fragmente oder lokale Mutationen.^[113] Diese Eigenschaften konnten jedoch durch geschicktes Assaydesign zur genomweiten Kartierung einiger Modifikationen genutzt werden. A-nach-I-Editierungen konnten beispielsweise durch chemische Derivatisierung von Inosin mit Acrylnitril detektiert werden, was Inosin in einen Block der reversen Transkription überführt.^[114] Die Verwendung von Sequenzierungstechniken der nächsten Generation (NGS) bei der Sequenzierung von Bisulfid-behandelter RNA hat zur genomweiten Identifizierung von m⁵C-Modifikationen geführt.^[115] Limbach und Mitarbeiter haben durch chemische Derivatisierung von Pseudouridin in Kombination mit Restriktion durch Endonukleasen und MALDI-TOF-MS Pseudouridinylierungen in komplexen RNA-Mischungen nachweisen können.^[116] Erst kürzlich wurde auch eine genomweite Karte von m⁶A-Modifikationen erstellt, basierend auf der Anreicherung m⁶A-haltiger Fragmente mit einem spezifischen Antikörper und anschließender massiver paralleler RNA-Sequenzierung (RNAseq).^[117] Diese Methoden ermöglichen die Detektion ausgewählter Modifikationen innerhalb der RNA. Demgegenüber konnte kürzlich eine Vielzahl unterschiedlicher Modifikationen auf einem präzise geschachtelten Mikroarray detektiert werden. Durch die vergleichende Analyse von Wildtyp-Hefe mit modifikationsdefizienten Mutanten wurden enzymespezifische und bislang unbekannte Modifikationsstellen entdeckt.^[118] Es ist jedoch weiterhin schwierig, die exakte RNA-Spezies oder gar die genaue Position einer bestimmten Modifikation zuzuordnen. Es werden weiterhin effiziente Methoden zur Isolierung einzelner RNA-Spezies aus komplexen Gewebeproben benötigt, die die anschließende Analyse mittels partiellem Verdau und Massenspektrometrie ermöglichen. Eine solche Methode könnte die positionsgenaue Bestimmung von Modifikationen in tRNA und anderen RNA-Spezies erheblich beschleunigen. Mit diesem Ziel wurde kürzlich eine Methode namens RCC (reciprocal circulating chromatography) entwickelt, die es erlaubt, einzelne RNA-Spezies aus einer komplexen Mischung mithilfe von komplementären DNA-Fängersonden, immobilisiert auf einer Festphase, zu isolieren. Der Vorteil der RCC-Technik ist ihre Anwendbarkeit auf Hochdurchsatz-Strategien.^[119] Ein anderer Ansatz mit spezifischen Sonden für bekannte Zielsequenzen wurde kürzlich für die Quantifizierung von m¹A in tRNA auf genomweiter Ebene genutzt. Hierbei wurden Primerverlängerung mit anschließender PCR-Amplifikation und Hybridisierung auf Mikroarrays verwendet.^[120]

Die Quantifizierung des Modifikationsgrades in Geweben und Zellen ist ebenfalls im Fokus intensiver Untersuchungen, und verschiedene Methoden wurden zu diesem Zweck entwickelt.^[112a,121] Zurzeit basieren die präzisesten Techniken auf

der Massenspektrometrie, genauer der LC-MS in Kombination mit Isotopenverdünnungsprotokollen^[122] oder Tandem-Massenspektrometrie.^[123] Obwohl sich LC-MS hervorragend für die Detektion individueller Modifikationen eignet und MSⁿ-Technologien zur Entschlüsselung der Struktur beitragen können, ist die Massenspektrometrie nur bedingt für die Quantifizierung geeignet, da die Signalintensität eines bestimmten Moleküls eher auf seinem Ionisierungspotential als auf seiner Konzentration im Analyt beruht. Mithilfe von isotopenmarkierten Standardsubstanzen, welche die gleichen Retentionszeiten und Ionisierungseigenschaften wie ihre natürlichen Pendanten besitzen, kann jedoch die Konzentration einer Modifikation sehr präzise bestimmt werden. Ihre Position innerhalb der RNA-Spezies kann jedoch nicht bestimmt werden. Basierend auf dieser Methode war es kürzlich möglich, einen breiten, quantitativen Einblick in die Verteilung von Modifikationen zwischen unterschiedlichen Geweben aus Maus und Schwein zu erhalten. Beispielhafte Daten sind in Abbildung 9 aufgeführt.^[124]

Wie in Abbildung 9 gezeigt, variieren die Mengen der modifizierten Nukleoside stark in den unterschiedlichen Geweben, was unterstreicht, dass der Sequenzcode in allen Zellen eines Organismus zwar identisch, der Modifikationsgrad der tRNA hingegen für jedes Gewebe spezifisch

ist. Dies könnte eine gewebsspezifische Modulation von Translation und Metabolismus widerspiegeln.

Andere Quantifizierungsansätze zeigten, dass Umweltbedingungen einen Einfluss auf den Modifizierungsgrad haben können. Eine kürzlich veröffentlichte Studie belegt, dass sich die Mengen spezifischer RNA-Modifikationen in *Saccharomyces cerevisiae* als Antwort auf verschiedene Stressbedingungen dynamisch verändern,^[123b] was darauf hindeutet, dass die Modifikationsgrade auch aufgrund von Umwelteinflüssen angepasst werden.

LC-MS-Analysen in Kombination mit Isotopenverdünnung wurden ebenfalls für die Quantifizierung von modifizierten Nukleosiden in tRNA bakterieller und eukaryotischer Spezies durchgeführt. Die gesammelten Quantifizierungsdaten variieren so stark zwischen den untersuchten Spezies, dass ein „phylogenetischer“ Stammbaum (Abbildung 10) konstruiert werden konnte, dessen Aufbau gut mit Stammbäumen auf klassischen, genomischen Datengrundlagen übereinstimmt.^[125] Diese Daten zeigen, dass zwar kleinere Schwankungen im Modifikationsgehalt aufgrund von Stress oder organspezifischen Aufgaben auftreten, die Gesamtheit der Modifikationen und ihre relativen Mengen aber vom phylogenetischen Ursprung einer Spezies bestimmt werden.

2.5. Die Rolle von RNA-Modifikationen in Krankheit und Diagnose

Obwohl es für die meisten Modifikationen bislang unbekannt ist, wie sie die Gesundheit eines Organismus beeinflussen, gibt es dennoch einige Beispiele, wie Veränderungen im Modifikationsstatus zur Entwicklung von Krankheiten beitragen. Hierbei ist es auffällig, dass die meisten Krankheiten im Zusammenhang mit Nukleosidmodifikationen im Bereich des Energiestoffwechsels zu finden sind (z.B. Mitochondriopathien, Diabetes mellitus Typ 2, Fettleibigkeit).

2.5.1. Die Rolle von RNA-Modifikationen bei menschlichen Erkrankungen

Eine besondere Modifikation, deren Abwesenheit direkt mit einem Krankheitsbild verknüpft ist, ist das Taurin, eine Modifikation der Wobble-Base Uridin in den mitochondrialen tRNA^{Leu}_{UUR} und tRNA^{Lys}.^[126] Untermodifizierte Wobble-Positionen in diesen tRNAs sind verantwortlich für die Ausprägung zweier bedeutender mitochondrialer Erkrankungen, MELAS (mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes) und MERRF (myoclonus epilepsy associated with

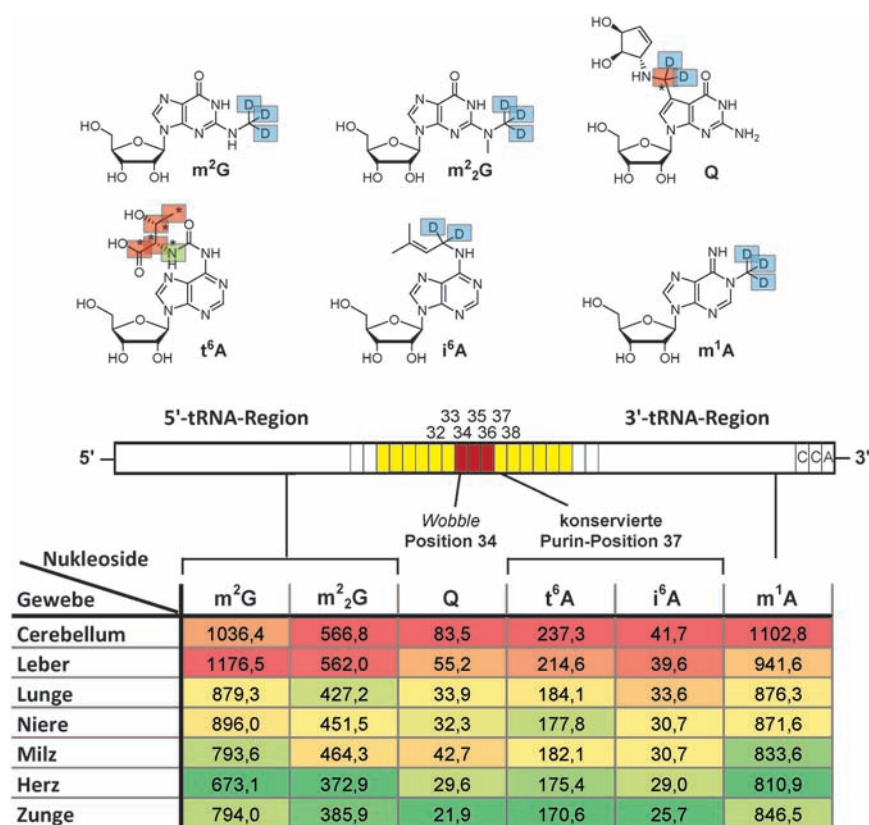


Abbildung 9. Darstellung von synthetischen, mit stabilen Isotopen markierten, modifizierten Nukleosiden und die ungefähre Position der natürlichen Analoga innerhalb der tRNA. Die farb-codierte Tabelle zeigt die Mengen der hergestellten Modifikationen in unterschiedlichen Geweben vom Schwein, was deutliche gewebsspezifische Unterschiede im Modifikationsgrad aufzeigt (die Werte beziehen sich auf 1000 tRNA-Moleküle, die Farben spiegeln den relativen Modifikationsgehalt wider; rot: hoch, grün: niedrig).

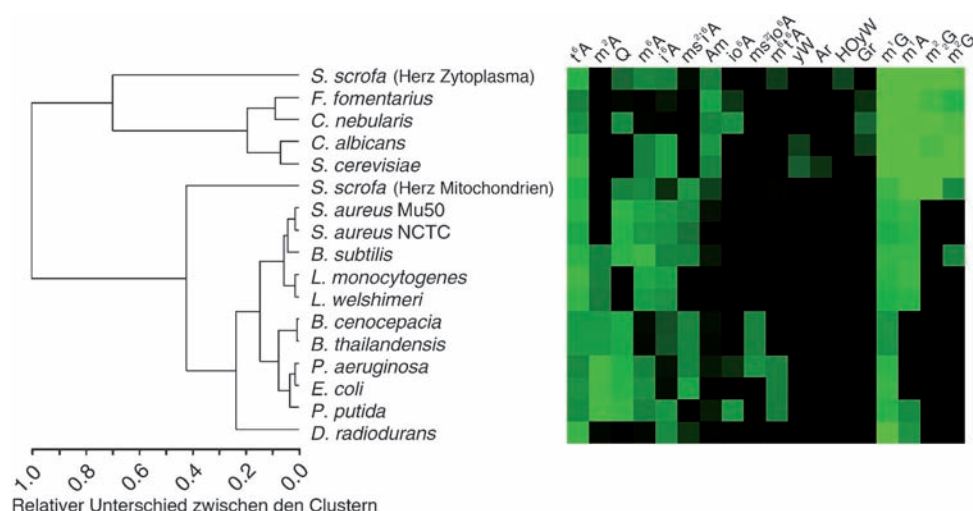


Abbildung 10. „Phylogenetischer“ Stammbaum auf der Grundlage von quantitativen Daten zur tRNA-Modifikation für einige Pro- und Eukaryoten. Die zunehmende Helligkeit des Grüntons steht für höhere Mengen der in der x-Achse aufgeführten Modifikationen in den in der y-Achse aufgeführten Spezies.

ragged-red fibers).^[126,127] Fehlende Modifikationen führen zu gestörter Dekodierung der Leucin- und Lysin-Wobble-Kodons, was wiederum zu Defiziten in der mitochondrialen Translation führt.

Eine andere Wobble-Modifikation, das Queuosin, wurde ebenfalls mit einer Reihe von Erkrankungen in Verbindung gebracht. So ist ein niedriger Q-Gehalt charakteristisch für bestimmte Tumortypen. Beispielsweise enthält Tumorgewebe der Lunge niedrigere Q-Mengen als gesundes Lungengewebe.^[128] Darüber hinaus wurden niedrige Q-Niveaus mit höherer Expressionsrate der tRNA-Guanin-Transglycosylase (TGT) bei Leukämie und Darmkrebszellen in Verbindung gebracht.^[129] Mehrere Krankheiten stehen in Verbindung mit der A-nach-I-Editierung durch RNA-spezifische Adenin-Desaminasen (ADAR). Desaminierung von Adenin führt zu Inosin, das bevorzugt mit Cytosin paart. Die meisten Editierungen wurden für Proteine des zentralen Nervensystems beschrieben. Bedeutende Beispiele sind die Familie der Glutamaterezeptoren und der Serotoninrezeptor 5-HT_{2C}. Zu den Krankheiten, die auf einem Defekt der ADAR-Enzyme beruhen, gehören neurodegenerative Erkrankungen, Diabetes, Krebs und AIDS (siehe den Übersichtsartikel von Gallo und Locatelli).^[130]

Kürzlich erweckten zwei weitere Modifikationen Aufmerksamkeit, nachdem sich ihre biosynthetischen Enzyme als bekannte genetische Faktoren für Fettleibigkeit (FTO) und Typ-2-Diabetes (CDKAL1) herausstellten. Während Allele beider Gene mit hoher statistischer Signifikanz schon früh für die Ausprägung einer Krankheit verantwortlich gemacht wurden,^[131] konnte die Funktion der Proteine erst kürzlich aufgeklärt werden. Die Gruppen um Fontecave und Tomizawa konnten schlüssig beweisen, dass das mit Typ-2-Diabetes (T2D) assoziierte Protein CDKAL1 (CDK5 regulatory subunit associated protein 1-like 1) eine S-Adenosylmethionin(SAM)-abhängige Methylthiotransferase ist, die für die Bildung von ms²t⁶A aus t⁶A in tRNA^{Lys}_{UUU} zuständig ist.^[132] Das Protein ist im endoplasmatischen Retikulum (ER) lokalisiert, und seine Ausschaltung in Mäusen beeinflusst si-

gnifikant die Insulinausschüttung und führt zur Dysfunktion der β -Zellen, wahrscheinlich verursacht durch ein fehlerhaftes Auslesen des Lys_{UUU}-Kodons.^[132b] Interessanterweise neigen CDKAL1-defiziente Mäuse zu erhöhtem Stress bei Ernährung mit hohem Fettgehalt, während das zweite neubewertete Protein FTO (fat mass and obesity associated protein) ein wohlbekannter Faktor bei Fettleibigkeit ist. Beide Gene sind zudem über eine Vielzahl von klinischen Studien zu Typ-2-Diabetes verbunden (siehe <http://string-db.org>). FTO wurde zunächst von Schofield und Mitarbeitern als Fe^{II}/ α -KG-abhängige Dioxygenase identifiziert, die in der Lage ist, die alkylierten DNA-Schäden m³dT, m¹dA und in geringerem Maße auch m³dC in einzelsträngiger DNA zu de-

methylieren.^[133] Später erweiterten He und Mitarbeiter das Substratspektrum auf m³U und RNA.^[134] So war das genaue Substrat von FTO längere Zeit unklar, bis die Gruppe um He kürzlich FTO als RNA-spezifische m⁶A-Demethylase beschrieb.^[135] m⁶A ist eine häufige Modifikation in mRNA, und eine Aktivität von FTO auf m⁶A-enthaltende Prolactin-mRNA konnte gezeigt werden. Die kürzlich publizierte genomweite Karte von m⁶A zeigt eine Anreicherung dieser Modifikation am Transkriptionsstart und nahe der Stopkodons. Es konnte ebenfalls gezeigt werden, dass m⁶A das Spleißen der mRNA beeinflusst.^[117] FTO wurde im Zellkern gefunden und lokalisiert dort in spezifischen Foci, die als potentielle Spleißstellen charakterisiert wurden.^[135]

Es ist in diesem Zusammenhang vielleicht interessant zu bemerken, dass ein Paralog von CDKAL1, CDK5RAP1 (CDK5 regulatory subunit associated protein 1), ein bekannter Repressor der gut untersuchten Kinase CDK5 ist.^[136] CDK5 ist nicht nur ein wichtiger Faktor bei der Regulation der Gehirnentwicklung, sondern auch an der Ausprägung neurogenerativer Erkrankungen wie Alzheimer und der Insulinsekretion beteiligt.^[137] Sein Inhibitor CDK5RAP1 ist das humane Ortholog des bakteriellen MiaB-Proteins und für die Methylthiolierung von i⁶A unter Bildung von ms²i⁶A in mitochondrialer tRNA und anderen RNA-Spezies zuständig.^[138] Alle hier erwähnten Krankheiten sind über die Funktion der Mitochondrien verbunden.^[139] Zusammengefasst ergeben sich Hinweise auf eine Vernetzung der Funktion von RNA-Modifikationen bei der Ausprägung der Zivilisationskrankheiten Adipositas und Typ-2-Diabetes.

2.5.2. RNA-Modifikationen als diagnostische Marker und therapeutische Zielmoleküle

Neben einer direkten Beteiligung von RNA-Modifikationen in Krankheitsbildern haben Methoden zur Quantifizierung spezifischer Modifikationen zur Entwicklung diagnostischer Verfahren geführt. Bereits 1978 stellten Nishimura und Mitarbeiter abnehmende Mengen von Queuosin in

Tumorgewebe fest.^[140] Dieser Fakt wurde später als Biomarker für Lungenkrebs entwickelt.^[141] Das Vorhandensein dieser Modifikation ist zudem für die Virulenz von *Shigella flexneri* verantwortlich, einem Shigellenruhr verursachenden Bakterium. Untersuchungen am Gen des TGT-Enzyms, das für den Austausch von Guanin mit preQ₁ zuständig ist, haben gezeigt, dass Mutationen zu einem signifikanten Verlust der Pathogenität von *S. flexneri* führen.^[142] Dies macht TGT zu einem interessanten Zielmolekül bei der Bekämpfung der Shigellenruhr.^[142a] Es wurden in der Medizinalchemie bereits Anstrengungen zur Inhibition von TGT unternommen.^[143] Insbesondere die Entwicklungen in der Massenspektrometrie zu immer niedrigeren Nachweisgrenzen hat zur Entwicklung von RNA-Modifikationen als Biomarker beigetragen. Die Analyse von Nukleosiden im Urin durch LC- oder CE-gekoppelte Massenspektrometrie wurde für die Diagnose unterschiedlicher Krebsarten genutzt.^[144]

3. Modifizierte Basen in der DNA

Im Vergleich zur großen Vielfalt an Modifikationen in der RNA und der komplexen Biosynthesemaschinerie, die zu ihrer Generierung sowohl einzeln agierende Enzyme als auch große Komplexe wie z.B. die snRNP erfordert, wurden in DNA verhältnismäßig wenige Modifikationen gefunden. Dies hat sicherlich mit der bedeutenden, aber eingegrenzten Funktion der DNA als Träger der genetischen Information zu tun. Als solche liegen die Moleküle der DNA meist in einer Doppelstrangstruktur vor, in der die Basen zum Zentrum der Helix zeigen und Wasserstoffbrücken zu ihren Gegenstücken ausbilden. RNA liegt dagegen meist einzelsträngig und in einer komplexen dreidimensionalen Faltung vor, in der nicht alle Basen Wasserstoffbrücken zu benachbarten Basen ausbilden. Wichtige Modifikationen in der DNA sind sicherlich die DNA-Schäden. Schäden sind modifizierte Basen, die aufgrund der Exposition mit schädlichen Agentien entstehen, seien es UV-Strahlung, reaktive Sauerstoffspezies oder aus der Umwelt, z.B. mit der Nahrung aufgenommene Substanzen wie aromatische Amine. Letztere werden in der Leber zu reaktiven Nitronium-Intermediaten aktiviert, die mit den nukleophilen Gruppen der DNA-Basen zu den sogenannten „bulky adducts“ reagieren können. All diese DNA-Schäden sind mutagen oder cytotoxisch. Daher besitzen Zellen dedizierte Reparatur- und Schadentoleranzsysteme, um den mutagenen oder auch oft letalen Effekten der modifizierten DNA-Basen zu entgegen. DNA-Schäden stellen keine biosynthetisch modifizierten Basen dar und sollen daher in diesem Aufsatz nicht berücksichtigt werden. Interessierte Leser verweisen wir auf weiterführende Literatur zu diesem wichtigen Forschungsgebiet.^[145]

Neben den DNA-Schäden existiert eine kleine Anzahl modifizierter DNA-Basen, die enzymatisch generiert werden (Abbildung 11). In spezifischen Sequenzen bakterieller Genome können Adenine und Cytosine an der N6- bzw. der C5- oder N4-Position methyliert werden. Beide Cytosinderivate, m⁵dC und m⁴dC, schützen die DNA des Bakteriums vor dem Abbau durch eigene Restriktionsenzyme. Diese Endonukleasen werden als Abwehrsystem z.B. gegen Infektion mit

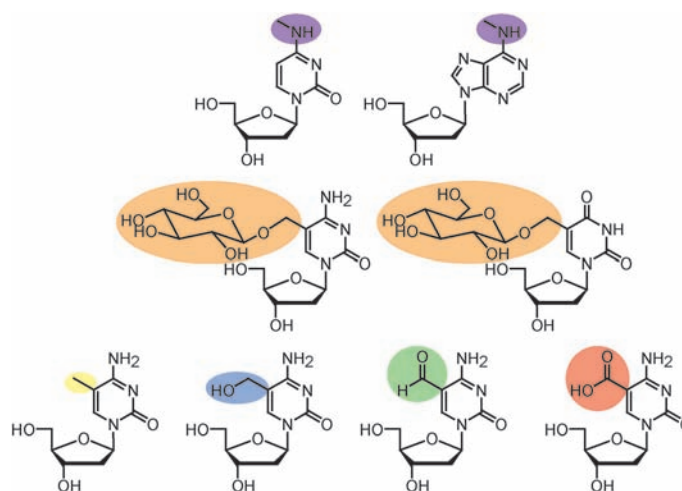


Abbildung 11. Oben: m⁴dC und m⁶dA. Mitte: β -D-glycosylierte Formen von 5-hmC/hm⁵dC (links) und 5-hmU/hm⁵dU (rechts). Unten: Abbildung der vier epigenetischen Basen, die eine zweite Ebene der genetischen Information darstellen. Diese Ebene enthält Informationen über die Genaktivität. In Anlehnung an die RNA-Nomenklatur sollten die modifizierten DNA-Basen mit m⁵dC, hm⁵dC, hm⁵dU, f⁵dC und ca⁵dC abgekürzt werden. In der DNA-Community haben sich jedoch die Abkürzungen 5-mC, 5-hmC, 5-fC und 5-caC oder die noch kürzere Darstellung mC, hmC, fC und caC eingebürgert, weshalb hier auf beide Formen zurückgegriffen wird.

Bakteriophagen, für die Zerstörung von nicht-methylierter Fremd-DNA bereitgehalten. m⁶dA ist zusätzlich zu dieser Funktion an der Regulation von Virulenz, der Reparatur von Fehlpaarungen, dem Timing der DNA-Replikation und der Kontrolle der Genexpression beteiligt.^[146] Weil während der Replikation gebildetes dA zunächst unmethyliert vorliegt, erlaubt es der Zelle für wenige Sekunden bis zur Remethylierung, den neugebildeten DNA-Strang von seinem Templatstrang zu unterscheiden. Diese Zeit erlaubt die Reparatur von Replikationsfehlern. Eben kopierte Replikationsursprünge werden durch die Bindung des SeqA-Proteins vor einer Remethylierung geschützt und bleiben daher länger im halbmethylierten Zustand. Weil halbmethylierte Replikationsursprünge inaktiv sind, wird die weitere Zellteilung unterbunden, bis die Remethylierung vollständig abgeschlossen ist.^[147]

Die modifizierten Basen 5-Hydroxymethyluracil (hmU/hm⁵dU, Nukleosid J) und 5-Hydroxymethylcytosin (hmC/hm⁵dC) wurden in der DNA der geradzahigen T-Bakteriophagen (hm⁵dC) und in dem parasitären Flagellaten *Trypanosoma brucei* (hm⁵dU) gefunden. In beiden Fällen wird die Hydroxymethylgruppe in das Triphosphat vor der Inkorporation in die DNA eingeführt und anschließend glycosyliert. Die Funktion dieser Basen ist der Schutz gegen Abbau durch die Restriktionsenzyme des Wirtes.^[13b] hmU steht zudem in Verdacht, die Transkription zu hemmen.

Die am besten untersuchte, nicht-kanonische Nukleobase ist jedoch das 5-Methyldeoxycytosin (m⁵dC). Es wird als fünfte Hauptkomponente des DNA-Moleküls bezeichnet. Die Existenz dieser DNA-Base, die genau eine Methylgruppe am C5 des Cytosin trägt, wurde bereits 1925 von Johnson und Cokel postuliert, die die Base im Tuberkelbazillus entdeck-

ten.^[148] Dieser Befund, zunächst nur als pure Spekulation abgetan, wurde 24 Jahre später von Wyatt bestätigt, der bewies, dass m⁵dC tatsächlich Bestandteil des genetischen Materials ist.^[149] Es gilt heute als bestätigt, dass diese Base an der Regulation der Transkription beteiligt ist. Während die Sequenz der DNA unverändert bleibt, kann die Transkriptionsaktivität eines Gens über die Methylierung der Cytosin-Basen beeinflusst werden, was das Forschungsgebiet der Epigenetik definiert. Damit ist m⁵dC in Pflanzen, Pilzen und Bakterien eine epigenetische Modifikation zur Kontrolle der Genaktivität, während es in Bakterien vor allem dem Schutz vor Infektionen durch Bakteriophagen dient (siehe oben).

3.1. Die Entdeckung der modifizierten DNA-Basen 5-Hydroxymethylcytosin, 5-Formylcytosin und 5-Carboxycytosin in höheren Organismen

Das Forschungsgebiet um das m⁵dC ist so gewaltig, dass es hier nicht abgehandelt werden kann. Stattdessen konzentrieren wir uns auf den Entdeckungsprozess, der kürzlich zur Beobachtung weiterer modifizierter Cytosinderivate in der DNA geführt hat.

2009 zeigten zwei Forschergruppen unabhängig voneinander die Existenz von 5-Hydroxymethylcytosin (hmC/hm⁵dC) in DNA aus Stammzellen der Maus und in Neuronen, zusätzlich zu dem bekannten 5-Methylcytosin (mC/m⁵dC).^[150] Tatsächlich handelte es sich hier um eine Wiederentdeckung, da hmC bereits 1972 von Yura und Mitarbeitern als Bestandteil tierischer DNA beschrieben wurde.^[151] In darauffolgenden Arbeiten konnte gezeigt werden, dass 5-Hydroxymethylcytosin besonders häufig in der DNA des zentralen Nervensystems (ZNS) vorkommt.^[152] Diese Arbeiten wurden durch die Synthese von isotopenmarkiertem mC und hmC und deren Nutzung in der quantitativen LC-Massenspektrometrie ermöglicht, wie bereits für die Quantifizierung von RNA-Modifikationen beschrieben.^[122a] Mit dieser Methode konnte gezeigt werden, dass der Gehalt an mC in allen Geweben der Maus bei konstant etwa 4,5 % aller Cytosine liegt, während die Mengen an hmC stark variierten. Wir wissen heute, dass diese Base in allen Geweben, jedoch in stark unterschiedlichen Mengen vorkommt. Dabei ist hmC vor allem im Gehirn stark angereichert, besonders in Regionen mit höherer kognitiver Funktion. In Tumoren des Gehirns sind diese Werte deutlich reduziert.^[153] Neben neuronalem Gewebe wurden signifikante Mengen von hmC auch in embryonalen Stammzellen gefunden. Diese Beobachtung führte zu der Idee, dass die hmC-Base eine Rolle bei der epigenetischen Programmierung des Genoms und insbesondere bei dem rätselhaften Prozess der aktiven Demethylierung spielt.^[154]

Anfängliche Versuche, die wahrscheinlichsten Oxidationsprodukte des hmC, nämlich 5-Formylcytosin (fC/f⁵dC) und 5-Carboxycytosin (caC/ca⁵dC), in Hirngewebe zu detektieren, schlugen fehl.^[155] Es wurde gezeigt, dass trotz großer Mengen hmC weder fC noch caC in auswertbaren Mengen im zentralen Nervensystem (ZNS) vorlag. Im weiteren Verlauf dieser Untersuchungen wurde fC kürzlich in der DNA von embryonalen Stammzellen der Maus gefunden.^[156] Darüber

hinaus wurde hier kurz darauf auch caC nachgewiesen.^[157] Es wurde berichtet, dass caC in größeren Mengen in Stammzellen auftritt, deren TDG-Enzym deletiert wurde. Diese Reparaturglycosylase ist bekannt, die glycosidische Bindung zwischen Desoxyribose und caC zu spalten, um diese Modifikation zu reparieren.^[157b] Diese Ergebnisse führten zu dem Schluss, dass caC über die Basenexzisionsreparatur aus der DNA entfernt wird. Eine Kristallstruktur von TDG mit gebundener, caC-haltiger DNA stützt diese Hypothese.^[158]

3.2. Die Rolle der modifizierten DNA-Basen 5-Hydroxymethylcytosin, 5-Formylcytosin und 5-Carboxycytosin

Obwohl die exakte Menge und Verteilung von fC und caC in Stammzellen und somatischen Geweben noch nicht klar ist, besteht die Frage nach ihrer Funktion. Die wichtige Base mC ist an einer Vielzahl biologischer Prozesse, wie Regulation der Genexpression, der genetischen Prägung, der Inaktivierung des X-Chromosoms und der Unterdrückung transposabler Elemente, beteiligt.^[159] Wenn ein Promotor signifikante Mengen an mC ausweist, ist die Expression des zugehörigen Gens meist unterdrückt. Es wird aktuell angenommen, dass hmC eine vergleichbare epigenetische Funktion einnimmt. Proteine, die für ihre Bindung an mC bekannt sind, können nicht mit hmC interagieren, was die transkriptionelle Aktivität der hydroxymethylierten Gene gegenüber der methylierten Gene beeinflussen könnte. Somit stellen hydroxymethylierte Promotoren womöglich bereits eine Reversion der durch Methylierung unterdrückten Genexpression dar.^[160] Darüber hinaus sind die Basen hmC, fC und caC womöglich Intermediate auf dem Weg, die epigenetische Base mC aus dem Genom zu entfernen und durch unmodifiziertes C zu ersetzen.^[161] Dies würde den Zellen eine flexible Reprogrammierung des epigenetischen Status eines Gens erlauben. Der tatsächliche Mechanismus hierfür ist bislang unklar, könnte jedoch prinzipiell auf zwei unterschiedliche Arten geschehen (Abbildung 12). Der erste Mechanismus basiert auf der Basenexzisionsreparatur.^[162] Hier geht man davon aus, dass die modifizierten Basen fC und caC von speziellen Glycosylasen, wie der Thymin-DNA-Glycosylase (TDG), erkannt und durch Spaltung der glycosidischen Bindung zwischen Zucker und Base herausgeschnitten werden. Dabei bleibt eine abasische Stelle zurück (Abbildung 12), die in der Folge mit unmodifiziertem dC repariert wird,^[163] was in der Konsequenz zu einem Austausch der epigenetischen Markierung mC mit unmodifiziertem dC führt. Ein ähnlicher Mechanismus geht hierbei von hmC aus. Allerdings hat sich gezeigt, dass hmC ein schlechtes Substrat für TDG ist. Enzyme wie AID/APOBEC können jedoch hmC zu hmU desaminieren, was wiederum ein gutes Substrat für die DNA-Glycosylasen SMUG1 und TDG ist. Somit wird ein Austausch von hmC nach C über den Umweg der Desaminierung möglich.^[164]

Der zweite, chemisch attraktive Weg zur aktiven Demethylierung wäre die Decarboxylierung von caC oder alternativ die Dehydroxymethylierung von hmC oder Deformylierung von fC. In diesem Szenario werden zur Demethylierung

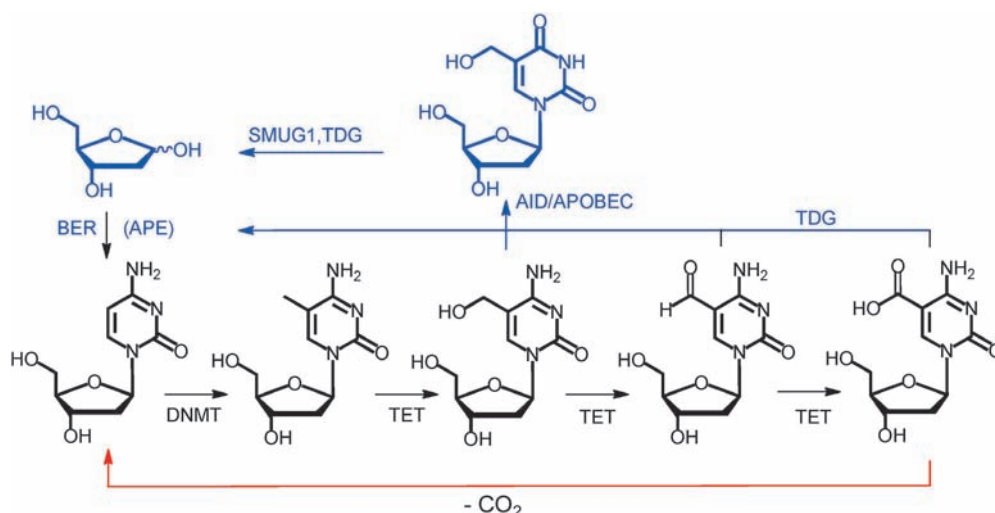


Abbildung 12. Mögliche Wege zur aktiven Demethylierung, basierend auf der Oxidation von mC. In Blau der Glycosylase/BER-Mechanismus, der zunächst durch DNA-Glycosylasen eine abasische Stelle erzeugt, diese durch A/P-Endonuklease (APE) in einen Strangbruch überführt und mittels Polymerase und Ligase die Lücke mit dC erneut schließt. hmC ist kein Substrat der TDG. Hier erfolgt zuerst eine Desaminierung zum 5-Hydroxymethyluracil (hmU) durch Enzyme der AID/APOBEC-Familie. Der mutmaßliche Decarboxylierungsweg ist in Rot gezeigt.

vorgesehene mC-Basen spezifisch schrittweise bis zum caC enzymatisch oxidiert, um sie anschließend einem der Prozesse zu unterziehen. An diesem Punkt scheint die Decarboxylierung des caC chemisch am attraktivsten, da ähnliche Reaktionen bereits von der Orotat- und der Isoorotatdecarboxylase bekannt sind, die an der Bildung von Uracil beteiligt sind.^[165]

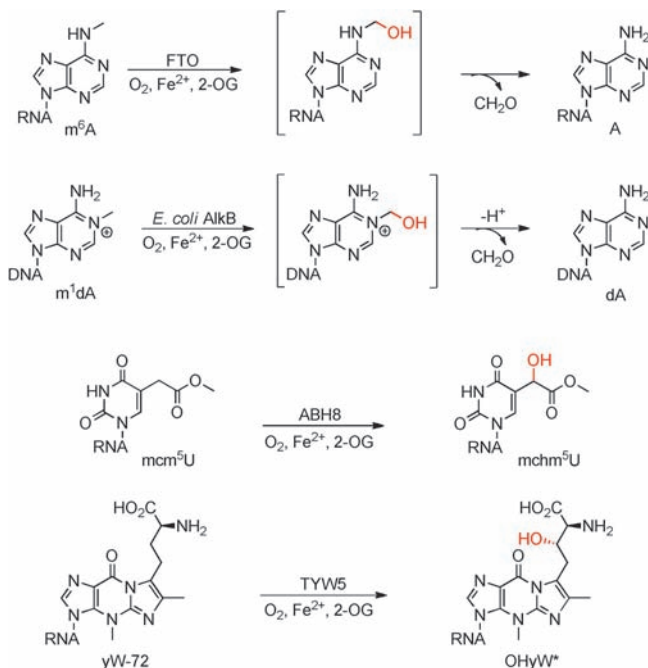
Obwohl bislang keine spezifische caC-Decarboxylase entdeckt wurde, konnte kürzlich gezeigt werden, dass Extrakte von Stammzellen die Fähigkeit zur Decarboxylierung von caC zu C besitzen und dass diese Aktivität chemisch durch Sättigung der C5-C6-Doppelbindung ermöglicht wird,^[166] wie bereits vorher postuliert.^[167]

Eine wichtige Entdeckung auf diesem Gebiet sind die Enzyme zur schrittweisen Oxidation von mC zu hmC, fC und caC. Dies sind die „ten-eleven-translocation“-Enzyme (TET1, 2 und 3). Diese Proteine sind α -Ketoglutarat-Dioxygenasen, die molekularen Sauerstoff in einem Fe^{II}-haltigen aktiven Zentrum für die radikalische Oxidation der Methylgruppe, zunächst in hmC, nutzen. Diese Enzyme sind ebenso in der Lage, hmC zu fC und weiter zu caC zu oxidieren.^[157,168] Letztere Eigenschaft wurde kürzlich zur basengenauen Sequenzierung von hmC entwickelt, für die hmC mithilfe von rekombinantem TET-Enzym zum caC oxidiert wurde, das ein unterschiedliches Verhalten in der Bisulfidsequenzierung zeigt.^[168] Eine alternative Methode mit demselben Ziel nutzt die chemische Oxidation von hmC zum fC mit Perruthenat, gefolgt von Bisulfidsequenzierung.^[169]

In diesem Aufsatz wurde die interessante Klasse der Fe^{II}/ α -KG-abhängigen Dioxygenasen bereits mehrere Male in unterschiedlichem Zusammenhang erwähnt. Diese Enzyme sind im Augenblick von großem Interesse, da sie in der Lage sind, die Methylgruppen von Nukleobasen zu funktionalisieren und zu entfernen. Die Entfernungreaktion kehrt den Effekt dieser speziellen Modifikationen um. Das FTO-Pro-

tein und die DNA-Reparaturenzyme der ALKB-Familie entfernen z. B. die Methylgruppen von m⁶A oder m¹dA. Sie oxidieren die Methylgruppe zu einem Hydroxymethylintermediat (Schema 1). Weil diese Methylgruppe sich auf einem Heteroatom befindet, entsteht eine leicht hydrolysierbare „Acetal“-Struktur, die letztlich zum vollständigen Verlust der Methylgruppe führt. Demgegenüber oxidieren die Enzyme ABH8 und TYW5 Methylgruppen zu Hydroxymethylgruppen in mchm⁵U und OHyW, welche nicht eliminiert werden, da sie mit einem weiteren Kohlenstoffatom verknüpft sind.

In dieser Situation ist die Hydroxymethylgruppe der Endpunkt der Biosynthese. Die kürzlich entdeckten TET-Enzyme haben nun die anspruchsvolle Aufgabe, die C5-Methylgruppe des Methylcytosins vollständig zu entfernen. Dies geschieht durch Aufoxidation zu fC und caC. Diese Basen sind bekannt, von der TDG-Glycosylase durch Spaltung der glycosidischen Bindung entfernt zu werden, was die schwierige Spaltung einer C-C-Bindung umgeht. Es sind nun Arbeiten gefordert, die die Entschlüsselung der Aktivität zum Ziel haben, die zur



Schema 1. Oxidationsreaktion verschiedener Nukleoside durch Mitglieder der Familie der Fe^{II}/ α -KG-Dioxygenasen.

Decarboxylierung führt, sei es ein enzymatischer oder anderweitig katalysierter Prozess.

4. Zusammenfassung

Es ist Lehrbuchwissen, dass die DNA und RNA aus jeweils vier kanonischen Nukleobasen zusammengesetzt ist. Darüber hinaus enthalten beide Polymere des Lebens eine Vielzahl modifizierte Basen, die ihren chemischen Informationsgehalt vergrößern. RNA ist dabei besonders reich an Modifikationen, was die vielfältigen Aufgaben der RNA-Moleküle in der Zelle widerspiegelt. RNA ist kein Informationsspeicher, sondern übersetzt die Information in eine Aminosäuresequenz. Zudem besitzt RNA regulatorische und gelegentlich auch katalytische Funktion, und sie ist in Form von Templatsträngen maßgeblich an der Modifikation ihrer selbst beteiligt. Die Vielfalt und große Zahl an Modifikationen in der RNA ist offensichtlich das Resultat der Anpassung an die komplexen und vielfältigen Aufgaben von RNA. Kürzlich lernten wir auch, dass die chemische Diversität der DNA größer ist als zunächst erwartet. Neben den vier kanonischen Basen und den lange bekannten Modifikationen m⁴dC, m⁵dC, m⁶dA, hm⁵dU/hmU und hm⁵dC/hmC wissen wir nun, dass DNA insgesamt vier epigenetische Basen (mC, hmC, fC und caC) enthält, von denen fC und caC eben erst entdeckt wurden. Wiederum dienen diese Modifikationen regulatorischen, epigenetischen Zwecken. Während alle Zellen, so unterschiedlich sie auch sein mögen, das gleiche genetische Material enthalten, so erfordert die Erfüllung der großen Zahl unterschiedlicher Aufgaben und Funktionen innerhalb eines komplexen höheren Organismus die spezifische Aktivierung und Inaktivierung Zelltyp-spezifischer Gene. Die Regulation der zugrundeliegenden Unterdrückungs- und Aktivierungsprozesse erfordert eine zweite, epigenetische Informationsebene, die augenscheinlich direkt mit erhöhter chemischer Diversität verbunden ist. Diese Diversität wird von den epigenetischen Basen mC, hmC, fC und caC bereitgestellt. Zusammenfassend kann man sagen, dass DNA und RNA zusätzlich zu ihrer ersten Informationsebene, bereitgestellt durch die Abfolge der kanonischen Watson-Crick-Basen, eine zweite Ebene auf der Basis chemischer Komplexität enthalten, die durch die gezielte Einführung von Modifikationen in DNA und RNA aufgebaut wird. Die Wissenschaft wird sich sicherlich in den kommenden Jahren vermehrt der Entschlüsselung der Biochemie dieser Substanzen, ihrer räumlichen Verteilung (Sequenz) sowie der Orchestrierung ihrer Einführung in DNA und RNA widmen. Die Forschung an DNA- und RNA-Modifikationen ist in den Fachgebieten der chemischen Synthese, der Biochemie und der Zellbiologie beheimatet, was den synthetisch orientierten Forschern der chemischen Biologie Gelegenheit gibt, zu dem Verständnis, wie Nukleosidmodifikationen das Leben beeinflussen, maßgeblich beizutragen.

Wir danken dem Exzellenzcluster Center for Integrated Protein Science (CiPS^M), dem SFB 749, dem SFB 646 und der DFG im Rahmen des Normalverfahrensprojekts (CA 275/8-4)

für die finanzielle Unterstützung. D.P. dankt der Alexander-von-Humboldt-Stiftung für ein Postdoktorandenstipendium.

Eingegangen am 13. Februar 2012,

veränderte Fassung am 7. Mai 2012

Online veröffentlicht am 28. Juni 2012

- [1] a) R. Dahm, *Hum. Genet.* **2008**, *122*, 565–581; b) R. Olby, *Nature* **2003**, *421*, 402–405.
- [2] E. Chargaff, *Experientia* **1950**, *6*, 201–209.
- [3] P. A. Levene, *J. Biol. Chem.* **1919**, *40*, 415–424.
- [4] a) A. D. Hershey, M. Chase, *J. Gen. Physiol.* **1952**, *36*, 39–56; b) M. McCarty, *Nature* **2003**, *421*, 406; c) O. T. Avery, C. M. MacLeod, M. McCarty, *J. Exp. Med.* **1944**, *79*, 137–158.
- [5] a) R. E. Franklin, R. G. Gosling, *Nature* **1953**, *171*, 740–741; b) M. H. Wilkins, A. R. Stokes, H. R. Wilson, *Nature* **1953**, *171*, 738–740.
- [6] a) J. D. Watson, F. H. Crick, *Nature* **1953**, *171*, 737–738; b) J. D. Watson, F. H. Crick, *Nature* **1953**, *171*, 964–967.
- [7] a) M. Guttman, J. L. Rinn, *Nature* **2012**, *482*, 339–346; b) L. He, G. J. Hannon, *Nat. Rev. Genet.* **2004**, *5*, 522–531; c) E. D. Wiklund, J. Kjems, S. J. Clark, *Epigenomics* **2010**, *2*, 823–840.
- [8] a) T. R. Cech, *Gene* **1993**, *135*, 33–36; b) S. Altman, *Biosci. Rep.* **1990**, *10*, 317–337.
- [9] J. S. Mattick, *EMBO J.* **2012**, *31*, 515–516.
- [10] S. Y. Ng, R. Johnson, L. W. Stanton, *EMBO J.* **2011**, *31*, 522–533.
- [11] a) M. E. Dinger, P. P. Amaral, T. R. Mercer, K. C. Pang, S. J. Bruce, B. B. Gardiner, M. E. Askarian-Amiri, K. Ru, G. Soldá, C. Simons, S. M. Sunkin, M. L. Crowe, S. M. Grimmond, A. C. Perkins, J. S. Mattick, *Genome Res.* **2008**, *18*, 1433–1445; b) M. Guttman, J. Donaghey, B. W. Carey, M. Garber, J. K. Grenier, G. Munson, G. Young, A. B. Lucas, R. Ach, L. Bruhn, X. Yang, I. Amit, A. Meissner, A. Regev, J. L. Rinn, D. E. Root, E. S. Lander, *Nature* **2011**, *477*, 295–300.
- [12] T. R. Mercer, D. J. Gerhardt, M. E. Dinger, J. Crawford, C. Trapnell, J. A. Jeddell, J. S. Mattick, J. L. Rinn, *Nat. Biotechnol.* **2011**, *30*, 99–104.
- [13] a) M. Ehrlich, G. G. Wilson, K. C. Kuo, C. W. Gehrke, *J. Bacteriol.* **1987**, *169*, 939–943; b) J. H. Gommers-Ampt, F. Van Leeuwen, A. L. de Beer, J. F. Vliegthart, M. Dizdaroglu, J. A. Kowalak, P. F. Crain, P. Borst, *Cell* **1993**, *75*, 1129–1136.
- [14] a) P. F. Agris, F. A. P. Vendeix, W. D. Graham, *J. Mol. Biol.* **2007**, *366*, 1–13; b) H. Grosjean, V. de Crécy-Lagard, C. Marck, *FEBS Lett.* **2010**, *584*, 252–264.
- [15] P. F. Agris, *EMBO Rep.* **2008**, *9*, 629–635.
- [16] Y. Motorin, M. Helm, *Biochemistry* **2010**, *49*, 4934–4944.
- [17] a) K. Phelps, A. Morris, P. A. Beal, *ACS Chem. Biol.* **2012**, *7*, 100–109; b) D. Iwata-Reuyl, *Curr. Opin. Chem. Biol.* **2008**, *12*, 126–133; c) E. M. Phizicky, A. K. Hopper, *Genes Dev.* **2010**, *24*, 1832–1860.
- [18] W. A. Cantara, P. F. Crain, J. Rozenski, J. A. McCloskey, K. A. Harris, X. Zhang, F. A. P. Vendeix, D. Fabris, P. F. Agris, *Nucleic Acids Res.* **2011**, *39*, D195–D201.
- [19] M. Charette, M. W. Gray, *IUBMB Life* **2000**, *49*, 341–351.
- [20] E. M. Phizicky, J. D. Alfonzo, *FEBS Lett.* **2010**, *584*, 265–271.
- [21] A. Czerwonec, S. Dunin-Horkawicz, E. Purta, K. H. Kamin-ska, J. M. Kasprzak, J. M. Bujnicki, H. Grosjean, K. Rother, *Nucleic Acids Res.* **2009**, *37*, D118–121.
- [22] Y. Motorin, M. Helm, *Wiley Interdiscip. Rev. RNA* **2011**, *2*, 611–631.
- [23] a) Y. Ikeuchi, S. Kimura, T. Numata, D. Nakamura, T. Yokogawa, T. Ogata, T. Wada, T. Suzuki, T. Suzuki, *Nat. Chem. Biol.* **2010**, *6*, 277–282; b) D. Mandal, C. Köhrer, D. Su, S. P. Russell, K. Krivos, C. M. Castleberry, P. Blum, P. A. Limbach, D. Söll,

- U. L. RajBhandary, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 2872–2877.
- [24] a) T. Muramatsu, K. Nishikawa, F. Nemoto, Y. Kuchino, S. Nishimura, T. Miyazawa, S. Yokoyama, *Nature* **1988**, *336*, 179–181; b) T. Muramatsu, S. Yokoyama, N. Horie, A. Matsuda, T. Ueda, Z. Yamaizumi, Y. Kuchino, S. Nishimura, T. Miyazawa, *J. Biol. Chem.* **1988**, *263*, 9261–9267.
- [25] a) J. A. Kowalak, E. Bruenger, J. A. McCloskey, *J. Biol. Chem.* **1995**, *270*, 17758–17764; b) A. M. Giessing, S. S. Jensen, A. Rasmussen, L. H. Hansen, A. Gondela, K. Long, B. Vester, F. Kirpekar, *RNA* **2009**, *15*, 327–336.
- [26] a) K. S. Long, J. Poehlsaard, C. Kehrenberg, S. Schwarz, B. Vester, *Antimicrob. Agents Chemother.* **2006**, *50*, 2500–2505; b) D. N. Wilson, *Ann. N. Y. Acad. Sci.* **2011**, *1241*, 1–16.
- [27] A. A. Van Aerschot, P. Mamos, N. J. Weyns, S. Ikeda, E. De Clercq, P. A. Herdewijn, *J. Med. Chem.* **1993**, *36*, 2938–2942.
- [28] a) C. Laezza, M. Notarnicola, M. G. Caruso, C. Messa, M. Macchia, S. Bertini, F. Minutolo, G. Portella, L. Fiorentino, S. Stingo, M. Bifulco, *FASEB J.* **2006**, *20*, 412–418; b) H. Sakakibara, *Ann. Rev. Plant Biol.* **2006**, *57*, 431–449.
- [29] T. Suzuki, T. Wada, K. Saigo, K. Watanabe, *Nucleic Acids Res. Suppl.* **2001**, *1*, 257–258.
- [30] a) Y. G. Chen, W. E. Kowtoniuk, I. Agarwal, Y. Shen, D. R. Liu, *Nat. Chem. Biol.* **2009**, *5*, 879–881; b) W. E. Kowtoniuk, Y. Shen, J. M. Heemstra, I. Agarwal, D. R. Liu, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 7768–7773.
- [31] H. Kasai, K. Nakanishi, R. D. Macfarlane, D. F. Torgerson, Z. Ohashi, J. A. McCloskey, H. J. Gross, S. Nishimura, *J. Am. Chem. Soc.* **1976**, *98*, 5044–5046.
- [32] D. W. Phillipson, C. G. Edmonds, P. F. Crain, D. L. Smith, D. R. Davis, J. A. McCloskey, *J. Biol. Chem.* **1987**, *262*, 3462–3471.
- [33] a) J. C. Salazar, A. Ambrogelly, P. F. Crain, J. A. McCloskey, D. Soll, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7536–7541; b) M. Blaise, H. D. Becker, J. Lapointe, C. Cambillau, R. Giege, D. Kern, *Biochimie* **2005**, *87*, 847–861.
- [34] D. Iwata-Reuyl, *Bioorg. Chem.* **2003**, *31*, 24–43.
- [35] N. Okada, S. Nishimura, *Nucleic Acids Res.* **1977**, *4*, 2931–2938.
- [36] F. Harada, S. Nishimura, *Biochemistry* **1972**, *11*, 301–308.
- [37] a) J. M. Gregson, P. F. Crain, C. G. Edmonds, R. Gupta, T. Hashizume, D. W. Phillipson, J. A. McCloskey, *J. Biol. Chem.* **1993**, *268*, 10076–10086; b) M. Sprinzl, T. Hartmann, J. Weber, J. Blank, R. Zeidler, *Nucleic Acids Res.* **1989**, *17*, R1–R12; c) T. Ohgi, T. Kondo, T. Goto, *J. Am. Chem. Soc.* **1979**, *101*, 3629–3633; d) F. Klepper, E. M. Jahn, V. Hickmann, T. Carell, *Angew. Chem.* **2007**, *119*, 2377–2379; *Angew. Chem. Int. Ed.* **2007**, *46*, 2325–2327; e) T. Brückl, F. Klepper, K. Gutmiedl, T. Carell, *Org. Biomol. Chem.* **2007**, *5*, 3821–3825.
- [38] N. Shindo-Okada, N. Okada, T. Ohgi, T. Goto, S. Nishimura, *Biochemistry* **1980**, *19*, 395–400.
- [39] a) U. Gündüz, J. R. Katze, *Biochem. Biophys. Res. Commun.* **1982**, *109*, 159–167; b) U. Gunduz, J. R. Katze, *J. Biol. Chem.* **1984**, *259*, 1110–1113.
- [40] a) G. A. Garcia, J. D. Kittendorf, *Bioorg. Chem.* **2005**, *33*, 229–251; b) N. Okada, S. Noguchi, H. Kasai, N. Shindo-Okada, T. Ohgi, T. Goto, S. Nishimura, *J. Biol. Chem.* **1979**, *254*, 3067–3073; c) B. Stengl, K. Reuter, G. Klebe, *ChemBioChem* **2005**, *6*, 1926–1939.
- [41] Y. Kuchino, H. Kasai, K. Nihei, S. Nishimura, *Nucleic Acids Res.* **1976**, *3*, 393–398.
- [42] a) B. W. Lee, S. G. Van Lanen, D. Iwata-Reuyl, *Biochemistry* **2007**, *46*, 12844–12854; b) R. M. McCarty, A. Somogyi, G. Lin, N. E. Jacobsen, V. Bandarian, *Biochemistry* **2009**, *48*, 3847–3852.
- [43] R. K. Slany, M. Bosl, P. F. Crain, H. Kersten, *Biochemistry* **1993**, *32*, 7811–7817.
- [44] B. Frey, J. McCloskey, W. Kersten, H. Kersten, *J. Bacteriol.* **1988**, *170*, 2078–2082.
- [45] Z. D. Miles, R. M. McCarty, G. Molnar, V. Bandarian, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 7368–7372.
- [46] M. Vinayak, C. Pathak, *Biosci. Rep.* **2010**, *30*, 135–148.
- [47] a) J. M. Skuzeski, L. M. Nichols, R. F. Gesteland, J. F. Atkins, *J. Mol. Biol.* **1991**, *218*, 365–373; b) K. Zerfass, H. Beier, *Nucleic Acids Res.* **1992**, *20*, 5911–5918.
- [48] F. Meier, B. Suter, H. Grosjean, G. Keith, E. Kubli, *EMBO J.* **1985**, *4*, 823–827.
- [49] M. Bienz, E. Kubli, *Nature* **1981**, *294*, 188–190.
- [50] V. H. Cowling, *Biochem. J.* **2010**, *425*, 295–302.
- [51] a) R. Reddy, R. Singh, S. Shimba, *Pharmacol. Ther.* **1992**, *54*, 249–267; b) F. Rottman, A. J. Shatkin, R. P. Perry, *Cell* **1974**, *3*, 197–199.
- [52] S. Farajollahi, S. Maas, *Trends Genet.* **2010**, *26*, 221–230.
- [53] R. Giegé, M. Sissler, C. Florentz, *Nucleic Acids Res.* **1998**, *26*, 5017–5035.
- [54] J. Urbonavicius, Q. Qian, J. M. B. Durand, T. G. Hagervall, G. R. Bjork, *EMBO J.* **2001**, *20*, 4863–4873.
- [55] a) N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz, *Science* **2000**, *289*, 905–920; b) F. Schlutzen, A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi, A. Yonath, *Cell* **2000**, *102*, 615–623; c) V. Ramakrishnan, *Angew. Chem.* **2010**, *122*, 4454–4481; *Angew. Chem. Int. Ed.* **2010**, *49*, 4355–4380.
- [56] a) F. V. Murphy, V. Ramakrishnan, A. Malkiewicz, P. F. Agris, *Nat. Struct. Mol. Biol.* **2004**, *11*, 1186–1191; b) L. B. Jenner, N. Demeshkina, G. Yusupova, M. Yusupov, *Nat. Struct. Mol. Biol.* **2010**, *17*, 555–560.
- [57] a) J. W. Stuart, Z. Gdaniec, R. Guenther, M. Marszalek, E. Sochacka, A. Malkiewicz, P. F. Agris, *Biochemistry* **2000**, *39*, 13396–13404; b) J. W. Stuart, K. M. Koshlap, R. Guenther, P. F. Agris, *J. Mol. Biol.* **2003**, *334*, 901–918; c) V. Dao, R. Guenther, A. Malkiewicz, B. Nawrot, E. Sochacka, A. Kraszewski, J. Jankowska, K. Everett, P. F. Agris, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 2125–2129; d) P. C. Durant, A. C. Bajji, M. Sundaram, R. K. Kumar, D. R. Davis, *Biochemistry* **2005**, *44*, 8078–8089.
- [58] A. Weixlbaumer, F. V. Murphy, A. Dziergowska, A. Malkiewicz, F. A. P. Vendeix, P. F. Agris, V. Ramakrishnan, *Nat. Struct. Mol. Biol.* **2007**, *14*, 498–502.
- [59] a) J. R. Sampson, O. C. Uhlenbeck, *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 1033–1037; b) K. B. Hall, J. R. Sampson, O. C. Uhlenbeck, A. G. Redfield, *Biochemistry* **1989**, *28*, 5794–5801; c) V. Perret, A. Garcia, J. Puglisi, H. Grosjean, J. P. Ebel, C. Florentz, R. Giegé, *Biochimie* **1990**, *72*, 735–743; d) P. Davanloo, M. Sprinzl, K. Watanabe, M. Albani, H. Kersten, *Nucleic Acids Res.* **1979**, *6*, 1571–1581.
- [60] a) A. Alexandrov, I. Chernyakov, W. Gu, S. L. Hiley, T. R. Hughes, E. J. Grayhack, E. M. Phizicky, *Mol. Cell* **2006**, *21*, 87–96; b) L. Kotelawala, E. J. Grayhack, E. M. Phizicky, *RNA* **2008**, *14*, 158–169.
- [61] P. C. Durant, D. R. Davis, *J. Mol. Biol.* **1999**, *285*, 115–131.
- [62] M. L. Wilkinson, S. M. Crary, J. E. Jackman, E. J. Grayhack, E. M. Phizicky, *RNA* **2007**, *13*, 404–413.
- [63] M. Helm, R. Giegé, C. Florentz, *Biochemistry* **1999**, *38*, 13338–13346.
- [64] a) E. Madore, C. Florentz, R. Giege, S.-i. Sekine, S. Yokoyama, J. Lapointe, *Eur. J. Biochem.* **1999**, *266*, 1128–1135; b) L. A. Sylvers, K. C. Rogers, M. Shimizu, E. Ohtsuka, D. Soll, *Biochemistry* **1993**, *32*, 3836–3841.
- [65] J. Pütz, C. Florentz, F. Benseler, R. Giege, *Nat. Struct. Biol.* **1994**, *1*, 580–582.
- [66] S. Kiesewetter, G. Ott, M. Sprinzl, *Nucleic Acids Res.* **1990**, *18*, 4677–4681.
- [67] R. Basavappa, P. B. Sigler, *EMBO J.* **1991**, *10*, 3105–3111.

- [68] C. Föhrster, K. Chakraborty, M. Sprinzl, *Nucleic Acids Res.* **1993**, *21*, 5679–5683.
- [69] S. U. Åström, A. S. Byström, *Cell* **1994**, *79*, 535–546.
- [70] V. I. Lim, *J. Mol. Biol.* **1994**, *240*, 8–19.
- [71] a) E. M. Gustilo, F. A. P. Vendeix, P. F. Agris, *Curr. Opin. Microbiol.* **2008**, *11*, 134–140; b) M. Yarus, *Science* **1982**, *218*, 646–652.
- [72] a) M. Srinivasan, P. Mehta, Y. Yu, E. Prugar, E. V. Koonin, A. W. Karzai, R. Sternglanz, *EMBO J.* **2011**, *30*, 873–881; b) B. El Yacoubi, B. Lyons, Y. Cruz, R. Reddy, B. Nordin, F. Agnelli, J. R. Williamson, P. Schimmel, M. A. Swairjo, V. de Crécy-Lagard, *Nucleic Acids Res.* **2009**, *37*, 2894–2909.
- [73] a) D. Fu, J. A. N. Brophy, C. T. Y. Chan, K. A. Atmore, U. Begley, R. S. Paules, P. C. Dedon, T. J. Begley, L. D. Samson, *Mol. Cell. Biol.* **2010**, *30*, 2449–2459; b) Y. Fu, Q. Dai, W. Zhang, J. Ren, T. Pan, C. He, *Angew. Chem.* **2010**, *122*, 9069–9072; *Angew. Chem. Int. Ed.* **2010**, *49*, 8885–8888; c) E. van den Born, C. B. Vågbo, L. Songe-Møller, V. Leihne, G. F. Lien, G. Leszczynska, A. Malkiewicz, H. E. Krokan, F. Kirpekar, A. Klungland, P. Ø. Falnes, *Nat. Commun.* **2011**, *2*, 172.
- [74] a) J. P. Aris, G. Blobel, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 931–935; b) R. P. Jansen, E. C. Hurt, H. Kern, H. Lehtonen, M. Carmo-Fonseca, B. Lapeyre, D. Tollervey, *J. Cell Biol.* **1991**, *113*, 715–729.
- [75] N. S. Heiss, S. W. Knight, T. J. Vulliamy, S. M. Klauck, S. Wiemann, P. J. Mason, A. Poustka, I. Doka, *Nat. Genet.* **1998**, *19*, 32–38.
- [76] X. Darzacq, B. E. Jady, C. Verheggen, A. M. Kiss, E. Bertrand, T. Kiss, *EMBO J.* **2002**, *21*, 2746–2756.
- [77] K. Nakanishi, L. Bonnefond, S. Kimura, T. Suzuki, R. Ishitani, O. Nureki, *Nature* **2009**, *461*, 1144–1148.
- [78] D. A. Melton, E. M. De Robertis, R. Cortese, *Nature* **1980**, *284*, 143–148.
- [79] a) M. Steiner-Mosonyi, D. Mangroo, *Biochem. J.* **2004**, *378*, 809–816; b) E. Lund, J. E. Dahlberg, *Science* **1998**, *282*, 2082–2085.
- [80] a) W.-K. Huh, J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson, J. S. Weissman, E. K. O'Shea, *Nature* **2003**, *425*, 686–691; b) T. Yoshihisa, K. Yunoki-Esaki, C. Ohshima, N. Tanaka, T. Endo, *Mol. Biol. Cell* **2003**, *14*, 3266–3279.
- [81] a) A. Takano, T. Endo, T. Yoshihisa, *Science* **2005**, *309*, 140–142; b) H. H. Shaheen, A. K. Hopper, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 11290–11295.
- [82] A. K. Hopper, H. H. Shaheen, *Trends Cell Biol.* **2008**, *18*, 98–104.
- [83] M. Thompson, R. A. Haeusler, P. D. Good, D. R. Engelke, *Science* **2003**, *302*, 1399–1401.
- [84] E. Bertrand, F. Houser-Scott, A. Kendall, R. H. Singer, D. R. Engelke, *Genes Dev.* **1998**, *12*, 2463–2468.
- [85] T. Ohira, T. Suzuki, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 10502–10507.
- [86] H.-Q. Jiang, Y. Motorin, Y.-X. Jin, H. Grosjean, *Nucleic Acids Res.* **1997**, *25*, 2694–2701.
- [87] G.-J. Arts, S. Kuersten, P. Romby, B. Ehresmann, I. W. Mattaj, *EMBO J.* **1998**, *17*, 7430–7441.
- [88] S. Kadaba, A. Krueger, T. Trice, A. M. Krecic, A. G. Hinnebusch, J. Anderson, *Genes Dev.* **2004**, *18*, 1227–1240.
- [89] M. L. Whitney, R. L. Hurto, H. H. Shaheen, A. K. Hopper, *Mol. Biol. Cell* **2007**, *18*, 2678–2686.
- [90] A. Ghavidel, T. Kislinger, O. Pogoutse, R. Sopko, I. Jurisica, A. Emili, *Cell* **2007**, *131*, 915–926.
- [91] a) Y. Suzuki, A. Noma, T. Suzuki, M. Senda, T. Senda, R. Ishitani, O. Nureki, *J. Mol. Biol.* **2007**, *372*, 1204–1214; b) Y. Suzuki, A. Noma, T. Suzuki, R. Ishitani, O. Nureki, *Nucleic Acids Res.* **2009**, *37*, 2910–2925; c) M. Kato, Y. Arais, A. Noma, A. Nagao, T. Suzuki, R. Ishitani, O. Nureki, *Nucleic Acids Res.* **2011**, *39*, 1576–1585; d) M. Umitsu, H. Nishimasu, A. Noma, T. Suzuki, R. Ishitani, O. Nureki, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 15616–15621.
- [92] a) L. Droogmans, H. Grosjean, *EMBO J.* **1987**, *6*, 477–483; b) G. R. Bjork, K. Jacobsson, K. Nilsson, M. J. O. Johansson, A. S. Bystrom, O. P. Persson, *EMBO J.* **2001**, *20*, 231–239.
- [93] A. P. Young, V. Bandarian, *Biochemistry* **2011**, *50*, 10573–10575.
- [94] a) W. F. Waas, V. de Crécy-Lagard, P. Schimmel, *J. Biol. Chem.* **2005**, *280*, 37616–37622; b) A. Noma, Y. Kirino, Y. Ikeuchi, T. Suzuki, *EMBO J.* **2006**, *25*, 2142–2154.
- [95] H. R. Kalhor, M. Penjwini, S. Clarke, *Biochem. Biophys. Res. Commun.* **2005**, *334*, 433–440.
- [96] A. Noma, R. Ishitani, M. Kato, A. Nagao, O. Nureki, T. Suzuki, *J. Biol. Chem.* **2010**, *285*, 34503–34507.
- [97] N. R. Rose, M. A. McDonough, O. N. F. King, A. Kawamura, C. J. Schofield, *Chem. Soc. Rev.* **2011**, *40*, 4364–4397.
- [98] J. R. Wagner, J. Cadet, *Acc. Chem. Res.* **2010**, *43*, 564–571.
- [99] J. S. Mattick, I. V. Makunin, *Hum. Mol. Genet.* **2006**, *15*, R17–29.
- [100] B. Yu, Z. Yang, J. Li, S. Minakhina, M. Yang, R. W. Padgett, R. Steward, *Science* **2005**, *307*, 932–935.
- [101] J. Li, Z. Yang, B. Yu, J. Liu, X. Chen, *Curr. Biol.* **2005**, *15*, 1501–1507.
- [102] a) Z. Yang, Y. W. Ebright, B. Yu, X. Chen, *Nucleic Acids Res.* **2006**, *34*, 667–675; b) Y. Zhao, Y. Yu, J. Zhai, V. Ramachandran, T. T. Dinh, et al., *Curr. Biol.* **2012**, *22*, 689–694.
- [103] a) T. Ohara, Y. Sakaguchi, T. Suzuki, H. Ueda, K. Miyauchi, T. Suzuki, *Nat. Struct. Mol. Biol.* **2007**, *14*, 349–350; b) Y. Kirino, Z. Mourelatos, *RNA* **2007**, *13*, 1397–1401.
- [104] a) L. M. Kamminga, M. J. Luteijn, M. J. den Broeder, S. Redl, L. J. T. Kaaij, E. F. Roovers, P. Ladurner, E. Berezikov, R. F. Ketting, *EMBO J.* **2010**, *29*, 3688–3700; b) S. Houwing, L. M. Kamminga, E. Berezikov, D. Cronembold, A. Girard, H. van den Elst, D. V. Flippov, H. Blaser, E. Raz, C. B. Moens, R. H. A. Plasterk, G. J. Hannon, B. W. Draper, R. F. Ketting, *Cell* **2007**, *129*, 69–82.
- [105] M. D. Horwich, C. Li, C. Matranga, V. Vagin, G. Farley, P. Wang, P. D. Zamore, *Curr. Biol.* **2007**, *17*, 1265–1272.
- [106] C. Lacoux, D. Di Marino, P. Pilo-Boyl, F. Zalfa, B. Yan, M. T. Ciotti, M. Falconi, H. Urlaub, T. Achsel, A. Mougin, M. Caziergues-Ferrer, C. Bagni, *Nucleic Acids Res.* **2012**, *40*, 4086–4096.
- [107] Y.-K. Kim, I. Heo, V. N. Kim, *Cell* **2010**, *143*, 703–709.
- [108] a) W. Yang, T. P. Chendrimada, Q. Wang, M. Higuchi, P. H. Seeburg, R. Shiekhattar, K. Nishikura, *Nat. Struct. Mol. Biol.* **2006**, *13*, 13–21; b) Y. Kawahara, B. Zinshteyn, T. P. Chendrimada, R. Shiekhattar, K. Nishikura, *EMBO Rep.* **2007**, *8*, 763–768.
- [109] J. Karijovich, Y.-T. Yu, *RNA Biol.* **2010**, *7*, 192–204.
- [110] N. K. Kim, C. A. Theimer, J. R. Mitchell, K. Collins, J. Feigon, *Nucleic Acids Res.* **2010**, *38*, 6746–6756.
- [111] X. Zhao, J. R. Patton, S. L. Davis, B. Florence, S. J. Ames, R. A. Spanjaard, *Mol. Cell* **2004**, *15*, 549–558.
- [112] a) M. Buck, M. Connick, B. N. Ames, *Anal. Biochem.* **1983**, *129*, 1–13; b) P. F. Crain, *Mass Spectrom. Rev.* **1990**, *9*, 505–554; c) J. A. McCloskey, *Acc. Chem. Res.* **1991**, *24*, 81–88; d) J. Ofengand, M. Del Campo, Y. Kaya, *Methods* **2001**, *25*, 365–373; e) I. Behm-Ansmant, M. Helm, Y. Motorin, *J. Nucleic Acids* **2011**, *2011*, 408053; f) S. Kellner, J. Burhenne, M. Helm, *RNA Biol.* **2010**, *7*, 237–247.
- [113] Y. Motorin, S. Muller, I. Behm-Ansmant, C. Branlant, *Methods Enzymol.* **2007**, *425*, 21–53.
- [114] M. Sakurai, T. Yano, H. Kawabata, H. Ueda, T. Suzuki, *Nat. Chem. Biol.* **2010**, *6*, 733–740.
- [115] J. E. Squires, H. R. Patel, M. Nusch, T. Sibbritt, D. T. Humphreys, B. J. Parker, C. M. Suter, T. Preiss, *Nucleic Acids Res.* **2012**, *40*, 5023–5033.

- [116] A. Durairaj, P. A. Limbach, *Rapid Commun. Mass Spectrom.* **2008**, *22*, 3727–3734.
- [117] D. Dominissini, S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, et al., *Nature* **2012**, *485*, 201–206.
- [118] S. L. Hiley, J. Jackman, T. Babak, M. Trocheset, Q. D. Morris, E. Phizicky, T. R. Hughes, *Nucleic Acids Res.* **2005**, *33*, e2.
- [119] K. Miyauchi, T. Ohara, T. Suzuki, *Nucleic Acids Res.* **2007**, *35*, e24.
- [120] M. Saikia, Y. Fu, M. Pavon-Eternod, C. He, T. Pan, *RNA* **2010**, *16*, 1317–1327.
- [121] a) C. W. Gehrke, K. C. Kuo, *J. Chromatogr.* **1989**, *471*, 3–36; b) J. D. Johnson, J. Horowitz, *Biochim. Biophys. Acta* **1971**, *247*, 262–279; c) X. L. Zhao, Y. T. Yu, *RNA* **2004**, *10*, 996–1002.
- [122] a) T. Brückl, D. Globisch, M. Wagner, M. Müller, T. Carell, *Angew. Chem.* **2009**, *121*, 8074–8077; *Angew. Chem. Int. Ed.* **2009**, *48*, 7932–7934; b) J. J. Dalluge, T. Hashizume, J. A. McCloskey, *Nucleic Acids Res.* **1996**, *24*, 3242–3245.
- [123] a) B. Addepalli, P. A. Limbach, *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1363–1372; b) C. T. Y. Chan, M. Dyavaiah, M. S. DeMott, K. Taghizadeh, P. C. Dedon, T. J. Begley, *PLoS Genet.* **2010**, *6*, e1001247.
- [124] T. Brückl, Dissertation, Ludwig-Maximilians-Universität, München, Germany, **2010**.
- [125] D. Globisch, D. Pearson, A. Hienzsch, T. Brückl, M. Wagner, I. Thoma, P. Thumbs, V. Reiter, A. C. Kneutinger, M. Müller, S. A. Sieber, T. Carell, *Angew. Chem.* **2011**, *123*, 9913–9916; *Angew. Chem. Int. Ed.* **2011**, *50*, 9739–9742.
- [126] T. Suzuki, T. Wada, K. Saigo, K. Watanabe, *EMBO J.* **2002**, *21*, 6581–6589.
- [127] Y. Kirino, T. Yasukawa, S. Ohta, S. Akira, K. Ishihara, K. Watanabe, T. Suzuki, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15070–15075.
- [128] a) G. Dirheimer, W. Baranowski, G. Keith, *Biochimie* **1995**, *77*, 99–103; b) C. Pathak, Y. K. Jaiswal, M. Vinayak, *RNA Biol.* **2005**, *2*, 143–148.
- [129] a) S. Ishiwata, J. Katayama, H. Shindo, Y. Ozawa, K. Itoh, M. Mizugaki, *J. Biochem.* **2001**, *129*, 13–17; b) S. Ishiwata, Y. Ozawa, J. Katayama, S. Kaneko, H. Shindo, Y. Tomioka, T. Ishiwata, G. Asano, S. Ikegawa, M. Mizugaki, *Cancer Lett.* **2004**, *212*, 113–119.
- [130] A. Gallo, F. Locatelli, *Biol. Rev. Cambridge Philos. Soc.* **2012**, *87*, 95–110.
- [131] a) K. A. Fawcett, I. Barroso, *Trends Genet.* **2010**, *26*, 266–274; b) J. R. Perry, T. M. Frayling, *Curr. Opin. Clin. Nutr. Metab. Care* **2008**, *11*, 371–377.
- [132] a) S. Arragain, S. K. Handelman, F. Forouhar, F.-Y. Wei, K. Tomizawa, J. F. Hunt, T. Douki, M. Fontecave, E. Mulliez, M. Atta, *J. Biol. Chem.* **2010**, *285*, 28425–28433; b) F.-Y. Wei, T. Suzuki, S. Watanabe, S. Kimura, T. Kaitsuka, A. Fujimura, H. Matsui, M. Atta, H. Michiue, M. Fontecave, K. Yamagata, T. Suzuki, K. Tomizawa, *J. Clin. Invest.* **2011**, *121*, 3598–3608.
- [133] T. Gerken, C. A. Girard, Y. C. Tung, C. J. Webby, V. Saudek, K. S. Hewitson, G. S. H. Yeo, M. A. McDonough, S. Cunliffe, L. A. McNeill, J. Galvanovskis, P. Rorsman, P. Robins, X. Prieur, A. P. Coll, M. Ma, Z. Jovanovic, I. S. Farooqi, B. Sedgwick, I. Barroso, T. Lindahl, C. P. Ponting, F. M. Ashcroft, S. O'Rahilly, C. J. Schofield, *Science* **2007**, *318*, 1469–1472.
- [134] G. Jia, C. G. Yang, S. Yang, X. Jian, C. Yi, Z. Zhou, C. He, *FEBS Lett.* **2008**, *582*, 3313–3319.
- [135] G. Jia, Y. Fu, X. Zhao, Q. Dai, G. Zheng, Y. Yang, C. Yi, T. Lindahl, T. Pan, Y.-G. Yang, C. He, *Nat. Chem. Biol.* **2011**, *7*, 885–887.
- [136] Y. P. Ching, A. S. Pang, W. H. Lam, R. Z. Qi, J. H. Wang, *J. Biol. Chem.* **2002**, *277*, 15237–15240.
- [137] F.-Y. Wei, K. Nagashima, T. Ohshima, Y. Saheki, Y.-F. Lu, M. Matsushita, Y. Yamada, K. Mikoshiba, Y. Seino, H. Matsui, K. Tomizawa, *Nat. Med.* **2005**, *11*, 1104–1108.
- [138] a) K. H. Kaminska, U. Baraniak, M. Boniecki, K. Nowaczyk, A. Czerwonec, J. M. Bujnicki, *Proteins* **2008**, *70*, 1–18; b) V. Reiter, D. M. Matschkal, M. Wagner, D. Globisch, A. C. Kneutinger, M. Müller, T. Carell, *Nucleic Acids Res.* **2012**, DOI: 10.1093/nar/gks240.
- [139] a) P. I. Moreira, S. M. Cardoso, C. M. Pereira, M. S. Santos, C. R. Oliveira, *CNS Neurol. Disord. Drug Targets* **2009**, *8*, 492–511; b) P. I. Moreira, A. I. Duarte, M. S. Santos, A. C. Rego, C. R. Oliveira, *J. Alzheimer's Dis.* **2009**, *16*, 741–761.
- [140] N. Okada, S. Noguchi, S. Nishimura, T. Ohgi, T. Goto, P. F. Crain, J. A. McCloskey, *Nucleic Acids Res.* **1978**, *5*, 2289–2296.
- [141] B. S. Huang, R. T. Wu, K. Y. Chien, *Cancer Res.* **1992**, *52*, 4696–4700.
- [142] a) U. Grädler, H. D. Gerber, D. M. Goodenough-Lashua, G. A. Garcia, R. Ficner, K. Reuter, M. T. Stubbs, G. Klebe, *J. Mol. Biol.* **2001**, *306*, 455–467; b) J. K. Hurt, S. Olgen, G. A. Garcia, *Nucleic Acids Res.* **2007**, *35*, 4905–4913.
- [143] a) S. R. Hörtnner, T. Ritschel, B. Stengl, C. Kramer, W. B. Schweizer, B. Wagner, M. Kansy, G. Klebe, F. Diederich, *Angew. Chem.* **2007**, *119*, 8414–8417; *Angew. Chem. Int. Ed.* **2007**, *46*, 8266–8269; b) B. Stengl, E. A. Meyer, A. Heine, R. Brenk, F. Diederich, G. Klebe, *J. Mol. Biol.* **2007**, *370*, 492–511.
- [144] a) S. Gamagedara, Y. Ma, *Bioanalysis* **2011**, *3*, 2129–2142; b) W.-Y. Hsu, W.-D. Lin, Y. Tsai, C.-T. Lin, H.-C. Wang, L.-B. Jeng, C.-C. Lee, Y.-C. Lin, C.-C. Lai, F.-J. Tsai, *Clin. Chim. Acta* **2011**, *412*, 1861–1866; c) A. Seidel, S. Brunner, P. Seidel, G. I. Fritz, O. Herbarth, *Br. J. Cancer* **2006**, *94*, 1726–1733.
- [145] a) E. Friedberg, G. Walker, W. Siede, R. Wood, R. Schultz, T. Ellenberger, *DNA Repair and Mutagenesis*, 2nd ed., ASM, Washington, **2006**; b) M. P. Stone, H. Huang, K. L. Brown, G. Shanmugam, *Chem. Biodiversity* **2011**, *8*, 1571–1615.
- [146] D. Ratel, J. L. Ravanat, F. Berger, D. Wion, *Bioessays* **2006**, *28*, 309–315.
- [147] B. R. Palmer, M. G. Marinus, *Gene* **1994**, *143*, 1–12.
- [148] T. B. Johnson, R. D. Coghill, *J. Am. Chem. Soc.* **1925**, *47*, 2838–2844.
- [149] G. R. Wyatt, *Nature* **1950**, *166*, 237–238.
- [150] a) S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930; b) M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930–935.
- [151] N. W. Penn, R. Suwalski, C. O'Riley, K. Bojanowski, R. Yura, *Biochem. J.* **1972**, *126*, 781–790.
- [152] M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmler, S. Michalakakis, M. Müller, M. Biel, T. Carell, *Angew. Chem.* **2010**, *122*, 5503–5505; *Angew. Chem. Int. Ed.* **2010**, *49*, 5375–5377.
- [153] a) T. F. Kraus, D. Globisch, M. Wagner, S. Eigenbrod, D. Widmann, M. Münzel, T. Pfaffeneder, B. Hackner, W. Feiden, U. Schüller, T. Carell, H. A. Kretzschmar, *Int. J. Cancer* **2012**, DOI: 10.1002/ijc.27429; b) H. Yang, Y. Liu, F. Bai, J. Y. Zhang, S. H. Ma, J. Liu, Z. D. Xu, H. G. Zhu, Z. Q. Ling, D. Ye, K. L. Guan, Y. Xiong, *Oncogene* **2012**, DOI: 10.1038/onc.2012.67.
- [154] K. I. Ladwein, M. Jung, *Angew. Chem.* **2011**, *123*, 12347–12349; *Angew. Chem. Int. Ed.* **2011**, *50*, 12143–12145.
- [155] D. Globisch, M. Münzel, M. Müller, S. Michalakakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS ONE* **2010**, *5*, e15367.
- [156] T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem.* **2011**, *123*, 7146–7150; *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012.
- [157] a) S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303; b) Y.-F. He, B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C.-X. Song, K. Zhang, C. He, G.-L. Xu, *Science* **2011**, *333*, 1303–1307.

- [158] L. Zhang, X. Lu, J. Lu, H. Liang, Q. Dai, G.-L. Xu, C. Luo, H. Jiang, C. He, *Nat. Chem. Biol.* **2012**, 8, 328–330.
- [159] a) H. Cedar, Y. Bergman, *Nat. Rev. Genet.* **2009**, 10, 295–304; b) M. M. Suzuki, A. Bird, *Nat. Rev. Genet.* **2008**, 9, 465–476; c) H. Sasaki, Y. Matsui, *Nat. Rev. Genet.* **2008**, 9, 129–140.
- [160] M. R. Branco, G. Ficuz, W. Reik, *Nat. Rev. Genet.* **2012**, 13, 7–13.
- [161] S. K. Ooi, T. H. Bestor, *Cell* **2008**, 133, 1145–1148.
- [162] a) C. Popp, W. Dean, S. Feng, S. J. Cokus, S. Andrews, M. Pellegrini, S. E. Jacobsen, W. Reik, *Nature* **2010**, 463, 1101–1126; b) M. Wossidlo, J. Arand, V. Sebastiano, K. Lepikhov, M. Boiani, R. Reinhardt, H. Schöler, J. Walter, *EMBO J.* **2010**, 29, 1877–1888; c) P. Hajkova, S. J. Jeffries, C. Lee, N. Miller, S. P. Jackson, M. A. Surani, *Science* **2010**, 329, 78–82; d) S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Sorprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, A. Bellacosa, *Cell* **2011**, 146, 67–79.
- [163] A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, 286, 35334–35338.
- [164] J. U. Guo, Y. Su, C. Zhong, G. L. Ming, H. Song, *Cell* **2011**, 145, 423–434.
- [165] a) A. Radzicka, R. Wolfenden, *Science* **1995**, 267, 90–93; b) J. A. Smiley, M. Kundracik, D. A. Landfried, V. R. Barnes, Sir, A. A. Axhemi, *Biochim. Biophys. Acta Gen. Subj.* **2005**, 1723, 256–264.
- [166] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, C. Hagemeyer, M. Truss, T. Carell, *Angew. Chem.* **2012**, DOI: 10.1002/ange.201202583; *Angew. Chem. Int. Ed.* **2012**, DOI: 10.1002/anie.201202583.
- [167] T. P. Jurkowski, A. Jeltsch, *ChemBioChem* **2011**, 12, 2543–2545.
- [168] M. Yu, G. C. Hon, K. E. Szulwach, C. X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park, J.-H. Min, P. Jin, B. Ren, C. He, *Cell* **2012**, 149, 1368–1380.
- [169] M. J. Booth, M. R. Branco, G. Ficuz, D. Oxley, F. Krueger, W. Reik, S. Balasubramanian, *Science* **2012**, 336, 934–937.