RNA Hypermodifications

Structural Elucidation, Biological Characterization and Manipulation

Dissertation
zur Erlangung des Grades
»Doktor der Naturwissenschaften«
im Promotionsfach Pharmazie

am Fachbereich Chemie, Pharmazie, Geographie und Geowissenschaften
der Johannes Gutenberg-Universität Mainz

Larissa Sabine Bessler
geboren in Öhringen

Mainz, April 2023

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Dekanin: ██████████████████

1. Berichterstatter: ███████████████

2. Berichterstatterin: ██████████████████

Datum der mündlichen Prüfung: 23. Mai 2023

D77 (Dissertation Mainz)

Die vorliegende Arbeit wurde am Institut für Pharmazeutische und Biomedizinische Wissenschaften der Johannes Gutenberg-Universität in Mainz zur Erlangung des Grades »Doktor der Naturwissenschaften« unter der Betreuung von ███████████████ angefertigt.

1. Berichterstatter

███████████████

Fachbereich Chemie, Pharmazie, Geographie und Geowissenschaften
Institut für Pharmazeutische und Biomedizinische Wissenschaften
Johannes Gutenberg-Universität Mainz

2. Berichterstatterin

███████████████████

Fachbereich Chemie, Pharmazie, Geographie und Geowissenschaften
Institut für Pharmazeutische und Biomedizinische Wissenschaften
Johannes Gutenberg-Universität Mainz

Hiermit versichere ich eidesstattlich :

1. Ich habe die jetzt als Dissertation vorgelegte Arbeit selbst angefertigt und alle benutzten Hilfsmittel (Literatur, Apparaturen, Material) in der Arbeit angegeben.

2. Ich habe oder hatte die jetzt als Dissertation vorgelegte Arbeit nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht.

3. Ich hatte weder die jetzt als Dissertation vorgelegte Arbeit noch Teile davon bei einer anderen Fakultät bzw. einem anderen Fachbereich als Dissertation eingereicht.

Ort, Datum	Unterschrift

Danksagung

Contents

Abstract

Ribonucleic acid (RNA) is a ubiquitous biopolymer with a versatility of functions in cellular processes, which can be fulfilled not least due to the implementation of a sequence-independent layer of information in the form of a multitude of chemically modified nucleosides. Understanding the complex network of RNA modifications and their involvement in physiological processes but also in pathological conditions is an active field of research and even if a plethora of RNA modifications is already known, increasingly sophisticated instrumentation still allows the discovery of new low abundant species. For both, known and new structures, biological roles have to be disentangled and knowledge of the entirety of non-canonical nucleoside structures within the cell is essential to capture the big picture. In this work, two potentially new RNA hypermodifications were investigated, based on previous work in the ███ group in the form of a wide-ranging screening for new RNA modifications in bacterial transfer RNA (tRNA) using liquid chromatography coupled tandem mass spectrometry (LC-MS/MS). For the first candidate, the structure of a new nucleoside that occurs in tRNA from *Escherichia coli* (*E. coli*) was unequivocally elucidated to be 2-methylthio-1,N6-ethenoadenosine (ms^{2}εA), based on the combination of LC-MS/MS analysis with stable isotope labeling, modification enzyme knockouts and organic synthesis. In contrast to RNA modifications that are understood to improve RNA function, experiments on the origin of ms^{2}εA showed that this putative modification might rather be an oxidatively damaged variant of the known RNA modification 2-methylthio-N6-isopentenyladenosine with a loss of function and, concomitantly, a detrimental character of this modified nucleoside. These findings open up for a new subfield of research within the world of RNA modifications, namely RNA modification damages and their consequences, but simultaneously urge caution in the discovery of new modifications and their classification as such. The context of RNA damage also came into focus with regard to the second candidate, as the more detailed characterization of this compound, in addition to its natural occurrence in *E. coli* tRNA, revealed a UV-induced formation of this dimeric structure *in vivo* and *in vitro*, which was used to gain further structural information. Furthermore, knockout experiments associated the native formation of this candidate with the sulfurtransferase ThiI and UV-induced formation of the candidate in dethiolated tRNA pointed to a favoring effect of 4-thiouridine in the UV-dependent mechanism. While this candidate's structure could not be fully elucidated, the results of this work help to shape a new perception of this candidate and

imply a formation of this structure via different mechanisms as well as a relation to RNA damage. The fact that the alteration of an intrinsically present hypermodified structure is not necessarily accompanied by impaired function on the molecular level was shown in another project within the scope of this work. In a deliberately manipulated transglycosylation reaction, an azide-containing synthetic precursor of the hypermodification queuosine was incorporated *in vitro* as well as *in vivo* in *E. coli*, *Schizosaccharomyces pombe* and human cells by the endogenous enzyme and characterized by exploiting the clickability of the azide moiety with respect to success of incorporation and distribution on different RNA species. An investigation of the physiological effects at the molecular level indicated a minimally invasive nature of the developed system, showing that the surrogate was not only functionally integrated into the translational process but also complied with queuosine's stimulating effect in the modification circuit with 5- methylcytosine at position 38 of tRNA[Asp].

Zusammenfassung

Ribonukleinsäure (RNA) ist ein ubiquitäres Biopolymer mit vielfältigen Funktionen in zellulären Prozessen, die nicht zuletzt durch die Implementierung einer sequenzunabhängigen Informationsebene in Form einer Vielzahl chemisch modifizierter Nukleoside erfüllt werden können. Das Verständnis des komplexen Netzwerks von RNA-Modifikationen und deren Beteiligung an physiologischen Prozessen, aber auch an pathologischen Zuständen, ist ein aktives Forschungsgebiet, und selbst wenn eine Vielzahl an RNA-Modifikationen bereits bekannt ist, ermöglichen immer ausgefeiltere Instrumente weiterhin die Entdeckung neuer, selten vorkommender Spezies. Sowohl für bekannte als auch für neue Strukturen müssen die biologischen Funktionen entschlüsselt werden und die Kenntnis der Gesamtheit der nicht-kanonischen Nukleosidstrukturen in der Zelle ist unerlässlich, um übergeordnete Zusammenhänge zu erfassen. In dieser Arbeit wurden auf Grundlage vorangegangener Arbeiten in der ▬▬ Gruppe in Form eines breit angelegten Screenings nach neuen RNA Modifikationen in bakterieller Transfer-RNA (tRNA) mittels Flüssigchromatographie gekoppelter Tandem-Massenspektrometrie (LC-MS/MS) zwei potenziell neue RNA-Hypermodifikationen untersucht. Für den ersten Kandidaten wurde die Struktur eines neuen Nukleosids, das in der tRNA von *Escherichia coli* (*E. coli*) vorkommt, eindeutig als 2-Methylthio-1,*N*6-ethenoadenosin (ms^{2}εA) aufgeklärt, basierend auf der Kombination von LC-MS/MS-Analyse mit Isotopenmarkierung, Knockouts von Modifikationsenzymen und organischer Synthese. Im Gegensatz zu RNA-Modifikationen, von denen man annimmt, dass sie die RNA-Funktion verbessern, zeigten Experimente zum Ursprung von ms^{2}εA, dass es sich bei dieser vermeintlichen Modifikation um eine oxidativ geschädigte Variante der bekannten RNA-Modifikation 2-Methylthio-*N*6-isopentenyladenosin handeln könnte, die ihre Funktion verliert und dem modifizierten Nukleosid somit einen schädlichen Charakter verleiht. Diese Erkenntnisse eröffnen ein neues Teilgebiet der Forschung im Bereich der RNA-Modifikationen, nämlich RNA-Modifikationsschäden und deren Folgen, mahnen aber gleichzeitig zur Vorsicht bei der Entdeckung neuer Modifikationen und deren Klassifizierung als solche. Ein Zusammenhang zu RNA-Schaden rückte auch bei dem zweiten Kandidaten in den Fokus, da die genauere Charakterisierung dieser Verbindung neben ihrem natürlichen Vorkommen in *E. coli* tRNA eine UV-induzierte Bildung dieser dimeren Struktur *in vivo* und *in vitro* ergab, die zur Gewinnung weiterer Strukturinformationen genutzt wurde. Darüber hinaus brachten Knockout-Experimente die native Bildung dieses Kandidaten

mit der Sulfurtransferase ThiI in Verbindung, und die UV-induzierte Bildung des Kandidaten in dethiolierter tRNA deutete auf eine Begünstigung des UV-bedingten Mechanismus durch 4-Thiouridin hin. Obwohl die Struktur dieses Kandidaten nicht vollständig aufgeklärt werden konnte, tragen die Ergebnisse dieser Arbeit dazu bei, eine neue Wahrnehmung dieses Kandidaten zu formen und implizieren eine Bildung dieser Struktur über verschiedene Mechanismen sowie einen Zusammenhang zu RNA-Schaden. Dass die Veränderung einer intrinsisch vorhandenen, hypermodifizierten Struktur nicht zwangsläufig mit einer Funktionsbeeinträchtigung auf molekularer Ebene einhergeht, wurde in einem weiteren Projekt im Rahmen dieser Arbeit gezeigt. In einer gezielt manipulierten Transglykosylierungsreaktion wurde ein azidhaltiger synthetischer Vorläufer der Hypermodifikation Queuosin sowohl *in vitro* als auch *in vivo* in *E. coli*, *Schizosaccharomyces pombe* und humanen Zellen durch das endogene Enzym eingebaut und unter Ausnutzung der Klickbarkeit des Azidrests im Hinblick auf den erfolgreichen Einbau und die Verteilung auf verschiedene RNA-Spezies charakterisiert. Eine Untersuchung der physiologischen Effekte auf molekularer Ebene wies auf einen minimal invasiven Charakter des entwickelten Systems hin, wobei gezeigt wurde, dass das Surrogat nicht nur funktionell in den Translationsprozess integriert wurde, sondern auch die stimulierende Wirkung von Queuosin im Modifikationskreislauf mit 5-Methylcytosin an Position 38 von tRNA^Asp erfüllen konnte.

Abbreviations

8oxoG	8-oxoguanosine
A	adenosine
ac^{4}C	N4-acetylcytidine
ACN	acetonitrile
Am	$2'$-O-methyladenosine
APS	ammonium persulfate
Asn	asparagine
Asp	aspartatic acid
ATP	adenosine triphosphate
BSA	bovine serum albumine
bTGT	bacterial tRNA guanine transglycosylase
C	cytidine
CDG	7-carboxy-7-deazaguanine
CFU	colony forming units
CID	collision induced dissociation
CPH$_4$	6-carboxy-5,6,7,8-tetrahydropterin
ct^{6}A	cyclic N6-threonylcarbamoyladenosine
CTP	cytidine triphosphate
CuAAC	copper(I)-catalyzed azide-alkyne cycloaddition
D	dihydrouridine
dA	$2'$-deoxyadenosine
dC	$2'$-deoxycytidine
DCP2	mRNA-decapping enzyme 2
DMAPP	dimethylallyl pyrophosphate
DMEM	Dulbecco's modified eagle medium
dMRM	dynamic multiple reaction monitoring
DNA	deoxyribonucleic acid
DPBS	Dulbecco's balanced salt solution
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
eGFP	enhanced green fluorescent protein

EM	electron multiplier
EMCV	encephalomyocarditis virus
ESI	electrospray ionization
eTGT	eukaryotic tRNA guanine transglycosylase
εA	1,N6-ethenoadenosine
Fast AP	fast alkaline phosphatase
FBS	fetal bovine serum
G	guanosine
GCH I	guanosine triphosphate cyclohydrolase I
Glu	glutamic acid
GTP	guanosine triphosphate
H_2NTP	7,8-dihydroneopterin triphosphate
HCD	higher-energy collisional dissociation
HED	high-energy dynode
HEK	human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	histidine
ho^5C	5-hydroxycytidine
HPLC	high performance liquid chromatography
HRMS	high-resolution mass spectrometry
I	inosine
i^6A	N6-isopentenyladenosine
IEDDA	inverse electron-demand Diels-Alder cycloaddition
IPTG	isopropyl β-D-1-thiogalactopyranoside
IRES	internal ribosome entry site
IVT	*in vitro* transcription
KO	knockout
LB	lysogeny broth
LC-MS/MS	liquid chromatography-tandem mass spectrometry
LPO	lipid peroxidation
m/z	mass-to-charge ratio
m^1A	1-methyladenosine
m^1G	1-methylguanosine
$m^1\Psi$	1-methylpseudouridine
m^3C	3-methylcytidine
m^5C	5-methylcytidine
m^5U	5-methyluridine
m^6A	N6-methyladenosine
m^6Am	N6,2′-O-dimethyladenosine
m^7G	7-methylguanosine
mcm^5s^2U	5-methoxycarbonylmethyl-2-thiouridine

miRNA	micro RNA
mRNA	messenger RNA
ms^2C	2-methylthiocytidine
$ms^2\varepsilon A$	2-methylthio-1,N6-ethenoadenosine
ms^2i^6A	2-methylthio-N6-isopentenyladenosine
ms^2t^6A	2-methylthio-N6-threonylcarbamoyladenosine
ms^4U	4-methylthiouridine
$msms^2i^6A$	2-methylthiomethylenethio-N6-isopentenyladenosine
N	nucleoside
NAIL-MS	nucleic acid isotope labeling coupled mass spectrometry
NLS	neutral loss scan
Nm	$2'$-O-methylnucleoside
NMR	nuclear magnetic resonance
oQ	epoxyqueuosine
pDNA	plasmid DNA
Phe	phenylalanine
PP_i	pyrophosphate
$preQ_0$	7-cyano-7-deazaguanine
$preQ_1$	7-aminomethyl-7-deazaguanine
Ψ	pseudouridine
Q	queuosine
q	queuine
QQQ	triple quadrupole
RNA	ribonucleic acid
ROS	reactive oxygen species
RP	reversed-phase
rRNA	ribosomal RNA
rT	ribothymidine
RT	retention time
S. cerevisiae	*Saccharomyces cerevisiae*
S. pombe	*Schizosaccharomyces pombe*
s^2C	2-thiocytidine
s^2U	2-thiouridine
s^4U	4-thiouridine
SAM	S-adenosyl methionine
SARS-CoV-2	severe acute respiratory syndrome coronavirus type 2
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
SOC	super optimal broth with catabolite
SPAAC	strain-promoted azide-alkyne cycloaddition

t^{6}A	*N*6-threonylcarbamoyladenosine
TAG	transglycosylation at guanosine
TBE	tris/borate/ethylenediaminetetraacetic acid
TEMED	*N,N,N',N'*-tetramethylethylenediamine
TGT	tRNA guanine transglycosylase
THPTA	tris(3-hydroxypropyltriazolylmethyl)amine
Thr	threonine
TOF	time-of-flight
tRNA	transfer RNA
Tyr	tyrosine
U	uridine
UTP	uridine triphosphate
UTR	untranslated region
UV	ultraviolet
yW	wybutosine

1 Introduction

1.1 Ribonucleic acid

1.1.1 Structure and function of RNA

As early as 1869, Friedrich Miescher discovered and characterized a hitherto unknown cell substance with high phosphorous content, pioneering the structural elucidation of nucleic acids and their integration in the central dogma of molecular biology.[1] The latter highlights the fundamental roles of nucleic acids, naturally occurring in the forms of deoxyribonucleic acid (DNA) which carries the genetic information and ribonucleic acid (RNA) which – amongst diverse other functions - transports this information to the site of protein biosynthesis and serves as a template.[2,3] RNA is a single-stranded polyribonucleotide and mainly consists of four building blocks, so-called nucleotides. The core structure of nucleotides is formed by a ribose sugar (2′-deoxyribose in DNA) with a phosphate residue attached to the hydroxyl group at C-5′ and only differs by one of the four nitrogenous nucleobases that is attached to C-1′ of the ribose via an *N*-glycosidic bond (Figure 1.1a). Of note, the combination of a ribose with a nucleobase is termed nucleoside which describes the biosynthetic precursors of nucleotides or RNA degradation products. The canonical nucleobases are classified in the pyrimidine nucleobases cytosine and uracil as well as the purine nucleobases adenine and guanine, whereas in DNA the structurally related thymine is used instead of uridine. The RNA strand is formed by connecting the building blocks via 3′-5′-phosphodiester bonds resulting in a characteristic sugar-phosphate backbone with information on sequence direction given in the form of the 5′ end terminating with a phosphate group at carbon-5′ and the 3′ end terminating with a hydroxyl group at carbon-3′ of the ribose (Figure 1.1b).[4] In contrast to DNA, which is known for its double-stranded helix structure, RNA is of single-stranded nature in the first place. Nevertheless, base pairing, an interaction of the complementary nucleobases adenine and uracil as well as guanine and cytosine (Watson-Crick base pairs) via hydrogen bonds, enables the formation of double-stranded regions and as a consequence of that highly complex secondary and tertiary structures.[4,5]
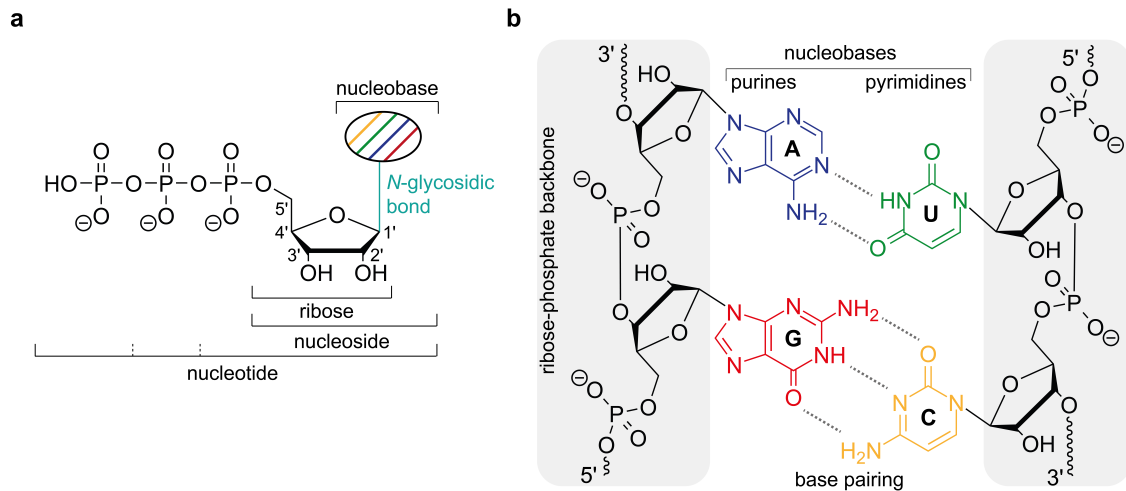
a

b

Figure 1.1: RNA structure. a Structural components of the RNA building blocks. Coloring according to the nucleobases in (b). **b** Phosphate groups that are attached to the nucleosides cytidine (C), uridine (U), guanosine (G) and adenosine (A) link the 5′ position of one nucleoside's ribose to the 3′ position of the ribose of the adjacent nucleoside and form a ribose-phosphate backbone, the structural basis of an RNA strand. Base pairing interactions between the complementary nucleobases via hydrogen bonds are depicted as dashed lines.

Depending on the specific sequence of nucleobases, the appearances of RNA molecules are as versatile as the roles they can fulfil in the cellular environment. The most prominent interplay of different RNA species is the sophisticated process of protein biosynthesis involving messenger RNA (mRNA) which provides the sequence information for the protein of interest in the form of base triplets, transfer RNA (tRNA) which delivers the required amino acids by direct interaction with the mRNA, and ribosomal RNA (rRNA) which together with ribosomal proteins forms ribosomes, the sites of protein biosynthesis, and has catalytic functions.[6–8] The ensemble of all RNA species that do not contain sequence information for building proteins and are therefore not translated are referred to as non-coding RNAs. These range from rRNA which is several thousand base pairs long, to significantly shorter RNA such as micro RNA (miRNA) and small interfering RNA (siRNA), which regulate gene expression and play an important role in gene silencing, but also small nuclear RNA (snRNA), which is responsible for the processing and modification of other RNA species.[9–11]

1.1.2 Transfer RNA

Its function as an adapter molecule that decodes an mRNA sequence into a protein, makes tRNA one of the most essential biomolecules and requires a certain degree of structural complexity and variability. Already in the early days of tRNA research, Holley *et al.* suggested that tRNA adopts a secondary structure in the form of a cloverleaf as a result of reversely complementary parts of the sequence.[12] Referring to either struc-

tural features or a specific function, the cloverleaf structure can be divided into five conserved sections: acceptor stem, D loop, anticodon loop, variable loop and TΨC loop (Figure 1.2a). True to its name, the variable loop allows the incorporation of additional nucleotides resulting in a variable length of different tRNA species between 76 and 95 nucleotides. Especially the single-stranded loop structures are prone to diverse post-transcriptional modifications which play an important role in both stability and function of a tRNA (see section 1.2.3).[13,14] In physiological conditions, interactions between the D loop and the TΨC loop promote correct folding and stabilization of a three-dimensional L-shaped structure, resulting in two major domains, namely the acceptor domain and the anticodon domain.[15,16]
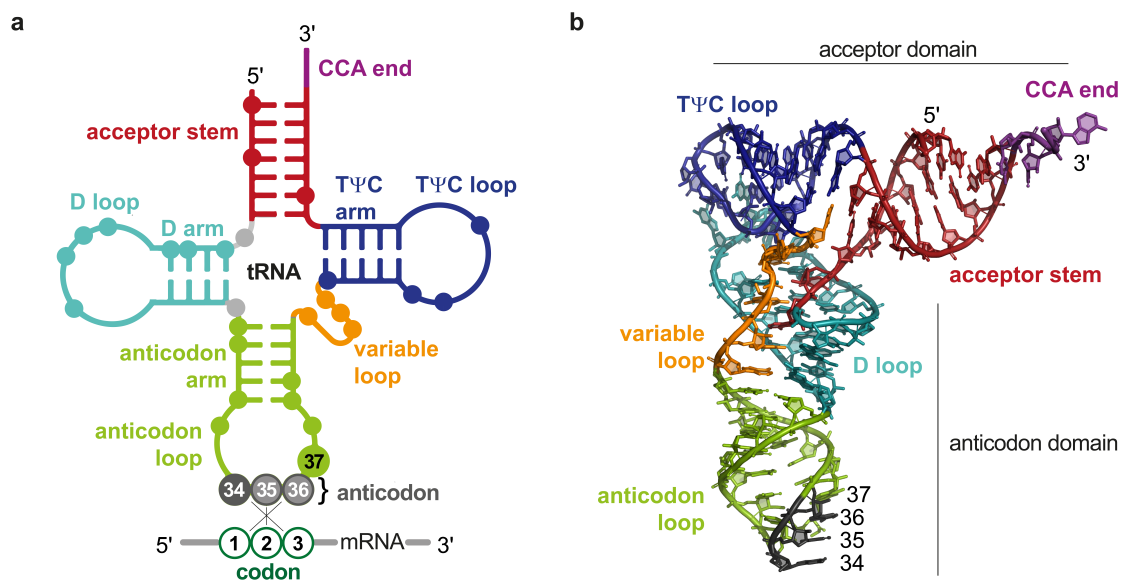

Figure 1.2: Secondary and tertiary structure of cytoplasmic tRNA. a Cloverleaf-like secondary structure of tRNA highlighting the five conserved sections in different colors as well as the hydrogen bond-based pairing between the anticodon (positions 34, 35 and 36 of the tRNA) with an mRNA codon (positions 1, 2 and 3). Positions marked with filled circles point to frequent, often conserved modification of nucleosides in these sites within eukaryotic tRNA, with positions 34 and 37 showing an incomparably high degree of modification. **b** L-shaped tertiary structure of tRNA with structural elements in the same color code as in (a), based on the Protein Data Bank entry for *Saccharomyces cerevisiae* tRNA[Phe] (1EHZ). Figure adapted.[16–18]

The acceptor stem contains a 5′ phosphate and invariably carries a single-stranded CCA sequence at the 3′ end to enable the attachment of the appropriate amino acid by formation of an ester bond, which is catalyzed by individual aminoacyl-tRNA synthases.[19] The incorporation of the correct amino acid into the growing polypeptide chain is ensured by base pairing between the mRNA codon, a specific sequence of three nucleobases, and its reversely complementary anticodon which is located in positions 34–36 of the anticodon loop (Figure 1.2a). While there are only 20 canonical amino acids, the combination of the four nucleobases to yield different triplets results in 64 unique combi-

nations, i.e. the genetic code is degenerate and several codons encode for the same amino acid.[20] From the tRNA perspective this means on the one hand that the incorporation of the same amino acid can be mediated by several tRNA species, so-called isoacceptors, which differ in their sequence including the anticodon. Intriguingly, the cellular tRNA pool in various eukaryotic organisms comprises 41 to 55 different tRNA isoacceptors, which on the other hand implies that some of them can interact with different mRNA codons.[21] This phenomenon is explained by the so-called wobble hypothesis which states that perfect base pairing according to Watson and Crick is pivotal between the first two positions of the mRNA codon and positions 36 and 35 of the tRNA anticodon, whereas the exact Watson-Crick counterpart of the nucleobase in the third position of the mRNA is not discriminatory at position 34 of the tRNA.[22] Apart from isoaccepting tRNAs, there are also isodecoders which share their anticodon sequence and thus also their amino acid but differ in other parts of the sequence.[21]

1.2 Chemical modification of RNA

In order to comply with its versatile functions, RNA harbors an additional layer of information beyond its sequence which is implemented by a broad spectrum of enzymatically installed, post-transcriptional modifications, collectively referred to as the epitranscriptome.[23] Since the characterization of the first chemical modification more than 60 years ago, about 170 distinct RNA modifications have been described in coding and non-coding RNA from all kingdoms of life and still new structures continue to be discovered.[24–28] The chemical variety among RNA modifications is remarkable, ranging from the rather simple attachment of a methyl group, isomerization, reduction and deamination to highly complex structures in which the plethora of building blocks and metabolites that nature has at hand, such as amino acids, sugars or isoprene residues, finds new purposes (Figure 1.3).

To describe these diverse structures, a special nomenclature evolved throughout the years (Figure 1.3a). Usually, chemical alterations are indicated in relation to one of the four main nucleosides cytidine (C), uridine (U), guanosine (G) or adenosine (A). In the case of nucleobase modifications, the abbreviation of the main nucleoside is prefixed by an abbreviated form of the attached group e.g. m for methylation, s for thiolation or i for an isopentenyl group, as well as the superscripted position of the modification within the ring according to the conventional labeling of pyrimidine and purine bases, i.e. 1-methylguanosine is described as m^1G. By lining up the abbreviations of the functional groups and, if necessary, the different positions within the original nucleobase, this nomenclature is also applicable to nucleosides that carry modified residues in various positions or a chain composed of different groups in one position, as for exam-

ple in the case of 2-methylthio-$N6$-threonylcarbamoyladenosine which is abbreviated as ms^{2}t^{6}A. Furthermore, to distinguish between nucleobase modifications and modifications of the ribose in $2'$-O position, the latter are indicated by the same acronyms added as a suffix after the main nucleoside. Due to historical reasons or complexity of certain structures that cannot be described with this code, some modifications were assigned new letters of which pseudouridine (Ψ), inosine (I) and queuosine (Q) are prominent examples.[13,29]
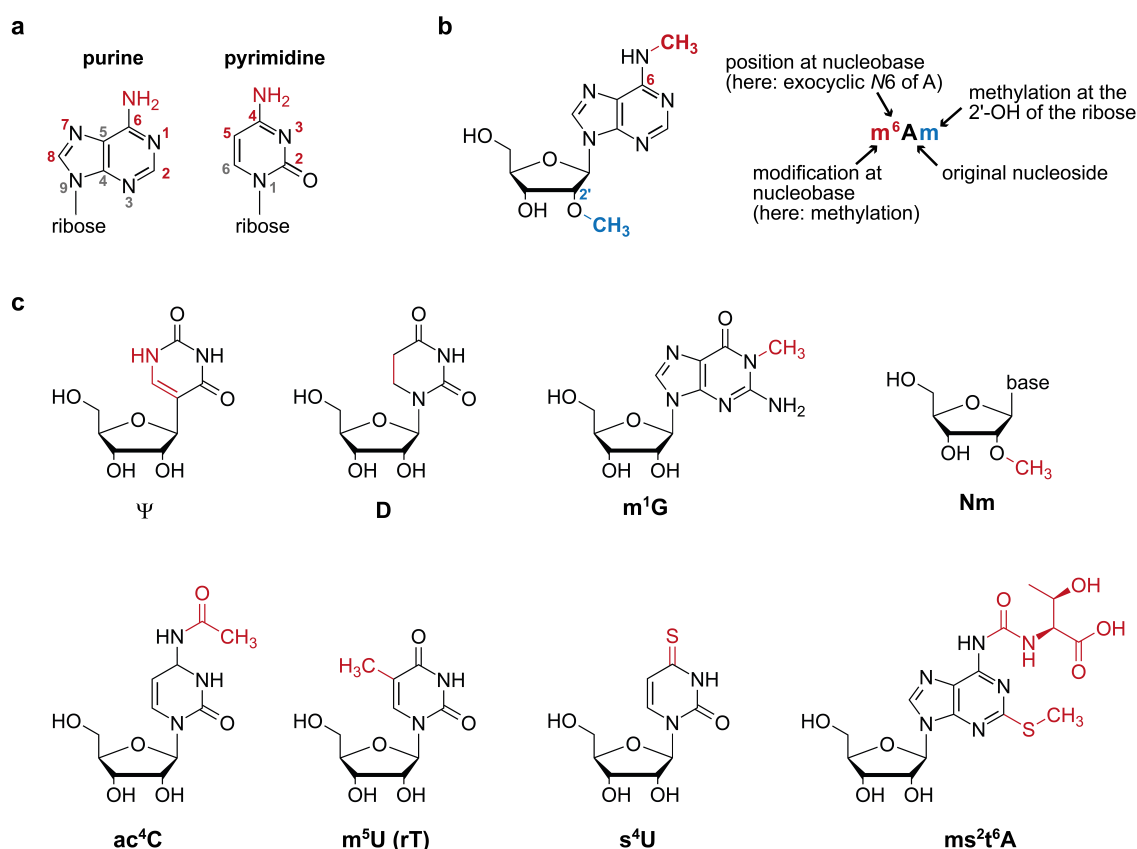

Figure 1.3: Nomenclature and structural diversity of RNA modifications. a Conventional numbering of purine and pyrimidine structures with nucleobase modification sites highlighted in color. Figure adapted.[29] **b** Nomenclature of modified nucleosides exemplified for $N6,2'$-O-dimethyladenosine (m^{6}Am). Figure adapted.[13,29] **c** Selection of RNA modifications including pseudouridine (Ψ), dihydrouridine (D), 1-methylguanosine (m^{1}G), $2'$-O-methylated nucleosides (Nm), $N4$-acetylcytidine (ac^{4}C), 5-methyluridine (m^{5}U, also known as ribothymidine (rT)), 4-thiouridine (s^{4}U) and 2-methylthio-$N6$-threonylcarbamoyladenosine (ms^{2}t^{6}A). Chemical alterations in relation to the original nucleoside are marked in red.

The following subsections provide an overview of the role of RNA modifications in the three major RNA species involved in protein synthesis, considering also chemical alterations of RNA that do not exclusively occur on purpose, i.e. with biological implications, but also in the context of RNA damage.

1.2.1 Modifications of ribosomal RNA

Ribosomal RNA makes up an essential part of the ribosome and plays an important role not only in RNA-RNA recognition for the decoding process, but also exerts the catalyzing peptidyl transferase activity.[30,31] In particular conserved regions of an rRNA that are involved in such important processes including for example the decoding site and the peptidyl transferase center are often decorated with modified nucleosides, resulting in a prevalence of 2% modified nucleosides in eukaryotic rRNA.[32–36] However, the chemical variety of rRNA modifications is rather limited (e.g. to 14 different RNA modification species in human) and mainly characterized by pseudouridines and ribose methylations but can also contain other modifications like methylation or acetylation of the nucleobase.[32,33] Intriguingly, pseudouridylation and ribose methylation are predominantly installed by small nucleolar ribonucleoprotein particles that are guided to the site of modification by small nucleolar RNA (snoRNA) while base modifications are introduced by stand-alone modification enzymes.[37,38] Modifications of rRNA are involved in biogenesis, functional folding of rRNA and stabilization of the ribosome structure as well as translation performance in terms of rate and accuracy.[39–43]

1.2.2 Modifications of messenger RNA

The core of a eukaryotic mRNA is its coding sequence which contains the amino acid sequence of the corresponding protein in the form of base triplets and has a distinct beginning and ending in the form of specific start and stop codons. The generic mRNA structure is complemented by an untranslated region (UTR) adjacent to the start and stop codon, respectively, which contains regulatory elements and a cap structure at the 5′ end as well as a poly(A) tail at the 3′ end.[44,45] Among various mRNA modifications that can occur throughout the mRNA strand (Figure 1.4), N6-methyladenosine (m^{6}A) is the most prominent and best characterized representative. Driven by the discovery of a m^{6}A demethylase, suggesting a dynamic nature of this modification, and precise m^{6}A mapping methods to analyze its transcriptome-wide distribution, m^{6}A ignited resurgent interest in the field of RNA modifications during the past years and decisively shaped the perception of an epitranscriptome with writers (RNA modifying enzymes), readers (enzymatic recognition of modifications) and erasers (removal of modified residues).[23,46–50] The m^{6}A modification occurs at an average of one to three accordingly modified sites in mammalian mRNAs and is enriched in the 3′ UTR and around stop codons.[49–51] Interestingly, the m^{6}A modification is installed co-transcriptionally, indicating an involvement of the modification already in the early steps of a transcript's life cycle in the nucleus.[52,53] Indeed, m^{6}A was shown to affect processing and alternative splicing of pre-mRNAs and to mediate nuclear export of several mRNAs by interaction with nuclear reader proteins.[54–59] Furthermore, alteration of mRNA stability and modulation of

translation initiation as well as translation efficiency have been reported to add to the wide-ranging effects of m^6A.[60–65] In line with the essential roles m^6A plays within the cellular environment, it is not surprising that numerous studies revealed a correlation between m^6A and a variety of disorders including cancer and neurodegeneration, but also autoimmune disorders, metabolic diseases among others.[66–71]

Another highly abundant modification in eukaryotic mRNA is 7-methylguanosine (m^7G), an integral element of the 5′ cap structure, which is attached to the transcript via a reverse 5′-5′-triphosphate bond and is frequently followed by one or two 2′-*O*-methylated nucleosides (Figure 1.4 a+b).[72–74] If the first position of the mRNA is a 2′-*O*-methyladenosine (Am), the nucleoside can be additionally methylated in the *N*6-position to form *N*6,2′-*O*-dimethyladenosine (m^6Am), a modification that is exclusively found in the cap and provides accordingly modified transcripts with resistance to the decapping enzyme DCP2, complementing the roles of the cap structure in transcript stability by protecting against 5′ degradation as well as recruitment of the translation initiation machinery.[46,74–76]
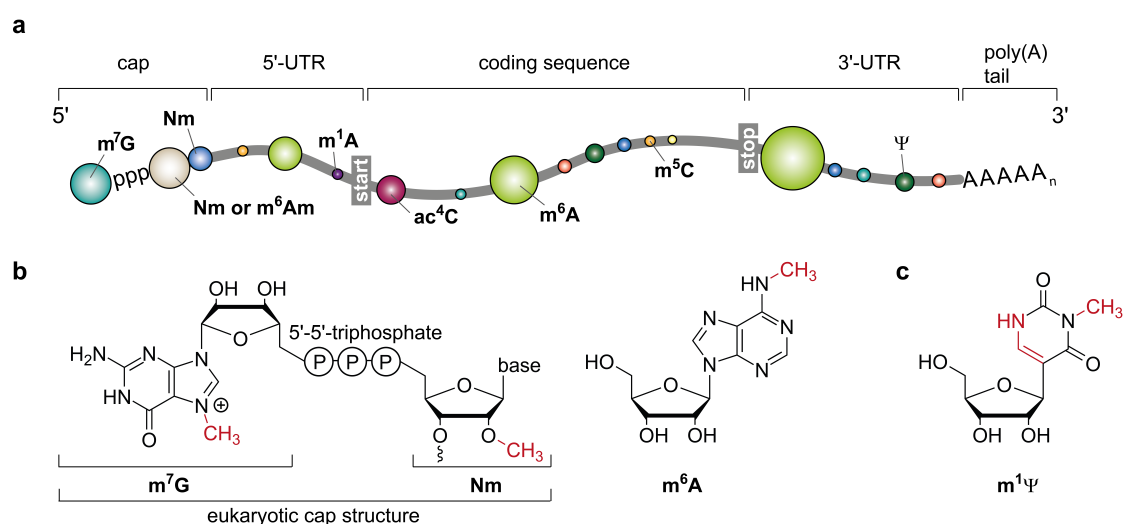

Figure 1.4: mRNA structure and modifications. a Generic structure of a eukaryotic transcript illustrating its different functional regions together with an overview of modification sites within these regions of an mRNA. The circle dimension corresponds to the abundance of the respective modification. For reasons of simplicity not all modifications were assigned. Figure adapted.[54] **b** The structures of highly abundant mRNA modifications are depicted, including 7-methylguanosine (m^7G) and 2′-*O*-methylated nucleosides (Nm) as parts of the cap structure as well as m^6A, the flagship among mRNA modifications. Further modifications that reportedly occur on mRNAs are *N*6,2′-*O*-dimethyladenosine (m^6Am), 1-methyladenosine (m^1A), *N*4-acetylcytidine (ac^4C) and 5-methylcytidine (m^5C). **c** The modification 1-methylpseudouridine ($m^1\Psi$) is an important feature of therapeutic mRNAs. Chemical alterations in relation to the original nucleoside are marked in red.

Additional nucleoside modifications that were detected throughout eukaryotic mRNA, but with a current level of knowledge far behind m^6A, include Ψ, *N*4-acetylcytidine (ac^4C), 1-methyladenosine (m^1A) and 5-methylcytidine (m^5C) as well as 2′-*O*-methylated

nucleosides, some of which are even controversially discussed in terms of abundance, distribution and biological function.[77–83] Nevertheless, the importance of being modified becomes also evident in the context of therapeutic mRNA approaches since in particular Ψ as well as its derivative 1-methylpseudouridine (m^{1}Ψ, Figure 1.4c) - alone or in combination with m^{5}C - have been associated with improved translation efficiency and reduced immunogenicity by evading activation of the innate immune system which allows cells to differentiate between self and non-self RNA species.[84–87] Impressive proof of concept is provided by the regulatorily approved mRNA vaccines from Biontech/Pfizer and Moderna, both of which contain m^{1}Ψ and are playing a key role in combating the severe acute respiratory syndrome coronavirus type 2 (SARS CoV-2) pandemic, thus showcasing the enormous potential of RNA based therapeutics.[88–90]

1.2.3 Modifications of transfer RNA

Simultaneously with the structural elucidation of the first tRNA, the first modified nucleosides were discovered, triggering intense research on (t)RNA modifications in diverse organisms, of which eukaryotes were found to generally harbor a much higher density of modifications compared to prokaryotes.[12,91] To date, with an average of eight modifications per cytoplasmic tRNA, tRNA is not only the RNA species with the highest frequency of modifications, but it also displays the greatest chemical variety among its structures.[14,92,93] A hotspot for so-called hypermodifications, which are characterized by exceptional diversity and complexity of their residues and whose insertion is often accomplished by successive activity of several enzymes, are found in the anticodon loop (Figure 1.5a).[14,91] Here, modification of the nucleoside in position 34, commonly known as the wobble position of the anticodon, directly affects the binding properties during interaction with the mRNA codon and thus the efficiency as well as the accuracy of the translational process.[94–96] One of the most elaborate structures found in this position is the guanosine-derived modification queuosine, for which a detailed section is provided at the end of this chapter. Apart from that, in particular U34 is almost invariably modified in every organism, forming the family of wobble uridines which include structures like 5-methoxycarbonylmethyl-2-thiouridine (mcm^{5}s^{2}U) and is associated with alternative decoding, giving an idea of how decoding of the degenerate code is controlled by tRNA modifications that either expand or restrict the codon recognition (Figure 1.5b).[17,97] In this context, the accurate positioning of the anticodon plays a fundamental role and is achieved by restraining the conformational dynamics of the anticodon loop via stacking interactions involving modifications in position 34 and position 37 as well as preventing unintended interactions within the anticodon loop and thus directing it into an ordered structure for stable anticodon-codon interactions.[98–100] Demonstrating that nature is without limit in compiling new structures based on diverse metabolites and other
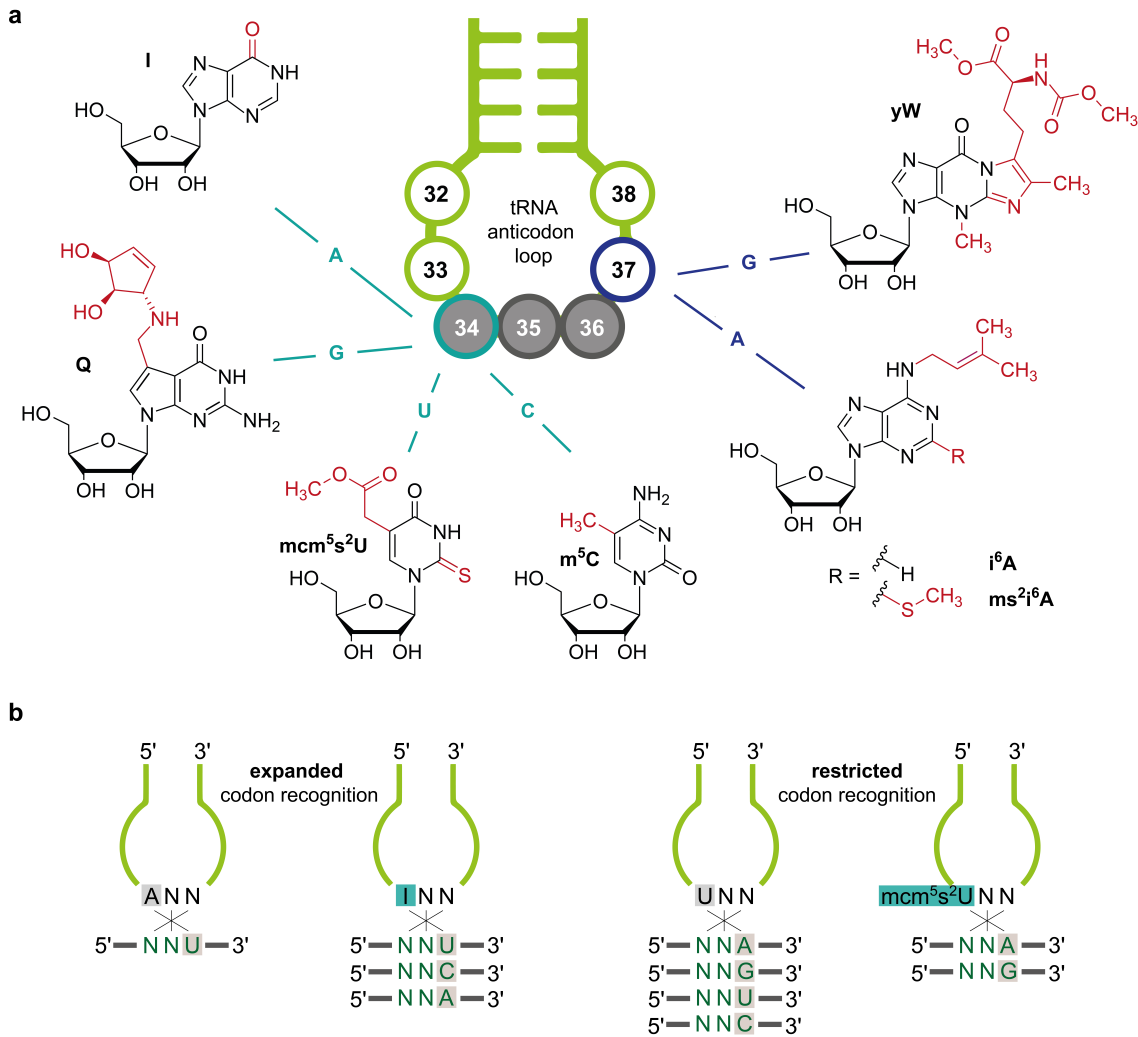
Figure 1.5: Modifications in the anticodon loop of eukaryotic tRNAs and their impact on decoding. a Selection of structurally versatile nucleoside modifications in positions 34 and 37 of the anticodon loop. Prominent modifications in position 34 are inosine (I), queuosine (Q), 5-methoxycarbonylmethyl-2-thiouridine (mcm⁵s²U) and 5-methylcytidine (m⁵C). Position 37 often harbors complex purine-derived modifications like (2-methylthio)-N6-isopentenyladenosine ((ms²)i⁶A) or wybutosine (yW). The original nucleoside prior to modification is indicated (C = cytidine, U = uridine, G = guanosine, A = adenosine) and structural alterations are marked in red, respectively. Figure adapted.[101] **b** Control of decoding by modifications in position 34 exemplified by the expanded codon recognition in the case of I34 or restricted codon-recognition in the case of mcm⁵s²U34, compared to the codon recognition in the presence of the unmodified nucleoside. Abbreviations are the same as in (a), N is generic for nucleoside. Figure adapted.[17]

building blocks within the cellular environment, the conserved purine in position 37, 3′ adjacent to the anticodon, often carries bulky residues, manifesting in adenosine hyper-modifications like $N6$-isopentenyladenosine (i⁶A) and $N6$-threonylcarbamoyladenosine (t⁶A) as well as their thiomethylated derivatives ms²i⁶A and ms²t⁶A or the guanosine modification m¹G which can undergo further modification steps to yield wybutosine (yW).[14,102] While the latter was associated with reading-frame maintenance, isopentenyladenosines in position 37 play key roles in codon-specific mRNA decoding, decisively affecting translation efficiency.[14,103,104]

Intriguingly, for some of the modifications located in the anticodon loop a certain inter-dependency was observed in a way that the presence of one or more modifications drives the formation of another, suggesting that the firstly introduced modification might serve as a recognition site for the enzyme introducing the second modification.[105,106] Like this, in *Schizosaccharomyces pombe* (*S. pombe*) and *Saccharomyces cerevisiae* (*S. cerevisiae*) ribose methylations of C32 and G34 have been shown to stimulate formation of m^{1}G37, initiating synthesis of yW, and i^{6}A or t^{6}A in position 37 have been identified as recognition sites and driving forces for efficient methylation of C32 to yield m^{3}C32.[107–109] Only recently, this very m^{3}C32 modification was revealed as a crucial feature of a subset of mitochondrial tRNAs, optimizing their structure as well as mitochondrial translation.[110,111] Consequently, imbalances in the methylation level of C32 of affected mitochondrial tRNAs have been linked to the enhanced respiratory chain activity in pancreatic cancer.[112]

Apart from modifications in the anticodon loop, there is another set of ubiquitous modifications which are highly conserved in both position and identity, and thus are namesake for certain substructures in the core region of the tRNA molecule. This refers to the D arm which is characterized by the prevalence of dihydrouridine (D) as well as the TΨC arm which comprises a conserved ribothymidine (equivalent to m^{5}U) in position 54 and a neighboring Ψ55, altogether contributing to the proper folding of the characteristic L-shaped tRNA structure and its stabilization, but also controlling conformational flexibility of the core region.[113–117] Of these modifications Ψ in particular can be found throughout the tRNA structure and is not only involved in the stabilization of the tertiary structure but also in anticodon recognition and translation efficiency.[118–120] Beyond that, a number of other modifications, mainly single or double methylations have been detected on tRNA and are associated with structural functions, guided degradation or prevention of cleavage, but also with fine-tuning of translation, the cellular stress response and thermal adaption.[121,122]

While hypomodified tRNAs lacking one or multiple modifications can be subject to rapid tRNA decay in eukaryotes and thus exonucleolytic degradation during tRNA maturation, the detection of a wide variety of hypomodified tRNAs in both healthy and pathogenic cells indicates dynamic regulation of the modification profile, depending on the cellular conditions.[17,123–125] Among the pathological consequences emerging from aberrant tRNA modification are mitochondrial diseases and neurological disorders as well as diabetes and cancer.[17,126]

1.2.3.1 Queuosine – a hypermodified nucleoside

With its cyclopentenediol moiety attached to a 7-deazaguanosine core via an amino-methyl linker, queuosine is one of the most intricate RNA hypermodifications.[127] Even

before the structure had been fully elucidated in 1975, queuosine was shown to be universally present in the wobble position 34 of tRNAs decoding asparagine (Asn), aspartate (Asp), histidine (His) and tyrosine (Tyr), all of them sharing a GUN anticodon.[128–130] Although queuosine occurs in both prokaryotes and eukaryotes with only few exceptions, it is noteworthy that only bacteria are equipped with the set of enzymes to synthesize queuosine *de novo* (Figure 1.6) while eukaryotes rely on environmental sources to salvage Q or the corresponding nucleobase queuine (q).[131,132] In *E. coli*, the initial steps of queuosine's biosynthesis, which are consecutively catalyzed by a cascade of GTP cyclohydrolase I (GCH I), QueD, QueE and QueC, lead to the conversion of guanosine triphosphate (GTP) to the precursor 7-cyano-7-deazaguanine (preQ$_0$) whose nitrile carbon is subsequently reduced to the methylene amine group by QueF.[133–137] The resulting 7-aminomethyl-7-deazaguanine (preQ$_1$) is substrate of the bacterial tRNA guanine transglycosylase (bTGT) which replaces the nucleobase guanine in position 34 of the aforementioned tRNAs with this precursor structure in a transglycosylation reaction, an exceptional mechanism of post-transcriptional RNA modification.[138,139] Once incorporated into tRNA, the preQ$_1$ nucleoside is subject to further modification by the action of QueA and QueG, ultimately yielding the queuosine structure.[140,141] While bTGT is a homodimeric enzyme which catalyzes the incorporation of precursor nucleobases but is inhibited by the fully modified nucleobase q, eukaryotic TGT (eTGT) evolved as a heterodimeric enzyme with substrate specificity broadened to the hypermodified nucleobase.[127,139,142–144] In order to provide this substrate, salvaged Q is hydrolyzed in an upstream process.[145]
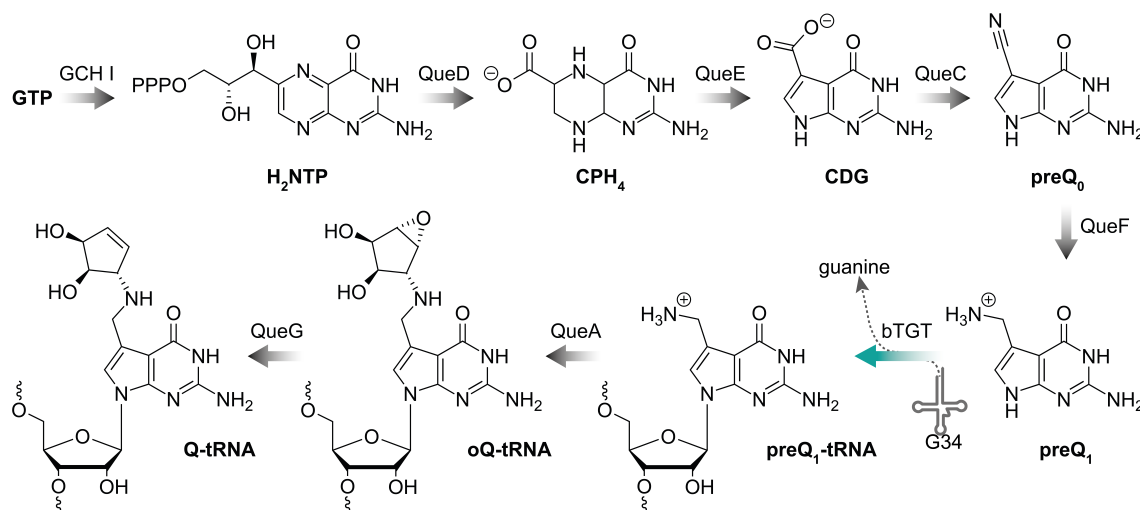

Figure 1.6: *De novo* synthesis of queuosine in *E. coli*. Illustration of the multi-step biosynthetic process, starting from guanosine triphosphate (GTP) with the enzymes being involved in each step indicated above the arrows and abbreviations for intermediates below respective structures. The transglycosylation step catalyzed by bacterial tRNA guanine transglycosylase (bTGT) is highlighted in color. Abbreviations: 7,8-dihydroneopterin triphosphate (H$_2$NTP), 6-carboxy-5,6,7,8-tetrahydropterin (CPH4), 7-carboxy-7-deazaguanine (CDG), 7-cyano-7-deazaguanine (preQ$_0$), 7-aminomethyl-7-deazaguanine (preQ$_1$), epoxyqueuosine (oQ), queuosine (Q).

Given the reported modification site in the wobble position of the anticodon, it is not surprising that Q was shown to affect the decoding process.[146,147] In this regard, computational modelling pointed to an important contribution of Q in terms of tRNA-mRNA interaction by establishing an extended network of hydrogen bonds which is also consistent with the conceptualized necessity to stabilize codon-anticodon interactions with modifications at position 34 in case of predominant base pairing interactions with less than three hydrogen bonds.[148,149] Concomitantly, several studies in *E. coli* associated Q with translational accuracy as well as enhanced viability under stress conditions, but major biological defects in the absence of Q failed to appear.[150,151] Nevertheless, in yeast and higher organisms, where Q was also linked to fidelity and speed of the translational process, there is accruing evidence for a correlation between aberrant Q modification and various pathological conditions including cancer, neuronal disorders and infections which confers Q metabolism with a certain therapeutic potential.[147,152–157] Beyond its role in translation, Q was shown to stimulate the Dnmt2 catalyzed methylation of C38 in *S. pombe* tRNA$^{\text{Asp}}$ to yield m^{5}C38 which not only prevents this tRNA from stress-induced cleavage by the ribonuclease angiogenin but was also shown to affect accurate protein synthesis, e.g. in hematopoiesis.[158–161] In the context of this modification circuit, the role of Q34 is suggestively attributed to a conformational change of the anticodon loop that might provide the methyltransferase with improved access to C38 and allow ideal positioning of its substrates on the one hand or stabilize the transition state on the other hand, in both cases supporting the catalytic efficiency of the enzyme.[159,162,163]

1.2.4 Unintended chemical modification: RNA damage

In the light of maintaining genomic integrity and association with various diseases, especially cancer, in case of any instability, damage to DNA and the conserved repair pathways to cope with it have been extensively studied for years.[164–167] Not surprisingly, RNA is also vulnerable to damage caused by endogenous and exogenous factors, including ultraviolet (UV) light, reactive oxygen species (ROS) and alkylating agents, but much less studied because the consequences of damage to transient RNA molecules were not expected to be as severe as the consequences of DNA damage.[168,169] However, even the smallest structural changes can lead to an impaired functionality of RNA, e.g. in terms of base pairing during protein synthesis, and such damage events were also linked to various neurodegenerative disorders and cancer.[168,170–172] Furthermore, deleterious factors have been observed to affect RNA more severely than DNA, probably due to the exposed Watson-Crick face of the nucleobases in single-stranded parts of several RNA species.[173–175] The overall stability of RNA also suffers from the combination of its single-stranded nature and the $2'$-hydroxyl group at the ribose which make RNA susceptible to alkaline hydrolysis.[176,177]

The exposure of RNA to UV light not only causes crosslinking, formation of photohydration and oxidation products but also affects the RNA modification landscape (Figure 1.7).[178–181] A special role in crosslinking of *E. coli* tRNA is played by the widespread s^4U in position 8, which can absorb radiation of comparatively high wavelengths and upon exposure to UV-A (310-400 nm) forms a crosslink with a spatially close cytidine in position 13, acting as a UV-A radiation sensor that triggers stress response.[182–186] Furthermore, a recent study unraveled an increased lability of tRNA modifications containing amino and oxy groups as well as sulfur in the oxidative environment that is generated under UV-A influence.[180,187]

While physiological concentrations of reactive oxygen species are involved in signaling pathways, increased levels of ROS, i.e. oxidative stress, cause oxidative damage to diverse cellular components including RNA.[188–192] Oxidized nucleosides (Figure 1.7) with their most abundant representative being 8-oxoguanosine (8oxoG) interfere with RNA-RNA interactions and lead to translation errors and ribosome stalling.[168,193,194] Accumulation of oxidatively damaged RNA is often associated with aging and neurodegeneration, thus questioning the efficiency of turnover mechanisms.[173,195–200]

Apart from direct reaction of nucleobases with the highly reactive hydroxyl radical or other ROS, simultaneous damage to other biomolecules can also affect nucleobases, e.g. formation of etheno (ε) adducts like 1,*N*6-ethenoadenosine arising from the reaction with lipid peroxidation products.[169,201,202] Another cause for the formation of such etheno adducts was traced to exposure to noxious chemicals such as vinyl chlorides whose metabolites possess a high alkylating potential.[203,204] Although the vast majority of alkylating agents has their origin outside the cell, there is some evidence that ubiquitous methyl group donors like *S*-adenosyl methionine (SAM) nonenzymatically methylate nucleic acids.[205,206] Apart from reaction with exocyclic amino or keto groups, alkylating damage most severely affects the nitrogen atoms within the heterocyclic ring of the nucleobase (Figure 1.7).[169] Here, in particular alkylation of *N*1 in the case of purines and *N*3 in the case of pyrimidines have grave consequences since the formation of hydrogen bonds in the context of base pairing is blocked which impairs RNA-RNA interactions and was linked to neurodevelopmental disorders and cancer.[171,172,193] The high importance of functional RNA-RNA interactions is emphasized by the repair of such lesions which is catalyzed by the oxidative demethylase AlkB, a versatile enzyme that emerged in the context of DNA repair at first, but meanwhile was shown to have various substrate sites on RNA, including m^1A and m^3C lesions.[207,208] Since methylation in particular is highly dynamic and does not only occur in the context of unspecific alkylation damage but also as a true RNA modification introduced by enzymes, the development of sophisticated methods is necessary to distinguish between those.[209] Although to date not studied in detail, there is an overlap between both fields as alkylation
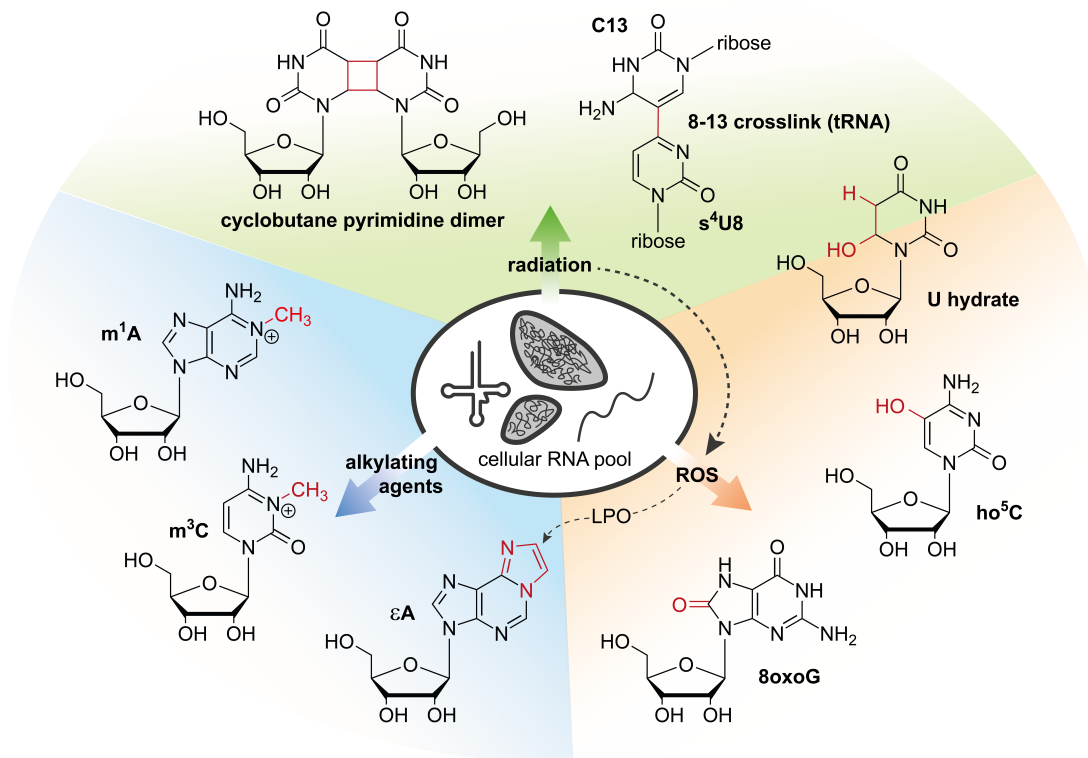
Figure 1.7: Overview of structures resulting from several forms of RNA damage. Direct consequences of irradiation of RNA include the formation of crosslinks such as cyclobutane pyrimidine dimers, a crosslink between a 4-thiouridine (s⁴U) in position 8 and a cytidine in position 13 of bacterial tRNA or photohydration products like uridine (U) hydrate. Oxidative damage products represented by 8-oxoguanosine (8oxoG) and 5-hydroxycytidine (ho⁵C) result from reaction with reactive oxygen species (ROS) that can be generated as a consequence of irradiation (indicated with dashed arrow) but also have other sources. The exposure to endogenous or exogenous alkylating agents leads to the formation of RNA lesions such as 1-methyladenosine (m¹A), 3-methylcytidine (m³C) or ethenoadenosine (εA). The latter can also be formed by reaction with products of lipid peroxidation (LPO) after oxidative stress. Structural alterations in relation to the original nucleoside are marked in red.

damage does not spare RNA modifications which was recently shown by generation of 2-methylthiocytidine (ms^{2}C) from 2-thiocytidine (s^{2}C) and 4-methylthiouridine (ms^{4}U) from s^{4}U upon treatment of cells with methylating agents.[210,211] However, to assess the effects of damaged RNA modifications within the complex cellular network, further research in this field is required.

1.3 Strategies for the detection and characterization of RNA modifications

In order to approach the bigger picture of RNA modifications and understand their functions, it is inevitable to have methods at hand that do not only allow to identify and quantify these structures but also to localize the sites of modification within a given RNA.[212] Throughout the years, high-throughput sequencing and liquid chromatography-tandem mass spectrometry (LC-MS/MS) evolved as the two pillars of RNA modification analytics that complement each other in these requirements.[213] While RNA sequencing methods are characterized by preserving the sequence information and pinpointing modification sites at nucleotide resolution, their applicability is limited to a small set of modified structures so far and prediction of modification sites from big data makes room for controversial discussions, which may be overcome in the near future by recent developments towards the use of nanopores for single-molecule sequencing.[82,83,214–216] In contrast to that, widely used LC-MS/MS methods rely on nucleoside-level analysis and enable discovery of new modifications, as well as unambiguous identification and simultaneous quantification of the entire spectrum of known RNA modifications, albeit at the expense of sequence information.[217–220] Moreover, metabolic and chemo-enzymatic labeling are valuable tools in structural elucidation of RNA modifications and evaluation of their dynamics and fate within the cellular environment as well as in creating basic prerequisites for further derivatization via so-called click chemistry.[221,222] Due to their special importance in the context of the present study, the following sections provide details on the LC-MS/MS analysis of nucleosides and diverse approaches to label RNA.

1.3.1 LC-MS/MS analysis of nucleosides

While antecedent success had been achieved in chromatographic separation of RNA hydrolysates using reversed-phase (RP) high performance liquid chromatography (HPLC), the advent of electrospray ionization (ESI) in the 1980s revolutionized mass spectrometric analysis of biomolecules.[223–226] Soon after, the compatibility of this gentle ionization technique with LC systems was integrated in new methods for nucleoside analysis and established the basis of contemporary LC-MS/MS approaches.[217] Prior to the actual analysis, RNA samples require complete enzymatic hydrolysis which is commonly achieved by the addition of an enzyme cocktail including endonucleases like nuclease P1 as well as a phosphodiesterase and in a first step yields nucleotides whose turnover by an alkaline phosphatase consecutively releases nucleosides.[227] With regard to modification analysis, the conditions of enzymatic hydrolysis must be chosen carefully, since labile modifications such as the cyclic form of t^6A (ct^6A) may be destroyed or disguised by artifacts whose formation also impedes the discovery and identification of new RNA

modifications.[228–230] After RNA hydrolysis, the nucleoside mixture is separated by liquid chromatography, before the eluate enters the coupled mass detector which is typically employed in positive ion mode but can differ in the type of mass detector depending on the research question.[220] While time-of-flight (TOF) analyzers and orbitraps are high-resolution mass spectrometers that provide exact masses, triple quadrupole (QQQ) mass spectrometer are characterized by outstanding sensitivity and are frequently used for the analysis of low abundant RNA modifications.[231] Generally, the set-up of a QQQ coupled to a LC system (Figure 1.8) starts with an ESI source to ensure simultaneous ionization and desolvation of the analytes from the surrounding liquid before entering the high vacuum area of the mass spectrometer. The positively charged ions (precursor ions) are then guided via the ion transfer capillary and the octopole, summarized as ion optics, into the first quadrupole and filtered according to their mass-to-charge ratio (m/z), i.e. only ions that comply with the predefined characteristics are allowed to move on to the collision cell. The latter is historically often called a quadrupole, although most modern mass spectrometers contain a hexapole in this place and is filled with collision gas. Collisions between the latter and the entering precursor ions in a process called collision induced dissociation (CID) generate fragmented ions (product ions) which are characterized by a reduced m/z compared to the precursor ions, representing a so-called mass transition. Subsequently, the product ions enter another quadrupole for further filtering according to programmed m/z values before the ions reach the detector. The high sensitivity of triple quadrupoles is a result of the double filtering for target analytes and concomitant noise reduction.[231–233] Different scan modes of the QQQ (Figure 1.8) allow to address different research questions including discovery and identification of new RNA modifications as well as quantification of known modifications. Three methods using different modes of the QQQ that are relevant for the present study are illustrated in Figure 8 and described below. Despite disparities in method programming and data output, most methods for nucleoside analysis exploit predominant fragmentation at the *N*-glycosidic bond, a feature that is shared by the vast majority of nucleoside structures with only very few exceptions and results in a characteristic loss of ribose (-132 Da) or 2′-*O*-methylated ribose (-146 Da).[217]

1.3.1.1 Neutral loss scan (NLS)

Generally, neutral loss scans are characterized by detecting all ions whose CID results in a predefined neutral loss which, advantageously, also applies to the aforementioned fragmentation at the *N*-glycosidic bond, opening up for a wide-ranging screening within the landscape of nucleoside structures.[27,234] During an NLS, precursor ions are initially filtered according to a predefined mass range in the first mass analyzer. After fragmentation of precursor ions in the collision cell, the second mass analyzer ensures that only

compounds exhibiting the mass transition of interest reach the detector, which is technically achieved by synchronization of both quadrupole mass analyzers.[235] On the one hand this allows generation of highly informative data sets that can even cover so far unknown nucleoside species and assist in their structural elucidation, but on the other hand the high number of simultaneous scans might negatively affect sensitivity, a case that becomes even more relevant when including further scan segments for nucleosides with aberrant fragmentation behavior, such as queuosine or pseudouridine.[217]
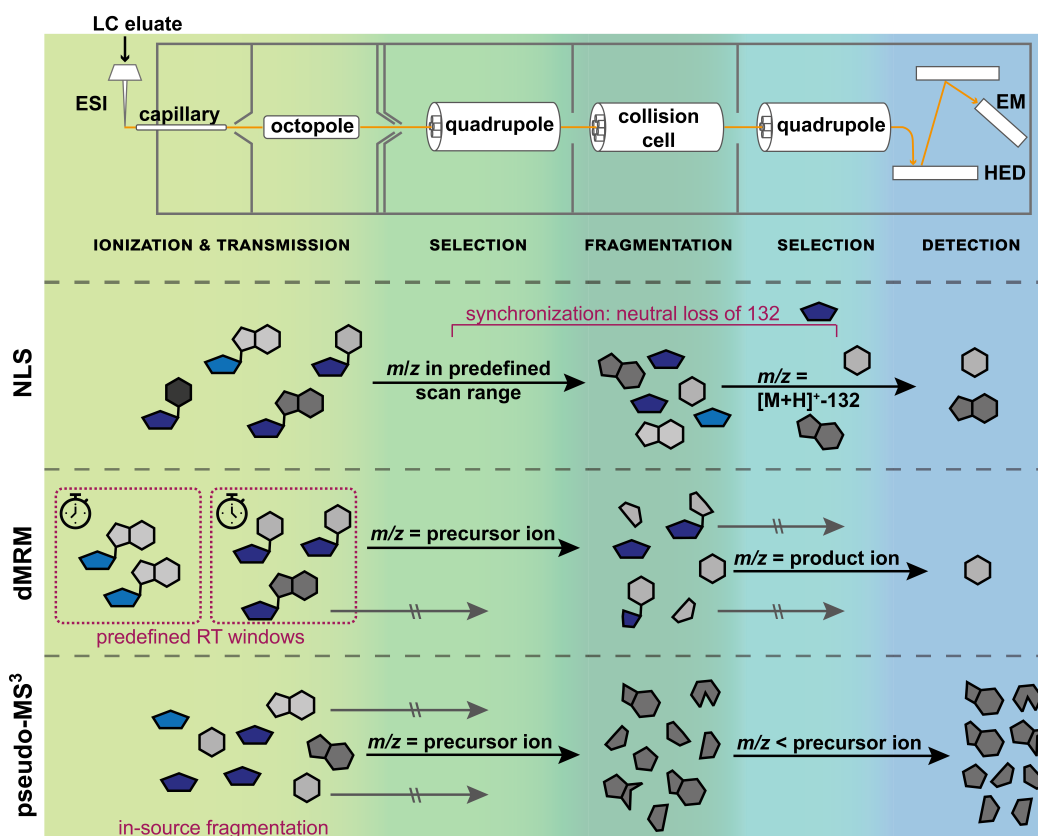

Figure 1.8: The QQQ mass spectrometer and its scan modes. The upper part shows the set-up for LC-MS/MS analysis with a QQQ mass spectrometer, illustrating the way of the ion beam (shown in yellow) that is generated from the entering LC eluate by ESI and directed via ion optics, including an octopole, into the first quadrupole. Here, only ions of certain mass-to-charge-ratios are selected and pass through to the collision cell, where these ions are fragmented. Prior to detection via a high-energy dynode (HED) and an electron multiplier (EM), another quadrupole ensures selection of the product ions of interest. The lower part illustrates a selection of different scan modes including the neutral loss scan (NLS), the dynamic multiple reaction monitoring (dMRM) and the pseudo-MS³ scan. Special features of the respective modes are highlighted in dark pink (RT = retention time). Sugar compounds are illustrated as blue pentagons with dark blue ones corresponding to an unmodified ribose (132 Da) and light blue ones corresponding to a 2′-O-methylated ribose (146 Da) while the nucleobases are depicted in the shape of their structure (gray).

1.3.1.2 Dynamic multiple reaction monitoring (dMRM)

In contrast to the NLS mode, dMRM methods exclusively detect predefined nucleoside species. While this limits data depth to a certain end, it enables the QQQ to reach its full potential of sensitivity to accurately quantify modified nucleosides at extremely low abundances, even in the subfemtomolar range. Within this mode, retention time windows for each nucleoside to be monitored are specified and set the period for detection of predefined mass transitions, optimally reducing concurrent monitoring events that might limit the sensitivity of the triple quadrupole and enabling simultaneous analysis of a broad spectrum of modifications within one biological sample. The required information on chromatographic as well as mass spectrometric characteristics is accessible via preceding analysis of reference substances which also help to circumvent the non-quantitative nature of mass spectrometry by calibration curves. Furthermore, the addition of internal standards, prevalently the stable isotope labeled forms of the nucleosides of interest (see section 1.3.2.1), that account for ion suppression and divergent ionization characteristics are indispensable in the context of absolute quantification.[218,219,231]

1.3.1.3 Pseudo-MS³ scan

The pseudo-MS³ scan is a multistage fragmentation approach exploiting fragmentation at the most labile bond already in the ionization source in a process called in-source fragmentation. This is achieved by application of increased voltages, conferring the ionization source with the characteristics of an additional mass analyzer and thus imitates an MS/MS/MS approach.[236,237] As a consequence of that and in contrast to the previously described methods, the ions entering the first mass analyzer are not at the stage of nucleoside ions but already at the stage of nucleobase ions. This enables the generation of characteristic nucleobase fragment ions in the collision cell, which result in a unique fragment spectrum, corresponding to a spectral fingerprint of one specific nucleoside species. Like this, such methods can demonstrate structural similarities to known modifications, provide evidence on the presence of certain functional groups by means of mass transitions or validate structure proposals for new modifications by comparison with the appropriate organically synthesized substance, emphasizing the high value of pseudo-MS³ scans for the structural elucidation of new RNA modifications.[27] The concept of such multistage fragmentation approaches is also applied in the context of high-resolution mass spectrometry where it is often referred to as higher-energy collisional dissociation (HCD) and can provide even more information on structural components of a new nucleoside species. Significant efforts in this field have been made by the Limbach group who did not only leverage this fingerprinting to circumvent the challenge of unambiguous identification of positional isomers as for example in the case of methylcytidines (3-, *N*4- and 5-methylcytidine), but also introduced a spectral network which

contains the fingerprints of 100 known nucleoside structures and is meant to facilitate future structural elucidation of novel RNA modifications, as exemplified within their work.[238,239]

1.3.2 RNA labeling

1.3.2.1 Metabolic stable isotope labeling of RNA

One of the first and still one of the most prominent experiments employing stable isotope labeling in the field of nucleic acids is the Meselson-Stahl experiment, which was published in 1958 and proved the semi-conservative nature of DNA replication.[240] Since then, the interest in methods based on stable isotope labeling, especially labeling with ^{13}C, ^{15}N, ^{34}S but also deuterium (^2H) has increased continuously and has become an integral part of modern RNA modification analysis by LC-MS/MS.[231,234] Here, stable isotope labeled nucleosides are especially predestined for use as internal standards in quantitative methods because they coelute with their unlabeled counterparts due to identical physicochemical properties, but exhibit a different mass transition.[218,241–243] An elegant way to prepare such an internal standard, which can accommodate the chemical diversity of RNA modifications but avoids the time-consuming organic synthesis of the corresponding compounds, is to isolate uniformly labeled RNA that is hydrolyzed for nucleoside LC-MS/MS.[218,244] Uniform labeling approaches rely on supplying the culture of simple organisms, such as *E. coli* or *S. cerevisiae*, in minimal medium containing only a sole source of respective atoms with stable isotope labeled nutrients, which are consequently incorporated into the biomolecules synthesized by these cells.[222,245] In addition to its value for the quantification of known RNA modifications, stable isotope labeling is also a key technology with regard to modification identification, in that conclusions can be drawn about the elemental composition of the compound of interest through corresponding mass shifts (Figure 1.9a). Within the past years, the combination of such approaches with LC-MS/MS contributed significantly to the identification and structural elucidation of hitherto unknown RNA modifications, including geranylated uridines, ms^{2}C and 2-methylthiomethylenthio-*N*6-isopentenyladenosine (msms^{2}i^{6}A).[27,210,234,246] The structural elucidation of the latter served as proof of principle for an NLS-based screening for new RNA modifications by the Helm group which resulted in a list of 37 potentially new modifications in *E. coli* and highlighted the existence of many more hitherto uncharacterized nucleoside structures, even in extensively studied model organisms, yet to be discovered.[27] Apart from complete labeling of RNA and its modifications, it is also possible to specifically use labeled bioavailable metabolites with a known role in RNA modifying processes to examine the presence of certain groups. For example, supplementation of the growth medium with the deuterated precursor of the ubiquitous methyl group donor SAM, methionine-CD$_3$, can reveal the presence of a methyl

or methylene group within a modified structure.[27,210,247,248] In view of the high price of labeled metabolites, but even more the unavailability of individual substances of interest, a remedy can be found by adding unlabeled metabolites to labeled growth medium which leads to differential labeling of the biomolecules that are synthesized from these metabolites and ultimately a difference in *m/z* that can be visualized by LC-MS/MS.[27] As a sophisticated application of differential labeling, the Kaiser lab established nucleic acid isotope labeled coupled mass spectrometry (NAIL-MS), a method that defies the transient nature of RNA and its modification status and whose hallmark is the accessibility of RNA modification dynamics, for example in different growth phases or upon exposure to an alkylating agent. In principle, the culture of a simple organism starts in a specifically labeled growth medium until the cells reach a certain optical density or are exposed to a pulse which is experimentally accompanied by replacement of the starting
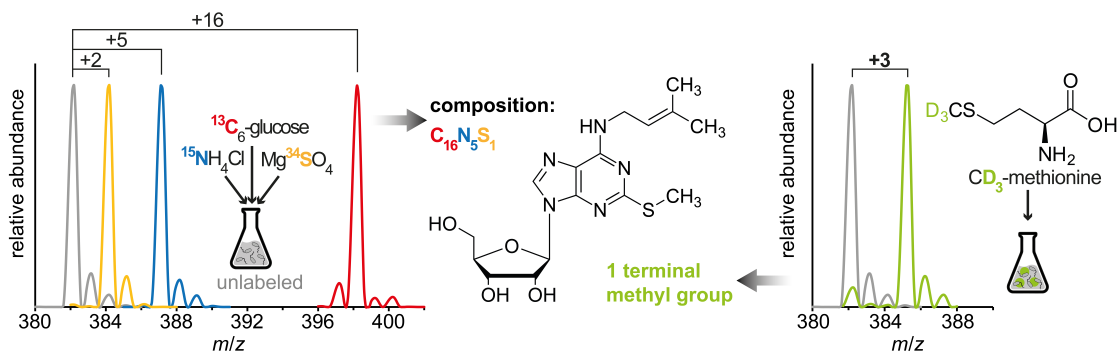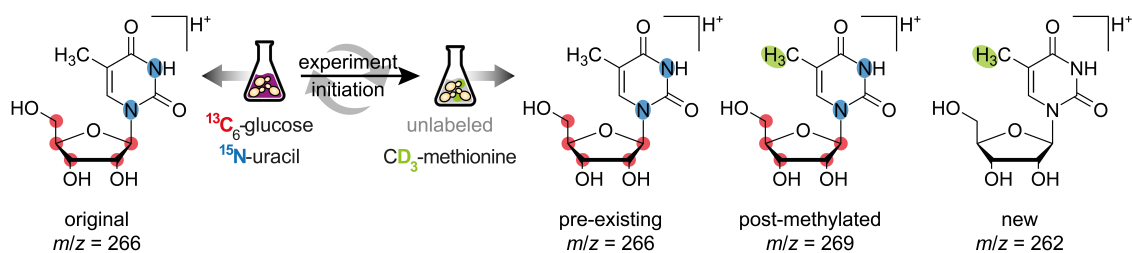

Figure 1.9: Combination of LC-MS/MS analysis and stable isotope labeling strategies. a Replacement of an exclusive source of certain atoms with its isotope labeled version (left side) or supplementation of an *E. coli* culture with labeled metabolite precursors, as for example methionine which eventually becomes part of the universal methyl group donor SAM (right side) results in uniform labeling which can provide information on the elemental composition and structural features of a nucleoside structure by respective *m/z* shifts as illustrated for the tRNA modification ms²i⁶A. **b** NAIL-MS approach to study the methylation of nucleosides in different growth phases of *S. cerevisiae*. Initially, a specifically labeled medium is used (the separately added but required amino acids including methionine are unlabeled), but is exchanged for a completely unlabeled medium supplemented with CD_3-methionine in order to target methylation as soon as the culture reaches the growth phase of interest and corresponding atoms are incorporated into RNA from this time point on. This results in altered mass-to-charge-ratios in subsequent LC-MS/MS analysis and thus allows to distinguish between methylated nucleosides present from the start on (pre-existing), original nucleosides that were methylated after the defined time point (post-methylated) and newly synthesized methylated nucleosides (new). Figure adapted.[249,250]

medium with a differently labeled medium (Figure 1.9b). Like this, in subsequent analysis, NAIL-MS goes past static modification analysis and allows not only to access growth-dependent modification dynamics, but also the fate of RNA undergoing specific pulses, as well as their effect on newly transcribed RNAs or the stimulation of demodification processes in the context of stress responses. Consequently, NAIL-MS also enables differentiation between modifications that are introduced natively by an enzymatic process or arise from deleterious effects.[209,210,222,250–253]

1.3.2.2 Functionalization of RNA and click chemistry

Long before the sibling fields of click chemistry and bioorthogonal chemistry entered the public stage in the context of awarding the pioneers in these fields with the Nobel Prize in chemistry in 2022, these methods were widely applied in all fields of life and material science research.[254–260] The term click chemistry was coined by Barry Sharpless to describe a type of chemistry that is as convenient and satisfying as a simple plug-in connection between two objects, whose linkage is acoustically confirmed by a "click" and whose declared flagship is the copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC) which results in the formation of a triazole (Figure 1.10a).[261–263] Apart from the use of copper(I) as a catalyst and the application-dependent addition of a stabilizing ligand, the basic prerequisite for this reaction is the presence of an azide in one reactant and the presence of a terminal alkyne in the other reactant, so-called bioorthogonal functionalities, which react in a highly specific and effective manner with each other, but are neither influenced by a complex biological environment nor hinder the physiological processes taking place in this environment.[264–267]

In addition to statistic incorporation of these biocompatible functionalities into RNA by utilization of accordingly modified building blocks during *in vitro* transcription, it is also possible to exploit the promiscuity of various enzymes such as methyltransferases or TGT for the purpose of site-specific modification of RNA molecules *in vitro* by using modified analogues as co-substrates (Figure 1.10b), or to similarly functionalize RNA *in vivo* or *in cellulo* by providing a cell culture or animals with appropriately functionalized metabolite-like substances (Figure 1.10c).[221,268,272,274–277] Once one of these functionalities is installed on RNA, there is a broad spectrum of possible applications, among them purification via affinity tags or facilitated visualization as well as tracking by fluorescence tags or determination of modification sites.[267,271,272,276,278–280] Nevertheless, the toxicity of copper towards living systems confers the CuAAC with a non-bioorthogonal nature and limits its applicability, thus alternate metal-free reactions like the strain-promoted azide-alkyne cycloaddition (SPAAC) that leverages a strained alkyne as a driving force or the inverse electron-demand Diels-Alder cycloaddition (IEDDA) between a tetrazine and an alkene have been developed and established as bioorthogonal labeling
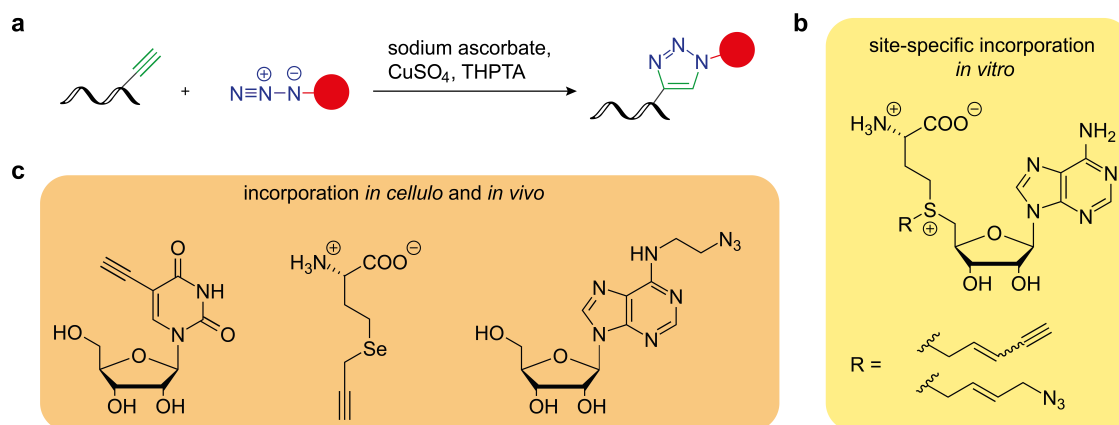
Figure 1.10: Click chemistry in the context of RNA. a Exemplary reaction scheme for the copper(I)-catalyzed azide-alkyne cycloaddition of an alkyne-functionalized RNA (green) with an azide (blue) that can be attached to a variety of molecules, represented by the red circle, including fluorescence and affinity tags. The catalytic Cu(I) is generated in situ via reduction of Cu(II) with ascorbate and then stabilized with water-soluble ligands as for example tris(3-hydroxypropyltriazolylmethyl)amine (THPTA). **b** Examples for co-substrates of methyltransferases bearing a clickable functionality that were site-specifically incorporated into different RNA species *in vitro*.[268−270] **c** Selection of metabolite-like substances that were successfully incorporated into cellular RNA in *in cellulo* and *in vivo* experiments.[271−273]

approaches that enable to track biological processes *in vivo*, including those associated with nucleic acids.[221,277,281−285]

Apart from these two-step covalent labeling approaches via click reaction, an alternative route for labeling of RNA with dyes, affinity tags or light-activatable groups is a direct one-step enzymatic incorporation of accordingly functionalized surrogates, despite their bulkiness and strong deviation from the natural substrates at first glance and, of course, in close dependency on the degree of promiscuity displayed by the corresponding enzyme.[221] An example for this kind of labeling is the RNA-TAG (transglycosylation at guanosine) method which was developed and used in many ways by the Devaraj group, leveraging the broad tolerance of bTGT to various substrate analogs *in vitro*, thus incorporating them reliably and site-specifically into RNA transcripts comprising the appropriate short recognition motif.[286−290]

The entirety of methods to label and functionalize RNA is also of interest with regard to therapeutic approaches, where especially the fate as well as the translation efficiency of mRNA and controllability of its pharmacological effect are of high interest. Consequently, these issues were addressed in several studies by labeling mRNA at its hallmarking structures at the 5′ and the 3′ end or at inserted structural recognition motifs to exploit enzymatic reactions, revealing promising results that emphasize the high potential of these methods in future studies in this emerging field.[277,289,291−296]

1.3.3 Previous work: a list of 37 potentially new RNA modifications

As mentioned before, the Helm group combined an NLS method for the ultrasensitive detection of new nucleoside structures with stable isotope labeling experiments and created a list of 37 potentially new RNA modifications in *E. coli*, one of which was already characterized in detail and identified as msms^{2}i^{6}A.[27,297,298] Due to the importance of this previous work for the present study, this chapter will provide some details on the existing data set of these "candidates", which are named according to their mass-to-charge ratio.

In order to warrant the authenticity of the detected compounds as true RNA modifications, rigorous filtering criteria were applied after the initial NLS screening, including exclusion of structures that were detectable in *in vitro* transcribed RNA that is supposed to be free of any modified nucleoside. Another important step was a comparison with the Modomics database,[24] allowing to identify and eliminate known modifications or isomers as well as adducts thereof. Furthermore, consistency of the candidate's MS signal in samples obtained from different sample preparation methods was mandatory to defend a position on the list. The first version of the data set included apart from the method-related *m/z* value and the retention time also information on the elemental composition (content of carbon, nitrogen and sulfur atoms) as well as first structural implications like the canonical nucleoside prior to modification, determined in a differential feeding experiment by supplying a ^{15}N-culture with an unlabeled nucleobase, or the presence of terminal methyl groups introduced by SAM as previously described. This initial data was complemented by HRMS data and the resulting calculated sum formula for a selection of candidates and a first screening (singly performed experiments) for biological characteristics including for example the occurrence in different organisms or altered abundance of the candidates in knockout (KO) strains of prominent modification enzymes. However, it is worth noting that the robustness of results from biological experiments that were carried out only once should be treated with caution.[27,297,298]

1.3.3.1 Candidate 338

For this candidate, LC-MS/MS analysis of stable isotope labeled samples implied an adenosine-based structure with an elemental composition of 13 carbon atoms, five nitrogen atoms and one sulfur atom, which was substantiated by the exact mass, establishing a sum formula of $C_{13}H_{15}N_5O_4S$.[297–299] Furthermore, a CD$_3$-methionine feeding experiment and the fragmentation pattern in a pseudo-MS^3 scan provided evidence for a terminal methyl group within this molecule. In addition, the sensitivity of candidate 338 in the knockout strains of MiaA and MiaB, as detected within the first screening

for biological characteristics, pointed to a possible structural relation to the hypermodification ms^{2}i^{6}A which is synthesized in a consecutive reaction of these two enzymes in *E. coli* (Figure 1.11a).[300–303] Altogether these characteristics led to a first structure proposal containing a thiomethyl group in the *C*2 and a terminal alkyne group in the *N*6-position of the original adenosine structure (Figure 1.11b), but attempts to detect this putative structure by click reaction failed.[299]
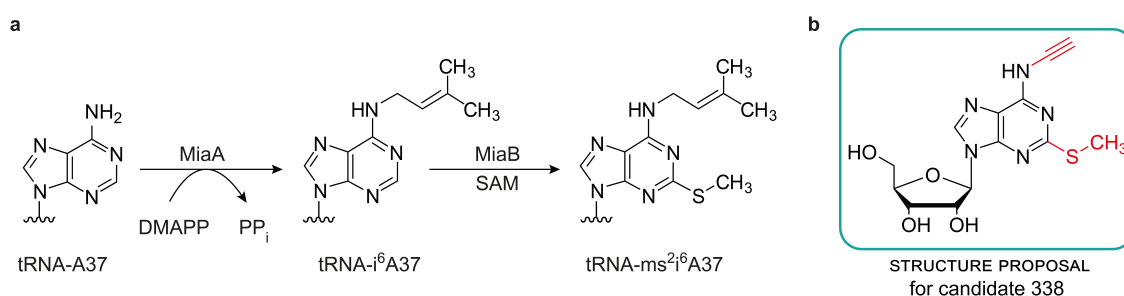

Figure 1.11: Introduction of ms^{2}i^{6}A in tRNA and first structure proposal for candidate 338. a Enzymatic reactions catalyzed by MiaA and MiaB which are involved in the modification of adenosines in position 37 (A37) of several tRNAs. MiaA catalyzes the transfer of the isopentenyl chain from dimethylally pyrophosphate (DMAPP) onto the *N*6-position of the adenosine to yield *N*6-isopentenyladeonsine (i^{6}A). Subsequent thiomethylation in the *C*2-position of the adenosine is mediated by the SAM-dependent ironsulfur-enzyme MiaB and results in 2-methylthio-*N*6-isopentenyladenosine (ms^{2}i^{6}A).[300–303] **b** Proposed structure for candidate 338 with structural modifications compared to the underlying adenosine marked in red.

1.3.3.2 Candidate 470

In the case of candidate 470, the NLS analysis of stable isotope labeled samples suggested a cytidine- or uridine-derived structure, bearing an exceptionally high content of 18 carbon atoms and five nitrogen atoms.[297,298] Of note, the differential labeling approach for determination of the main nucleoside species prior to modification is limited in the distinction between the two pyrimidine-based structures since uridine serves as a precursor in the biogenesis of cytidine.[304–306] Based on the exact mass, the sum formula was calculated as $C_{18}H_{23}N_5O_{10}$, which together with the pyrimidine-based nature of the compound pointed to a dimeric structure. This was further reinforced by a pseudo-MS^3 scan that indicated an additional loss of a second ribose moiety after the initial loss.[299] However, since neither the identity of the monomers nor the type of their connection was elucidated so far, a schematic illustration (Figure 1.12a) will be used to depict candidate 470 within this thesis. Information obtained from the first biological screening included occurrence of this putatively dimeric structure in *E. coli* while it was substantially reduced in eukaryotic samples like HeLa cells and absent in samples from higher mammals, e.g. mouse liver. Moreover, the abundance of candidate 470 was affected in knockout strains of two different sulfurtransferases, namely MnmA and ThiI, which introduce sulfur in position 2 or 4 of a uridine, respectively (Figure 1.12b).[297,307]
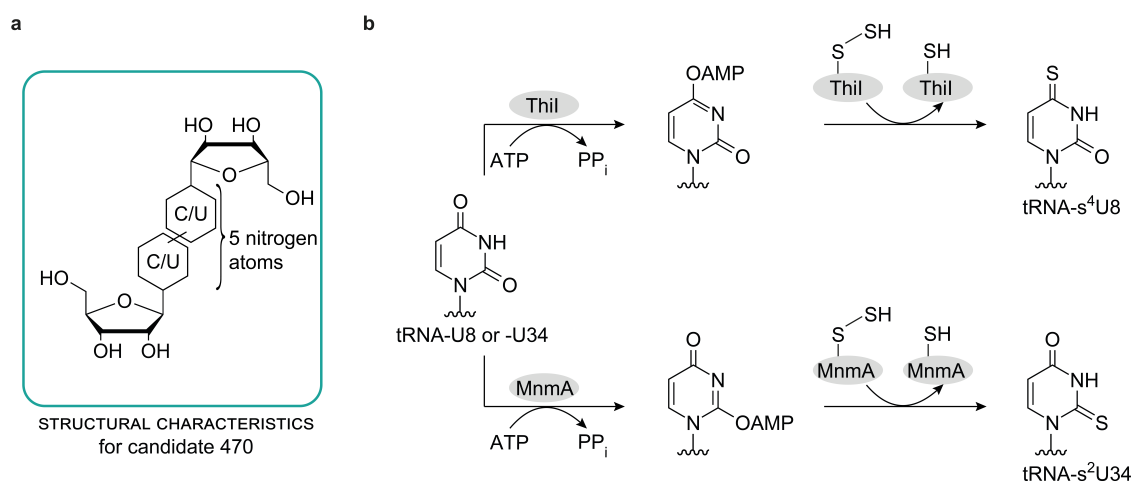
Figure 1.12: Introduction of s^{4}U and s^{2}U in tRNA and structural characteristics for candidate 470.
a Schematic illustration of structural characteristics for candidate 470 depicting the dimeric character, the possibly underlying main nucleosides and the amount of nitrogen atoms. **b** Enzymatic reactions catalyzed by ThiI and MnmA which are sulfurtransferases that target different sites on tRNA and introduce sulfur in different positions of uridines. In a first step both enzymes use adenosine triphosphate (ATP) to activate the carbon atom in the respective position of the uridine and subsequently a persulfidic sulfur is inserted in the respective position. ThiI catalyzes formation of s^{4}U in position 8 of several tRNAs while MnmA-mediated 2-thiolation of uridines in position 34 yields a transient s^{2}U which is the first step in the formation of complex wobble uridine modifications.[307] Abbreviations: AMP = adenosine monophosphate; PP$_i$ = pyrophosphate.

2 Motivation and Objectives

In recent years, research in the field of RNA modifications was reignited by increasingly sophisticated instrumentation and concomitant breakthroughs in deciphering the biological roles of modified nucleoside structures. Both, the application of modification-containing mRNA vaccines in the fight against the SARS CoV-2 pandemic as well as accumulating reports linking individual modification species to pathological processes are recent circumstances, underlining the essentiality of RNA modifications in the regulation of cellular processes and in therapeutic approaches. To date, more than 170 modified structures are known and not only the number of known RNA modifications in all domains of life is ever growing, but also their chemical diversity. In particular, tRNA excels in the number of newly discovered structures as well as in complexity of modified nucleosides and is the site for hypermodifications to occur. The structural elucidation and biological characterization as well as the targeted manipulation and concomitant biological implications of compounds belonging to this intriguing group of nucleoside structures, together with the development and refinement of the methods required for this purpose have been the focus of this thesis.

The first part of this work deals with the structural elucidation and characterization of new modifications in tRNA, based on previous work in the Helm group which has established a list of potentially new modifications including the elucidation of the new hypermodification msms^{2}i^{6}A as a proof of principle and first glimpses on sum formulas and structural as well as biological features of other modification candidates.[27,297–299] Starting with the selection of two particularly promising candidates from the list, namely candidate 338 and candidate 470, both implying characteristics of potential hypermodifications, the present work aimed at their structural elucidation and detailed investigation of their occurrence and origin. This task ought to be addressed by combination of diverse biochemical experiments including stable isotope labeling and highly sensitive LC-MS/MS analysis as well as organic synthesis of proposed structures as reference substances and their alignment with the naturally occurring compound. Application of the aforementioned methods on samples with diverse biological backgrounds as for example modification enzyme knockouts or metabolic engineering approaches were supposed to contextualize modification candidates with a potential biological function. In addition to the circumstances of the formation of the modification candidates and possibly in-

volved enzymes and due to the outlined roles of hypermodifications in tRNA, it was also of interest to evaluate the functionality of these structures in the fundamental process of translation.

Another major objective of this work focused on the known hypermodification queuosine. Despite its early discovery in the 1970s, queuosine is one of the modifications for which biological implications are scarce, but due to the unique nature of its introduction into RNA via a transglycosylation reaction and the demonstrated promiscuity of the TGT enzyme in *in vitro* experiments, it offers tempting opportunities to be manipulated *in vivo*. First studies reported a therapeutic effect after suggestive but only indirectly proved incorporation of a substrate analogue in an animal model, albeit with disregard to the effects of such a substitution at the molecular level.[155] Therefore, a workflow that allows on the one hand to demonstrate the successful incorporation and the distribution of such derivatives *in vivo* and at the same time to assess the integration into cellular processes, among them translation and modification circuits, was in demand. In order to meet these requirements and provide a molecular basis for potential therapeutic applications in the future, a combination of surrogate labeling experiments with modern and sensitive detection methods such as bioconjugation of RNA in a click reaction, sequencing or LC-MS/MS analysis was aimed at in this work.

3 List of Publications

Publications included within this thesis

1) **A New Bacterial Adenosine-Derived Nucleoside as an Example of RNA Modification Damage**

   Bessler, L., Vogt, L.-M., Lander, M., Dal Magro, C., Keller, P., Kühlborn, J., Kampf, C. J., Opatz, T., Helm, M., *Angewandte Chemie International Edition* **2023**, 62, e202217128.

   Author contributions: L. Bessler planned and performed the majority of the experimental work including diverse LC-MS/MS experiments and corresponding sample preparation and analysis for structural elucidation, based on preliminary work by ██████████ and ███████. Furthermore, L. Bessler synthesized the $ms^2\varepsilon A$ reference, verified the identity of the native and the synthetic structure and developed and applied a combination of a known metabolic engineering approach and stable isotope labeling to determine the biological origin of the modification. ███████ and █████████ provided tRNA samples from polysome preparations. █████████ and ███████ performed NMR analysis of the synthetic reference. █████████ performed HRMS measurements. The manuscript was written by L. Bessler and ████████.

2) **Functional Integration of a Semi-synthetic Azido-queuosine Derivative into Translation and a tRNA Modification Circuit**

   Bessler, L.*, Kaur, N.*, Vogt, L.-M.*, Flemmich, L., Siebenaller, C., Winz, M.-L., Tuorto, F., Micura, R., Ehrenhofer-Murray, A., Helm, M., *Nucleic Acids Research* **2022**, 50, 10785-10800.
   *These authors contributed equally and are listed in alphabetical order.

   Author contributions: L. Bessler, ███████ and █████████ performed the majority of the experimental work. L. Bessler established the *in vivo* feeding experiments in *E. coli* as well as detection via gel electrophoresis with subsequent click reaction and performed LC-MS/MS analysis of different samples. Investigation on

phenotypic consequences of analog feeding *in vivo* (growth) and polysome preparations were performed by ▮▮▮▮▮▮▮. ▮▮▮▮▮▮ performed all *in vitro* experiments and *in vivo* experiments in *S. pombe* and their subsequent analysis. ▮▮▮▮▮▮ and ▮▮▮▮▮▮ provided the clickable compounds. ▮▮▮▮▮▮ performed HeLa culture and polysome preparations. ▮▮▮▮▮▮▮ and ▮▮▮▮▮▮▮ helped with *E. coli* polysome preparations. The manuscript was written by L. Bessler, ▮▮▮▮▮ and ▮▮▮▮▮▮.

Further publications (not included within this thesis)

3) **RNA Marker Modifications Reveal the Necessity for Rigorous Preparation Protocols to Avoid Artifacts in Epitranscriptomic Analysis**

Richter, F., Plehn, J., Bessler, L., Hertler, J., Jörg, M., Cirzi, C., Tuorto, F., Friedland, K., Helm, M., *Nucleic Acids Research* **2022**, 50, 4201-4215.

Contribution as co-author: L. Bessler analyzed and interpreted LC-MS/MS data for absolute quantification of RNA modifications.

4) **Bacterial tRNA 2′-*O*-methylation is Dynamically Regulated under Stress Conditions and Modulates Innate Immune Response**

Galvanin, A., Vogt, L.-M., Grober, A., Freund, I., Ayadi, L., Bourguignon-Igel, V., Bessler, L., Jacob, D., Eigenbrod, T., Marchand, V., Dalpke, A., Helm, M., Motorin, Y., *Nucleic Acids Research* **2020**, 48, 12833-12844.

Contribution as co-author: In teamwork with ▮▮▮▮▮▮, L. Bessler optimized and performed isolation of individual tRNA species by affinity column chromatography.

4 Results and Discussion

4.1 Structural elucidation and characterization of 2-methylthio-1,*N*6-ethenoadenosine

Although more than 170 RNA modifications are known to date, the ever-expanding number of newly identified structures, even in well-characterized model organisms such as *E. coli*, implies that an end is not yet in sight.[24,210,246,308] In line with this, former members of the Helm group established a pipeline for the identification and structural elucidation of new RNA modifications based on LC-MS/MS analysis and stable isotope labeling experiments which, after rigorous filtering steps among candidate compounds, resulted in a list of potentially new RNA modifications.[27,234,297,298] This preliminary work by ███████████████ and ████████████ provided first structural features like elemental composition and parent nucleoside of the potentially new RNA modifications, including a compound with a mass-to-charge ratio of 338. Further work to characterize this candidate within my master thesis provided an exact mass (measurements performed by ███████████████, Johannes Gutenberg-University Mainz, Germany) which allowed calculation of the candidate's sum formula and confirmed the presence of a terminal methyl group that had been suggested after a CD_3-methionine labeling experiment and, together with sensitivity in a corresponding knockout in a single experiment pointed to a thiomethyl group in the C2 position of adenosine.[297,299] Starting from this point, I re-investigated the possibly involved enzymes which established the thiomethyl group in the *C*2 position. Subsequently, the instability of *N*-ethynylamines known from literature as well as a deuterium exchange experiment visualizing the amount of fast exchanging, i.e. heteroatom bound, protons provided the crucial clue that the underlying adenosine was complemented by an additional ring structure, shaping the structure proposal of the hypermodified 2-methylthio-1,*N*6-ethenoadenosine (ms^{2}εA).[309] In order to unequivocally verify the identity of the naturally occurring compound, I synthesized ms^{2}εA (corresponding NMR analysis was performed by ███████████████, ██████ group, Johannes Gutenberg-University Mainz, Germany) and collated the LC-MS/MS behavior of the natural and the synthetic substance which finally confirmed the suspected structure. Due to the association of such etheno-adducts with damage to nucleic acids, it was necessary to critically question whether this compound

is an intentionally synthesized modification with a biological function or rather a damaged variant of another modification. Since $ms^2\varepsilon A$ showed parallels in structural and metabolic characteristics to the known hypermodification ms^2i^6A, we hypothesized that the ring structure might be generated from an initially present isopentenyl residue and thus originate from the isoprene pathway. Direct proof of this hypothesis via labeling was hampered by the poor bioavailability of isoprene precursors when supplied to an *E. coli* culture. Consequently, I applied a metabolic engineering approach to introduce an artificial pathway for isoprenoid biosynthesis decoupled from native hemiterpene metabolism and combined it with stable isotope labeling.[310] This allowed me to trace back the origin of the two carbon atoms forming the etheno bridge to a former isopentenyl moiety and together with the sensitivity in the corresponding enzyme knockout strains implied that ms^2i^6A is the precursor of this new structure. While the hypermodified adenosine structure ms^2i^6A is regularly present in position 37 of certain tRNAs and plays a key role in codon-specific mRNA decoding and thus influences translation efficiency, absolute quantification showed a very low abundance for $ms^2\varepsilon A$. Additionally, analysis of polysomal tRNA (samples provided by ▮▮▮▮▮▮▮▮▮ and ▮▮▮▮▮▮▮▮ from the ▮▮▮▮ group), i.e. tRNAs actively involved in translation, indicated that tRNAs carrying $ms^2\varepsilon A$ are functionally impaired. These factors ultimately led to the classification of $ms^2\varepsilon A$ as RNA modification damage, a newly introduced term to describe chemical alterations of RNA modifications that are accompanied by a loss of function and are detrimental to the cell.

RNA Modifications   Hot Paper

A New Bacterial Adenosine-Derived Nucleoside as an Example of RNA Modification Damage

Larissa Bessler, Lea-Marie Vogt, Marc Lander, Christina Dal Magro, Patrick Keller, Jonas Kühlborn, Christopher J. Kampf, Till Opatz, and Mark Helm*

Abstract: The fields of RNA modification and RNA damage both exhibit a plethora of non-canonical nucleoside structures. While RNA modifications have evolved to improve RNA function, the term RNA damage implies detrimental effects. Based on stable isotope labelling and mass spectrometry, we report the identification and characterisation of 2-methylthio-1,*N*6-ethenoadenosine (ms²εA), which is related to 1,*N*6-ethenoadenine, a lesion resulting from exposure of nucleic acids to alkylating chemicals in vivo. In contrast, a sophisticated isoprene labelling scheme revealed that ms²εA biogenesis involves cleavage of a prenyl moiety in the known transfer RNA (tRNA) modification 2-methylthio-*N*6-isopentenyladenosine (ms²i⁶A). The relative abundance of ms²εA in tRNAs from translating ribosomes suggests reduced function in comparison to its parent RNA modification, establishing the nature of the new structure in a newly perceived overlap of the two previously separate fields, namely an RNA modification damage.

The ensemble of post transcriptional modifications of cellular RNA, dubbed the epitranscriptome, features a large chemical variety. About 170 distinct chemical structures are known from various species,[1] and new structures emerge at high frequency.[2–4] Among RNA species, tRNA modifications contribute most to the chemical diversity and newly discovered structures. Increasingly sophisticated instrumentation has driven the discovery of new species occurring at substoichiometric quantities. Yet further non-canonical nucleoside structures are known from the field of nucleic acid damage, although many more of these structures are

[*] L. Bessler, L.-M. Vogt, M. Lander, Dr. C. Dal Magro, Dr. P. Keller, Prof. Dr. M. Helm
Institute of Pharmaceutical and Biomedical Sciences (IPBS), Johannes Gutenberg University Mainz
Staudingerweg 5, 55128 Mainz (Germany)
E-mail: mhelm@uni-mainz.de

Dr. J. Kühlborn, Dr. C. J. Kampf, Prof. Dr. T. Opatz
Department of Chemistry, Johannes Gutenberg University Mainz
Duesbergweg 10–14, 55128 Mainz (Germany)

being intensively scrutinised in DNA than they are in RNA. Most of these structures derive from radiation, oxidative or alkylating agents.[5,6] Rather than investigating damage to canonical nucleosides, a few recent reports focus on irradiation[7] and alkylating damage[3,8] to post-transcriptional modifications.

Given that tRNAs are the most highly modified RNA species, they are also the most probable place for such "RNA modification damage" to occur. Considering the high functional importance of certain RNA modifications, e.g. in the tRNA anticodon, their damage might arguably cause more severe harm than random damage to canonical nucleosides. Consequently, one might suspect the existence of repair mechanisms akin to those described for DNA and RNA.[9] Indeed, 2-methylthiocytidine (ms²C), an alkylating damage to 2-thiocytidine in *Escherichia coli (E. coli)* was shown to be reverted by AlkB,[3] an enzyme from a family with reported function in RNA and DNA repair.[6,10] On the other hand, in the case of 4-methylthiouridine (ms⁴U), repair was AlkB-independent.[8]

While a biological origin of alkylating agents was traced to lipid peroxidation,[11] nucleic acid alkylation from human-made agents such as vinyl halides was extensively studied early on in vitro and in vivo.[12–14] A relevant point in case was investigated upon emergence of unusual liver tumours in the 1970s. These occurred in workers from the polymer industry, with a causal connection to exposure to vinyl chloride.[15] Vinyl chloride was shown to be metabolised to 2-chloroacetaldehyde (CAA),[16] a highly mutagenic reagent that reacts with nucleobases, resulting in the formation of etheno (ε)-adducts such as 1,*N*6-ethenoadenine in DNA[12,17] and in RNA.[13,18] We here report the detection and structural elucidation of a new ε-adenosine, namely ms²εA, in bacterial tRNA. Surprisingly, the etheno fragment was not the result of alkylating damage, but oxidatively derived from the isoprene moiety in ms²i⁶A. An analysis of the distribution between the cellular tRNA pool and tRNAs on actively translating polysomes indicated decreased functionality, causing us to classify ms²εA as RNA modification damage.

We previously reported a list of unknown ribonucleosides consisting of potentially new RNA modifications, that were identified by combination of various stable isotope labelling experiments with a so-called neutral loss scan (NLS) by liquid chromatography-tandem mass spectrometry (LC-MS/MS) using a triple quadrupole (QQQ) mass spectrometer.[2] One elucidated structure was msms²i⁶A (2-methylthiomethylenethio-*N*6-isopentenyladenosine), a

GDCh

Communications
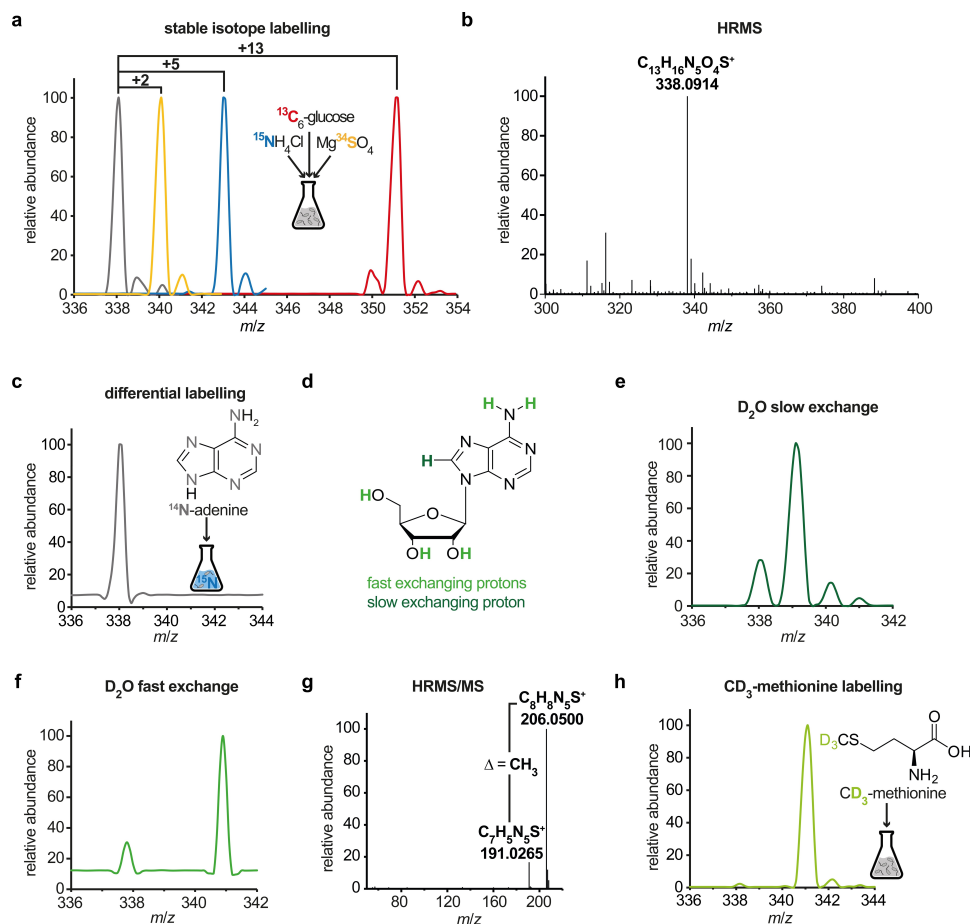
Angewandte International Edition *Chemie*

Figure 1. Structural characterisation of candidate 338 by LC-MS/MS. Unless otherwise stated, abundances were set in relation to the highest signal in the respective mass spectrum (relative abundance). a) Overlay of QQQ mass spectra at 27.7 min corresponding to candidate 338 from a NLS of digested labelled (^{13}C = red, ^{15}N = blue, ^{34}S = yellow) and unlabelled (grey) *E. coli* tRNA. Mass shifts indicating the number of respective atoms within the molecule are assigned. b) Mass spectrum and ion formula of candidate 338, calculated from the exact mass detected by HRMS analysis of digested tRNA enriched for candidate 338. c) QQQ NLS mass spectrum of tRNA isolated from a ^{15}N-labelled *E. coli* culture supplemented with ^{14}N-adenine at 27.7 min. d) Overview of slow (dark green) and fast (light green) exchanging hydrogen atoms in adenosine. e) QQQ NLS mass spectrum after incubation of RNA samples with D$_2$O (slow exchange). f) QQQ NLS mass spectrum after the use of a D$_2$O LC-buffer (fast exchange). g) HRMS fragmentation pattern of candidate 338, as well as calculated ion formulas of the generated fragments which differ by a methyl group. h) QQQ NLS mass spectrum of candidate 338 in tRNA isolated from *E. coli* cells supplemented with CD$_3$-methionine.

highly lipophilic derivative of the hypermodification ms^{2}i^{6}A found at position 37 of certain bacterial tRNAs. The structural elucidation established "inverse" feeding experiments as highly useful, where bacteria were raised in "heavy" media containing e.g. ^{13}C-glucose as only carbon source. Incorporation into RNA of any metabolite consisting of "light" ^{12}C-carbon added to the culture was thereby easily detectable as a mass difference corresponding to the

number of carbons incorporated. While application to nucleobases was uncomplicated, compounds that were not naturally taken up from the medium included prenylation reagents and precursors. Thus, while we could indirectly demonstrate the presence of an isopentenyl residue by mass spectrometry and the use of the knockout strain of the prenyltransferase MiaA, there was no direct way to monitor prenylation. This was of interest in the context of another
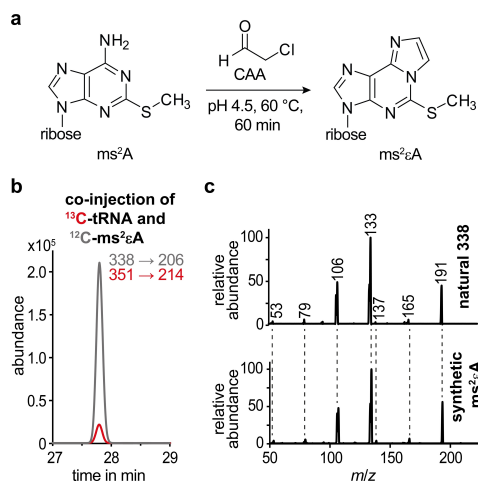
Communications

Figure 2. Structural confirmation of ms²εA by LC-MS/MS analysis of the synthesised compound. a) Reaction scheme displaying the synthesis of ms²εA, starting from 2-methylthioadenosine (ms²A). b) QQQ LC-MS/MS analysis of a digested ¹³C-labelled tRNA sample (red) spiked with the synthetic ¹²C-compound (grey). An overlay of the extracted ion chromatograms at a mass transition of *m/z* 338→206 and *m/z* 351→214 is shown. c) Comparison of fragmentation patterns of candidate 338, recorded from digested native tRNA (upper panel), and the synthesised reference ms²εA (lower panel) in a QQQ pseudo-MS³ scan.

candidate nucleoside species from the list, which displayed a mass-to-charge ratio of 338 (herein referred to as candidate 338). It eluted at 27.7 min, significantly after *N6,N6*-dimethyladenosine (m⁶·⁶A) but before *N6*-isopentenyladenosine (i⁶A), from the reversed-phase column. Stable isotope labelling revealed that this compound was elementally constituted of 13 carbon atoms, 5 nitrogen atoms and one sulphur atom which, in combination with the results from high-resolution mass spectrometry (HRMS), suggested a sum formula of $C_{13}H_{15}N_5O_4S$ (Figure 1a+b). Inverse feeding identified adenosine as parent nucleoside prior to modification (Figure 1c).

Comparison of adenosine to the constitution of candidate 338 showed a difference of C_3H_2S and two double-bond equivalents. Given that the initial NLS had provided the fragmentation pattern of an unsubstituted ribose, the above formula implied that at least some of the nitrogens of the parental nucleoside carried substituents for the original hydrogens. To locate the modification sites, we performed slow and fast deuterium exchange experiments, which provide insights into the substitution of the C8-position and the substitution of heteroatoms, respectively (Figure 1d).[2,19] Interestingly, while the 8-position was suggested to be unmodified (Figure 1e), all nitrogens bound to fast exchanging hydrogen atoms in the parent adenosine were suggested to carry substituents in candidate 338 (Figure 1f). HRMS fragmentation experiments indicated the loss of a methyl

group after loss of the ribose (Figure 1g). The presence of a methyl group was further substantiated by a corresponding +3 *m/z* shift in a CD₃-methionine labelling experiment (Figure 1h).

Given the presence of a sulphur atom in conjunction with a methyl group in a modified adenosine structure, we suspected a thiomethylation in position 2 (as for example in ms²i⁶A), which would typically be catalysed by the iron-sulphur enzyme MiaB. The absence of candidate 338 in the MiaB knockout established a thiomethyl moiety in the adenine-C2 position (Figure 1g and S1). However, this only left two further carbon atoms to be assigned to the structure of candidate 338. Considering the double-bond equivalents and the fact that none of the nitrogen atoms was bound to fast exchanging protons, the most plausible structure proposal contained an additional ring structure, i.e. an etheno bridge linking the exocyclic *N6* with *N1* of the nucleobase. A theoretically conceivable triple bond at the exocyclic *N6*-position would have been incompatible with the instability of *N*-ethynylamines[20] and the D₂O fast exchange results (Figure 1f). With only one plausible structure left, we synthesised ms²εA, adapting a protocol from Bhatt et al. (Figure 2a).[21] LC-MS/MS analysis of the ¹³C-labelled *E. coli* tRNA digest, spiked with the synthetic compound, showed co-elution of the synthetic compound and its isotopologue, confirming the hypothetical structure (Figure 2b). Ultimately, the perfectly matching pseudo-MS³ spectra of candidate 338, enriched from digested *E. coli* tRNA, and the synthesised ms²εA demonstrated identity of the molecular structures of the naturally occurring candidate 338 and the synthesised compound ms²εA (Figure 2c).

The unusual tricyclic structure of ms²εA evoked the question on the metabolic origin of the two carbon atoms of the etheno bridge. Based on the involvement of MiaB and the metabolic as well as structural proximity to ms²i⁶A, we speculated that this bridge might be the vestigial relic of an originally present *N6*-isopentenyl moiety. We hence tested RNA isolated from a knockout of the MiaA enzyme. The latter catalyses RNA prenylation of the *N6*-position of an adenosine, resulting in i⁶A, using dimethylallyl pyrophosphate (DMAPP), an isoprene cofactor that is provided via the methylerythritol-phosphate (MEP) pathway. We did not detect any ms²εA in the MiaA knockout (Figure S1), suggesting that the carbon atoms of the etheno bridge might indeed originate from the MEP pathway. Formation of ms²εA would involve ms²i⁶A as a hypothetical precursor, with its isoprene moiety being the carbon source of the etheno bridge (Figure 3a). To verify this hypothesis, we isolated 30 tRNA species and found ms²εA only in those 9 tRNAs known to carry ms²i⁶A (Figure S2).

For direct clarification, we needed to label the isoprene moiety in i⁶A/ms²i⁶A by supplying a "light" C5 compound (i.e. composed of ¹²C) to the isoprene pathway in an inverse feeding experiment (see above). To circumvent uptake problems of the negatively charged isoprene cofactors DMAPP and isopentenyl pyrophosphate by *E. coli*, we turned to a metabolic engineering approach that required inhibiting the native hemiterpene biosynthesis pathway by addition of fosmidomycin. To supply a "light" C5 surrogate,
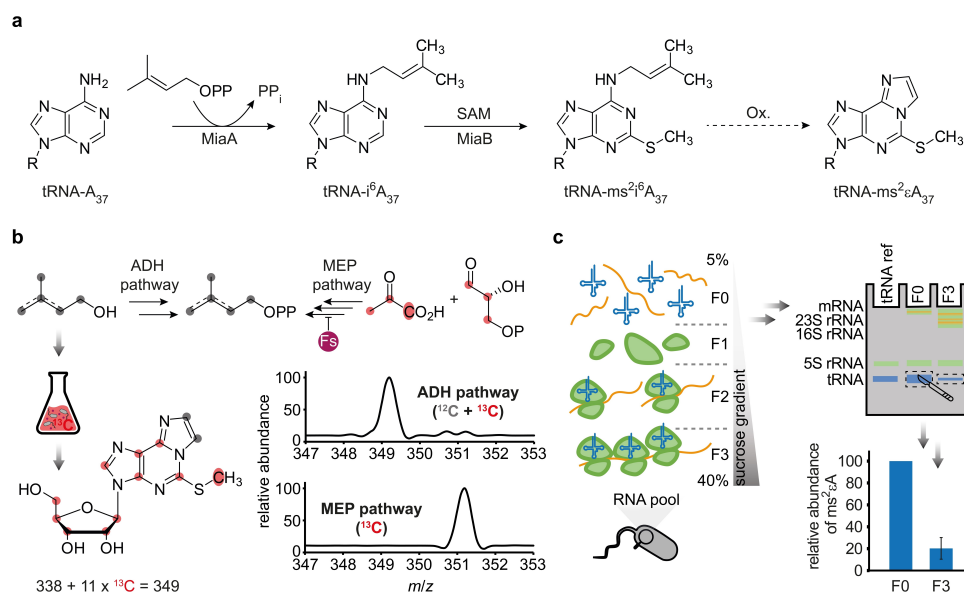
Figure 3. Biological origin and impact of ms²εA. a) Biosynthesis of ms²i⁶A and proposed generation of ms²εA. b) Inverse isoprene labelling experiment and mass spectra of subsequent QQQ NLS analysis: red (¹³C) and grey (¹²C) markers in the structure formula illustrate the mass-to-charge ratio upon inverse feeding (ADH pathway). In the untreated control sample, the native MEP pathway leads to the completely ¹³C-labelled structure. Fs: fosmidomycin, PP$_i$: pyrophosphate, SAM: *S*-adenosylmethionine. c) Constitution of the cellular RNA pool with its most abundant representatives tRNA, messenger RNA (mRNA) and ribosomal RNA (rRNA) and their distribution upon sucrose gradient fractionation. Relative LC-MS/MS quantification of ms²εA in tRNA purified from F0 corresponding to the cellular tRNA pool and F3 corresponding to actively translating tRNAs via gel elution. Normalised peak areas were related pairwise to the respective F0 fraction (set to 100%) and the average of three independent biological replicates was calculated.

we adapted a protocol to introduce an artificial alcohol-dependent hemiterpene (ADH) pathway for isoprenoid biosynthesis into *E. coli* and combined it with stable isotope labelling.[22] Bacteria cultured in ¹³C-labelled medium were supplemented with the bioavailable ¹²C-alcohol precursors dimethylallyl alcohol and isopentenol (detailed description in Supporting Information 1.1.3). LC-MS/MS analysis of prenylated nucleosides from isolated tRNA was then conducted comparing a "heavy" ¹³C-culture with a similar culture supplemented with the ¹²C-alcohols. A mass difference in the signals of i⁶A and ms²i⁶A indicated incorporation of the C5 alcohols into the hemiterpene pathway (Figure S3). Importantly, a mass shift of −2 units (m/z 349 versus m/z 351) was observed for the ms²εA signal (Figure 3b). This shows that the etheno bridge originates from the MEP pathway, and thereby that ms²εA is derived from ms²i⁶A.

Synthetic ms²εA (Figure 2) as calibration standard permitted absolute quantification e.g. in purified tRNA$^{Leu}_{(BAA)}$, showing that only 0.06% of these tRNA molecules contained ms²εA. This low abundance might signify that the presence of ms²εA in tRNAs is a spurious side product that might be non-functional or even detrimental to the cell. We investigated this hypothesis by comparing the ms²εA content of the cellular tRNA pool with that of

tRNAs from actively translating polysomes, which are considered functional "by definition". The observed significant reduction of ms²εA in polysomal tRNA relative to free tRNA (Figure 3c), was interpreted to mean that tRNAs carrying this derivative of ms²i⁶A cannot fully perform their designated function and are excluded from translation by yet unknown mechanisms. This makes ms²εA an RNA modification damage of ms²i⁶A.

The new nucleoside structure elucidated here is already the second one derived from ms²i⁶A.[2] Previous identification of msms²i⁶A demonstrated a presumed side reaction of the iron-sulfur cluster enzyme MiaB, involving radical chemistry. In the formation of ms²εA from ms²i⁶A, the type of chemistry catalysed is truly remarkable, involving oxidative cleavage of the isoprene double bond in analogy to an ozonolysis. We are currently pursuing identification of the responsible enzyme, an endeavour that is hampered by the low abundance of the structures involved and by the experimental challenge of isotopic labelling of the isoprene moieties in RNA. In the present case, this was achieved after a significant effort devoted to funnelling exogenous C5 alcohols into bacterial hemiterpene synthesis. This approach, adopted from the work of Lund et al.,[22] allowed to show that the etheno bridge within the third heterocycle derives

Communications

Angewandte
Chemie

from the isoprene moiety in ms²i⁶A. While successful in the case at hand, its application is sophisticated in practice and in need of detailed optimisation before becoming a tool for routine investigations. Most biological implications of our findings remain to be discovered, but the nucleoside quantification data of polysome-bound tRNA are strongly suggestive of damage to tRNA functionality and a detrimental character of ms²εA and warrant its designation as RNA damage.

While older examples of damaging reactions on tRNA modifications have received the low attention of isolated findings,[23] we conclude that such events are indeed more frequent than hitherto assumed. The proposed term "RNA modification damage" would not only cover structural alteration of intact RNA modifications as a consequence of exposure to chemical or physical stress,[3,7,8] but also the enzymatic side reactions leading to dysfunctional modified nucleosides. Whether such damage is primarily caused by promiscuity of RNA modification enzymes, as it is the case for msms²i⁶A, is not clear for ms²εA.

Conflict of Interest

Mark Helm is a consultant for Moderna Inc.

Data Availability Statement

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

[1] P. Boccaletto, F. Stefaniak, A. Ray, A. Cappannini, S. Mukherjee, E. Purta, M. Kurkowska, N. Shirvanizadeh, E. Destefanis, P. Groza, G. Avşar, A. Romitelli, P. Pir, E. Dassi, S. G. Conticello, F. Aguilo, J. M. Bujnicki, *Nucleic Acids Res.* **2022**, *50*, D231–D235.

[2] C. Dal Magro, P. Keller, A. Kotter, S. Werner, V. Duarte, V. Marchand, A. Ignarski, A. Freiwald, R.-U. Müller, C. Dieterich, Y. Motorin, F. Butter, M. Atta, M. Helm, *Angew. Chem. Int. Ed.* **2018**, *57*, 7893; *Angew. Chem.* **2018**, *130*, 8019.

[3] V. F. Reichle, D. P. Petrov, V. Weber, K. Jung, S. Kellner, *Nat. Commun.* **2019**, *10*, 5600.

[4] T. Ohira, K. Minowa, K. Sugiyama, S. Yamashita, Y. Sakaguchi, K. Miyauchi, R. Noguchi, A. Kaneko, I. Orita, T. Fukui, K. Tomita, T. Suzuki, *Nature* **2022**, *605*, 372.

[5] a) J. H. J. Hoeijmakers, *N. Engl. J. Med.* **2009**, *361*, 1475; b) J. Cadet, J. R. Wagner, *Cold Spring Harbor Perspect. Biol.* **2013**, *5*, a012559; c) Q. Kong, C.-L. G. Lin, *Cell. Mol. Life Sci.* **2010**, *67*, 1817; d) B. Nawrot, E. Sochacka, M. Düchler, *Cell. Mol. Life Sci.* **2011**, *68*, 4023.

[6] F. Drabløs, E. Feyzi, P. A. Aas, C. B. Vaagbø, B. Kavli, M. S. Bratlie, J. Peña-Diaz, M. Otterlei, G. Slupphaug, H. E. Krokan, *DNA Repair* **2004**, *3*, 1389.

[7] C. Sun, M. Jora, B. Solivio, P. A. Limbach, B. Addepalli, *ACS Chem. Biol.* **2018**, *13*, 567.

[8] C. Borek, V. F. Reichle, S. Kellner, *ChemBioChem* **2020**, *21*, 2768.

[9] a) N. Chatterjee, G. C. Walker, *Environ. Mol. Mutagen.* **2017**, *58*, 235; b) B. Sedgwick, P. A. Bates, J. Paik, S. C. Jacobs, T. Lindahl, *DNA Repair* **2007**, *6*, 429; c) Z. Li, J. Wu, C. J. Deleo, *IUBMB Life* **2006**, *58*, 581; d) C. L. Simms, H. S. Zaher, *Cell. Mol. Life Sci.* **2016**, *73*, 3639; e) E. J. Wurtmann, S. L. Wolin, *Crit. Rev. Biochem. Mol. Biol.* **2009**, *44*, 34.

[10] a) B. I. Fedeles, V. Singh, J. C. Delaney, D. Li, J. M. Essigmann, *J. Biol. Chem.* **2015**, *290*, 20734; b) G. Zheng, Y. Fu, C. He, *Chem. Rev.* **2014**, *114*, 4602.

[11] a) F. L. Chung, H. J. Chen, R. G. Nath, *Carcinogenesis* **1996**, *17*, 2105; b) F. el Ghissassi, A. Barbin, J. Nair, H. Bartsch, *Chem. Res. Toxicol.* **1995**, *8*, 278; c) C. E. Vaca, J. Wilhelm, M. Harms-Ringdahl, *Mutat. Res. Rev. Genet. Toxicol.* **1988**, *195*, 137.

[12] K. Bergman, *Arch. Toxicol.* **1982**, *49*, 117.

[13] R. J. Laib, H. M. Bolt, *Toxicology* **1977**, *8*, 185.

[14] H. Ottenwälder, R. J. Laib, H. M. Bolt, *Arch. Toxicol.* **1979**, *41*, 279.

[15] J. L. Creech, M. N. Johnson, *J. Occup. Med.* **1974**, *16*, 150.

[16] a) A. Barbin, H. Brésil, A. Croisy, P. Jacquignon, C. Malaveille, R. Montesano, H. Bartsch, *Biochem. Biophys. Res. Commun.* **1975**, *67*, 596; b) C. Malaveille, H. Bartsch, A. Barbin, A. M. Camus, R. Montesano, A. Croisy, P. Jacquignon, *Biochem. Biophys. Res. Commun.* **1975**, *63*, 363.

[17] R. J. Laib, L. M. Gwinner, H. M. Bolt, *Chem.-Biol. Interact.* **1981**, *37*, 219.

[18] W. J. Krzyzosiak, J. Biernat, J. Ciesiołka, K. Gulewicz, M. Wiewiórowski, *Nucleic Acids Res.* **1981**, *9*, 2841.

[19] a) F. J. Bullock, O. Jardetzky, *J. Org. Chem.* **1964**, *29*, 1988; b) J. Tropp, A. G. Redfield, *J. Am. Chem. Soc.* **1980**, *102*, 534; c) M. P. Schweizer, S. I. Chan, G. K. Helmkamp, P. O. P. Ts'o, *J. Am. Chem. Soc.* **1964**, *86*, 696.

[20] H.-W. Winter, C. Wentrup, *Angew. Chem. Int. Ed. Engl.* **1980**, *19*, 720; *Angew. Chem.* **1980**, *92*, 743.

[21] D. P. Bhatt, X. Chen, J. D. Geiger, T. A. Rosenberger, *J. Chromatogr. B* **2012**, *889–890*, 110.

[22] S. Lund, R. Hall, G. J. Williams, *ACS Synth. Biol.* **2019**, *8*, 232.

[23] a) A. Favre, E. Hajnsdorf, K. Thiam, A. Caldeira de Araujo, *Biochimie* **1985**, *67*, 335; b) J. Biernat, J. Ciesiołka, P. Górnicki, R. W. Adamiak, W. J. Kryzosiak, M. Wiewiórowski, *Nucleic Acids Res.* **1978**, *5*, 789.

Angewandte
Chemie

Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

A New Bacterial Adenosine-Derived Nucleoside as an Example of RNA Modification Damage

L. Bessler, L.-M. Vogt, M. Lander, C. Dal Magro, P. Keller, J. Kühlborn, C. J. Kampf,
T. Opatz, M. Helm*

WILEY-VCH

SUPPORTING INFORMATION

Table of Contents

SUPPORTING INFORMATION

1 Experimental Procedures

1.1 Sample preparations

1.1.1 Unlabelled and isotopically labelled RNA from *E. coli* MC4100

For unlabelled RNA, *E. coli* (strain MC4100) cells were cultured in standard M9-Medium composed of 6.8 g/L Na_2HPO_4 (Carl Roth, Germany), 3 g/L KH_2PO_4 (Carl Roth, Germany), 0.5 g/L NaCl (Carl Roth, Germany), 1 g/L NH_4Cl (Merck-Millipore, Germany), 2 mM $MgSO_4$ (Carl Roth, Germany), 0.1 mM $CaCl_2$ (Grüssing GmbH, Germany) and 0.4% glucose (Carl Roth, Germany). In order to prepare ^{13}C-, ^{15}N- or ^{34}S-labelled RNA, the isotopically labelled form of glucose (for ^{13}C-M9), NH_4Cl (for ^{15}N-M9) or $MgSO_4$ (for ^{34}S-M9) (all Sigma-Aldrich, Germany), respectively, was used in these cultures.

For the differential labelling the ^{15}N-M9 medium additionally contained one of the canonical ^{14}N nucleobases uracil (Carl Roth, Germany), guanine (Carl Roth, Germany) or adenine (Sigma-Aldrich, Germany) in a final concentration of 300 μM.[1]

In the case of methionine feeding a final concentration of 0.5 mM methionine-(methyl-D_3) (Sigma-Aldrich, Germany) was added to standard M9 medium.

The *E. coli* MC4100 cells were grown in the corresponding M9 medium to an OD_{600} of 1.8 and harvested (10 min, 8 000 g, 4 °C). Subsequently, RNA was extracted from the cell pellets using TRI reagent® (Sigma-Aldrich, Germany) according to the manufacturer's protocol.

1.1.2 RNA from knockout strains

For *E. coli* knockout experiments *ΔmiaA* (JW4129) and *ΔmiaB* (JW0658) strains from the *E. coli* Keio knockout collection (GE Healthcare (Dharmacon™), United Kingdom) were used as well as the corresponding parent strain (BW25113). The cells were cultured in standard M9 medium and RNA was isolated as described above.

1.1.3 Isoprene labelling Experiment

Adapting a protocol published by Lund *et al.*,[2] in a first step *E. coli* NovaBlue (DE3) competent cells (Sigma-Aldrich, Germany) were transfected with a pET-Duet-1 plasmid encoding for an archaeal isopentenyl phosphate kinase (IPK) in multi-cloning site 2 as well as the class-A nonspecific acid phosphatase from *Shigella flexneri* (PhoN) in multi-cloning site 1 (GenScript Biotech Corporation, USA; gene sequences listed in Table S1) according to the manufacturer's protocol and selected by resistance to antibiotics (12.5 μg/mL tetracycline and 100 μg/mL ampicillin (both Carl Roth, Germany). *E. coli* NovaBlue (DE3) cells containing the pET-Duet-1 plasmid were grown overnight (37 °C, 190 rpm) in 15 mL of ^{13}C-labelled OD2 medium (Silantes, Germany) supplemented with 12.5 μg/mL tetracycline and 100 μg/mL ampicillin. The next day, two cultures were prepared by supplementing 100 mL of ^{13}C-labelled OD2 medium with 12.5 μg/mL tetracycline and 100 μg/mL ampicillin and inoculating these cultures to an OD_{600} of 0.1 with the overnight culture. One culture was grown as a control without further supplementations. A final concentration of 50 μM fosmidomycin (optimised concentration for the purpose of this experiment) was added to the other culture directly at the beginning and as soon as an OD_{600} of 0.2 was reached, IPTG, dimethyl allyl alcohol (dissolved in DMSO) and isopentenol (dissolved in DMSO) were added to final concentrations of 1 mM, 5 mM and 5 mM, respectively. After the induction, the cells were grown for another 24 h and then harvested (20 min, 10 000 g, 4 °C). Total tRNA was isolated from the cell pellets using TRI reagent® (Sigma-Aldrich, Germany) according to the manufacturer's protocol and qualitatively analysed by LC-MS/MS as described below.

Table S1. Codon-optimised gene sequences of proteins used for isoprene labelling.

Protein	Sequence (5' → 3')
IPK from *Thermoplasma acidophilum* codon-optimised for *E. coli*	ATGATGATTTTGAAGATTGGTGGGAGTGTCATTACGGACAAGTCTGCTTACCGTACCGCACGTACCTATGCCATCCGTTCGATCGTGAAAGTCCTTTCGGGGATCGAGGATCTTGTTTGTGTAGTACATGGCGGCGGCAGCTTCGGTCACATTAAAGCGATGGAATTTGGGCTTCCAGGCCCGAAGAATCCGCGCTCATCGATCGGATACTCTATCGTTCATCGTGACATGGAGAACTTGGACTTAATGGTCATTGATGCTATGATTGAGATGGGCATGCGCCCAATTTCAGTACCGATCAGCGCGCTGCGTTACGATGGCCGTTTCGACTACACACCACTTATTCGTTACATTGACGCAGGATTTGTTCCCGTATCTTATGGCGACGTCTATATTAAGGACGAGCATTCATACGGAATCTATTCGGGGGACGACATTATGGCCGACATGGCCGAACTTTTGAAACCAGACGTGGCAGTGTTCCTTACAGATGTGGACGGTATCTATTCCAAAGACCCTAAGCGCAACCCAGATGCGGTACTGCTGCGTGATATTGATACGAACATCACCTTCGATCGTGTCCAGAACGATGTAACAGGCGGAATTGGGAAGAAATTCGAGTCCATGGTAAAGATGAAATCTTCCGTCAAAAATGGGGTGTATCTGATTAACGGTAACCATCCCGAGCGTATCGGAGACATCGGCAAGGAATCCTTCATCGGCACGGTAATCCGC
PhoN from *Shigella flexneri* codon-optimised for *E. coli*	ATGAAGCGTCAGTTGTTCACGCTTTCGATTGTGGGGGTATTTTCGTTGAATACTTTCGCGTCGATTCCACCGGGCAACGATGTTACCACGAAGCCCGACTTATATTATCTTACGAATGACAACGCAATCGACAGTCTTGCGTGCTTACCACCTCCTCCCCAAATTGGCTCGATCGCTTTTCTGAATGACCAAGCCATGTACGAAAAAGGGCGTTTGCTTCGTAATACTGAACGTGGTAAACTGGCCGCGGAGGATGCAAACCTGTCAAGCGGCGGGGTTGCAAATGTGTTCTCGGCGGCCTTCGGGAGTCCTATCACAGCAAAGGACTCACCAGAATTGCA

3

SUPPORTING INFORMATION

```
CAAATTGCTTACAAATATGATCGAGGACGCCGGGGACCTGGCGACGCGTTCTGCGAAGGAGTACTATATGCGTATCCGCCCCTTT
GCCTTCTATGGAGTTTCTACTTGCAACACCAAGGAACAAGATACACTGAGTCGCAATGGAAGTTATCCATCAGGACACACAAGTAT
CGGTTGGGCGACAGCATTAGTACTGAGCGAAATTAACCCGGCCCGTCAGGATACGATTTTGAAGCGCGGATACGAATTGGGTGA
TTCCCGCGTCATCTGCGGTTACCACTGGCAATCCGACGTAGATGCCGCACGCATTGTTGGAAGCGCGATTGTTGCCACCCTGCA
TTCAAACCCTGTATTTCAAGCGCAGTTGCAAAAAGCTAAAGACGAGTTCGCCAATAATCAGAAGAAA
```

1.1.4 Isolation of specific tRNAs

The isolation of various specific tRNAs from the total tRNA pool of *E. coli* MC4100 was achieved by hybridization of a complementary DNA probe (IBA GmbH, Germany; listed in Table S2) which was fluorescein-tagged at the 3'-end and biotin-labelled at the 5'-end. The latter allowed to separate the target tRNA from the pool by immobilisation on streptavidin-coated magnetic beads (Dynabeads® MyOne™, Thermo Fisher Scientific, Germany). In a total volume of 100 µL, 400 pmol of the respective complementary DNA probe were mixed with 800 µg of total tRNA in 5X saline-sodium citrate (SSC) buffer. The mixture was first incubated for 3 min at 90 °C (denaturation step) and then incubated for 10 min at 65 °C (hybridisation step). While the samples cooled down to room temperature, the magnetic beads were prepared according to the manufacturer's protocol. For immobilisation, the hybridised samples were incubated with 200 µL beads solution (25 °C, 650 rpm). After 30 min of incubation the samples were put in the magnetic rack, the supernatant was removed and the remaining beads (with immobilised RNA/DNA hybrids) were washed with 1X SSC-buffer once and three times with 0.1X SSC buffer. After resuspending the beads in 50 µL MilliQ-H$_2$O, the RNA was removed from the beads by heating for 3 min at 75 °C. The supernatant was transferred to a new tube and incubated with 200 U DNase I (RNase-free 50 U/µL, Thermo Fisher Scientific, Germany) and DNase I 10X buffer (final concentration of 1X; Thermo Fisher Scientific, Germany) for 2 h at 37 °C. The target tRNA was purified via PAGE and subsequent gel elution (1.1.6).

Table S2. List of complementary DNA probes used for the isolation of specific *E. coli* tRNAs. The targeted tRNAs (defined by abbreviation of the amino acid they carry together with their anticodon) are listed in the first column. Modifications within the anticodon are shown according to the one letter code.[3]

tRNA	DNA probe sequence (5' → 3')
Ala$^{(VGC)}$	AAATGGTGGGTGATGACGGGATCGAACCGCCGACCCCCTCCTTGTAAGGGAGGTGCTCTCCCAGCTGAGCTAATCACCC
Arg$^{((CU1)}$	AAATGGTACCCGGAGCGGGACTTGAACCCGCACAGCGCGCAACGCCGAGGGATTTTAAATCCCTTGTGTCTACCGATTCCACCAT CCGGGC
Arg$^{((CU2)}$	AAATGGTGCCCGGACTCGGAATCGAACCAAGGACACGGGGATTTTCAATCCCCTGCTCTACCGACTGAGCTATCCGGGC
Arg$^{(CCG)}$	AAATGGAGGCGCGTTCCGGAGTCGAACCGGACTAGACGGATTTGCAATCCGCTACATAACCGCTTTGTTAACGCGCC
Arg$^{(ICG)}$	AAATGGCGGAGAGAGGGGGATTTGAACCCCCGGTAGAGTTGCCCCTACTCCGGTTTTCGAGACCGGTCCGTTCAGCCGCTCCG GCATCTCTCC
Asn$^{(QUU)}$	AAATGGCGGAAGCGCAGAGATTCGAACTCTGGAACCCTTTCGGGTCGCCGGTTTTCAAGACCGGTGCCTTCAACCGCTCGGCCA CACTTCC
Cys$^{(GCA)}$	AAATGGTGGTGGGGGAAGGATTCGAACCTTCGAAGTCGATGACGGCAGATTTACAGTCTGCTCCCTTTGGCCGCTCGGGAACCC CACC
Gly$^{(CCC)}$	AAATGGCGCGCCCGACAGGATTCGAACCTGAGACCTCTGCCTCCGGAGGGCAGCGCTCTATCCAGCTGAGCTACGGGCGC
Gly$^{(GCC)}$	AAATGGCGCGCCCTGCAGGATTCGAACCTGCGGCCCACGACTTAGAAGGTCGTTGCTCTATCCAACTGAGCTAAGGGCGC
Gly$^{(NCC)}$	AAATGGTGTCCCCTGCAGGAATCGAACCTGCAATTAGCCCTTAGAAGGGGCTCGTTATATCCATTTAACTAAGAGGAC
His$^{(QUG)}$	AAATGGCTCCTCTGACTGGACTCGAACCAGTGACATACGGGATTAACAGTCCGCCGTTCTACCGACTGAACTACAGAGGA
Ile$^{((AU)}$	AAATGGCTGGGGTACGAGGATTCGAACCTCGGAATGCCGGAATCAGAATCCGGTGCCTTACCGCTTGGCGATACCCCA
Ini$^{(CAU)}$	AAATGGCGTCCCCTAGGGGATTCGAACCCCTGTTACCGCCGTGAAAGGGCGGTGTCCTGGGCCTCTAGACGAAGGGGAC
Leu$^{(JAA)}$	AAATGGAGCGGGCGAAGGGAATCGAACCCTCGTATAGAGCTTGGGAAGCTCTCGTTCTACCATTGAACTACGCCCGC
Leu$^{(BAA)}$	AAATGGAGCGGGAAACGAGACTCGAACTCGCGACCCCGACCTTGGCAAGGTCGTGCTCTACCAACTGAGCTATTCCCGC
Leu$^{(CAG)}$	AAATGGAGCGGGCAGCGGGAATCGAACCCGCATCATCAGCTTGGAAGGCTGAGGTAATAGCCATTATACGATGCCCGC
Leu$^{(GAG)}$	AAATGGGGTGGCTAATGGGATTCGAACCCACGACAACTGGAATCACAATCCAGGGCTCTACCAACTGAGCTATAGCCACC
Lys$^{(SUU)}$	AAATGGTAGGCCTGAGTGGACTTGAACCACCGACCTCACCCTTATCAGGGGTGCGCTCTAACCACCTGAGCTACAAGCCT
Met$^{(MAU)}$	AAATGGTGGCCCCTGCTGGACTTGAACCAGCGACCAAGCGATTATGAGTCGCCTGCTCTAACCACTGAGCTAAGGGGCC
Phe$^{(GAA)}$	AAATGGTTGCGGGGGCCGGATTTGAACCGACGATCTTCGGGTTATGAGCCCGACGAGCTACCAGGCTGCTCCACCCCGCG

4

SUPPORTING INFORMATION

Pro(CGG)	AAATGGTGCCGAAGGCCGGACTCGAACCGGCACGTATTTCTACGGTTGATTTTGAATCAACTGTGTCTACCGATTTCGCCACTTC GGC
Ser(CGA)	AAATGGTGGCTACGACGGGATTCGAACCTGTGACCCCATCATTATGAGTGATGTGCTCTAACCAACTGAGCTACGTAGCC
Ser(GCU)	AAATGGCGGAGAGAGGGGGATTTGAACCCCCGGTAGAGTTGCCCCTACTCCGGTTTTCGAGACCGGTCCGTTCAGCCGCTCCG GCATCTCTCC
Ser(GGA)	AAATGGTCGGTGATAGAGGATTCGAACCTCCGACCCCTTCGTCCCGAACGAAGTGCGCTACCAGGCTGCGCCAATCACCG
Ser(UGA)	AAATGGTGGGTCGTGCAGGATTCGAACCTGCGACCAATTGATTAAAAGTCAACTGCTCTACCAACTGAGCTAACGACCC
Thr(GGU)	AAATGGCGGTGAGGGGGGGATTCGAACCCCCGATACGTTGCCGTATACACACTTTCCAGGCGTGCTCCTTCAGCCACTCGGACA CCTCACC
Trp(CCA)	AAATGGCGGTGAGGCGGGGATTCGAACCCCGGATGCAGCTTTTGACCGCATACTCCCTTAGCAGGGGAGCGCCTTCAGCCTCTC GGCCACCTCACC
Tyr(QUA)	AAATGGTGCTGATAGGCAGATTCGAACTGCCGACCTCACCCTTACCAAGGGTGCGCTCTACCAACTGAGCTATATCAGC
Val(GAC)	AAATGGTGGGTGATGACGGGATCGAACCGCCGACCCCCTCCTTGTAAGGGAGGTGCTCTCCCAGCTGAGCTAATCACCC
Val(VAC)	AAATGGCAAGATCACAGGAGTCGAACCTGCCCGGGACCGCGCGGCCCCAACTGGATTTGAAGTCCAGCCGCCTCACCGGAGAC GACGATCTT

1.1.5 RNA from polysome preparations

E. coli cells were grown in LB medium in 800 mL culture volume. When an OD_{600} of 0.6-0.7 was reached, chloramphenicol (Carl Roth, Germany) was added to a final concentration of 100 µg/mL and the mixture was incubated for 3 min. Subsequently, the cells were harvested (10 min, 10 000 g, 4 °C) and the cell pellets resuspended in lysis buffer (100 mM NH_4Cl (Merck-Millipore, Germany), 10 mM $MgCl_2$, 20 mM Tris, pH 7.5 (both Carl Roth, Germany)). After the addition of lysozyme (Carl Roth, Germany), freeze-thaw cycles were performed in liquid nitrogen and 10% deoxycholate (Sigma-Aldrich, Germany) was added. Another centrifugation step (12 000 g, 10 min, 4 °C) separated the remaining cell wall debris from the cell lysate which was then loaded on top of sucrose (Sigma-Aldrich, Germany) gradients (5% to 40%) that were generated using a Biocomp gradient station model 108 (settings: time 1.24 min, angle 81.5 °, speed 21 rpm). The gradients were then subjected to ultracentrifugation (Beckman Ultracentrifuge Optima LE-80K, SW40 Ti rotor from Beckman Coulter) at 150 000 g and 4 °C for 2.5 h and subsequently fractionated by measuring the absorbance at 260 nm (Biocomp Gradient Station model 108 with a Gilson Fraction Collector FC203B). Total RNA was isolated from these fractions using TRI reagent® (Sigma-Aldrich, Germany) according to the manufacturer's protocol and purified by gel elution.

1.1.6 Purification of specific RNA species by PAGE and gel elution

In order to purify either specifically isolated tRNA or different RNA species from polysomal fractions, RNA samples were separated on a 10% denaturing PAGE gel and stained with GelRed (Biotium). Using a Typhoon 9400, the bands were visualised at an excitation wavelength of 532 nm. The resulting image was used as a template to locate and excise the bands of interest from the gel. The gel cuts were mashed with a scalpel, frozen for 1 h and 300 µL of 0.5 M ammonium acetate (Sigma Aldrich, Germany) were added. After overnight shaking (750 rpm, 25 °C) the mixture was filtered using NanoSep® centrifugal filters (VWR, Germany) and precipitated by addition of three volumes of 100% ethanol.

1.1.7 RNA hydrolysis

Up to 10 µg of total tRNA was digested to nucleosides by incubation with 0.6 U nuclease P1 from *P. citrinum* (Sigma-Aldrich, Germany), 0.2 U snake venom phosphodiesterase from *C. adamanteus* (Worthington, USA), 2 U FastAP (Thermo Fisher Scientific, Germany), 10 U benzonase (Sigma-Aldrich, Germany), 200 ng Pentostatin (Sigma-Aldrich, Germany) and 500 ng Tetrahydrouridine (Merck-Millipore, Germany) in 25 mM ammonium acetate (pH 7.5; Sigma-Aldrich, Germany) overnight at 37 °C.

1.1.8 Deuterium exchange experiments

Deuterium exchange experiments were performed as previously described.[4,5] To facilitate the subsequent qualitative LC-MS/MS analysis (1.2.1), candidate 338 was enriched by collecting the fraction of 50 µg digested and chromatographically separated *E. coli* RNA around 27.7 min three times for samples to be used for the slow exchange treatment and five times for samples to be used in the fast exchange experiment. The HPLC separation was performed as detailed in 1.1.9.

For the slow exchange experiment, the samples were lyophilized, resuspended together in D_2O (Sigma-Aldrich) and incubated for 1 h at 90 °C.
For the fast exchange experiment the lyophilised samples were resuspended in D_2O and the solvents used for chromatographic separation in the LC-MS/MS analysis were adjusted as detailed in 1.2.1.

5

SUPPORTING INFORMATION

1.1.9 Synthesis of 2-methylthio-1,*N*6-ethenoadenosine

Adapting the protocol for synthesis of ethenoadenosine described by Bhatt *et al.*,[6] a freshly prepared mixture of 11.2 µL 50% chloroacetaldehyde and 138.8 µL 1 M sodium acetate buffer (pH 4.5) was added to 50 µL of 2.5 mM 2-methylthioadenosine (0.125 µmol, Carbosynth Ltd, United Kingdom) in an Eppendorf tube. After vortexing, the samples were centrifuged at 450 g for 2 min at 22 °C and subsequently incubated for 60 min at 60 °C. In order to stop the reaction, the tubes were placed on ice and extracted with water-saturated ether. The aqueous phase was separated on an Agilent 1100 series HPLC using a Synergi Fusion RP-C18 column (250 × 2.0 mm, 4 µM, 80 Å; Phenomenex, Germany) with the following gradient (A = 5 mM NH_4OAc, pH 5.3; B = HPLC grade acetonitrile): 0 min 100% A, linear increase to 10% B in 20 min followed by another linear increase to 100% B at 50 min, hold for 3 min, then back to 100% A in 7 min, followed by re-equilibration of the column for 5 min. The product peak which was monitored by UV signal was collected and the solvent was evaporated. For NMR analysis, the synthesis and purification procedure were repeated several times and the products were pooled.

NMR spectra were recorded in CD_3OD on a Bruker Avance III spectrometer (Bruker Corporation, USA). Chemical shifts were reported in parts per million (ppm) relative to tetramethylsilane and internally referenced to CD_3OD (3.31 ppm for ^{1}H-NMR and 49.0 ppm for ^{13}C-NMR). Signals were assigned based on ^{1}H-^{1}H COSY, ^{1}H-^{13}C HSQC and ^{1}H-^{13}C HMBC spectra. Coupling constants *J* are given in Hz and for multiplicities the following abbreviations were used: s = singlet, d = doublet, dd = double doublet, t = triplet, q = quartet.

^{1}H-NMR (600 MHz, CD_3OD): δ = 8.46 (s, 1H; H-2), 7.82 (d, *J* = 1.7 Hz, 1H; H-7), 7.59 (d, *J* = 1.7 Hz, 1H; H-8), 6.21 (d, *J* = 5.1 Hz, 1H; H-1'), 4.76 (t, *J* = 5.1 Hz, 1H; H-2'), 4.41 (t, *J* = 4.9 Hz, 1H; H-3'), 4.12 (q, *J* = 3.8 Hz, 1H; H-4'), 3.88 (dd, *J* = 3.3 Hz, *J* = 12.1 Hz, 1H; H-5'), 3.78 (dd, *J* = 3.8, *J* = 12.1 Hz, 1H; H-5'') 2.90 (s, 3H; CH_3) ppm.

^{13}C-NMR (151 MHz, CD_3OD): δ = 149.3 (C-5), 142.4 (C-9a), 141.1 (C-3a), 140.3 (C-2), 133.3 (C-8), 121.2 (C-9b), 112.0 (C-7), 90.2 (C-1'), 86.7 (C-4'), 76.1 (C-2'), 71.7 (C-3'), 62.7 (C-5'), 14.4 (CH_3) ppm.

Impurities in the spectra can be traced to residual water and the acetate used during the purification procedure.[7]

1.2 LC-MS/MS analysis

As indicated in the detailed description below, all LC-MS/MS measurements were performed on an Agilent 1260 Infinity (II) series HPLC connected to either an Agilent 6460A triple quadrupole (QQQ) mass spectrometer, an Agilent 6470B QQQ mass spectrometer or a high-resolution Agilent 6545 quadrupole time-of-flight (Q-ToF) mass spectrometer, each of them interfaced with an Agilent Jet Stream electrospray ionisation source (ESI) (all devices: Agilent Technologies, Germany). For separation either a YMC-Triart C8 column (150 mm x 3 mm, 3 µm, 120 Å; YMC Europe GmbH, Germany) or a Synergi Fusion RP-C18 column (250 × 2.0 mm, 4 µM, 80 Å; Phenomenex, Germany) were used at a flow of 0.35 mL/min and a temperature of 35 °C. In both cases, the solvents consisted of 5 mM ammonium acetate buffer (pH 5.3; solvent A) and LC-MS grade acetonitrile (solvent B; Honeywell). The LC method started with 100% solvent A. Solvent B was linearly increased to 10% at 20 min and to 100% at 50 min (hold for 3 min). Finally, initial conditions were restored in 7 min and the column re-equilibrated for 5 min prior to next injection. When the initial method (C8 column) was transferred to the C18 column, the identity of candidate 338 was ensured by comparing the results of stable isotope labelled samples and the retention time relative to known modifications. The main nucleosides were monitored at 254 nm (UV trace) with a diode array detector (DAD). All mass spectrometers were operated in the positive ion mode and ESI parameters are detailed in table S3.

Table S3. ESI settings of the different mass spectrometers.

parameters	6460A QQQ	6470B QQQ	6545 Q-ToF
gas temperature	350	300	325
gas flow (L/min)	5	7	10
nebulizer pressure (psi)	35	60	33
sheath gas temperature (°C)	350	400	325
sheath gas flow (L/min)	12	12	10
capillary voltage (V)	3500	3000	3500

WILEY-VCH

SUPPORTING INFORMATION

1.2.1 Qualitative analysis

Unless otherwise stated, qualitative LC-MS/MS analysis was performed with the 6460A QQQ. Qualitative analysis included identification of candidate 338 as a potentially new ribonucleoside as previously described and investigation of mass shifts as a consequence of stable isotope labelling.[5,8] For this type of analysis, 10 μg of digested RNA sample were injected into the LC-MS system and separated on the C8 column. The Agilent MassHunter software (version B.05.00) was employed in the neutral loss scan (NLS) mode, detecting the loss of an uncharged moiety as defined in the method, usually a loss of 132 Da in the case of ribonucleosides, corresponding to the loss of the ribose. If the analysed samples had a stable isotope labelling background affecting the ribose, the neutral loss was adjusted to 137 Da (^{13}C-labelled samples) or to 135 Da (D_2O fast exchange samples). NLS settings of the 6460A QQQ were as follows: fragmentor voltage 80 V, collision energy 15 eV, cell accelerator voltage 2 V, scan time 555 ms and mass range *m/z* 300-460.

To investigate fast exchanging species within candidate 338 (D_2O fast exchange experiment), for this specific LC-MS/MS run solvent A was prepared with D_2O instead of H_2O.

In case of the co-injection of digested ^{13}C-labelled RNA and the synthetic ms^{2}εA, the RNA sample was spiked with 100 fmol of ms^{2}εA and both the loss of the ^{12}C-ribose (-132 Da) as well as the loss of the ^{13}C-ribose (-137 Da) were included as a segment of the method. For this analysis the C18 column was used in combination with the 6470B QQQ (Agilent MassHunter software, version 10.0). NLS settings were similar to the 6460A QQQ.

For the analysis of samples obtained from the isoprene labelling experiment, 20 μg of digested RNA were analysed using the C18 column in combination with the 6470B QQQ (Agilent MassHunter software, version 10.0) regarding a neutral loss of 137. NLS settings were similar to the 6460A QQQ.

1.2.2 Relative quantification

For relative quantification of defined nucleoside species, the Agilent MassHunter software was employed in the dynamic multiple reaction monitoring (dMRM) mode. The mass transitions for the predefined nucleosides were as follows: ms^{2}εA (*m/z* 338 → 206), i^{6}A (*m/z* 336 → 204) and ms^{2}i^{6}A (*m/z* 382 → 250). Further dMRM settings were a fragmentor voltage of 80 V, a collision energy of 15 eV and a cell accelerator voltage of 2 V.

For this type of analysis up to 5 μg of digested RNA sample were injected into the LC-MS system. The data obtained from these measurements was analysed using Agilent MassHunter Qualitative Analysis software. Here, integration of the peak of interest in extracted ion chromatograms yielded the corresponding area under the curve which is often called abundance within this study. These abundances were then normalised to the peak area of adenosine in the UV trace of the respective sample.

In case of the analysis of ms^{2}εA on isolated RNA species, the C8 column was used in combination with the 6460A QQQ (Agilent MassHunter software, version B.05.00). The relative quantification of ms^{2}εA in polysomal fractions as well as the occurrence of ms^{2}εA in tRNA isolated from different knockout cells was performed with the C18 column and the 6470B QQQ (Agilent MassHunter software, version 10.0).

1.2.3 Absolute quantification of ms^{2}εA in isolated tRNA Leu^{BAA}

For absolute quantification the method described in 1.2.2 was applied in combination with external calibration, as previously described.[9] Briefly, for the main nucleosides, a calibration curve was created in the range of 1-1000 pmol, based on the UV signal of calibration solutions containing cytidine, uridine, guanosine and adenosine in defined concentrations which was recorded at 254 nm. This allowed calculation of the amount of the respective nucleoside in the sample which was used together with the known sequence of Leu^{BAA} to calculate the amount of analysed tRNA molecules.[10] For external calibration of ms^{2}εA, a calibration curve in the range of 1-5000 fmol was created using the synthesised ms^{2}εA as a standard in different calibration solutions. The inclusion of an internal standard (e.g. from ^{13}C-labelled *E. coli* RNA) was not possible, due to the compound's low abundance. The mass signal detected for ms^{2}εA in tRNA Leu^{BAA} was within the linear range of the calibration curve and was not affected by saturation effects.

1.2.4 Pseudo-MS^3 analysis (fragmentation pattern)

To obtain the fragmentation patterns of the naturally occurring and the synthetic ms^{2}εA, the C18 column was used in combination with the 6460A QQQ and the Agilent MassHunter software (version B.05.00) being employed in the product ion scan mode with an increased fragmentor voltage of 200 V. The consequence of this increase in the fragmentor voltage is an in-source fragmentation of the nucleosides to already yield the nucleobases in the MS^1 detector which filters for the predefined precursor ion *m/z* 206. Further fragmentation of the nucleobase with a collision energy of 40 eV then yielded the product ion spectrum. Compared to the NLS settings, the scan time was changed to 500 ms and the MS^2 scan range was set from *m/z* 15 to *m/z* 220.

Due to the lower sensitivity of this method, the naturally occurring ms^{2}εA had to be enriched. This was achieved by collecting the fraction of 50 μg digested and chromatographically separated (see section 1.1.9) *E. coli* RNA around 27.7 min five times. The solvent was evaporated and the fractions were pooled when the dried residue was resolved in MilliQ water. In case of the synthetic ms^{2}εA 1000 fmol were used for the analysis.

SUPPORTING INFORMATION

1.2.5 High-resolution mass spectrometry analysis

The Agilent 6545 Q-ToF mass spectrometer was used to perform high-resolution mass spectrometry. Since the Q-ToF is less sensitive compared to the QQQ, the naturally occurring ms^2_6A was enriched by collecting the fraction of 50 µg digested and chromatographically separated *E. coli* RNA around 27.7 min ten times. The HPLC separation was performed as detailed in 1.1.9. The solvent was evaporated and the fractions were pooled when the dried residue was resolved in MilliQ water. The enriched samples were chromatographically separated on the C8 column.

To identify the exact mass of candidate 338, the Agilent MassHunter software (version 10.0) was employed in the MS mode from m/z 100 to m/z 500 at a scan rate of 1 spectrum/s. Subsequent data analysis and use of the embedded sum formula calculator allowed calculation of the exact mass on the experimentally detected mass.
HRMS (ESI): m/z calcd for $C_{13}H_{16}N_5O_4S^+$: 338.0918 [M+H]$^+$; found: 338.0914.

In order to obtain the exact masses of the fragments (determined by pseudo-MS^3 analysis), the Agilent MassHunter software (version 10.0) was used in the targeted MS^2 mode with the precursor ion m/z 338.0900. Data was acquired in the centroid mode and method settings were as follows: MS m/z 100-1000, MS scan rate 1 spectrum/s, MS/MS m/z 30-500, MS/MS scan rate 1 spectrum/s, max time between MS 5 s and use of fixed collision energies (35 eV).
HRMS (ESI): m/z calcd for $C_8H_8N_5S^+$: 206.0495 [M+H]$^+$; found: 206.0500; m/z calcd for $C_7H_5N_5S^+$: 191.0260 [M+H]$^+$; found: 191.0265

SUPPORTING INFORMATION

2 Results and Discussion

2.1 Analysis of tRNA isolated from *E. coli ΔmiaA* or *ΔmiaB* strains
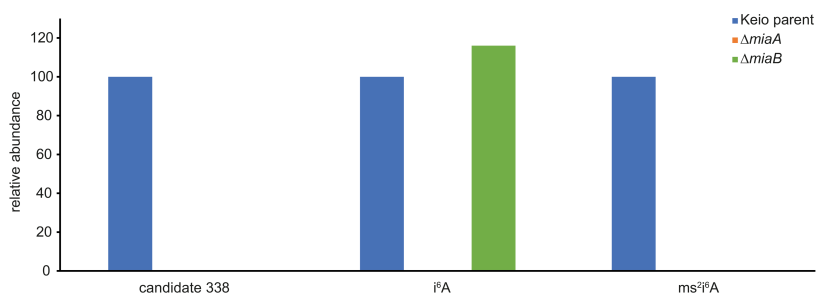
Figure S1. LC-MS/MS analysis of tRNA isolated from i⁶A- and ms²i⁶A-related knockout strains. Relative abundance of candidate 338, i⁶A and ms²i⁶A in tRNA from *ΔmiaA* (orange) and *ΔmiaB* (green) as well as in tRNA from the corresponding Keio parent strain (blue). First, the abundance of the modification was normalised to the UV-signal of adenosine in the respective sample, to account for different RNA amounts, and then set into relation to the amount of the respective modification in the parent strain (set to 100%).

2.2 Analysis of candidate 338 in specific tRNA species
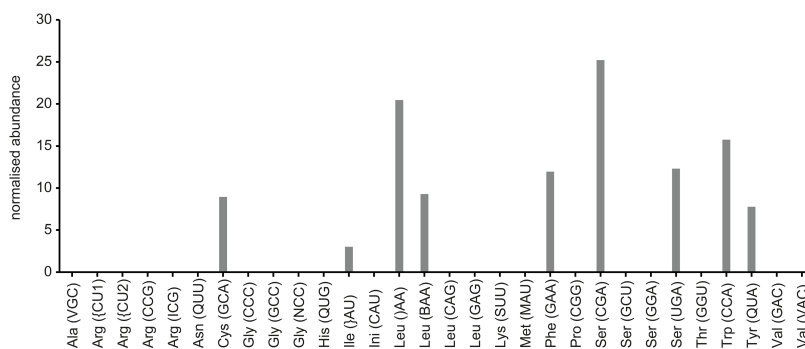
Figure S2. Occurrence of candidate 338 in specific tRNA species. LC-MS/MS analysis of candidate 338 in single tRNA species isolated from the tRNA pool. The abundance of candidate 338 in each tRNA was normalised to the UV signal of adenosine. The different tRNA species are displayed by the abbreviation of the amino acid they carry together with their anticodon (in brackets). Modifications within the anticodon are shown according to the one letter code.[3]

9

2.3 Analysis of isoprene-related modifications in tRNA from the isoprene labelling experiment
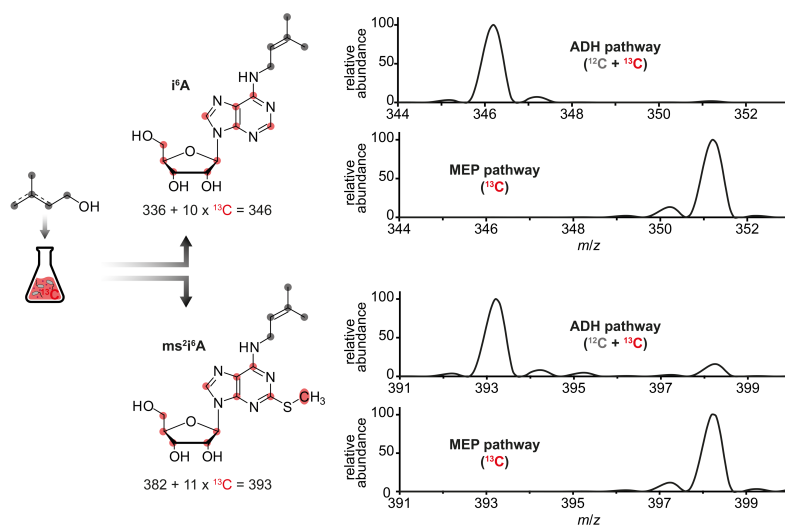
Figure S3. LC-MS/MS analysis of i⁶A and ms²i⁶A in tRNA from the isoprene labelling experiment. The upper part of the figure shows the mass spectra of i⁶A while the lower part details the results for ms²i⁶A. In both cases a mass spectrum at the retention time of the respective modification is shown, obtained from the analysis of a digested tRNA sample. The latter was either isolated from a ^{13}C-culture of prepared *E. coli* cells supplemented with ^{12}C-dimethylallyl alcohol and ^{12}C-isopentenol to be used in the artificial alcohol dependent hemiterpene (ADH) pathway or from a ^{13}C-culture without supplementation, i.e. the native non-mevalonate pathway (MEP pathway) generates ^{13}C-isoprene precursors. The red (^{13}C) and grey (^{12}C) markers in the structure formulas illustrate the occurrence of the observed mass-to-charge ratios upon inverse feeding. The relative abundance was calculated by setting every signal in the respective spectrum in relation to the highest signal in this spectrum (set to 100%).

3 NMR spectra

¹H-NMR spectrum of ms²εA (600 MHz, CD₃OD).

¹³C-NMR spectrum of ms²εA (151 MHz, CD₃OD).

11
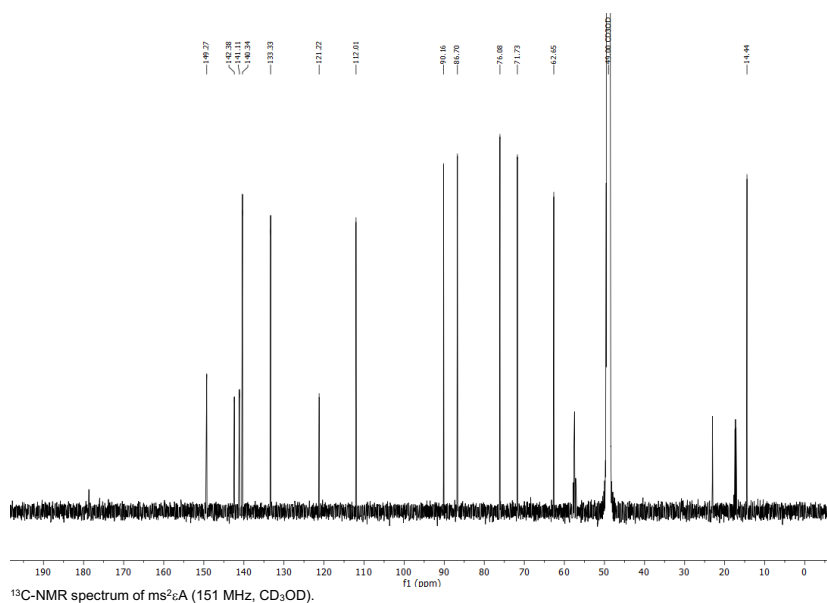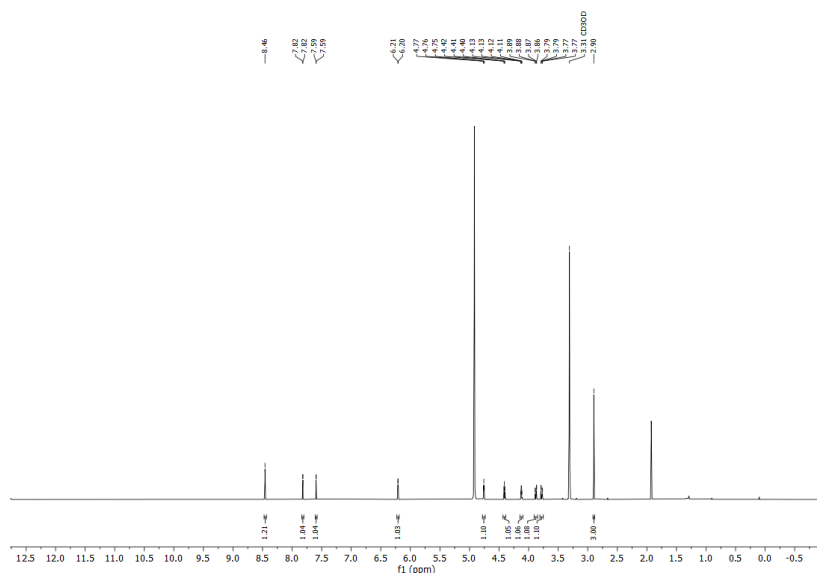
SUPPORTING INFORMATION

4 References

[1]   R. I. Christopherson, L. R. Finch, *European journal of biochemistry* **1978**, *90*, 347.
[2]   S. Lund, R. Hall, G. J. Williams, *ACS Synth. Biol.* **2019**, *8*, 232.
[3]   P. Boccaletto, F. Stefaniak, A. Ray, A. Cappannini, S. Mukherjee, E. Purta, M. Kurkowska, N. Shirvanizadeh, E. Destefanis, P. Groza, G. Avşar, A. Romitelli, P. Pir, E. Dassi, S. G. Conticello, F. Aguilo, J. M. Bujnicki, *Nucleic Acids Res.* **2022**, *50*, D231-D235.
[4]   a) J. Tropp, A. G. Redfield, *J. Am. Chem. Soc.* **1980**, *102*, 534; b) M. P. Schweizer, S. I. Chan, G. K. Helmkamp, P. O. P. Ts'o, *J. Am. Chem. Soc.* **1964**, *86*, 696.
[5]   C. Dal Magro, P. Keller, A. Kotter, S. Werner, V. Duarte, V. Marchand, A. Ignarski, A. Freiwald, R.-U. Müller, C. Dieterich, Y. Motorin, F. Butter, M. Atta, M. Helm, *Angew. Chem. Int. Ed. Engl.* **2018**, *57*, 7893; *Angew. Chem.* **2018**, *130*, 8019.
[6]   D. P. Bhatt, X. Chen, J. D. Geiger, T. A. Rosenberger, *J. Chromatogr. B: Analyt. Technol. Biomed. Life Sci.* **2012**, *889-890*, 110.
[7]   H. E. Gottlieb, V. Kotlyar, A. Nudelman, *The Journal of organic chemistry* **1997**, *62*, 7512.
[8]   S. Kellner, J. Neumann, D. Rosenkranz, S. Lebedeva, R. F. Ketting, H. Zischler, D. Schneider, M. Helm, *Chemical communications (Cambridge, England)* **2014**, *50*, 3516.
[9]   S. Kellner, A. Ochel, K. Thüring, F. Spenkuch, J. Neumann, S. Sharma, K.-D. Entian, D. Schneider, M. Helm, *Nucleic Acids Res.* **2014**, *42*, e142.
[10]   F. Jühling, M. Mörl, R. K. Hartmann, M. Sprinzl, P. F. Stadler, J. Pütz, *Nucleic Acids Res.* **2009**, *37*, D159-62.

Author Contributions

L. Bessler performed the majority of the experimental work based on preliminary work by C. Dal Magro and P. Keller. L.-M. Vogt and M. Lander performed polysome preparations. J. Kühlborn and T. Opatz performed NMR analysis. C. J. Kampf performed HRMS measurements. All authors discussed the results.

L. Bessler and M. Helm wrote the manuscript which was edited by all authors. M. Helm designed and supervised the work.

4.2 Structural elucidation and characterization of a pyrimidine-dimer

Among the potentially new RNA modifications in the list of former members of the ▮▮▮▮ group was candidate 470, whose first structural characterization pointed to a dimeric structure, containing two pyrimidine bases and a total of 5 nitrogen atoms, giving the appearance of a complex hypermodification.[297–299] In singly performed biological experiments, a predominant occurrence in *E. coli* as well as altered levels of candidate 470 in sulfurtransferase knockouts provided an interesting basis for further characterization of this compound in the context of this work.[297]

Of note, a lot of the experimental work described within this chapter was performed by ▮▮▮▮▮▮▮ within her internship and ▮▮▮▮▮▮ within his master thesis under my supervision and is marked accordingly, as are the contributions of other persons.

4.2.1 Validation of candidate 470 as authentic RNA modification

In view of accumulating reports on artifacts hampering identification of new RNA modifications, in particular putatively dimeric structures require critical reflection and examination of whether the compound of interest is a naturally occurring substance or whether the dimeric structure is formed during isolation of RNA from biological samples or during sample preparation including RNA hydrolysis for analysis.[230] To assess this question, *E. coli* cells that were grown separately in ^{14}N-medium or ^{15}N-medium were mixed after harvest by centrifugation and the mixed sample was subjected to RNA isolation, hydrolysis and subsequent LC-MS/MS analysis. Pure samples served as a control (Figure 4.1).

The exclusive formation of dimers with a uniform distribution of nitrogen atoms, i.e. either ^{14}N- or ^{15}N-atoms, while a mixed species containing both ^{14}N and ^{15}N in one structure was absent, indicated that candidate 470 is not formed during sample preparation in an intermolecular manner or between nucleoside species during or after tRNA hydrolysis, but might occur in its putatively dimeric form *in vivo*. However, with this experiment it cannot be excluded that candidate 470 results from an intramolecular reaction during RNA isolation.
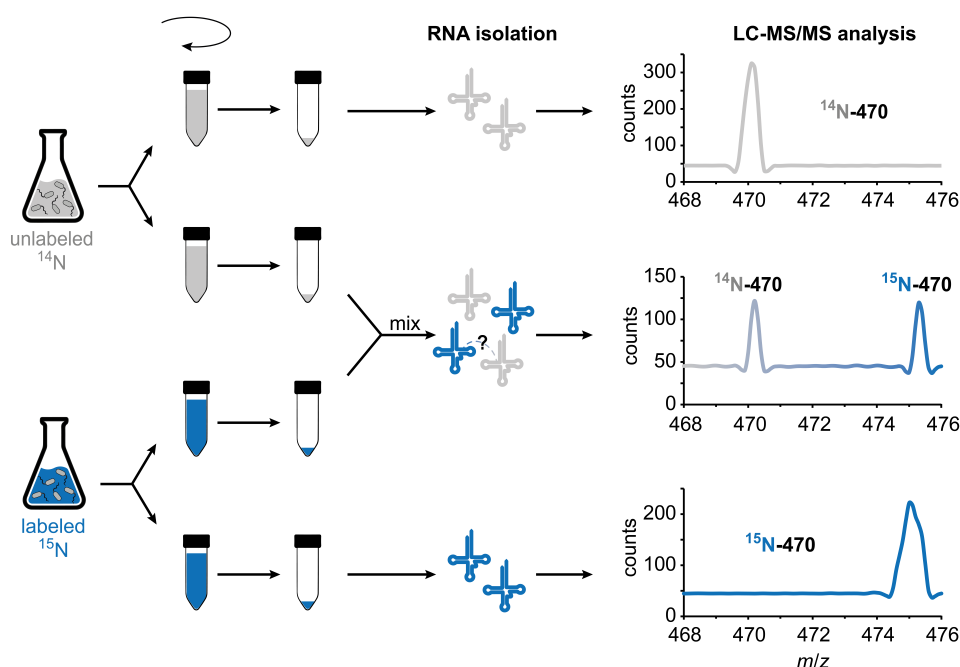
Figure 4.1: Validation of candidate 470 as authentic RNA modification. Workflow to investigate potential formation of the dimer as an artifact during RNA isolation or sample preparation. Results of the LC-MS/MS analysis for respective samples are shown on the right side, illustrated by the mass spectrum at the corresponding retention time extracted from the neutral loss scan (-132 Da) data. ^{14}N-470 describes the mass transition of unlabeled candidate 470 (m/z 470 $\rightarrow$ 338), ^{15}N-470 corresponds to the mass transition of the uniformly labeled compound (m/z 475 $\rightarrow$ 343).

4.2.2 Native and induced occurrence of candidate 470

Singly performed experiments by ▮▮▮▮▮▮▮▮▮▮▮▮ indicated that candidate 470 was clearly present in *E. coli* while it was significantly reduced in RNA from eukaryotic cell lines.[297] In order to choose the most suitable system for investigations on the native formation of candidate 470, the prevalence of this compound in tRNA from *E. coli* as well as from eukaryotic cell lines represented by HEK cells was re-examined by qualitative LC-MS/MS analysis. In this analysis of native tRNA, candidate 470 was exclusively detected in *E. coli* samples, while it was reproducibly absent in eukaryotic cell samples (Figure 4.2), indicating to choose *E. coli* as the model organism for further biological characterization and to consider an involvement of RNA compounds that are not available in eukaryotic cells in the biosynthetic process.

Since candidate 470 exhibited strong characteristics of a dimeric structure and the formation of dimeric structures in nucleic acids is reported to mainly occur in the context of exposure to UV radiation, investigating the influence of such a treatment on the occurrence of candidate 470 was a logical consequence.[168,311–313] To examine a possible UV-induced formation of candidate 470, a sample of *in vitro* transcribed mRNA, supposedly free of known RNA modifications and any candidate from the list of poten-

tially new modified structures, was irradiated with UV light at 254 nm and analyzed via LC-MS/MS regarding the presence of candidate 470. This experiment was first performed by ▮▮▮▮▮▮▮ in the context of RNA stability studies under UV irradiation and later reproduced by ▮▮▮▮▮▮▮ under my supervision. Of note, initial experiments were performed with a UV handlamp and UV energy doses were controlled in a time-dependent manner, while later on a UV crosslinker was used to expose samples to precisely defined UV energy doses and ensure their accurate reproducibility.
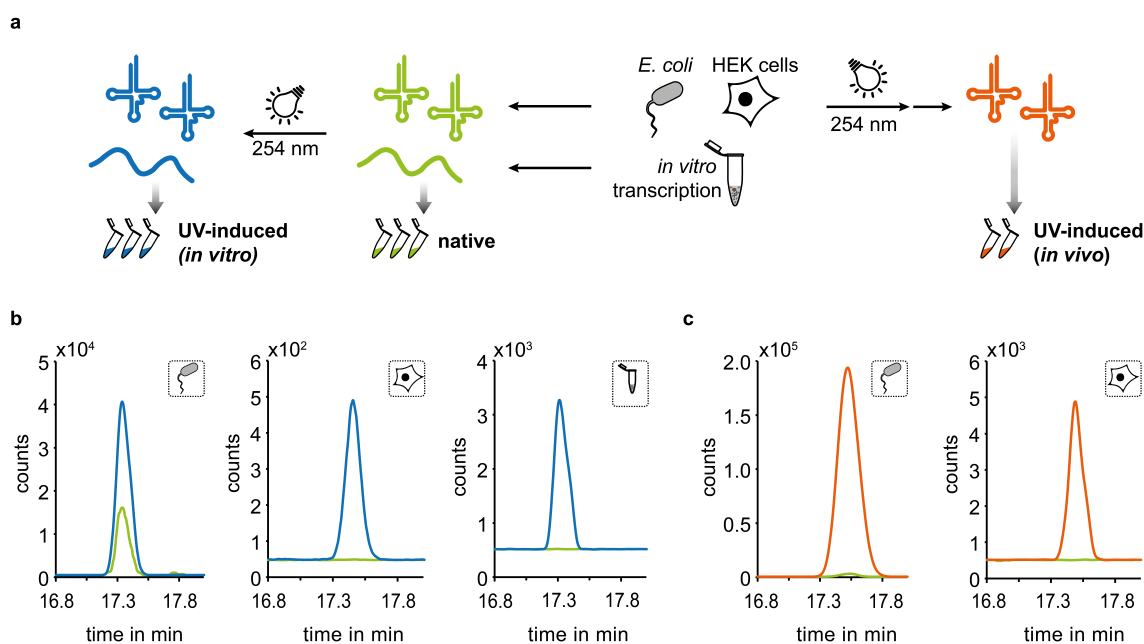
Figure 4.2: Native and induced occurrence of candidate 470. **a** Illustration of RNA sampling for LC-MS/MS analysis of candidate 470 occurrence from different biological backgrounds. For native RNA samples (green), tRNA was extracted from *E. coli* or HEK cells and mRNA was synthesized via *in vitro* transcription. Subsequent exposure of these RNA samples to UV light (254 nm) yielded RNA potentially containing candidate 470 that was UV-dependently formed *in vitro* (blue). In order to investigate the UV-induced formation of candidate 470 *in vivo*/*in cellulo* (orange), tRNA was isolated from *E. coli* or HEK cells after irradiation of the respective culture. Different RNA species are displayed as their respective secondary structures. **b** Qualitative LC-MS/MS analysis of the candidate's native (green) and UV-induced *in vitro* (blue) occurrence in RNA originating from different biological backgrounds, as indicated by respective icons. For the UV treatment, RNA was irradiated with a UV handlamp for 2 h at 254 nm. Extracted ion chromatograms of the mass transition m/z 470 → 338 are shown. **c** Qualitative LC-MS/MS analysis of the candidate's native (green) and UV-induced *in vivo*/*in cellulo* (orange) occurrence in RNA from *E. coli* or HEK cells. For the UV treatment, cell cultures were exposed to a UV energy dose of 1 J/cm^2. Extracted ion chromatograms of the mass transition m/z 470 → 338 are shown. For native control samples and their corresponding treated sample equal amounts of RNA were used and analysis was performed within the same measurement. Signal abundances between samples of different biological backgrounds cannot be compared due to analysis in different measurements, which also accounts for slight variances in the retention time.

In accordance with the filtering criteria during creation of the list of potentially new RNA modifications, the untreated *in vitro* transcribed mRNA did not contain any candidate 470, whereas, intriguingly, the compound was detected after exposure to UV light, sug-

gesting that this structure cannot only result from a biological process *in vivo* as shown for tRNA from *E. coli* but also from a photochemical reaction *in vitro* (Figure 4.2b). Importantly, the formation of candidate 470 from *in vitro* transcribed mRNA also provided first evidence that the photochemical reaction does not necessarily rely on the presence of other modified nucleoside structures or a tRNA structure, as it is the case for the previously mentioned dimer that is UV-dependently formed between s^4U8 and C13 of several tRNAs and considered a UV sensor of tRNA. Still, it is important to mention that the structural characteristics of this dimer, i.e. elemental composition, would meet the characteristics of candidate 470.[313,314] Extending experiments on the UV-induced formation *in vitro* on samples of diverse biological origins by irradiation of tRNA isolated from *E. coli* or HEK cells, showed an increase in the abundance of candidate 470 in the bacterial RNA and also led to the emergence of candidate 470 in tRNA from HEK cells, providing further evidence for a photochemical process taking place *in vitro*.

The findings on *in vitro* level demanded for an investigation of the candidate's occurrence in living cells upon UV treatment, in particular in eukaryotic cells (*in cellulo*), whose native tRNA was shown to be free of candidate 470, but also in *E. coli* (*in vivo*), to assess if the natively present levels of candidate 470 change upon this treatment. To this end, HEK cells and *E. coli* were irradiated with UV light at 254 nm during cell culture and RNA extracted from these cells was analyzed by LC-MS/MS for the occurrence of candidate 470. Indeed, UV-dependent formation of candidate 470 was observed *in cellulo*, indicating that in addition to its native formation in a biosynthetic process in *E. coli*, the dimeric structure might also occur in the context of UV damage of RNA from other organisms. The formation of candidate 470 as a UV damage product *in vivo* was also emphasized by the substantially increased formation of this compound after UV treatment of *E. coli*. With regard to the sensitivity of candidate 470 in sulfurtransferase knockout strains, as postulated by ▮▮▮▮▮▮▮▮▮▮ within her thesis, and the occurrence of s^4U in position 8 of the vast majority of *E. coli* tRNAs it has to be mentioned, that Ramabhadran *et al.* also observed formation of the crosslink between s^4U8 and C13 upon irradiation of *E. coli* cultures *in vivo*.[297,315]

Experiments regarding the native occurrence of candidate 470 in *in vitro* transcribed mRNA and in tRNA from eukaryotic cells as well as experiments regarding the UV-induced formation of candidate 470 *in vitro* were performed by ▮▮▮▮▮▮▮▮ and ▮▮▮▮▮▮▮▮▮ under my supervision.

4.2.3 Characterization of native candidate 470 formation *in vivo* in *E. coli*

In singly performed knockout experiments for the sulfurtransferases MnmA and ThiI, introducing s^2U and s^4U, respectively, █████████████ observed a reduction of candidate 470 abundance in both cases by at least 80%.[297] In order to validate these findings, the corresponding single gene knockout strains, taken from the Keio collection, were cultivated in biological triplicates, extracted RNA was analyzed by LC-MS/MS and compared to the level of candidate 470 in the wild-type Keio parent strain (Figure 4.3a). Intriguingly, the analysis revealed an equal occurrence of candidate 470 in RNA extracted from the parent strain and from the *mnmA* KO, indicating that the native generation of candidate 470 *in vivo* does not depend on the action of MnmA. While this was in contrast to the previous results, the sensitivity of candidate 470 in the *thiI* KO was confirmed. Here, the complete absence of candidate 470 provided evidence for an involvement of ThiI in the native formation of candidate 470 *in vivo*, in addition to its known role in the 4-thiolation of uridines in position 8 of *E. coli* tRNA. Of note, an involvement of ThiI would also be reasonable in view of the absence of native candidate 470 in tRNA from eukaryotic cells since the prevalence of s^4U and its corresponding modification enzyme is limited to bacterial species.[24]

Although ThiI was shown to act on all tRNAs in *E. coli*, some species like tRNA^{Phe} have been reported to be better ThiI substrates and consequently to contain a higher level of s^4U, than other species like tRNA^{Thr} or tRNA^{Glu}.[316,317] Based on the above findings, a comparison of candidate 470 occurrence on these isolated tRNA species might further substantiate the first indications of a correlation between ThiI and the native formation of candidate 470 *in vivo*. To this end, single tRNA species were purified from extracted total tRNA of the Keio parent strain or the *thiI* knockout strain (negative control), employing a hybridization-based approach with complementary DNA oligos. In subsequent LC-MS/MS analysis, the levels of s^4U and candidate 470 were evaluated.

In line with literature, among the different isolated tRNA species, tRNA^{Phe} contained the highest amount of s^4U, emphasizing the preference of ThiI towards this tRNA substrate (Figure 4.3b). Likewise, candidate 470 was most abundant in tRNA^{Phe} and its level notably reduced in the two other tRNA species, supporting a role of ThiI in the biosynthesis of candidate 470.

Despite accumulating evidence for a correlation between the occurrence of candidate 470 and the action of the sulfurtransferase ThiI, a proof for the direct dependency of candidate 470 on ThiI would plausibly be demonstrated by the use of this enzyme in an *in vitro* assay or by a rescue experiment, i.e. by examining if reintroduction of the *thiI* gene into the *thiI* knockout strain via transformation is accompanied by a reinstallation
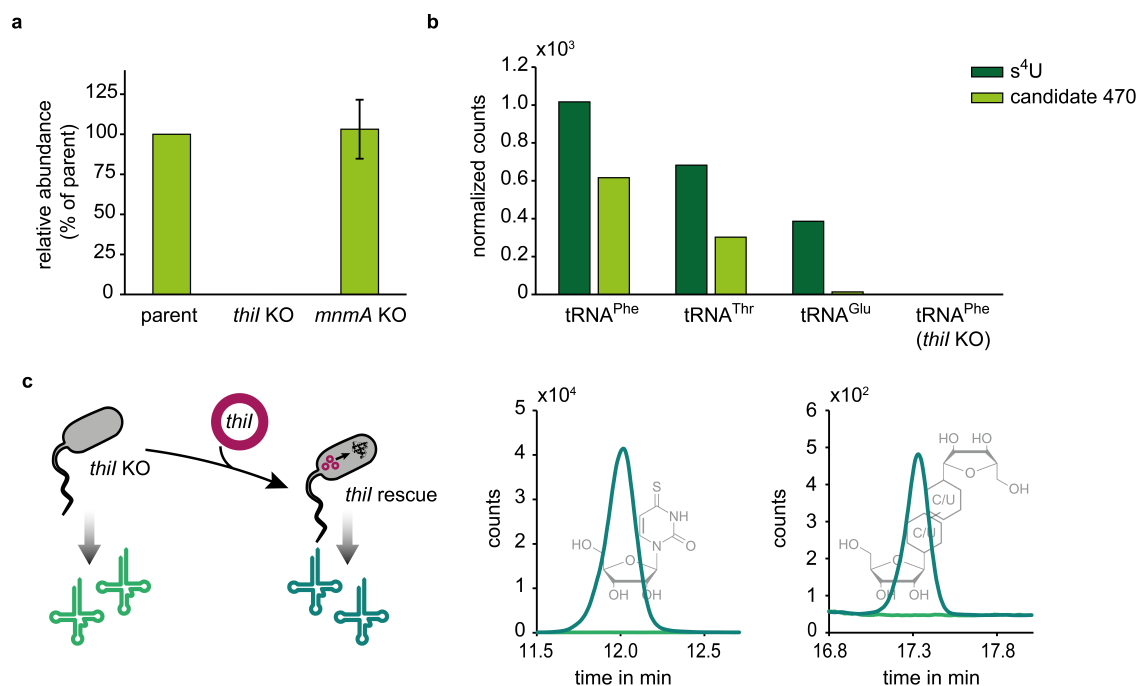
Figure 4.3: Influence of sulfurtransferases on the native occurrence of candidate 470. a Relative LC-MS/MS quantification of candidate 470 in tRNA isolated from the Keio parent, *thiI* KO or *mnmA* KO strain, respectively. Peak areas were normalized to the UV signal of adenosine and set in relation to the amount of candidate 470 in tRNA from the parent strain. The average of three independent biological replicates is shown. **b** LC-MS/MS analysis of s^4U and candidate 470 in single tRNA species isolated from the tRNA pool of *E. coli*. Abundances were normalized to the UV signal of adenosine. **c** Left: Scheme of *thiI* rescue experiment illustrating introduction of a plasmid containing the *thiI* gene into chemically competent *thiI* KO cells and consequent protein synthesis. Right: Qualitative LC-MS/MS analysis of s^4U (m/z 261 $\rightarrow$ 129) and candidate 470 (m/z 470 $\rightarrow$ 338) as illustrated by the respective (schematic) structure in the background in tRNA isolated from either *thiI* KO or *thiI* rescue cells after induction of *thiI* gene expression.

of candidate 470 on tRNA. In order to put the latter into practice, a plasmid containing a codon-optimized *thiI* sequence was designed and purchased from GenScript Biotech. After transformation of chemically competent *thiI* knockout cells with this plasmid and induction of *thiI* gene expression, RNA was extracted from these cells for LC-MS/MS analysis to investigate the presence of s^4U and candidate 470 in RNA from transformed bacteria compared with the similarly treated chemically competent cells.

In accordance with the previous results for *thiI* knockout cells, tRNA from the chemically competent *thiI* KO strain did not contain any of the compounds of interest. In contrast, after re-introduction of the *thiI* gene, detection of recovered s^4U indicated a successful expression of the enzyme after transformation and recurrence of native candidate 470 confirmed the direct dependency of this compound on the action of ThiI *in vivo*.

The experimental work on which this chapter is based was carried out by ██████████ under my supervision.

4.2.4 Effects of ultraviolet radiation on the occurrence of candidate 470 *in vitro*

4.2.4.1 Irradiation of nucleosides and nucleotides for structural characterization

In terms of a further structural characterization of candidate 470, the discovery of the UV-induced formation of candidate 470 in *in vitro* transcribed mRNA pointed out that the dimeric structure can be formed from canonical nucleosides and offered the promising opportunity to produce an amount of candidate 470 that would allow NMR analysis and thus elucidation of the structure. In this context, ██████████ performed a series of experiments in which the nucleosides putatively involved in the formation of candidate 470, i.e. cytidine and/or uridine, were irradiated in analogy to the treatment of *in vitro* transcribed mRNA and subsequently analyzed by LC-MS/MS (Figure 4.4). Since the initial irradiation of the pyrimidine-nucleoside mixtures did not lead to the formation of candidate 470 (representatively shown in Appendix, Figure A1), also adenosine and guanosine were added to the mixtures, considering that they were also present in the *in vitro* transcribed mRNA and might assist the formation of candidate 470 through guiding effects or other interactions. Additionally, in order to counteract heavy dilution and a concomitant distance of the reaction partners, the concentration of nucleosides was increased to the highest possible concentration for all four nucleosides to be completely dissolved in water. Furthermore, an increased temperature during the irradiation procedure was tested to accelerate a potential reaction. The fact that none of the described conditions for irradiation of nucleoside mixtures resulted in UV-induced formation of
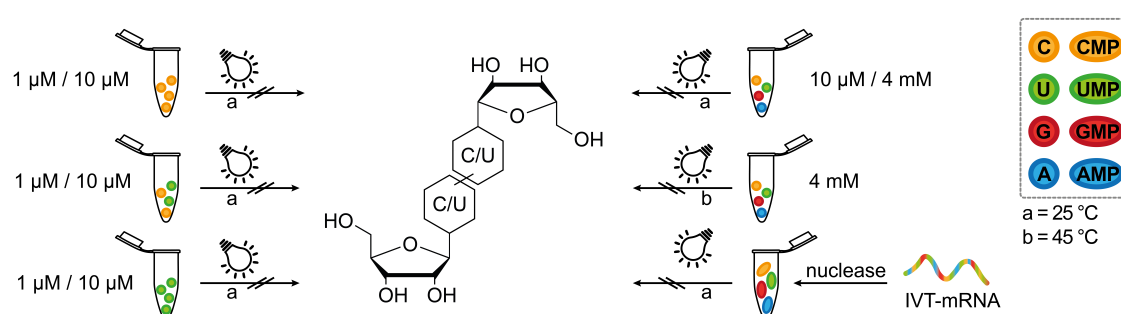
Figure 4.4: Attempts to form candidate 470 from nucleosides and nucleotides. Irradiation of nucleosides with a UV handlamp at 254 nm for 2 h in various combinations of different concentrations and at different temperatures, as indicated. The monomers are illustrated by different colors according to the color scheme on the right. Circles depict nucleosides while ellipses represent mononucleotides that resulted from hydrolysis of *in vitro* transcribed mRNA (IVT-mRNA) by nuclease treatment prior to irradiation. Coloring of the IVT-mRNA indicates a composition of all four canonical nucleosides but does not reflect the exact sequence. Crossed out reaction arrows point out that candidate 470 was not detected in the LC-MS/MS analysis subsequent to these experiments and thus was not formed from the according compounds.

candidate 470 suggested that an intact strand structure or at least a phosphate residue may be required for the dimerization.

To evaluate which case applies, *in vitro* transcribed mRNA, that, according to the previous results, was supposed to contain all building blocks to form candidate 470, was hydrolyzed to nucleotide level by nuclease treatment and exposed to UV light. The irradiated nucleotide mixture was further hydrolyzed to nucleosides and in the subsequent LC-MS/MS analysis, once more candidate 470 was not detectable, indicating that the strand structure may be necessary to create a spatial proximity of both reaction partners to form candidate 470.

4.2.4.2 Irradiation of RNA oligonucleotides for structural characterization

Considering the necessity of a strand structure for the photochemical formation of candidate 470 *in vitro*, insights on the strand context and the nucleoside species forming the dimeric structure may be gained from irradiation of RNA oligonucleotides of a certain composition. To this end, commercial RNA oligonucleotides with a sequence consisting of a combination of only two nucleoside species of which at least one was pyrimidine-based, were irradiated and analyzed by LC-MS/MS (Figure 4.5).
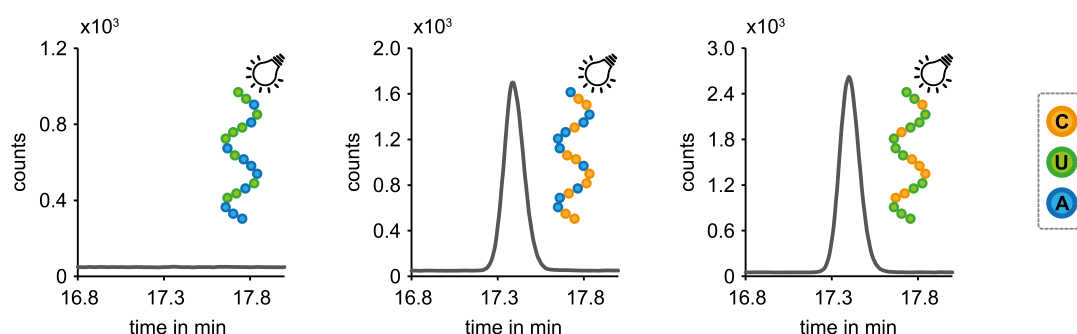
Figure 4.5: Formation of candidate 470 in oligonucleotides of a certain composition. Extracted ion chromatograms for the mass transition corresponding to candidate 470 (*m/z* 470 → 338) obtained from the qualitative LC-MS/MS analysis after irradiation of oligonucleotides with a UV handlamp at 254 nm for 2 h. Similar amounts of oligonucleotide were analyzed. Sequences (5′ → 3′) of the respectively analyzed oligonucleotides are illustrated schematically according to the color code on the right.

On the one hand the oligonucleotide exclusively containing uridine as a pyrimidine-based nucleoside did not exhibit formation of candidate 470, which matches with results from stable isotope labeling that indicated an amount of 5 nitrogen atoms within the molecule and would require an additional reaction partner in case of two uridines forming the dimeric structure. On the other hand, candidate 470 was formed after irradiation of oligonucleotides consisting of cytidine and uridine or cytidine and adenosine, indicating that the presence of cytidine is mandatory for the formation of candidate 470

in a photochemical reaction *in vitro*. Although these findings provide strong evidence for a dimer that is formed in a photochemical reaction between two cytidines including the loss of a nitrogen-containing group, assumingly in an exposed position, i.e. the exocyclic amino group, it cannot be completely excluded that the same structure might also arise from a reaction between a cytidine and a uridine which would require the loss of an oxygen-containing group. Furthermore, it is important to mention that Rhoades and Wang discovered the formation of a cytidine homoadduct after irradiation of poly-cytidylic acid which, although formed by other mechanisms, was reported to have the same structure as the s^4U8-C13 crosslink.[314,318,319]

Following up the finding, that candidate 470 can be formed from two cytidines, another question to be addressed by using oligonucleotides of a certain sequence was the relative position of the two cytidines required to form the dimeric structure. Considering a spatial proximity of both reaction partners (see above), an intramolecular reaction seemed plausible. Assessing if this kind of reaction requires a direct neighborhood of two cytidines, an oligonucleotide containing several cytidines but none directly adjacent to another cytidine was irradiated and analyzed by LC-MS/MS (Figure 4.6a). The detection of candidate 470 in this sample implied that a direct adjacency of cytidines within an RNA molecule may not necessarily be required and raised the question if the formation of candidate 470 might also be of intermolecular nature. This issue was addressed by irradiation of an oligonucleotide with only one cytidine site and subsequent LC-MS/MS analysis (Figure 4.6b). Interestingly, candidate 470 was detected in this sample, providing
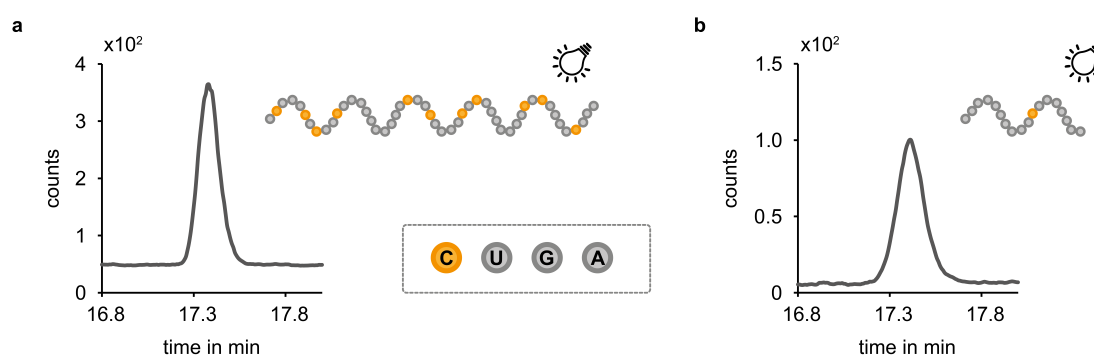
Figure 4.6: Formation of candidate 470 in oligonucleotides containing cytidines in certain positions. Following the depicted color code, uridines, guanosines and adenosines in the sequences ($5' \rightarrow 3'$) of the respectively analyzed oligonucleotides are displayed uniformly to highlight the distribution of cytidines within these sequences. The exact sequences of these oligonucleotides are listed in the materials and methods section (see 6.1.4.1). **a** Extracted ion chromatogram for the mass transition corresponding to candidate 470 (*m/z* 470 $\rightarrow$ 338) obtained from the qualitative LC-MS/MS analysis after irradiation of an oligonucleotide containing no directly adjacent cytidines with a UV handlamp at 254 nm for 2 h. **b** Extracted ion chromatogram for the mass transition corresponding to candidate 470 (*m/z* 470 $\rightarrow$ 338) obtained from the qualitative LC-MS/MS analysis after irradiation of an oligonucleotide containing only one cytidine with a UV energy dose of 1 J/cm^2.

evidence that the dimeric structure may indeed be formed in an intermolecular reaction if formation of candidate 470 exclusively occurs between two cytidines, assuming that uridines are not involved.

4.2.4.3 Influence of different UV energy doses on the formation of candidate 470

In order to further characterize the UV-dependent formation of candidate 470, the correlation between the applied UV irradiation energy and the formation of candidate 470 was examined. Therefore, RNA from HEK cells, qualified by the absence of native candidate 470 and an easy accessibility compared to *in vitro* transcribed RNA, was treated with increasing UV energy doses at 254 nm and subsequently analyzed by LC-MS/MS. Compliance and reproducibility of defined UV energy doses was ensured by use of a UV crosslinker that was employed in the energy mode, i.e. UV light was emitted by the light bulbs in the sample chamber until the built-in UV sensor perceived the defined dose of UV energy. To start with, the RNA samples were exposed to UV energy doses between 0 and 10 J/cm^2.

This initial experiment showed a consistently linear correlation between the applied dose and the formation of candidate 470 in RNA from HEK cells without any sign of saturation (Figure 4.7a). To take the UV-dependent formation of candidate 470 to extremes and assess if there might be any factors limiting the generation of the dimeric structure, in a follow-up experiment, RNA from HEK cells was irradiated with even higher doses of UV energy (14-38 J/cm^2) and analyzed by LC-MS/MS. Here, a dose of up to 25 J/cm^2 maintained the linear dependence of the candidate 470 level, while higher doses led to a flattening of the curve and saturation seemed to be reached upon irradiation with a dose of approximately 32 J/cm^2 (Figure 4.7b).

Considering that at least one cytidine and one more pyrimidine-derived nucleoside were shown to be involved in the UV-induced formation of candidate 470 *in vitro*, the availability of cytidine and uridine during the course of experiments with increasing UV energy doses may provide important information in the context of limiting factors. To this end, the levels of cytidine and uridine in samples irradiated at different doses were investigated on the basis of the UV signal of these main nucleosides in LC-MS/MS analysis. While the amount of uridine showed an exponential decrease in correlation with the UV energy dose, the response of cytidine to increasing UV energy doses was more linear in nature and not as drastic as in the case of uridine. For both pyrimidine-based nucleosides, the formation of lesions in the form of dimers or photohydrates upon UV irradiation was described in several studies, likewise noting an unequal behavior of cytidine and uridine which was attributed to a slower reaction rate of cytidine and a rapid reversal of cytidine hydrate formation.[168,179,312,320,321] The generation of such structures
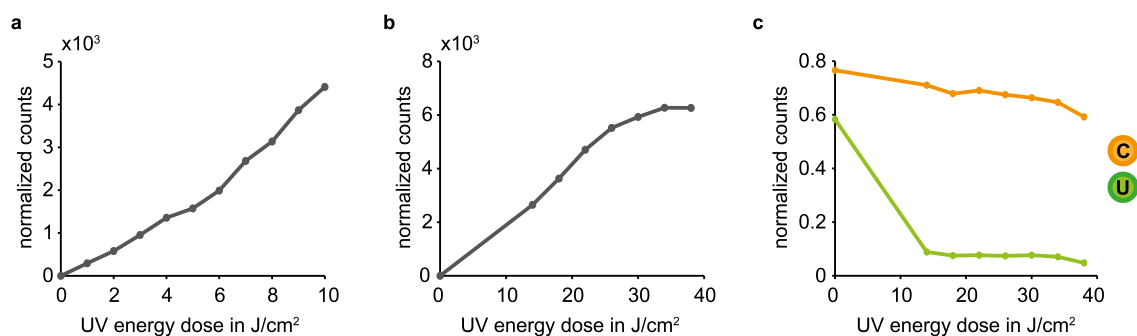
Figure 4.7: Influence of different UV energy doses on the occurrence of candidate 470 and pyrimidine-based nucleosides. a Abundance of candidate 470 in differently irradiated HEK RNA samples as detected in LC-MS/MS analysis and normalized to the UV signal of adenosine. **b** Abundance of candidate 470 in HEK RNA samples as detected in LC-MS/MS analysis and normalized to the UV signal of adenosine plotted against the UV energy dose used for irradiation. Absolute values are not comparable to (a) since measurements were recorded on different days and MS performance is a daily varying parameter. **c** Abundances of cytidine and uridine as detected in the UV traces recorded during LC-MS/MS analysis were normalized to the respective adenosine signal (nearly constant across all samples) plotted against the UV energy dose the HEK RNA samples were exposed to.

might presumably be the reason for the pronounced reductions in the cytidine and uridine levels, since the detected amounts of candidate 470 on MS level are incommensurate with the observed changes in the main nucleoside levels in the UV trace, suggesting that the formation of candidate 470 could be just one of several reactions that can occur under UV influence. As mentioned previously, in particular cytidine is a discriminatory monomer for the UV-dependent formation of candidate 470. However, since the UV energy dose leading to saturation of candidate 470 levels was not accompanied by a severe cytidine deficiency, it remains unclear if the availability of intact cytidines is the only limiting factor in the previously observed saturation behavior or if other factors come to play. Nevertheless, based on the results observed here, it can be assumed that a certain level of natural cytidines must be present in the RNA strand to induce the formation of candidate 470 by UV irradiation of RNA samples.

4.2.4.4 Extent of UV-induced formation of candidate 470 *in vitro* in different RNA samples

While the native occurrence of candidate 470 in RNA of different origins *in vivo* was clearly defined, the UV-induced formation of candidate 470 *in vitro* did not seem very particular at first sight, apart from the mandatory presence of cytidine. To assess if there are any differences in the UV-dependent formation of candidate 470 in tRNAs from different biological backgrounds, the occurrence of candidate 470 in total tRNA extracted from HEK cells was compared to its occurrence in total tRNA from two different *E. coli* strains, namely commercial tRNA from the MRE600 strain and freshly extracted tRNA

from the Keio parent strain (K-12), upon irradiation with UV energy doses between 0 and 20 J/cm^2 (Figure 4.8a).

In line with previous results, the irradiation of HEK tRNA with increasing UV energy doses showed a linear increase of candidate 470 levels. A similar behavior emerged for the irradiation of commercial tRNA from MRE600, although the relative amounts of candidate 470 were slightly increased compared to the HEK samples. In contrast to these linear dependencies, the course of candidate 470 formation in freshly extracted tRNA from the *E. coli* K-12 strain showed a high increase at the first sampling which flattened later on in the experiment to reach a similar slope compared to the other two RNA samples, but constantly displayed significantly higher levels of candidate 470. While differences between the tRNA from bacterial and mammalian cells were expected, not least because of the native occurrence of candidate 470, the heterogenous behavior of both *E. coli* samples implied that the freshly isolated tRNA from the K-12 strain might have contained some components, that actively favored the formation of candidate 470 upon UV irradiation on the one hand, but might be used up at some point on the other hand. Consequently, the UV-dependent formation of candidate 470 might be forced to follow the same reaction pathway as in the HEK and MRE600 sample at higher UV doses and might by then involve main nucleosides only. Considering the long storage period of the commercial tRNA from MRE600, one plausible explanation for the different behavior of the *E. coli* samples might be traced to the presence or absence of certain RNA modifications that are prone to oxidative depletion over time. A typical class of modifications that are known to be susceptible to such successive depletion processes are thiolated modifications, including s^{4}U, which was of particular interest in the present context, not only due to its literature-known role as UV sensor, but also due to the *in vivo* experiments performed within this study which pointed to a direct correlation between the s^{4}U-introducing enzyme ThiI and the native occurrence of candidate 470. Following a protocol of Watanabe *et al.*, the putative role of s^{4}U in the UV-induced formation of candidate 470 *in vitro* was examined by treating the freshly isolated *E. coli* K-12 tRNA with hydrogen peroxide for different time spans and thus creating an oxidative environment that was shown to result in dethiolation of RNA, i.e. a conversion of s^{4}U into unmodified uridines.[322] Subsequently, the samples with presumably different severities of dethiolation, were irradiated with a UV energy dose of 10 J/cm^2 and analyzed by LC-MS/MS for their candidate 470 content. Furthermore, to evaluate the degree of dethiolation, the s^{4}U levels in the samples with different times of exposure to hydrogen peroxide without further UV treatment were analyzed by LC-MS/MS (Figure 4.8b).

Proving the success of the hydrogen peroxide treatment in terms of dethiolation of the RNA samples, the analysis of s^{4}U levels showed a substantial reduction after hydrogen peroxide treatment compared to the untreated control, even at the shortest treatment
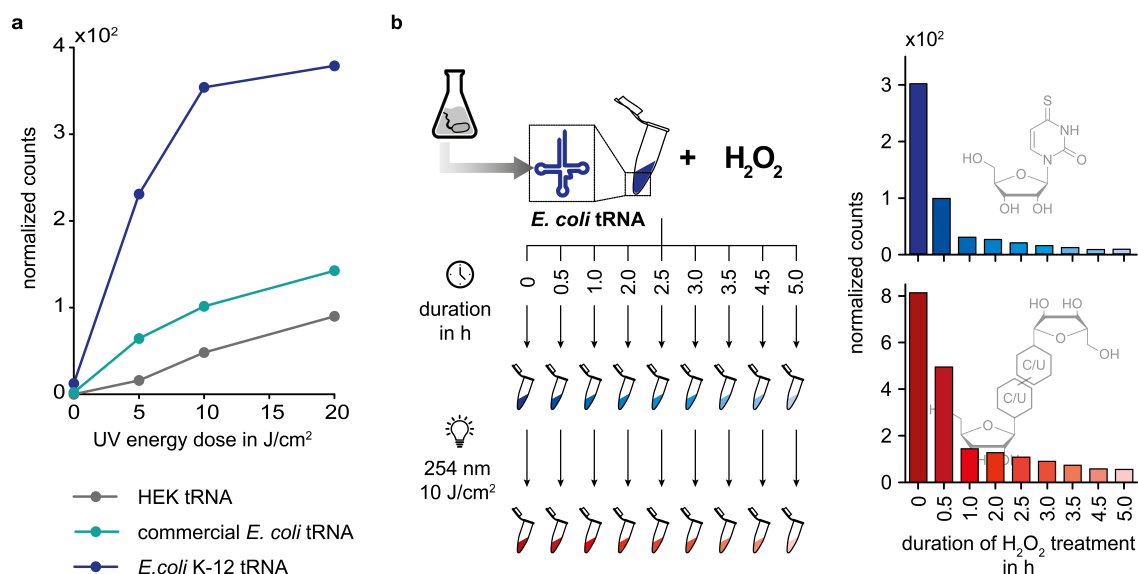
Figure 4.8: UV-induced formation of candidate 470 in vitro in different RNA samples. a Abundance of candidate 470 in differently irradiated RNA samples of various origins as detected in LC-MS/MS analysis and normalized to the UV signal of adenosine. **b** Left: Scheme to illustrate the dethiolation experiment in which freshly prepared *E. coli* K-12 tRNA was treated with H_2O_2 for different times (first sampling, blue) and subsequently irradiated with a UV energy dose of 10 J/cm^2 at 254 nm (second sampling, red). Right: Abundances of s^{4}U in dethiolated samples and candidate 470 in dethiolated and irradiated samples as detected in LC-MS/MS analysis and normalized to the UV signal of adenosine were plotted against the duration of hydrogen peroxide treatment. The (schematic) structure in the background of the diagram illustrates the analyzed compound, respectively.

duration. This effect was even more severe for increased exposure times, leaving only a minimal amount of s^{4}U in these RNA samples. A similar picture emerged for the levels of candidate 470 formed upon irradiation of the dethiolated RNA samples. Here, a progressing degree of dethiolation yielded significantly less candidate 470 compared to the s^{4}U-containing RNA sample, pointing to an entanglement of s^{4}U in the UV-induced formation of candidate 470 *in vitro*. In this context, the association of higher s^{4}U levels and higher candidate 470 levels provided evidence for a favoring effect of s^{4}U. This would also be consistent with previous results of comparatively low candidate 470 levels upon irradiation of *E. coli* RNA that was stored for a longer time period and thus was prone to successive s^{4}U oxidation as well as HEK RNA, which, if at all, only contains negligibly small amounts of s^{4}U (Figure 4.8a). However, as shown in previously described experiments with s^{4}U-free oligonucleotides (Figures 4.5 and 4.6) and *in vitro* transcribed RNA (Figure 4.2), the dimeric structure can also be formed from the main nucleosides and a presence of s^{4}U in the RNA strand is not pivotal for the UV-induced formation of candidate 470 *in vitro*. Given the structural identity of the dimer resulting from a UV-induced crosslink of s^{4}U8 and C13 in *E. coli* tRNA and the dimer obtained after irradiation of polycytidylic acid, these findings support the possibility that candidate 470 may be this same structure.[313,314,319]

This chapter (4.2.4) is based on experimental work by ▆▆▆▆▆▆▆▆ (subchapters 4.2.4.1 and 4.2.4.2) which was complemented by ▆▆▆▆▆▆▆▆ (subchapter 4.2.4.2: experiment with an oligonucleotide containing only one cytidine in its sequence and subchapters 4.2.4.3 as well as 4.2.4.4) under my supervision, respectively.

4.2.5 Structure proposals for candidate 470 and first synthetic approaches

4.2.5.1 Structure proposal based on the native formation of candidate 470 *in vivo*

Considering the native occurrence of candidate 470 and the identification of ThiI as a central player in the biosynthetic formation of candidate 470 *in vivo*, a closer look at the different steps of the catalyzed reaction might provide clues on the type of connection between the two pyrimidine-nucleosides. One biosynthetically plausible route to yield a dimer-forming bond would involve a nucleophilic attack of the nitrogen atom within the exocyclic amino group of a cytidine at position 4 of the uridine-8 in tRNA after its initial activation via adenylation of the oxygen atom, assuming a temporal offset of sulfur insertion. These assumptions shaped the proposal of two pyrimidine-based nucleoside structures, whose respective position 4 would be attached to each other via an amino linker (Figure 4.9a). However, since it was only shown that the occurrence of candidate 470 strictly depended on the presence of ThiI and this is constantly accompanied by generation of s^{4}U, it cannot be excluded, that the dimer might be formed after completion of the ThiI reaction between s^{4}U and a cytidine. Furthermore, it is not possible to predict if the additional cytidine which might putatively be involved in the native formation of candidate 470 in *E. coli* is present as a monomeric species or bound within an RNA strand.

Generally, the unambiguous verification of such structure proposals for newly identified and elucidated modification structures demands for NMR analysis, which is hampered by the low abundance of the native compounds and a lack of methods to sufficiently enrich them. A remedy can be found by organic synthesis of the proposed structure as a reference substance and its alignment with the naturally occurring compound in LC-MS/MS analysis. In the present case, this was realized in a collaboration with ▆▆▆▆▆▆▆▆ from the ▆▆▆▆ group (Johannes Gutenberg-University Mainz, Germany) who synthesized the proposed structure (Figure 4.9a) in multiple steps. Simultaneous LC-MS/MS analysis of native candidate 470 in a ^{13}C-labeled *E. coli* tRNA hydrolysate (retention time: 17.4 min) and the synthetic compound (retention time: 19.9 min) unveiled significant differences in the chromatographic behavior, providing strong evidence that the proposed structure, which was considered the most plausible structure from the biosynthetic point of view, is not identical with the actual structure of candidate 470. Therefore, the structure pro-

posal linking the respective 4-position of the two pyrimidine-based monomers can be excluded from the list of potential structures for candidate 470.
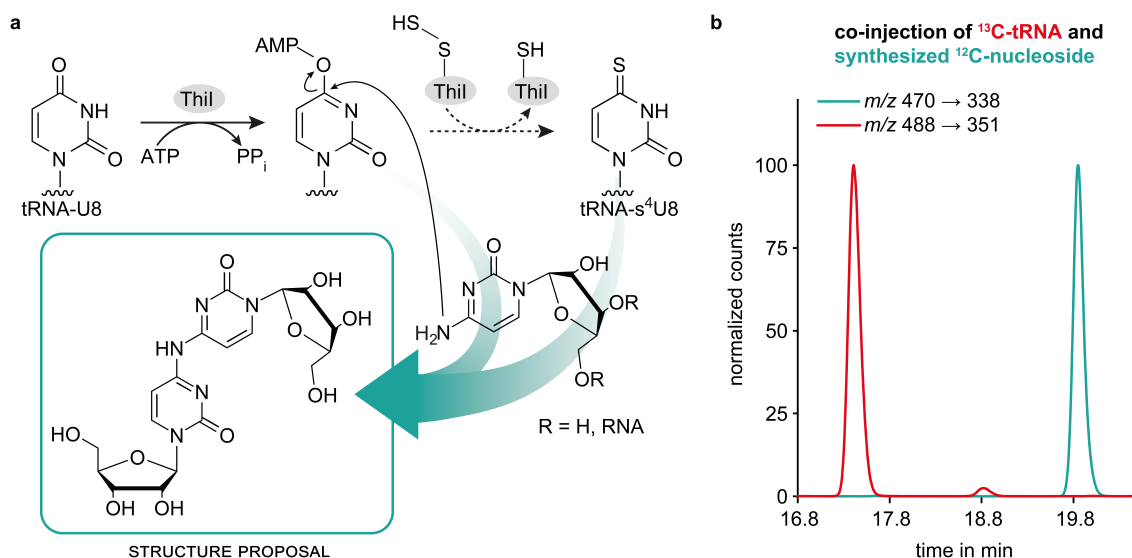
Figure 4.9: Evaluation of a biosynthetically plausible structural proposal. a Plausible pathways for a ThiI-related biosynthesis of candidate 470 in *E. coli* which involve either the intermediate of the enzymatic reaction and a cytidine (conceivable mechanism indicated) or the ThiI-product s^{4}U and a cytidine, establishing the structure proposal in the box. The reacting cytidine may either be a free nucleoside or integrated in an RNA strand.**b** LC-MS/MS analysis of hydrolyzed ^{13}C-labeled tRNA (red) spiked with the synthesized dimer shown as structure proposal in (a) (teal). An overlay of extracted ion chromatograms at a mass transition of *m/z* 470 → 338 and *m/z* 488 → 351 is shown. Abundances were normalized to the highest peak in the respective chromatogram (set to 100%).

4.2.5.2 Possible structures for candidate 470

For further structural elucidation of candidate 470, a wide variety of differently linked structures have to be taken into account because they meet characteristics like molecular mass and elemental composition (Figure 4.10). Due to the number of five nitrogen atoms within candidate 470 as well as the required presence of cytidine for dimer formation in oligonucleotide experiments it is probable that one pyrimidine-based monomer might be a cytidine which is marked with an "X" in the structures to better distinguish between the monomers for discussion while the other binding partner is marked with a "Y" and might either be derived from a cytidine or a uridine. Since an isomer of uridine frequently occurs in native RNA in the form of pseudouridine, an involvement of pseudouridine would also be a theoretically feasible option, however, this seems highly unlikely since there was no sensitivity in any of the Ψ-related modification enzyme KOs in the initial biological screening and a dimer containing this modified nucleoside would exhibit one mass transition of 132 Da and one mass transition of 36 Da (specific for Ψ) instead of the double loss of ribose in LC-MS/MS analysis.[297] Therefore, such structures
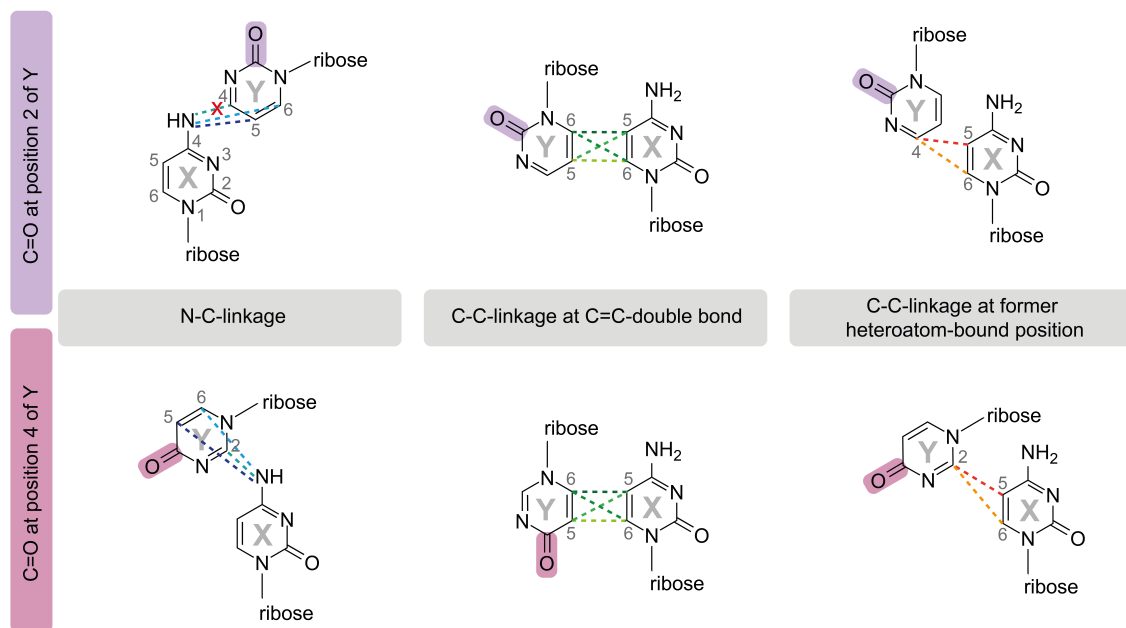
Figure 4.10: Overview of possible structures for candidate 470. To allow better description, the two different monomers were marked with "X" and "Y". Potential dimer-forming bonds are depicted as colored dashed lines. Complete numbering of positions within pyrimidines is recapped at the example of monomer X in the upper left (gray numbers) while in the following only positions involved in potential linkages are marked with their number according to nomenclature. Structures were sorted according to the location of the carbonyl group in pyrimidine Y (highlighted in corresponding color) and the type of dimer-forming bond. The red cross indicates that this linkage was already shown to be not the correct one. The 5-4 bipyrimidine adduct (depicted by the dashed red bond in the upper right) corresponds to the structure resulting from the s4U8-C13 crosslink in *E. coli* tRNA as well as the structure that emerged upon irradiation of polycytidylic acid.[314,319]

were excluded from further considerations. Among the list of potential structures for candidate 470, a first decisive feature for evaluation and classification of the structure proposals is the position of the carbonyl group in monomer Y. While a carbonyl group in position 4 of pyrimidine Y is theoretically conceivable, based on the findings that candidate 470 can be formed from two cytidines (carbonyl group at position 2) and that its formation is not related to the s2U-introducing enzyme MnmA, which would activate the 2-position for reaction and leave the carbonyl group in position 4 unaffected, it seems more likely that the carbonyl group within monomer Y might be located in position 2. Another criterion for classification is the type of linkage connecting the monomers X and Y among which are nitrogen-carbon-linkages involving the exocyclic amino group of the cytidine X, and diverse carbon-carbon-linkages either connecting carbon atoms in respective positions 5 and/or 6 (C=C double bond) of each monomer or connecting one carbon atom of the double bond to the carbon atom in position 4. In all described cases a loss of the heteroatom and possibly associated hydrogen atoms is required, which, from the biosynthetic point of view, seems the most likely if formation of the bond linking both monomers involves position 4 of monomer Y. However, since synthesis of one structure

proposal fulfilling this criterion, namely the structure proposal linking the respective 4 position of the two pyrimidine-based monomers, exhibited disparity to the natively occurring candidate 470, the structure proposals linking one carbon atom of the C=C double bond of monomer X to position 4 of monomer Y come to the fore. Importantly, the structure linking position 5 of monomer X to position 4 of monomer Y corresponds to the structure that is considered a UV sensor of tRNA forming a crosslink between s^{4}U8 and C13 in *E. coli* tRNA as well as the structure that emerged upon irradiation of polycytidylic acid.[313,314,319] With regard to the *in vitro* experiments performed with de-thiolated tRNA and the observed correlation of the formation of candidate 470 and the presence of s^{4}U, along with the central role of cytidine in oligonucleotide experiments, this seems a promising structure proposal, albeit other linkages cannot be completely ruled out. However, the UV-induced *in vivo* formation of candidate 470 in HEK cells (Figure 4.2) indicates that, in the context of UV damage *in vivo*, at least part of candidate 470 does not involve s^{4}U in its formation but might originate from other mechanisms as for example a reaction of two cytidines. This was further emphasized by a first experiment on the *in vivo* occurrence of candidate 470 after irradiation of the *E. coli thiI* KO strain which also showed formation of candidate 470 in its s^{4}U-free tRNA (Appendix, Figure A2). Despite these considerations, both confirmation and rejection of any potential structure from the list rely on organic synthesis, an endeavor that is pursued by the ▮▮▮▮ group, and subsequent alignment with the native compound in LC-MS/MS analysis as demonstrated for the first organically synthesized compound (Figure 4.9).

4.2.6 Occurrence of a candidate 470 analog in DNA

Since most photochemical reactions that occur in the context of nucleic acids arise from the nucleobases and this can also be assumed for the UV-induced formation of candidate 470 in RNA, it would be conceivable that an analogous dimer would be also formed from two 2′-deoxycytidines (dC) in DNA upon UV irradiation *in vitro*. Obviously, this derivative would contain two 2′-deoxyribose moieties instead of the ribose moieties (Figure 4.11a) and is therefore termed dideoxy-470 (d470) in the following. It is worth mentioning that, in the course of their studies on the formation of cytidine dimers from irradiation of polycytidylic acid in aqueous solution and irradiation of cytidine in frozen aqueous solution (solid state), Rhoades and Wang also observed formation of an equivalent dimer from deoxycytidine under the same conditions which consequently corresponds to the deoxy-derivative of the s^{4}U8-C13 crosslink structure.[318,319]

In order to evaluate a potential UV-dependent formation of d470 in DNA *in vitro*, in a first approach DNA extracted from *E. coli* was exposed to UV energy doses between 0 and 10 J/cm^2 and subsequently analyzed by LC-MS/MS regarding the neutral loss of a deoxyribose. In contrast to the findings for candidate 470 in *E. coli* RNA, there was

no indication for d470 in the non-irradiated sample, suggesting that d470 does not natively occur in cellular DNA. Nevertheless, the LC-MS/MS analysis of the irradiated DNA samples revealed a compound with the appropriate mass transition at a retention time of approximately 22.3 min. As a consequence of the slightly altered structure and physicochemical properties, i.e. the reduced polarity of d470 compared to candidate 470, due to the missing hydroxyl groups in the $2'$-positon of $2'$-deoxyribose, the shift towards a later elution from the reversed-phase column seemed plausible. Furthermore, the occurrence of this compound directly depended on the UV treatment and its abundance increased upon exposure to a higher UV energy dose, sustaining the hypothesis, that UV treatment might indeed induce the formation of d470 in DNA *in vitro* (Figure 4.11b).
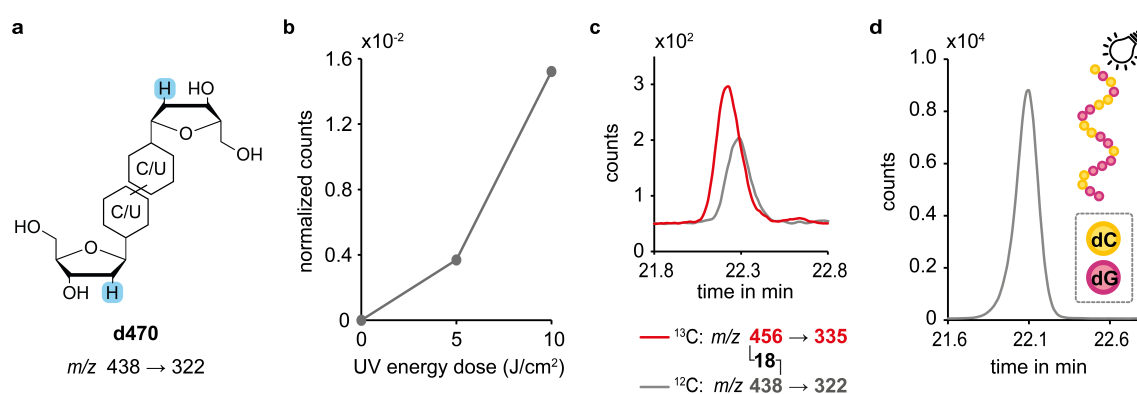
Figure 4.11: UV-induced formation of the candidate 470 DNA analog d470 *in vitro*. a Structure model illustrating the molecule sites altered in the DNA analog (compared to the RNA candidate 470) and its mass transition in LC-MS/MS analysis. **b** Abundance of putative d470 signal (at 22.3 min) in *E. coli* DNA samples as detected in LC-MS/MS analysis and normalized to the UV signal of $2'$-deoxyadenosine, plotted against the UV energy dose used for irradiation. **c** Uniformly labeled ^{13}C-DNA and unlabeled ([12]C-)DNA were extracted from *E. coli* and analyzed by LC-MS/MS after exposure to a UV energy dose of 5 J/cm². The extracted ion chromatograms displayed here correspond to the mass transitions of ^{13}C-d470 in the irradiated ^{13}C-labeled DNA sample and the ^{12}C-d470 in the irradiated unlabeled DNA sample, illustrating the presence of 18 carbon atoms within this compound. **d** Extracted ion chromatogram for the mass transition corresponding to d470 (*m/z* 438 → 322) obtained from the qualitative LC-MS/MS analysis after irradiation of a DNA oligonucleotide with the schematically shown sequence ($5' \rightarrow 3'$) containing only $2'$-deoxycytdines and $2'$-deoxyguanosines with a UV energy dose of 10 J/cm².

In order to substantiate the assumption of the detected compound to correspond to d470, it was inevitable to assess a similar composition compared to candidate 470. To this end, *E. coli* were grown in ^{13}C-labeled medium and the accordingly labeled DNA was extracted for subsequent *in vitro* irradiation and LC-MS/MS analysis. Importantly, the observed mass shift of +18 in comparison to an unlabeled control sample confirmed an equal number of carbon atoms in candidate 470 and the supposed d470 (Figure 4.11c).

Since *in vitro* experiments performed with RNA indicated that the exclusive presence of cytidines as pyrimidine-based building block within an oligonucleotide is sufficient for formation of the dimeric structure, this dependance was also assessed for formation of

the DNA analog d470 by irradiating a DNA oligonucleotide consisting of 2′-deoxyguanosine and 2′-deoxycytidine. In the subsequent LC MS/MS analysis, the signal assigned to d470 was clearly apparent and pointed out, that the putative DNA analog d470 can equally be formed from two 2′-deoxycytidines (Figure 4.11d).

Assessing if the putative d470, similar to its RNA analog, might be formed *in vivo* upon UV irradiation in the context of UV damage to nucleic acids, in some first experiments *E. coli* cells were exposed to increasing doses of UV energy (0-5 J/cm^2). Interestingly, the DNA extracted from these cells did not exhibit any signal for the putative d470 in the subsequent LC-MS/MS analysis (Appendix, Figure A3), indicating that a UV-dependent formation of d470 does not take place *in vivo* and its occurrence is limited to the *in vitro* level.

Although the findings within this chapter point to the existence of a DNA analog of candidate 470, this can ultimately only be proven by synthesis of the DNA analog after elucidation and validation of the structure of the RNA species.

The experimental work within this chapter was performed by ▮▮▮▮▮▮▮▮ under my supervision with the exception of the isotope labeling experiment which was performed by myself.

4.3 Investigation of a manipulation of the Q modification on molecular level

In recent years, a general trend has emerged in the field of epitranscriptomics to investigate the correlation of RNA modifications and diseases and, consequently, their therapeutic potential.[17,126] This also applies to the hypermodification Q, for whose occurrence in tRNA in both prokaryotes and eukaryotes the catalysis of a base exchange reaction by the organism-specific TGT is inevitable.[127] In an animal model of multiple sclerosis, the use of a surrogate instead of the natural substrate showed promising results for a potential therapeutic application, although incorporation was only indirectly demonstrated and the molecular mechanisms and effects were not studied.[155] In order to track the processes on the molecular level, our work aimed at the incorporation of azide-containing, i.e. clickable, derivatives of the nucleobase precursor preQ$_1$, which differed in the length and type of their side chain attached to the amino linker and were synthesized by the ▮▮▮▮▮ group (University of Innsbruck, Austria). While our group focused on the incorporation of these preQ$_1$ analogs *in vivo* in *E. coli* and established the detection via click chemistry and LC-MS/MS analysis (experimental work performed by Larissa Bessler) as well as the physiological consequences of such an incorporation (experimental work performed by ▮▮▮▮▮▮▮▮), we teamed up with the group of ▮▮▮▮▮▮▮▮▮▮▮▮▮▮ (Humboldt-Universität zu Berlin, Germany) who have profound expertise in *in vitro* experiments with bacterial and human TGT and in working with *S. pombe*, a eukaryotic model organism frequently used for studies in the context of queuosine.

As a first step, the tolerance of bTGT regarding the preQ$_1$ analogs *in vitro* was examined and it was demonstrated by click reaction that all preQ$_1$ analogs were successfully incorporated and consequently resulted in a semisynthetic nucleoside structure (experiments performed by ▮▮▮▮▮▮▮). In order to prevent the surrogate from competing with the natural preQ$_1$ during TGT-mediated incorporation *in vivo* in bacteria, I determined an appropriate knockout strain for feeding experiments. After supplementation of the selected strain with one of the preQ$_1$ derivatives, I was able to demonstrate a congruent dose-dependent incorporation of the preQ$_1$ analog into RNA *in vivo* by an intensity measurement of fluorescence after tagging via click reaction as well as by LC-MS/MS analysis of the semisynthetic nucleoside species. Taking advantage of the clickability of the introduced analog for affinity purification, our collaboration partners also investigated the RNA substrates of the TGT in *S. pombe* and were able to demonstrate that the TGT retains its four tRNA substrates and no other RNA species were modified. Since Q is mainly associated with speed and accuracy of translation as well as a modification circuit with m^{5}C38 in tRNA^{Asp}, we also investigated the involvement of surrogate-containing

tRNA in the translational process and the degree of stimulation of C38 methylation to assess the biological functionality of the incorporated analog.[146,147,152,159] The former was initially addressed by comparing the levels of Q and the semisynthetic nucleoside in polysomal tRNA in relation to the distribution in the free tRNA pool in *E. coli* (contribution of ███████████ from the ████ group). Although reduced in comparison to Q, an equal distribution of the derivative between the tRNAs actively involved in translation and the cytoplasmic pool, and thus an integration in the translational process was demonstrated. In view of a possible long-term therapeutic application, we joined forces with ███████████ (Medical Faculty Mannheim of Heidelberg University, Germany) to evaluate the situation in human cells and the analysis of these samples likewise implied recognition of the tRNAs bearing the analog by the translating ribosome *in cellulo*. Furthermore, bisulfite sequencing (performed by our collaboration partners) also demonstrated a functional integration of the analog into the aforementioned modification circuit, i.e. a dose-dependently enhanced methylation of tRNA$^{\mathrm{Asp}}$ C38 in *S. pombe*. These experiments suggested that the incorporated analog cannot only replace the native substrate of bacterial and eukaryotic TGT, but can also fulfill the function of Q to a certain extent. Considering the exclusive incorporation into the selected tRNA substrates and the maintenance of the functionality of Q, the developed approach is a minimally invasive method that paves the way for a detailed investigation of the biological effects of Q at the protein level under both normal and pathological conditions in different cell lines and provides a basis for further studies on the therapeutic potential of a targeted manipulation of Q.

Functional integration of a semi-synthetic azido-queuosine derivative into translation and a tRNA modification circuit

Larissa Bessler[1],[†], Navpreet Kaur[2],[†], Lea-Marie Vogt[1],[†], Laurin Flemmich[3], Carmen Siebenaller[4], Marie-Luise Winz [1], Francesca Tuorto [5], Ronald Micura [3], Ann E. Ehrenhofer-Murray [2],[*] and Mark Helm [1],[*]

[1]Institute of Pharmaceutical and Biomedical Sciences, Johannes Gutenberg-University Mainz, 55128 Mainz, Germany, [2]Institute of Biology, Humboldt-Universität zu Berlin, 10117 Berlin, Germany, [3]Department of Organic Chemistry, University of Innsbruck, 6020 Innsbruck, Austria, [4]Department of Chemistry – Biochemistry, Johannes Gutenberg-University Mainz, 55128 Mainz, Germany and [5]Division of Biochemistry, Mannheim Institute for Innate Immunoscience (MI3), Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

ABSTRACT

Substitution of the queuine nucleobase precursor preQ$_1$ by an azide-containing derivative (azido-propyl-preQ$_1$) led to incorporation of this clickable chemical entity into tRNA *via* transglycosylation *in vitro* as well as *in vivo* in *Escherichia coli*, *Schizosaccharomyces pombe* and human cells. The resulting semi-synthetic RNA modification, here termed Q-L1, was present in tRNAs on actively translating ribosomes, indicating functional integration into aminoacylation and recruitment to the ribosome. The azide moiety of Q-L1 facilitates analytics *via* click conjugation of a fluorescent dye, or of biotin for affinity purification. Combining the latter with RNAseq showed that TGT maintained its native tRNA substrate specificity in S*. pombe* cells. The semi-synthetic tRNA modification Q-L1 was also functional in tRNA maturation, in effectively replacing the natural queuosine in its stimulation of further modification of tRNA^{Asp} with 5-methylcytosine at position 38 by the tRNA methyltransferase Dnmt2 in *S. pombe*. This is the first demonstrated *in vivo* integration of a synthetic moiety into an RNA modification circuit, where one RNA modification stimulates another. In summary, the scarcity of queuosinylation sites in cellular RNA, makes our synthetic q/Q system a 'minimally invasive' system for placement of a non-natural, clickable nucleobase within the total cellular RNA.

INTRODUCTION

Post-transcriptional modification of tRNAs is a ubiquitous yet idiosyncratic feature with versatile chemical structures contributing to stability and folding, as well as fidelity of decoding and translational control (1–3). The largest variety of chemical structures in RNA is found in the anticodon loop which directly interacts with the mRNA during decoding in the translating ribosome. The chemical variety of the more than 170 modifications known to date is dominated by tRNA anticodon modifications occurring at positions 34 and 37, ranging from simple methylations to highly complex structures of which queuosine (Q) is a particular case (4,5). In both, prokaryotes and eukaryotes, this hypermodified 7-deazaguanosine is exclusively found in the anticodon wobble position 34 of tRNAs containing a $G_{34}U_{35}N_{36}$ motif and therefore specific for a selected group of four tRNAs, namely tRNA^{Asn}, tRNA^{Asp}, tRNA^{His} and tRNA^{Tyr} (6,7). In an intricate multi-step process involving various enzymes and co-factors, *Escherichia coli* and other prokaryotes are capable of first synthesising the modified precursor base 7-aminomethyl-7-deazaguanine (preQ$_1$) *de novo*. GTP is converted to preQ$_1$ *via* five enzymatic steps successively catalysed by the GTP cyclohydrolase I (GCH I), QueD, QueE, QueC and QueF (8–11). As a rare type of post transcriptional modification, the noncanonical nucleobase structure is then introduced into tRNA in an exchange reaction. During this transglycosylation step, the bacterial tRNA guanine transglycosylase (bTGT) replaces

[*]To whom correspondence should be addressed. Tel: +49 6131 39 25731; Fax: +49 6131 39 20373; Email: mhelm@uni-mainz.de
Correspondence may also be addressed to Ann E. Ehrenhofer-Murray. Tel: +49 30 2093 49630; Fax: +49 30 2093 49641; Email: ann.ehrenhofer-murray@hu-berlin.de
[†]The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

a guanosine in the anticodon wobble position of cognate tRNAs with the precursor base preQ$_1$ (12,13) which is then further enzymatically modified by QueA and QueG to yield the final queuosine structure (14,15). In contrast to prokaryotes, eukaryotes salvage the nucleoside queuosine and the corresponding nucleobase queuine (q) from environmental sources including the gut microbiota (reviewed in (5)). Queuosine is hydrolyzed by a queuosine nucleoside glycosylase to release q (16). The incorporation of the salvaged q into tRNA is catalysed by eukaryotic TGT (eTGT), which is a heterodimeric enzyme (17–19) composed of a catalytic queuine tRNA-ribosyltransferase subunit 1 (QTRT1) and a noncatalytic queuine tRNA-ribosyltransferase subunit 2 (QTRT2) (20).

Despite its suggestive positioning at position 34 of the tRNA anticodon, molecular details of the physiological relevance of Q remain scarce. It is generally accepted that Q impacts the decoding process on the translating ribosome, with cumulative evidence pointing to pivotal interactions at the A-site. The specific occurrence of Q in GUN anticodons is consistent with a general concept by Grosjean and Westhof (21), wherein modifications at position 34 compensate for the lower stability of codon-anticodon interactions including 2 or more base pairs with less than three hydrogen bonds. This concept receives support from computational modelling, which characterised a stabilising effect of Q on the overall tRNA-mRNA complex involving additional hydrogen bonds (22). *In vivo* and *in cellulo* studies did not reveal any strong phenotypes under Q deficiency or in strains lacking TGT. However, the presence of Q improved viability under stress conditions and affected translation accuracy in *E. coli* (23,24). *In vivo* studies in eukaryotes likewise reported an impact on the decoding process, enabling decoding of synonymous codons by wobble base pairing (22,25,26), and affecting translation speed and accuracy (27,28). Queuosine's multifaceted involvement in the cellular machinery was reported to be associated with cancer (29–32), neuronal disorders (33–35) as well as bacterial and parasitic infection (36,37). Consequently, the perception of therapeutic potential associated with its biogenesis has consistently increased, in keeping with a general trend in epitranscriptomics

So far, the only demonstrated molecular interaction affected by Q outside the ribosome is a so-called modification circuit with 5-methylcytidine (m^{5}C) in position 38 of *Schizosaccharomyces pombe* tRNA^{Asp}, stimulating its formation by the Dnmt2 homologue Pmt1 (38,39). Structural analysis suggested that the presence of Q34 leads to optimal positioning of the interacting substrates in the active site of Dnmt2, enhancing the catalytic efficiency of the methyltransferase (40).

Arguably, approaches to a deeper understanding of the molecular action of Q in living cells would need to involve manipulations of details of the structure of Q, e.g. *via* an incorporation of q-derivatives through transglycosylation. Apart from their natural substrates, both bTGT and eTGT have been shown to tolerate a certain variety of synthetic analogues harbouring large functional groups *in vitro* (41,42). Leveraging the short hairpin recognition motif of the bTGT installed on different RNA transcripts,

Devaraj and co-workers developed a method called RNA-TAG (transglycosylation at guanosine), allowing to site-specifically incorporate analogues *in vitro*, which contained large fluorophores or affinity labels for pull-down experiments. This method was also applied to visualize mRNA transcripts containing the recognition motif in a fixed cell environment in a direct one-step-reaction (42) and extended to the development of a light-activated mRNA translation system (43). Furthermore, RNA-TAG was used on modified mRNAs in a two-step-approach, incorporating a preQ$_1$-derivative bearing a bioorthogonal tetrazine moiety in the first step, and thus enabling further derivatization by IEDDA click chemistry in a second step (44). However, labelling with click-competent compounds *in vivo* or *in cellulo* has not yet been achieved in the queuosine field. Indeed, there is strongly suggestive, albeit indirect evidence of successful *in vivo* incorporation of a non-natural q-analogue as published by Kelly and co-workers in the context of an animal model of multiple sclerosis (34). In addition to concerns about cell permeability of a q-derivative, important aspects to determine for *in vivo* labelling studies would include the physiological impact of an artificial chemical structure in a functioning tRNA, which would primarily be expected on the level of translation.

In this study, we metabolically label tRNA with a preQ$_1$ derivative functionalized with an azide group, allowing for further derivatization by click reaction and thus facilitating the proof of successful incorporation as well as the isolation of accordingly tagged RNAs. The latter was combined with RNAseq, in order to re-investigate the RNA substrates of the TGT, which turned out to be specific for the previously reported tRNAs Asn, Asp, His and Tyr. While in previous studies the transglycosylation step was performed in a fixed cell environment, we herein focus on the incorporation of the analogues by the natively expressed TGT *in vivo* and the physiological consequences in the natural environment. Polysome preparations revealed an enrichment of Q-containing tRNAs in the polysomal fraction, indicating a targeted selection for modified tRNA to be integrated in the translational process. Moreover, our data demonstrate that the semi-synthetic tRNA modification replaces Q34 and is functionally integrated into the translational process, as well as in the modification circuit with m^{5}C38 in tRNA^{Asp} in *S. pombe*.

MATERIALS AND METHODS

S. pombe strains used in this study are given in Supplementary Table S1, Plasmids used in this study are given in Supplementary Table S2, oligonucleotides used in this study are given in Supplementary Table S3. The names and versions of all software used are provided in Supplementary Table S4.

Synthesis of preQ$_1$-L1, preQ$_1$-L2 and preQ$_1$-L3

The preQ$_1$-ligands were synthesized as previously described (45).

Recombinant expression and purification of bTGT

The pASK-IBA13plus vector expressing the *Zymomonas mobilis* TGT (bTGT) with a N-terminal Strep-tag II was kindly provided by Prof. Dr Klaus Reuter (Philipps-University, Marburg). Expression and purification were carried out as previously described with minor changes (46). Briefly, the TGT was expressed in *E. coli* BL21-CodonPlus (DE3)-RIPL cells, grown in 2× YT medium and protein production was induced using anhydrotetracycline to a final concentration of 0.2 mg/l. After growing the cells for 14 h at 15°C, cells were harvested and the cell pellets were stored at –80°C until further processing. To purify the bacterial TGT, cells were thawed in lysis buffer (100 mM Tris pH 7.8, 150 mM NaCl, 1 mM EDTA pH 8.0, 2 mM PMSF, 1 μg/ml leupeptin, 1 μg/ml aprotinin, 1 μg/ml pepstatin and 25 U of DNase I and RNase I, respectively). After sonication (60% amplitude, 6 min, 0.5 s on, 2 s off; Sonifier 250 D, Branson), soluble proteins were isolated by centrifugation at 20 000 g for 1 h, 4°C. Affinity chromatography was then used to purify the Strep II-tagged TGT. For this purpose, the lysate was incubated with Strep-Tactin® Superflow Plus resin (Qiagen) for 3 h at 4°C, 15 rpm. After washing with washing buffer (100 mM Tris pH 7.8, 150 mM NaCl, 1 mM EDTA pH 8.0), the protein complex was eluted in 100 mM Tris pH 7.8, 150 mM NaCl, 1 mM EDTA pH 8.0 and 2.5 mM desthiobiotin. Further purification was achieved by Superdex S200 (GE Healthcare) size exclusion chromatography (10 mM Tris pH 7.8, 150 mM NaCl, 1 mM EDTA pH 8.0). The purified bTGT was stored at –80°C in 10 mM Tris pH 7.8, 150 mM NaCl, 1 mM EDTA pH 8.0 with 50% glycerol.

E. coli strains and growth conditions

The *E. coli* Keio parent strain (BW25113) and the knockout strains for QueD, QueC, QueE, QueF and TGT were obtained from the *E. coli* Keio knockout collection (GE Healthcare (Dharmacon™), England) and grown in standard M9 medium (6.8 g/l Na_2HPO_4, 3 g/l KH_2PO_4, 0.5 g NaCl, 1 g/l NH_4Cl, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4% glucose) at 37°C and 190 rpm. Growth medium of knockout strains was additionally supplemented with kanamycin (25 μg/ml). Synthetic preQ$_1$-derivatives were added to final concentrations of 0.1, 1, 5 or 10 μM to the culture, respectively.

Isolation of total tRNA from *E. coli*

To isolate total tRNA, *E. coli* cells were grown to an OD$_{600}$ of 1 in 50 ml cultures and harvested by centrifugation (10 min, 10 000 g, 4°C). The RNA was extracted by using the RNA isolation reagent TRI Reagent© (Sigma-Aldrich, Germany) following the manufacturer's instructions and dissolved in MQ-water.

Polysome preparations from *E. coli*

For polysome preparations the *E. coli* cells were grown in M9 medium in 150 ml culture volume until they reached an OD of 0.6, chloramphenicol was added to final concentration of 100 μg/ml and after further incubation of 3 min the cells were harvested by centrifugation (10 min, 10 000 g, 4°C). For cell lysis, cell pellets were resuspended in buffer (100 mM NH_4Cl, 10 mM $MgCl_2$, 20 mM Tris, pH 7.5), lysozyme was added and freeze-thaw cycles in liquid nitrogen were performed. Subsequent to this 10% deoxycholate was added to complete lysis, remaining cell wall debris were separated by centrifugation (12 000 g, 10 min, 4°C). Sucrose gradients from 5 to 40% were generated using a Biocomp gradient station model 108 (settings: time 1.24 min, angle 81.5°, speed 21 rpm) and lysate was loaded on top of the gradient. After ultracentrifugation (Beckman Ultracentrifuge Optima LE-80K, SW40 Ti rotor from Beckman Coulter) at 150 000 g and 4°C for 2.5 h, gradients were fractionated by measuring the absorbance at 280 nm (Biocomp Gradient Station model 108 in combination with Gilson Fraction Collector FC203B). Total RNA was isolated from the respective fractions using TRI reagent© (Sigma-Aldrich).

Purification of total tRNA from collected fractions by gel elution

Total RNA extracted from polysomal fraction was separated on a 10% denaturing PAGE gel, stained with GelRed (Biotium) and the bands were visualized on Typhoon 9400 at an excitation wavelength of 532 nm. According to the resulting image, bands of interest were excised from the gel and mashed with a scalpel. The mashed gel pieces were frozen for 1 h and 300 μl of 0.5 M ammonium acetate were added. Subsequently, the samples were shaken at 25°C and 750 rpm overnight. The gel suspension was filtered through NanoSep® centrifugal filters and the filtrate was precipitated with three volumes of 100% ethanol.

S. pombe strains, plasmids and growth conditions

The *S. pombe* strains and plasmids used in this study are shown in Supplementary Table S1. Cells were cultured in YES (5 g/l yeast extract, 30 g/l glucose, 250 mg/l adenine, 250 mg/l histidine, 250 mg/l leucine, 250 mg/l uracil, 250 mg/l lysine) which did not contain queuosine or queuine. Synthetic queuine (kindly provided by Hans-Dieter Gerber and Gerhard Klebe (Universität Marburg) (47)) and preQ$_1$ derivatives were added to 0.1 μM to the culture.

Isolation of total RNA and small RNAs from *S. pombe*

To isolate total RNA, *S. pombe* cells were grown to an optical density at 600 nm (OD$_{600}$) of 1 in 50 ml cultures. 50 OD of cells were harvested and 1 ml of phenol, glass beads were added. After vigorous shaking for 5 min, samples were centrifuged at 20 000 g for 5 min to clear the cell debris. Equal volume of phenol/chloroform/isoamylalcohol was added to the aqueous phase and centrifuged at 20 000 g for 5 min. After mixing the upper phase with an equal volume of chloroform followed by centrifugation at 20 000 g for 5 min, the RNA was precipitated at –80°C for 1 h using 0.7 volume of isopropyl alcohol. Following precipitation, total RNA was washed with 70% ethanol and eluted in DEPC-treated water.

Isolation of small RNAs was performed using the PureLink™ miRNA Isolation Kit (Invitrogen) according to the manufacturer's instructions. Yeast cells were grown to an OD_{600} of 1 in 5 ml cultures. After harvesting 1 OD of cells, RNAs were isolated using 1 ml TriFast reagent (Peqlab), 0.2 ml chloroform and glass beads. After vigorous shaking for 2 min, samples were centrifuged at 16 000 g, 4°C for 15 min. After adding 215 μl ethanol to the aqueous phase, the samples were transferred to a spin cartridge followed by centrifugation at 12 000 g for 1 min. 700 μl ethanol was added to the flow-through and the sample was transferred to a new spin cartridge. Following centrifugation at 12 000 g for 1 min, the cartridge was washed and small RNAs were dissolved in DEPC-treated water. Northern blot–acryloyl aminophenylboronic acid (APB) gels were performed as previously described (28).

Removal of ribosomal RNA

Depletion of ribosomal RNA was performed as previously described (48). Oligonucleotides specific for 5.8S and 5S rRNA were ordered with a 5'-biotin tag from Metabion (see Supplementary Table S3). The oligonucleotides were diluted to 100 μM each in nuclease-free water and equal volumes of the 100 μM stock were combined to generate the rRNA depletion mix.

For hybridization, 8 μg of small RNAs were incubated with 9.92 μl of the 100 μM rRNA depletion mix in reaction buffer (10 μl of formamide, 2.5 μl of 20× SSC (3 M NaCl, 0.3 M sodium citrate, pH 7.0) and 5 μl of 0.005 M EDTA, pHnbsp;8). Reactions were carried out in a total volume of 50 μl with the following thermocycling: 80°C for 5 min, ramp down to 25°C at intervals of 5°C per minute. Following hybridization, 2 μl of RNase-OUT (Invitrogen) and 50 μl of 1× SCC containing 20% formamide were added. Removal of rRNA/oligonucleotide hybrids was performed using Dynabeads™ MyOne™ Streptavidin C1 (ThermoFisher) according to the manufacturer's instructions. 500 μl streptavidin coated magnetic beads were washed as instructed for immunoprecipitation of RNA and added to the hybridization reaction. After incubation for 15 min at room temperature with mild agitation and bead separation on a magnetic rack, the supernatant was once more incubated with 500 μl of washed beads for 15 min at room temperature under mild agitation followed by bead separation. Subsequently, the supernatant containing the 5S/5.8S rRNA-depleted RNA was precipitated with 1/10 volume of ammonium acetate and three volumes of 100% ethanol.

RNA substrates for *in vitro* modification

The *S. pombe* tRNA^{Asp} substrate was prepared as previously described (49). Briefly, the pJET1 vector carrying the tRNA^{Asp} sequence was linearized with NcoI, and 2.5 μg of the linear vector was used for *in vitro* transcription using the TranscriptAid T7 High Yield Transcription Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. Following an 8 h incubation at 37°C with nucleotides and the T7 RNA polymerase and subsequent DNase I treatment,

the respective tRNA was purified from the reaction using phenol/chloroform extraction followed by gel filtration with Sephadex G50 (GE Healthcare).

Recombinant expression and purification of hTGT

The pCDF-Duet1 vector co-expressing the human TGT (hTGT) heterodimer QTRT1 and QTRT2 with a cleavable N-terminal 6xHis tag to QTRT1 was kindly provided by Prof. Dr. Ralf Ficner (GZMB, Göttingen). Expression and purification were carried out as previously described with minor changes (20). Briefly, the heterodimer QTRT1/QTRT2 was co-expressed in *E. coli* (DE3) Rosetta cells, and protein production was induced using autoinduction. After growing the cells for 50 h at 18°C, cells were harvested and the cell pellets were stored at –80°C until further processing. To purify the human TGT, cells were thawed in lysis buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 10 mM imidazole, 2 mM PMSF, 1 μg/ml leupeptin, 1 μg/ml aprotinin, 1 μg/ml pepstatin and 25 U of DNase I and RNase I, respectively). After sonification (60% amplitude, 6 min, 0.5 s on, 2 s off; Sonifier 250 D, Branson), soluble proteins were isolated by centrifugation at 20 000 g for 1 h. Affinity chromatography was then used to purify the 6xHis tagged QTRT1/QTRT2 complex. For this purpose, the lysate was incubated with Talon® Superflow™ resin (Cytiva) for 3 h at 4°C, 15 rpm. After washing with washing buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 10 mM imidazole and 1 M LiCl), the protein complex was eluted in 50 mM HEPES pH 7.5, 100 mM NaCl and 500 mM imidazole. Further purification was achieved by Superdex S200 (GE Healthcare) size exclusion chromatography (20 mM HEPES pH 7.5, 100 mM NaCl). The purified hTGT was stored at –80°C in 20 mM HEPES pH 7.5, 100 mM NaCl with 50% glycerol.

In vitro labelling of tRNA with preQ$_1$ derivatives

For *in vitro* labelling of tRNA with the preQ$_1$ derivatives, 10 μM of *in vitro* transcribed tRNAs or alternatively 10 μg of total RNA from *S. pombe* was incubated with 200 nM hTGT (QTRT1:QTRT2) and 5 μM queuine in reaction buffer (50 mM Tris–HCl pH 7.5, 20 mM NaCl, 5 mM MgCl$_2$ and 2 mM dithiothreitol) for 5 h at 37°C. The RNA was purified using phenol/chloroform extraction and precipitated with 1/10 volume of ammonium acetate and three volumes of 100% ethanol.

HeLa cells growth conditions and *in vivo* labelling with preQ$_1$-L1

HeLa cell lines were obtained from ATCC and authenticated by multiplex human cell line authentication test (Multiplexon). Cells were grown in Dulbecco's modified Eagle's medium (DMEM) (Thermo Fisher Scientific). The cultures were supplemented with 10% heat-inactivated FBS, 2 mM L-glutamine and a commercial cocktail of antibiotics (Thermo Fisher Scientific). For minus-Q conditions, ultraculture serum-free medium (Lonza) was supplemented with 2 mM L-glutamine and 100 units/ml Penicillin/Streptomycin. PreQ$_1$-L1 derivative was added at a concentration of 0.1 μM for 72 h to the culture.

HeLa cell polysome profiling

10^7 cells were treated with 100 µg/ml cycloheximide for 5 min at RT to stabilize existing polysomes before washing with ice-cold PBS and harvesting by scraping in 400 µl polysome lysis buffer (20 mM Tris–HCl, pH 7.4, 5 mM $MgCl_2$, 150 mM NaCl, 1 mM DTT, 1% Triton X-100, 100 µg/ml cycloheximide, 1 × Complete Protease Inhibitors (Roche)). Lysates were rotated end-over-end for 10 min at 4°C and cleared by at 10 000 rpm for 10 min at 4°C. 40 µl of supernatant lysate was saved as input before loading the lysates to linear 17.5 to 50% sucrose gradients in 20 mM Tris–HCl (pH 7.4), 5 mM $MgCl_2$, 150 mM NaCl. Centrifugation was carried out at 35 000 rpm for 2.5 h at 4°C in a Beckmann SW60 rotor. Gradients were eluted with an ISCO UA-6 gradient fractionator, and polysome profiles were recorded by continuously monitoring the absorbance at 254 nm using PeakTrak software. During gradient elution, fractions of ∼300 µl were collected every 14 s. For RNA isolation, 300 µl urea buffer (10 mM Tris, pH 7.5, 350 mM NaCl, 10 mM EDTA, 1% SDS, and 7 M urea) and 300 µl phenol:chloroform:isoamylalcohol (25:24:1) were added to each fraction. After phase separation, RNA was isolated from the aqueous phase and precipitated using isopropanol and GlycoBlue (Thermo Fisher Scientific).

CuAAC click reaction

Chemical clicking was performed as previously described (50). Briefly, up to 10 µg of RNA was incubated in reaction buffer containing 50% (v/v) DMSO, 5 mM Tris ((1-hydroxy-propyl-1*H*-1,2,3-triazol-4-yl)methyl) amine (THPTA), 5 mM sodium ascorbate, 0.5 mM $CuSO_4$ and 50 µM ligand alkyne under light-protection for 2 h at 25°C. The ligand alkynes used were AlexaFluor 594 alkyne (Thermo Fisher Scientific) or biotin alkyne (PEG4 carboxamide-Propargyl biotin; Thermo Fisher Scientific). RNA was precipitated with 1/10 volume of ammonium acetate and three volumes of 100% ethanol.

Detection of queuine and preQ$_1$ modification of RNAs

Labelled RNA that had been CuAAC-clicked with AlexaFluor 594 alkyne was analyzed by denaturing PAGE. Up to 10 µg of labelled RNA was separated in 10% polyacrylamide gels (acrylamide/ bisacrylamide (19:1), urea 8 M in 1× TBE buffer). Detection was carried out on the Typhoon 9500 (GE Healthcare) using 532 nm for excitation. As a loading control, gels were stained with Sybr Gold nucleic acid gel stain (Thermo Fisher Scientific) or GelRed (Biotium) for 10 min followed by detection using 495 or 532 nm, respectively, for excitation.

To detect the queuine modification, 300 ng of total RNA or small RNAs from *S. pombe* WT and *qtr2Δ* strains were separated in a 10% polyacrylamide gel (acrylamide/ bisacrylamide (19:1), urea 8 M) supplemented with 5 mg/ml 3-(acrylamido)-phenylboronic acid as described previously (51). The separation was performed at room temperature in 1× TBE. The electrophoresed gels were transferred to a Biodyne B Nylon membrane (0.45 µM). Selected RNAs were detected using a 5′-biotin-labeled probe at a final concentration of 0.3 µM and the Chemiluminescence Nucleic Acid Detection Module Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The first blocking step was carried out using the DIG Easy Hyb buffer (Roche), and hybridization was performed overnight at 45°C.

Detection of Q and Q-L1 by LC–MS/MS analysis

Up to 5 µg of total tRNA was digested to nucleoside level using 0.6 U nuclease P1 from *P. citrinum* (Sigma-Aldrich), 0.2 U snake venom phosphodiesterase from *C. adamanteus* (Worthington), 2 U FastAP (Thermo Fisher Scientific), 10 U benzonase (Sigma-Aldrich), 200 ng Pentostatin (Sigma-Aldrich) and 500 ng Tetrahydrouridine (Merck-Millipore) in 25 mM ammonium acetate (pH 7.5; Sigma-Aldrich) overnight at 37°C. 1 µg of total tRNA was analyzed *via* LC–MS using an Agilent 1260 series LC with a Synergi Fusion column (4 µM particle size, 80 Å pore size, 250 × 2.0 mM; Phenomenex) and an Agilent 6460 Triple Quadrupole mass spectrometer equipped with an electrospray ion source (ESI). The elution started with 100% solvent A (5 mM ammonium acetate buffer, pH 5.3) with a flow rate of 0.35 ml/min at 35°C, followed by a linear gradient to 8% solvent B (LC–MS grade acetonitrile; Honeywell) at 10 min and 40% solvent B after 20 min. Initial conditions were regenerated with 100% solvent A for 10 min. The UV signal at 254 nm was recorded *via* a multiple wavelength detector (MWD) detector at 254 nm to monitor the main nucleosides. The following ESI parameters were defined for the measurement: gas temperature 350°C, gas flow 8 l/min, nebulizer pressure 50 psi, sheath gas temperature 350°C, sheath gas flow 12 l/min, capillary voltage 3000 V, nozzle voltage 0 V. The MS was operated in the positive ion mode using Agilent MassHunter software in the dynamic MRM (multiple reaction monitoring) mode.

Identification of preQ$_1$-L1-modified RNAs by HTS

Metabolically labelled and biotin-clicked RNA was purified from total RNA or isolated small RNAs using Dynabeads™ MyOne™ Streptavidin C1 (Thermo Fisher Scientific) according to the manufacturer's instructions. Streptavidin coated magnetic beads were washed as instructed for immunoprecipitation of RNA. 20 µg of biotin-labelled RNA was incubated with the beads for 1 h at room temperature with light agitation. After washing the beads, they were resuspended in nuclease-free water, and bound RNA was dissolved from the beads by incubating the samples at 95°C for 10 min. Library preparation of immunoprecipitated RNAs for deep sequencing was done using the NEBNext Small RNA Library Prep Set for Illumina (Multiplex Compatible; New England Biolabs). 300 ng of RNA per library as starting material was used, and ligation was performed with undiluted adaptors. Adaptor-ligated cDNA was amplified with 15 cycles of PCR reaction using barcoded primers and purified using the Monarch PCR & DNA Cleanup Kit (5 µg) (New England Biolabs). Libraries were eluted in nuclease-free water, multiplexed in equimolar ratios and sequenced on one lane of the Illumina MiSeq platform using paired-end 150 bp sequencing.

RT-qPCR quantification of tRNA^{Asp}, snoR38 and snoR69

For quantification of preQ$_1$-L1-labelled tRNA^{Asp}, snoR38 and snoR69 from metabolically labelled and immunoprecipitated (IPed) RNAs, quantitative RT-PCR was performed using a stem-loop primer (see Supplementary Table S3). cDNA was synthesized using IPed RNAs from *S. pombe* WT and *qtr2Δ* and a sequence specific stem-loop primer. First strand synthesis was carried out using the SuperScript™ III First-Strand Synthesis System (Invitrogen) according to the manufacturer's protocol. Synthesized cDNA was subsequently used for qPCR using the PerfeCTa SYBR Green SuperMix (QuantaBio). 4 μl of cDNA was used in a reaction containing 12.5 μl Master Mix (Quanta, 2×) and 250 nm primers. Reactions were carried out in a total volume of 25 μl with the following thermocycling: 95°C for 2 min, followed by 40 cycles of 95°C for 10 s, 58°C for 15 s and 72°C for 20 s.

RNA bisulfite sequencing

Bisulfite sequencing of tRNA^{Asp} was performed as previously described (38). Briefly, bisulfite-treated tRNAs were reverse transcribed using tRNA^{Asp} 3'-specific stem-loop primer followed by amplification with primers binding only to the deaminated sequence at their 5' end. Primer sequences are listed in Supplementary Table S3. Library preparation of the PCR products was performed with the NEXTflex^R qRNA-Seq™ Kit v2–Set C (Bioo Scientific) according to the manufacturer's instructions and sequenced on a MiSeq platform using paired-end 150 bp sequencing. Reads were processed using in-house R scripting and the Bioconductor package ShortRead (52). Following the processing, including trimming of PCR primers, selection of high-quality reads and sorting of the reads based on the sequence in the degenerate region of the RT-primer, the reads were analyzed for bisulfite conversion using BISMA (53).

HTS data processing

The sequencing data was adapter-trimmed using Skewer version 0.2.2 (54) and aligned to *S. pombe* non-coding RNAs (main and mitochondrial) excluding rRNA sequences from Pombase (https://www.pombase.org/) using Salmon version 14.0 (55) and HISAT2 version 2.1.1 (56), as a splice-site sensitive alignment program. The conversion of sam to bam files was performed using SAMtools (57). Aligned sequences were analyzed using custom R scripts and the Bioconductor package DESeq2 (58). Parameters were set to analyze only regions with a minimum of 10 reads and the adjusted *P*-value was set to <0.1. Additionally, independent hypothesis weighting was conducted using the Bioconductor package IHW (59,60) with an adjusted *P*-value of <0.1. Furthermore, peak calling was performed using the Bioconductor package exomePeak2 (61). Plots were generated using the integrative genomics viewer version 2.11.1 (IGV) (62).

RESULTS

In vitro incorporation of synthetic preQ$_1$ analogues in prokaryotes

To assess the substrate properties of synthetic preQ$_1$-ligands, their incorporation into tRNA by bacterial TGT (bTGT) was tested *in vitro* (Figure 1). For this purpose, preQ$_1$-ligands 1–3 (preQ$_1$-L1-3, Figure 1A), each harbouring an azide group, were incubated with tRNA^{Asp} in the presence of recombinant bTGT from *Z. mobilis*. Taking advantage of the terminal azide group, the successful *in vitro* incorporation of preQ$_1$-ligands was visualized by copper(I)-catalyzed azide alkyne cycloaddition (CuAAC) click reaction of tRNA^{Asp} with the fluorescent AlexaFluor 594-alkyne in the presence of CuSO$_4$, sodium ascorbate and THPTA (tris-((1-benzyl-1H-1,2,3-triazol-4-yl)methyl)amine) (Figure 1B and Supplementary Figure S1). As shown by fluorescence scan, all of the tested preQ$_1$-ligands were incorporated to the same extent (Figure 1C), indicating that the side chains attached to the preQ$_1$ structure do not hinder the recognition and turnover by bTGT. This indicates a tolerance of the bTGT active site for large ligands, similar to what was previously described for eTGT *in vitro* (42).

In vivo incorporation of synthetic preQ$_1$ analogues in prokaryotes

After the successful *in vitro* application of synthetic preQ$_1$-ligands with bTGT, we proceeded to metabolic labelling of RNAs *in vivo* in *E. coli*. First experiments were performed with the smallest preQ$_1$ ligand in the series, i.e. preQ$_1$-L1. In a feeding experiment, where an *E. coli* wild-type (WT) strain was grown in medium supplemented with preQ$_1$-L1, total tRNA was isolated and enzymatically digested to the nucleoside level for separation on an RP-C18 HPLC column and subsequent analysis of the Q levels by MS/MS. Of note, queuosine exhibits a fragmentation pattern differing from the standard nucleosides. Instead of the exclusive fragmentation at the *N*-glycosidic bond, cleavage of the ribose in combination with cleavage of the amino linker with a mass shift m/z 410 to m/z 163 was determined as the most abundant product ion (Supplementary Figure S2a) eluting at a retention time of 12.2 min in the WT sample. Using a fragmentation pattern for the incorporated synthetic nucleoside (Q-L1) that was inferred from that of native queuosine, additional signals for the expected transitions were detected at 16.9 min (Figure 2). Since the product ion m/z 163 was the most prevalent species, it was chosen as diagnostic ion in subsequent LC–MS/MS experiments. Monitoring this product ion produced a strong signal for queuosine and only a weak signal for Q-L1 (Supplementary Figure S3c). We concluded that preQ$_1$-L1 was indeed incorporated, but also that it was a weak competitor against the endogenous bacterial preQ$_1$. Consequently, we reasoned that abrogating preQ$_1$ biosynthesis would facilitate the incorporation of the supplied preQ$_1$-ligands. Considering the various steps of Q *de novo* synthesis in *E. coli* (Figure 2A), four different gene deletions, namely *ΔqueD*, *ΔqueE*, *ΔqueC* and *ΔqueF*, were tested for generation of preQ$_1$ by monitoring the presence of
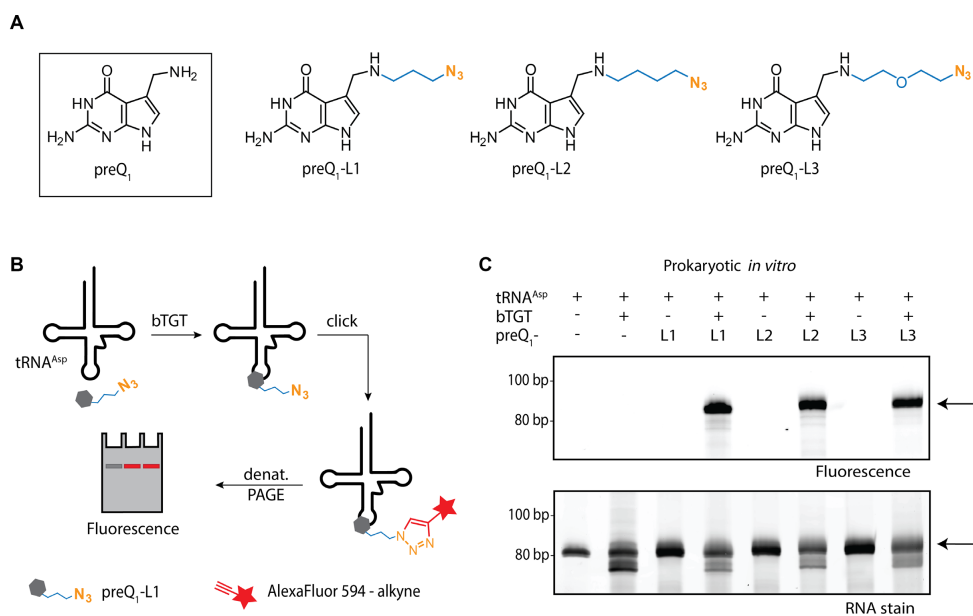
Figure 1. *In vitro* incorporation of preQ$_1$ analogues in tRNA^{Asp}. (**A**) Natural preQ$_1$ and synthetic preQ$_1$-ligands 1 (L1, 3-azidopropyl-preQ$_1$), 2 (L2, 4-azidobutyl-preQ$_1$) and 3 (L3, 2-(2-azidoethoxy)ethyl-preQ$_1$) containing side chains of different length and constitution (blue) but identical terminal azide groups (orange) designated for click chemistry. (**B**) Scheme of the *in vitro* experiment: incubation of tRNA^{Asp} with preQ$_1$ ligands (exemplarily shown for preQ$_1$-L1) in the presence of bacterial TGT (bTGT) and subsequent click reaction with AlexaFluor 594-alkyne (red), allowing the detection of tRNA with incorporated preQ$_1$ ligand *via* fluorescence scan. (**C**) Analysis of the tRNA^{Asp} click product after bTGT-catalysed incorporation of preQ$_1$-ligands L1–L3 by denaturing PAGE and following visualization by fluorescence scan for AlexaFluor 594 (excitation: 532 nm, emission: 610 nm). A loading control was obtained by RNA staining with GelRed. In both scans, tRNA^{Asp} is indicated by an arrow. Untreated tRNA^{Asp} and tRNA^{Asp} incubated with bTGT or preQ$_1$-ligands L1–L3, respectively, served as controls.

Q at position 34 of tRNAs. The deletion strains showed no significant growth defects compared to the wild type (Figure 2C). To validate the absence of Q *de novo* biosynthesis, tRNA was isolated and analyzed by LC–MS as before; none of the deletion strains generated measurable levels of Q (Supplementary Figure S2b–e). Since all deletions were on a par regarding growth and absence of Q in the isolated total tRNA, *ΔqueD* was chosen for further experiments, given that the absence of QueD prevents *de novo* synthesis of Q in its earliest stages and avoids synthesis of any precursor form (e.g. preQ$_0$ in *ΔqueF*) that was reported to be incorporated by the TGT and might thus compete with preQ$_1$-L1 (Supplementary Figure S3c) (13).

To determine a suitable feeding concentration, the *ΔqueD* cells were supplemented with increasing amounts of preQ$_1$-L1 (Figure 2B), which did not impair the bacterial growth compared to the control without preQ$_1$-L1 feeding (Supplementary Figure S3b). Total tRNA isolated from thus treated *ΔqueD* cells was labelled *via* CuAAC click chemistry and subsequently analysed by denaturing urea PAGE. The fluorescence scan revealed clearly visible fluorescent bands after feeding with preQ$_1$-L1 in a dose-dependent manner, providing evidence for

the enzymatic incorporation of the analogue into *E. coli* tRNA *in vivo* (Figure 2D). The signal intensity was quantified using ImageJ software and plotted against the feeding concentration of preQ$_1$-L1 (Figure 2E). To validate these observations, total tRNA isolated from the *ΔqueD* strain treated with preQ$_1$-L1 was further analysed *via* LC–MS/MS. The MS-based analysis of total tRNA isolated from the *ΔqueD* fed with increasing concentrations of preQ$_1$-L1 (0.1–10 μM) confirmed the results obtained from denaturing urea PAGE analysis and related quantification of the fluorescence signal (Figure 2E). Both plots indicate a beginning saturation around 10 μM. Based on the above, a ligand concentration of 5 μM for further feeding experiments was identified as a viable compromise between ligand material consumption and labelling efficiency. Supplementation of a *Δtgt* strain with the optimized concentration of 5 μM preQ$_1$-L1 resulted in no signal for Q-L1 in LC–MS/MS measurements, thus confirming that the incorporation was catalysed by bTGT (Supplementary Figure S3c).

Since preQ$_1$-L1 was successfully incorporated into RNA in bacteria, the *in vivo* experiments were expanded to preQ$_1$-L2 and preQ$_1$-L3. However, in contrast to the previously
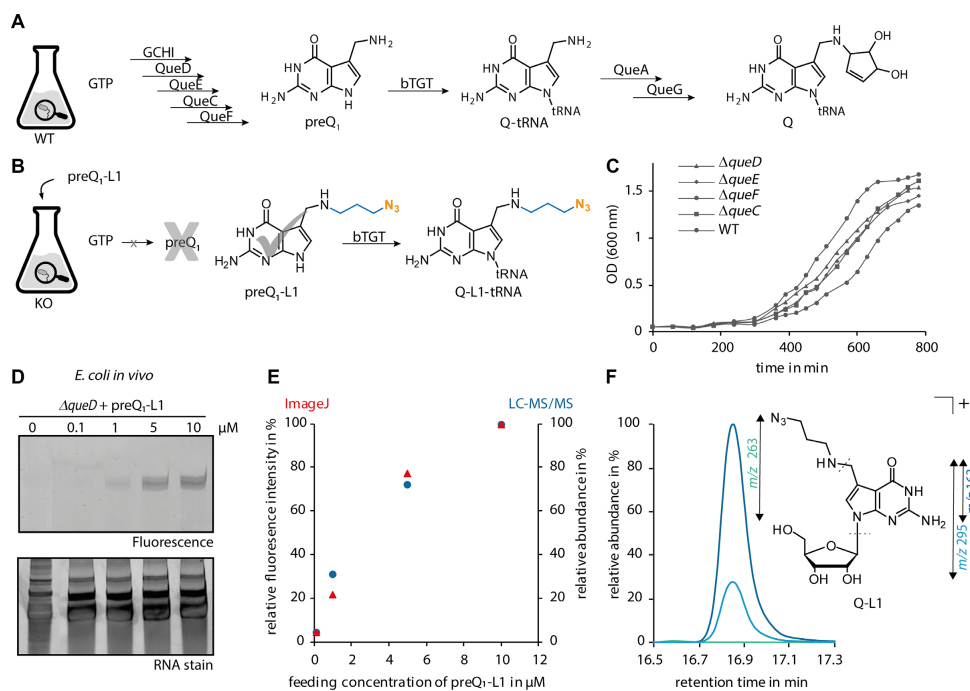
Figure 2. *De novo* biosynthesis of Q and induced incorporation of preQ₁-L1 in bacteria. (**A**) Biosynthesis of Q in position 34 of tRNAs (Q34-tRNA) *via* insertion of preQ₁ into tRNA, which is catalysed by the bacterial tRNA guanine transglycosylase (bTGT). (**B**) Treatment of *E. coli* mutant cells unable to synthesize preQ₁ with preQ₁-L1 and concomitant bTGT-catalysed incorporation of this analogue into tRNA. (**C**) Growth of the *E. coli* wild-type (WT) strain compared to the growth of several strains with deletions in genes encoding enzymes for Q *de novo* synthesis. (**D**) Analysis of total tRNA from *ΔqueD* cells grown with the indicated concentrations of preQ₁-L1 after click reaction by denaturing PAGE and subsequent scanning for fluorescence of AlexaFluor 594 (excitation: 532 nm, emission: 610 nm). (**E**) Merged diagram displaying the quantification of the dose-dependent fluorescence signal obtained from (**D**) by ImageJ software (Wayne Rasband, NIH) (shown as red triangles) and relative quantification of Q-L1 levels by LC–MS/MS (blue dots). Peak areas of Q-L1 (m/z 163) were normalized to the UV signal of adenosine and set in relation to the peak area of the highest feeding concentration (10 μM). (**F**) Extracted ion chromatograms displaying the fragmentation pattern of the incorporated synthetic nucleoside Q-L1 (m/z 395) in LC–MS/MS experiments, normalized to the highest peak area (m/z 163). Product ions are assigned in the structure of Q-L1.

described *in vitro* experiments, neither feeding preQ₁-L2 nor preQ₁-L3 at the optimized concentration of 5 μM or at higher concentrations (10 μM for preQ₁-L2 and 20 μM preQ₁-L3) led to a clear fluorescence signal in the clicked total tRNA samples (Supplementary Figure S3a), indicating that the incorporation efficiency of preQ₁-L2 and preQ₁-L3 into tRNA *in vivo* was drastically lower compared to preQ₁-L1. Since the *in vitro* results indicate indifference of the TGT enzyme towards the alkyl-modified preQ₁-ligands, the low incorporation *in vivo* suggests lower bioavailability of preQ₁-L2 and preQ₁-L3 for the bacteria.

In vivo interactions of synthetic preQ₁ analogues in prokaryotes

To investigate possible changes in the ensemble of molecular interactions undergone by Q-L1-carrying tRNA under physiological conditions, we turned to the analysis of polysomes. Given that these consist of actively translating ribosomes, their components, including tRNA, can be considered functional in interactions with essential molecular factors involved in translation. We thus aimed at determining the ratio of Q-L1-carrying tRNAs from polysomes versus that in the remainder of tRNAs.

For this purpose, cell lysates from *E. coli* WT and *ΔqueD* cells supplemented with 10 μM preQ₁-L1 were applied to a sucrose gradient (5–40%), enabling the separation of different fractions according to their size after ultracentrifugation. As schematically shown in Figure 3A, free RNAs including tRNAs and some mRNAs were located in fraction F0 at the top of the gradient (5% sucrose), while polysomes accumulated in fraction F3 at a sucrose concentration of ∼40%. This separation was monitored by UV absorbance at 260 nm, and the different fractions were collected. Subsequent to fractionation, total RNA was extracted from these fractions and applied to denaturing PAGE for purification of tRNA *via* gel elution. Digested tRNA samples were subjected to LC–MS/MS
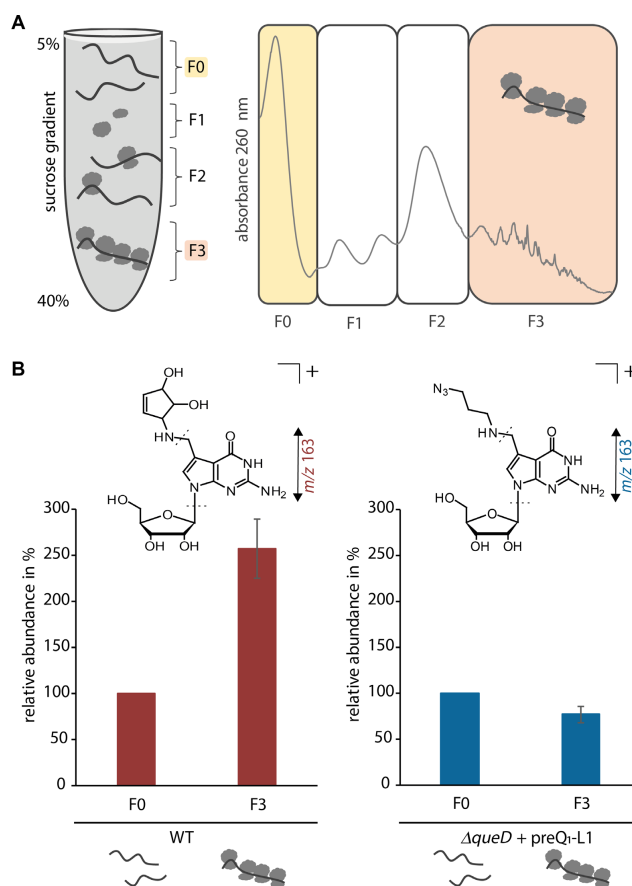
Figure 3. *E. coli* polysome preparation and analysis of isolated tRNA obtained from these samples by LC–MS/MS. (**A**) Schematic distribution of fractions F0-F3 from a cell lysate after sucrose gradient (5–40 %) fractionation and ultracentrifugation and representative UV trace at 260 nm (representing RNA) across the sucrose gradient. (**B**) Relative quantification of Q (m/z 410 → 163, red) and Q-L1 (m/z 395 → 163, blue) in tRNA purified from fractions F0 and F3 of WT and $\Delta queD$ cells supplemented with 10 μM preQ$_1$-L1 *via* LC–MS/MS. Peak areas were normalized to the UV signal of adenosine and related pairwise to the respective F0 fraction which was set to 100%. The average of normalized and related fractions F0 and F3 of three independent biological replicates are shown.

analysis, and the respective abundances of Q and Q-L1 were compared between the free RNA fraction F0 and the polysomal fraction F3 (Figure 3B). Interestingly, in WT cells, endogenous Q was more abundant in tRNAs isolated from the polysomal fraction compared to fraction F0. This suggests that queuosinylated tRNAs are selectively enriched in polysomes that are actively engaged in translation.

In contrast, the Q-L1 level in polysomal tRNA (F3 fraction) from preQ$_1$-L1 fed $\Delta queD$ cells reached a similar amount compared to its level in the respective F0 fraction. This may reflect either a deficit in the aforementioned selection, or a cumulation of minor detrimental effects at the different steps of translation. However, the data clearly illustrate that Q-L1-containing tRNAs actively engage in protein biosynthesis and are able to sustain it at a high enough level to not cause any perceivable growth phenotype.

In vitro incorporation of synthetic preQ$_1$ analogues in eukaryotes

In a next step, the investigations were extended from bacteria to eukaryotes. Of note, eukaryotes do not possess the enzymes to synthesize queuosine *de novo,* but salvage it from external sources for incorporation into tRNA (16). *S. pombe* is a particularly well-suited single cell eukaryotic

model organism, because salient features of queuosine have already been elaborated in this yeast, and queuosine levels can easily be manipulated by supplementation of the growth medium with queuine (38).

We next tested the ability of eTGT to incorporate the preQ$_1$-ligands into RNA *in vitro*. As substrates for this reaction, total RNA was isolated from *S. pombe* wild-type cells or *qtr2Δ* cells cultured in the presence of queuine. In WT cells, this results in Q-modification of the tRNAs, whereas *qtr2Δ* cells lack the essential Qtr2 subunit of *S. pombe* eTGT, therefore maintaining a guanosine in position 34 of the respective tRNAs. Total RNA preparations of these strains were incubated with the preQ$_1$ ligands in presence of hTGT, and subsequently labelled by click reaction. The incorporation of all three ligands into tRNA from both *S. pombe* strains was measured by fluorescence scan (Figure 4A and Supplementary Figure S4b). In comparison to the fluorescence signals of the tRNA from WT cells, the respective signals of the *qtr2Δ* tRNAs showed significantly higher intensities. This indicates that more tRNAs unmodified at position G34 are available for *in vitro* modification with preQ$_1$-L1 in the *qtr2Δ* sample. In contrast, in WT cells only guanosines that were not replaced by Q despite the presence of a functional enzyme remained for the *in vitro* reaction. Unlike observed for the bTGT, the hTGT incorporated the preQ$_1$-ligands to differing degrees, indicating a higher ability to distinguish between these analogues in accordance with previously published results by Kelly and co-workers (41). Additionally, incubation of *in vitro* transcribed tRNAs Asp, His, Tyr and Asn with human TGT and preQ$_1$-L1 showed successful incorporation of the analogue into all of the four tRNAs (Supplementary Figure S4a).

In vivo incorporation of synthetic preQ$_1$ analogues in eukaryotes

Subsequent to the successful *in vitro* experiment, the *in vivo* incorporation of the synthetic preQ$_1$ analogues was examined in *S. pombe*. To this end, *S. pombe* WT and *qtr2Δ* cells (as a control), were cultured in the presence of preQ$_1$-ligands in medium that otherwise lacked Q or q, and RNA was isolated and analysed as before. After click reaction, a fluorescence signal was detected in the RNA isolated from the WT cells treated with preQ$_1$-L1, but not *qtr2Δ* (Figure 4B), showing that the presence of Q-L1 in tRNA *in vivo* depended on functional TGT. As in bacteria, preQ$_1$-L1 did not negatively affect cell growth (Supplementary Figure S4c), and no labelling was observed with preQ$_1$-L2 and -L3, again indicating that their derivatives are not bioavailable for incorporation into tRNAs *in vivo*.

Collectively, the above experiments indicate that preQ$_1$-L1 can readily be employed as a proxy for Q from the perspective of synthetic biology. To further develop this compound for the investigation of the epitranscriptome, we made use of the click chemistry feature of preQ$_1$-L1 to identify RNAs into which it was incorporated *in vivo* by eTGT.

For this purpose, total RNA isolated from *S. pombe* wild-type or *qtr2Δ* cells that were cultured in the presence of preQ$_1$-L1 was bio-conjugated *in vitro* with alkyne-

functionalized biotin. Subsequent to affinity purification using streptavidin-coated magnetic beads, the biotin-labelled RNA was subjected to reverse transcription and high-throughput sequencing (Figure 5A, termed Q-RIP-Seq). The analysis showed that the known cytosolic Q-tRNAs tRNA^{Asn}, tRNA^{Asp}, tRNA^{His} and tRNA^{Tyr} were significantly enriched from WT, but not *qtr2Δ* cells ($n = 3$, $P_{adj} < 0.1$, Figure 5B, C and Supplementary Figure S5). Other enriched signals from snoR38 and snoR69 were scrutinised as potential substrates of TGT-mediated incorporation of preQ$_1$-L1. However, neither APB Northern blotting nor quantification by q-RT-PCR substantiated this hypothesis (Supplementary Figure S6). Interestingly, mitochondrial tRNA^{Asn}, when analysed for q content by APB-northern blot, was queuosinylated to about 50% (Supplementary Figure S5b). The fact that no mitochondrial tRNA sequences were found in Q-RIP-Seq could mean that they are too low in abundance. An alternative explanation would be that preQ$_1$-L1 is not incorporated into mitochondrial tRNA. The above findings indicate that the four known Q-tRNAs are the only cytosolic RNAs that are Q-modified in *S. pombe*, which is congruent with crosslinking-based studies in human cells (41). These results establish that any major metabolic influence resulting from feeding preQ$_1$-L1 would be mediated through the four classical tRNA substrates of TGT. It should, however, be noted that an early study reported preQ1-modification *in vitro* of larger RNA species in *E. coli* (63).

In vivo interactions of synthetic preQ$_1$ analogues in eukaryotes

Having established that preQ$_1$-L1 is actively incorporated into native tRNAs, we next investigated a particularly interesting effect of queuosine, namely a so-called tRNA modification circuit, where the formation of one modification is enhanced by the presence of another modification (64). The particular circuit involving queuosine was first identified in *S. pombe*. We had shown earlier by RNA bisulfite sequencing that the formation of m^{5}C38 in tRNA^{Asp} by the Dnmt2 tRNA methyltransferase is strongly enhanced by the presence of queuosine at position 34 (Figure 6A) (38,49).

We therefore asked whether Q-L1 can serve as a biologically active surrogate for queuosine in this circuit. Figure 6B shows the m^{5}C38 levels of tRNA^{Asp} in response to increasing concentrations of preQ$_1$-L1 in medium otherwise free of queuosine derivatives. A clear dose-dependent increase of the C38 methylation level was observed, indicating that the incorporated preQ$_1$-L1 is functionally integrated into this modification circuit, efficiently replacing queuosine in its capability of triggering Dnmt2 activity in *S. pombe*. Considering the direct functional connection of Q/Q-L1 and m^{5}C38, the increase of the C38 methylation level from 15% in non-treated culture up to 60% in cultures supplemented with 100 nm preQ$_1$-L1 points to its incorporation in significant amounts in *S. pombe*. However, it is important to mention that the effect of preQ$_1$-L1 incorporation on tRNA^{Asp} methylation is less efficient compared to the known effect of Q under
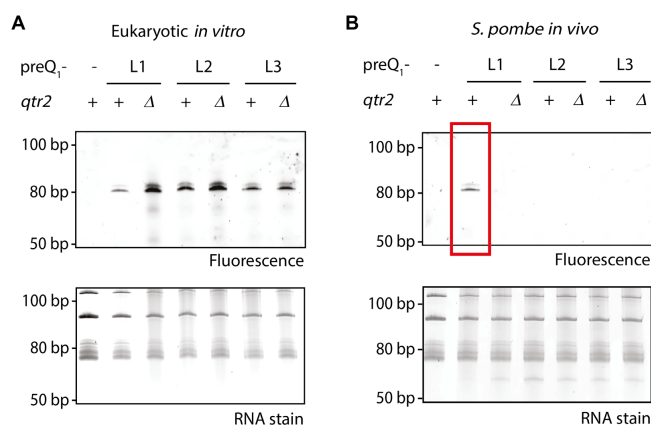
Figure 4. *In vitro* and *in vivo* incorporation of preQ$_1$-ligands in *S. pombe* tRNA. (**A**) Analysis of the total RNA click product after human tRNA guanine transglycosylase (hTGT)-catalysed incorporation of preQ$_1$-ligands L1–L3 into RNA from *S. pombe* by denaturing PAGE and visualization by fluorescence scan for AlexaFluor 594 (excitation: 532 nm, emission: 610 nm). Total RNA was extracted from *S. pombe* WT cells containing functional TGT (+) and *qtr2Δ* cells that lack functional TGT (*Δ*), which were both cultured in the presence of queuine. The incubation of total RNA from WT cells without preQ$_1$-ligand (-) served as a negative control. A loading control was obtained by RNA staining with SybrGold. (**B**) Analysis of total tRNA from *S. pombe* WT (+) and *qtr2Δ* (*Δ*) cells that were cultured in the presence of 0.1 μM of the respective preQ$_1$-ligand after click reaction by denaturing PAGE and subsequent visualization as described above. Total RNA from *S. pombe* WT cells supplemented with 0.1 μM queuine (–) instead of preQ$_1$-ligands was used as a negative control.

normal conditions, as we previously reported (38). The incorporation efficiency in *E. coli* can only be gauged even more indirectly, namely by comparison of fluorescent signals after click (Supplementary Figure S4d).

Lastly, we were also able to demonstrate successful incorporation of preQ$_1$-L1 in HeLa human cells deprived of q (Supplementary Figure S7a). In analogy to the earlier presented analysis of *E. coli* polysomes, we also investigated the levels of Q-L1 in tRNA purified from F0 and F3 of accordingly treated HeLa cultures. Similar to our observations in *E. coli*, the amount of Q-L1 detected in the polysomal tRNA (F3 fraction) from preQ$_1$-L1 fed HeLa cells was comparable to its level in the respective F0 fraction, indicating that Q-L1-containing tRNAs actively engage in protein biosynthesis *in cellulo* (Supplementary Figure S7b). This result indicates relevance of our investigations with respect to biomedical considerations, e.g. potential therapeutic interventions.

DISCUSSION

Interest in concepts for the incorporation of modified and/or non-natural derivatives of metabolites into nucleic acids has been steadily increasing, boosted in part by a surge in RNA modification research, and, more recently, in mRNA-based vaccines. Post-synthetic derivatization of RNA *in vitro*, e.g. by methyltransferases has been exploited for labelling in conjunction with click chemistry (65–69). In the queuosine field, a number of q-derived compounds, including clickable tetrazine derivatives, have been incorporated into native RNA preparation *in vitro* using recombinant TGT, and applied to fluorescent labelling, affinity purification, and interactome research (42–44,70,71). In a previous study, Brooks *et al.* reported that azide congeners of preQ$_1$ lacking the methylene amine were not incorporated by the TGT which they traced to the necessity of this structural element for a successful binding to the enzyme forming hydrogen bonds between aminoacid residues Leu231 and Met260 of the enzyme (72,73). Although, as mentioned, strong indirect evidence (34) suggested that incorporation of nonnatural q derivatives should be feasible in principle, no *in vivo* labelling of Q-tRNAs with clickable q-derivatives has been demonstrated so far.

Overall, concepts and applications in the RNA field currently move from *in vitro* (74) to metabolic feeding approaches *in cellulo* and *in vivo*. Here, the use of noncanonical nucleoside structures has opened up new experimental avenues in the community. As an example, in RNA modification research feeding of methionine analogues featuring e.g. propargyl residues, has enabled their incorporation into RNA *in lieu* of methyl groups. Subsequent derivatization by click chemistry was exploited for determination of modification sites (75,76,66).

An important progress featured in our work is that we demonstrate low toxicity of the labelling compound and provide corresponding data at the molecular level. Elsewhere in the field, little attention is paid to the physiological impact of surrogate feeding. In most cases, a moderate survival rate in cell culture is sufficient to conduct e.g. -omics type analyses after incorporation (75–77). However, in the next steps of its development, the field might conceivably move to applications in model organisms. Here, by the latest, one will need
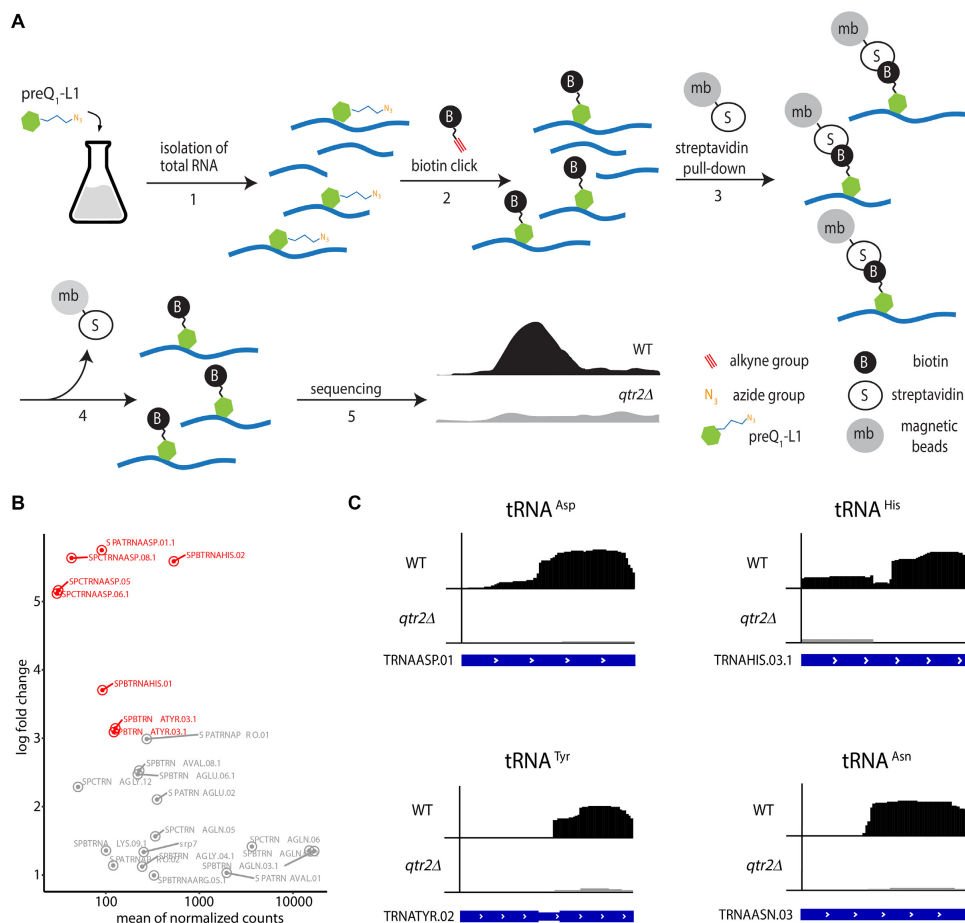
Figure 5. *In vivo* identification of Q-modified RNAs in *S. pombe* based on metabolic labelling with preQ$_1$-L1 and high-throughput sequencing (Q-RIP-Seq). (**A**) Concept of metabolic labelling and immunoprecipitation of Q-modified RNAs. *S. pombe* was cultured in the presence of 0.1 μM preQ$_1$-L1, leading to incorporation into otherwise Q-modified RNAs. Total RNA was extracted (1) and bio-conjugated *in vitro* with alkyne-functionalized biotin (2). Biotin-labelled RNAs were subsequently affinity-purified using streptavidin-coated magnetic beads (3), reverse-transcribed and subjected to high-throughput sequencing (5). As a control, metabolic labelling was performed in an *S. pombe* strain lacking TGT (*qtr2Δ*). (**B**) Log2 fold change of normalized read counts of RNAs from WT compared to *qtr2Δ* determined by exomePeak2. Red: tRNA^{Asp}, tRNA^{His} and tRNA^{Tyr}; (three independent replicates). (**C**) Q-RIP-Seq of tRNA^{Asp}, tRNA^{His}, tRNA^{Tyr} and tRNA^{Asn} after metabolic labelling with preQ$_1$-L1 in *S. pombe* WT and *qtr2Δ* cells. Coverage of the tRNA sequences from modified (WT, black) and unmodified *(qtr2Δ,* grey) samples is shown. The transcript architecture is shown below with thin and thick parts representing introns and mature tRNA sequences. Replicate 1 of three independent experiments is shown. Plots were generated using IGV.

to adopt concepts from medicinal chemistry, such as cell permeability, and toxicity. In this respect, the work presented here pioneers the combination of metabolic feeding of clickable surrogates with investigations into their physiological molecular impact after cellular uptake and their usage for the enrichment and identification of RNA species that were labelled *in vivo* by endogenous TGT. Apart from the observation of growth inhibition of q

derivatives in eukaryotic cell culture, which are somewhat suggestive (78), there is strong indirect evidence for the actual incorporation of a q-derivative by TGT *in vivo* in mouse (34), however without direct analysis of the tRNA. Significantly, said case features a background of medicinal chemistry, and the compound used is structurally related to our preQ1-L series used here. It does, however, feature a lipophilic phenylpropyl sidechain which is likely causative
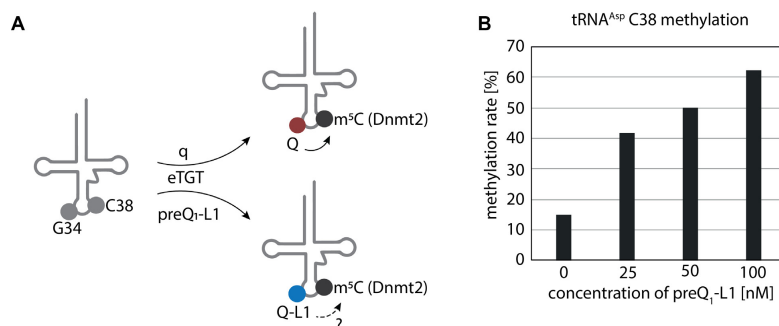
Figure 6. Incorporation of preQ$_1$-L1 in tRNA^{Asp} in *S. pombe* stimulates C38 methylation by Dnmt2. (**A**) Incorporation of q or preQ$_1$-L1 by eukaryotic tRNA guanine transglycosylase (eTGT) affects Dnmt2 activity in *S. pombe* (Pmt1). (**B**) Determination of tRNA^{Asp} methylation levels at C38 in total RNA from *S. pombe* WT supplemented with the indicated concentrations of preQ$_1$-L1 by RNA bisulfite sequencing combined with high throughput sequencing.

of, or enhancing the compound's cell permeability and biodistribution.

In the present work, we have developed the azido-propyl-derivative preQ$_1$-L1 as a bioactive surrogate for preQ$_1$ *in vivo*. preQ$_1$-L1 is taken up into unicellular prokaryotes as well as into eukaryotes, and incorporated into the known tRNA substrates of TGT. The resulting nucleoside is semi-synthetic in that its sugar moiety is native, while its nucleobase is synthetic. Its azide moiety can be employed to metabolically label and isolate Q-modified RNAs by affinity purification after conjugation by click chemistry. We used this feature to confirm similar data from human cells, obtained after UV-crosslinking (41). Taken together, this means that the single most important molecular interaction for a physiological impact of q (or preQ$_1$-L1) is mediated through position 34 in the anticodons of the four known TGT substrate tRNAs.

Known molecular interactions issuing from this nucleobase are mostly restricted to tRNA aminoacylation and mRNA decoding, which we have interrogated by investigating the amount of Q-L1 carrying tRNAs on polysomes. While Q-L1 was less abundant there than was native Q, it was clearly present, featuring an equal distribution between actively translating tRNAs and the cytoplasmic pool in both bacterial and human cell preparations.

One other known effect of Q was also faithfully emulated by Q-L1, namely the stimulation of m^{5}C38 formation by Dnmt2 in the anticodon stem of tRNA^{Asp}, representing a so-called modification loop. Technically speaking, we report the first-ever manipulation of a modification loop by atomic mutagenesis *in vivo*.

In spite of numerous described Q-dependent implications in various diseases, starting from cancer (29–32) to neurological and neuropsychiatric disorders, such as multiple sclerosis, schizophrenia and Parkinson (79,80,33,34), a defined mechanism explaining the role of Q in these pathologies is still missing (28). Recently, we discovered a direct connection between Q, accuracy and the speed of codon-biased translation (27,28), which promotes protein folding and prevents the accumulation of misfolded

proteins. The fact that Q-L1 is functionally involved in the translational process in a 'minimally invasive' system, opens the possibility to study the roles of Q34 modifications in protein translation in normal and pathogenic human cell lines, directly combining click chemistry or LC–MS/MS with polysome profiling.

In summary, the combination of very few queuosinylation sites and the effective functional replacement of Q by Q-L1 on the molecular level, makes the q/Q system uniquely suited for a 'minimally invasive' placement of a non-natural nucleobase within the total cellular RNA.

DATA AVAILABILITY

HTS data for Q-RIP-Seq experiments are available in the NCBI GEO database, record GSE210404. All data needed to evaluate the conclusions in the paper are present in the paper and/or Supplementary Data. Additional data related to this paper may be requested from the authors.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. El Yacoubi,B., Bailly,M. and de Crécy-Lagard,V. (2012) Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu. Rev. Genet.*, **46**, 69–95.
2. Lorenz,C., Lünse,C.E. and Mörl,M. (2017) tRNA modifications: impact on structure and thermal adaptation. *Biomolecules*, **7**, 35.
3. Motorin,Y. and Helm,M. (2010) tRNA stabilization by modified nucleotides. *Biochemistry*, **49**, 4934–4944.
4. Boccaletto,P., Stefaniak,F., Ray,A., Cappannini,A., Mukherjee,S., Purta,E., Kurkowska,M., Shirvanizadeh,N., Destefanis,E., Groza,P. *et al.* (2022) MODOMICS: a database of RNA modification pathways. 2021 update. *Nucleic Acids Res.*, **50**, D231–D235.
5. Fergus,C., Barnes,D., Alqasem,M.A. and Kelly,V.P. (2015) The queuine micronutrient: charting a course from microbe to man. *Nutrients*, **7**, 2897–2929.
6. Harada,F. and Nishimura,S. (1972) Possible anticodon sequences of tRNA his tRNA asm and tRNA asp from escherichia coli b. Universal presence of nucleoside q in the first postion of the anticondons of these transfer ribonucleic acids. *Biochemistry*, **11**, 301–308.
7. Kasai,H., Oashi,Z., Harada,F., Nishimura,S., Oppenheimer,N.J., Crain,P.F., Liehr,J.G., Minden,D.L.von and McCloskey,J.A. (1975) Structure of the modified nucleoside q isolated from escherichia coli transfer ribonucleic acid. 7-(4,5-cis-Dihydroxy-1-cyclopenten-3-ylaminomethyl)-7-deazaguanosine. *Biochemistry*, **14**, 4198–4208.
8. Phillips,G., El Yacoubi,B., Lyons,B., Alvarez,S., Iwata-Reuyl,D. and Crécy-Lagard,V.de (2008) Biosynthesis of 7-deazaguanosine-modified tRNA nucleosides: a new role for GTP cyclohydrolase I. *J. Bacteriol.*, **190**, 7876–7884.
9. McCarty,R.M., Somogyi,A. and Bandarian,V. (2009) Escherichia coli QueD is a 6-carboxy-5,6,7,8-tetrahydropterin synthase. *Biochemistry*, **48**, 2301–2303.
10. McCarty,R.M., Somogyi,A., Lin,G., Jacobsen,N.E. and Bandarian,V. (2009) The deazapurine biosynthetic pathway revealed: in vitro enzymatic synthesis of preq(0) from guanosine 5'-triphosphate in four steps. *Biochemistry*, **48**, 3847–3852.
11. van Lanen,S.G., Reader,J.S., Swairjo,M.A., Crécy-Lagard,V.de, Lee,B. and Iwata-Reuyl,D. (2005) From cyclohydrolase to oxidoreductase: discovery of nitrile reductase activity in a common fold. *Proc. Nat. Acad. Sci. U.S.A.*, **102**, 4264–4269.
12. Okada,N. and Nishimura,S. (1979) Isolation and characterization of a guanine insertion enzyme, a specific tRNA transglycosylase, from escherichia coli. *J. Biol. Chem.*, **254**, 3061–3066.
13. Okada,N., Noguchi,S., Kasai,H., Shindo-Okada,N., Ohgi,T., Goto,T. and Nishimura,S. (1979) Novel mechanism of post-transcriptional modification of tRNA. Insertion of bases of q precursors into tRNA by a specific tRNA transglycosylase reaction. *J. Biol. Chem.*, **254**, 3067–3073.
14. Slany,R.K., Bösl,M. and Kersten,H. (1994) Transfer and isomerization of the ribose moiety of adomet during the biosynthesis of queuosine tRNAs, a new unique reaction catalyzed by the QueA protein from escherichia coli. *Biochimie*, **76**, 389–393.
15. Miles,Z.D., McCarty,R.M., Molnar,G. and Bandarian,V. (2011) Discovery of epoxyqueuosine (oQ) reductase reveals parallels between halorespiration and tRNA modification. *Proc. Nat. Acad. Sci. U.S.A.*, **108**, 7368–7372.
16. Patel,B.I., Heiss,M., Samel-Pommerencke,A., Carell,T. and Ehrenhofer-Murray,A.E. (2022) Queuosine salvage in fission yeast by Qng1-mediated hydrolysis to queuine. *Biochem. Biophys. Res. Commun.*, **624**, 146–150.
17. Farkas,W.R., Jacobson,K.B. and Katze,J.R. (1984) Substrate and inhibitor specificity of tRNA-guanine ribosyltransferase. *Biochim. Biophys. Acta*, **781**, 64–75.
18. Boland,C., Hayes,P., Santa-Maria,I., Nishimura,S. and Kelly,V.P. (2009) Queuosine formation in eukaryotic tRNA occurs via a mitochondria-localized heteromeric transglycosylase. *J. Biol. Chem.*, **284**, 18218–18227.
19. Chen,Y.-C., Kelly,V.P., Stachura,S.V. and Garcia,G.A. (2010) Characterization of the human tRNA-guanine transglycosylase: confirmation of the heterodimeric subunit structure. *RNA*, **16**, 958–968.
20. Johannsson,S., Neumann,P. and Ficner,R. (2018) Crystal structure of the human tRNA guanine transglycosylase catalytic subunit QTRT1. *Biomolecules*, **8**, 81.
21. Grosjean,H. and Westhof,E. (2016) An integrated, structure- and energy-based view of the genetic code. *Nucleic Acids Res.*, **44**, 8020–8040.
22. Morris,R.C., Brown,K.G. and Elliott,M.S. (1999) The effect of queuosine on tRNA structure and function. *J. Biomol. Struct. Dyn.*, **16**, 757–774.
23. Noguchi,S., Nishimura,Y., Hirota,Y. and Nishimura,S. (1982) Isolation and characterization of an escherichia coli mutant lacking tRNA-guanine transglycosylase. Function and biosynthesis of queuosine in tRNA. *J. Biol. Chem.*, **257**, 6544–6550.
24. Manickam,N., Joshi,K., Bhatt,M.J. and Farabaugh,P.J. (2016) Effects of tRNA modification on translational accuracy depend on intrinsic codon-anticodon strength. *Nucleic Acids Res.*, **44**, 1871–1881.
25. Meier,F., Suter,B., Grosjean,H., Keith,G. and Kubli,E. (1985) Queuosine modification of the wobble base in tRNAHis influences 'in vivo' decoding properties. *EMBO J.*, **4**, 823–827.
26. Zaborske,J.M., DuMont,V.L.B., Wallace,E.W.J., Pan,T., Aquadro,C.F. and Drummond,D.A. (2014) A nutrient-driven tRNA modification alters translational fidelity and genome-wide protein coding across an animal genus. *PLoS Biol.*, **12**, e1002015.
27. Müller,M., Legrand,C., Tuorto,F., Kelly,V.P., Atlasi,Y., Lyko,F. and Ehrenhofer-Murray,A.E. (2019) Queuine links translational control in eukaryotes to a micronutrient from bacteria. *Nucleic Acids Res.*, **47**, 3711–3727.
28. Tuorto,F., Legrand,C., Cirzi,C., Federico,G., Liebers,R., Müller,M., Ehrenhofer-Murray,A.E., Dittmar,G., Gröne,H.-J. and Lyko,F. (2018) Queuosine-modified tRNAs confer nutritional control of protein translation. *EMBO J.*, **37**, e99777.
29. Zhang,J., Lu,R., Zhang,Y., Matuszek,Ż., Zhang,W., Xia,Y., Pan,T. and Sun,J. (2020) tRNA queuosine modification enzyme modulates the growth and microbiome recruitment to breast tumors. *Cancers*, **12**, 628.
30. Sebastiani,M., Behrens,C., Dörr,S., Gerber,H.-D., Benazza,R., Hernandez-Alba,O., Cianférani,S., Klebe,G., Heine,A. and Reuter,K. (2022) Structural and biochemical investigation of the heterodimeric murine tRNA-Guanine transglycosylase. *ACS Chem. Biol.*, **17**, 2229–2247.
31. Chen,Y.L. and Wu,R.T. (1994) Altered queuine modification of transfer RNA involved in the differentiation of human K562 erythroleukemia cells in the presence of distinct differentiation inducers. *Cancer Res.*, **54**, 2192–2198.
32. Dirheimer,G., Baranowski,W. and Keith,G. (1995) Variations in tRNA modifications, particularly of their queuine content in higher eukaryotes. Its relation to malignancy grading. *Biochimie*, **77**, 99–103.
33. Richard,P., Kozlowski,L., Guillorit,H., Garnier,P., McKnight,N.C., Danchin,A. and Manière,X. (2021) Queuine, a bacterial-derived hypermodified nucleobase, shows protection in in vitro models of neurodegeneration. *PLoS One*, **16**, e0253216.
34. Varghese,S., Cotter,M., Chevot,F., Fergus,C., Cunningham,C., Mills,K.H., Connon,S.J., Southern,J.M. and Kelly,V.P. (2017) In vivo modification of tRNA with an artificial nucleobase leads to full disease remission in an animal model of multiple sclerosis. *Nucleic Acids Res.*, **45**, 2029–2039.
35. Rakovich,T., Boland,C., Bernstein,I., Chikwana,V.M., Iwata-Reuyl,D. and Kelly,V.P. (2011) Queuosine deficiency in eukaryotes compromises tyrosine production through increased tetrahydrobiopterin oxidation. *J. Biol. Chem.*, **286**, 19354–19363.
36. Kulkarni,S., Rubio,M.A.T., Hegedüsová,E., Ross,R.L., Limbach,P.A., Alfonzo,J.D. and Paris,Z. (2021) Preferential import of queuosine-modified tRNAs into trypanosoma brucei

mitochondrion is critical for organellar protein synthesis. *Nucleic Acids Res.*, **49**, 8247–8260.

37. Hurt,J.K., Olgen,S. and Garcia,G.A. (2007) Site-specific modification of shigella flexneri virF mRNA by tRNA-guanine transglycosylase in vitro. *Nucleic Acids Res.*, **35**, 4905–4913.

38. Müller,M., Hartmann,M., Schuster,I., Bender,S., Thüring,K.L., Helm,M., Katze,J.R., Nellen,W., Lyko,F. and Ehrenhofer-Murray,A.E. (2015) Dynamic modulation of Dnmt2-dependent tRNA methylation by the micronutrient queuine. *Nucleic Acids Res.*, **43**, 10952–10962.

39. Ehrenhofer-Murray,A.E. (2017) Cross-Talk between dnmt2-dependent tRNA methylation and queuosine modification. *Biomolecules*, **7**, 14.

40. Johannsson,S., Neumann,P., Wulf,A., Welp,L.M., Gerber,H.-D., Krull,M., Diederichsen,U., Urlaub,H. and Ficner,R. (2018) Structural insights into the stimulation of s. pombe dnmt2 catalytic efficiency by the tRNA nucleoside queuosine. *Sci. Rep.*, **8**, 8880.

41. Fergus,C., Al-Qasem,M., Cotter,M., McDonnell,C.M., Sorrentino,E., Chevot,F., Hokamp,K., Senge,M.O., Southern,J.M., Connon,S.J. *et al.* (2021) The human tRNA-guanine transglycosylase displays promiscuous nucleobase preference but strict tRNA specificity. *Nucleic Acids Res.*, **49**, 4877–4890.

42. Alexander,S.C., Busby,K.N., Cole,C.M., Zhou,C.Y. and Devaraj,N.K. (2015) Site-Specific covalent labeling of RNA by enzymatic transglycosylation. *J. Am. Chem. Soc.*, **137**, 12756–12759.

43. Zhang,D., Zhou,C.Y., Busby,K.N., Alexander,S.C. and Devaraj,N.K. (2018) Light-Activated control of translation by enzymatic covalent mRNA labeling. *Angew. Chem. Int. Ed. Engl.*, **57**, 2822–2826.

44. Ehret,F., Zhou,C.Y., Alexander,S.C., Zhang,D. and Devaraj,N.K. (2018) Site-Specific covalent conjugation of modified mRNA by tRNA guanine transglycosylase. *Mol. Pharmaceutics*, **15**, 737–742.

45. Neuner,E., Frener,M., Lusser,A. and Micura,R. (2018) Superior cellular activities of azido- over amino-functionalized ligands for engineered preQ1 riboswitches in E.coli. *RNA Biol.*, **15**, 1376–1383.

46. Jakobi,S., Nguyen,T.X.P., Debaene,F., Metz,A., Sanglier-Cianférani,S., Reuter,K. and Klebe,G. (2014) Hot-spot analysis to dissect the functional protein-protein interface of a tRNA-modifying enzyme. *Proteins*, **82**, 2713–2732.

47. Gerber,H.-D. and Klebe,G. (2012) Concise and efficient syntheses of preQ1 base, q base, and (ent)-Q base. *Org. Biomol. Chem.*, **10**, 8660–8668.

48. Kraus,A.J., Brink,B.G. and Siegel,T.N. (2019) Efficient and specific oligo-based depletion of rRNA. *Sci. Rep.*, **9**, 12281.

49. Becker,M., Müller,S., Nellen,W., Jurkowski,T.P., Jeltsch,A. and Ehrenhofer-Murray,A.E. (2012) Pmt1, a dnmt2 homolog in *Schizosaccharomyces pombe*, mediates tRNA methylation in response to nutrient signaling. *Nucleic Acids Res.*, **40**, 11648–11658.

50. Schmid,K., Adobes-Vidal,M. and Helm,M. (2017) Alkyne-Functionalized coumarin compound for analytic and preparative 4-thiouridine labeling. *Bioconjug. Chem.*, **28**, 1123–1134.

51. Yuan,Y., Hutinet,G., Valera,J.G., Hu,J., Hillebrand,R., Gustafson,A., Iwata-Reuyl,D., Dedon,P.C. and Crécy-Lagard,V.de (2018) Identification of the minimal bacterial 2'-deoxy-7-amido-7-deazaguanine synthesis machinery. *Mol. Microbiol.*, **110**, 469–483.

52. Morgan,M., Anders,S., Lawrence,M., Aboyoun,P., Pagès,H. and Gentleman,R. (2009) ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics*, **25**, 2607–2608.

53. Rohde,C., Zhang,Y., Reinhardt,R. and Jeltsch,A. (2010) BISMA—fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. *BMC Bioinf.*, **11**, 230.

54. Jiang,H., Lei,R., Ding,S.-W. and Zhu,S. (2014) Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinf.*, **15**, 182.

55. Patro,R., Duggal,G., Love,M.I., Irizarry,R.A. and Kingsford,C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*, **14**, 417–419.

56. Kim,D., Langmead,B. and Salzberg,S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*, **12**, 357–360.

57. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

58. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.

59. Ignatiadis,N., Klaus,B., Zaugg,J.B. and Huber,W. (2016) Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat. Methods*, **13**, 577–580.

60. Ignatiadis,N. and Huber,W. (2021) Covariate powered cross-weighted multiple testing. *J. R. Stat. Soc. Series B*, **83**, 720–751.

61. Meng,J., Cui,X., Rao,M.K., Chen,Y. and Huang,Y. (2013) Exome-based analysis for RNA epigenome sequencing data. *Bioinformatics*, **29**, 1565–1567.

62. Robinson,J.T., Thorvaldsdóttir,H., Winckler,W., Guttman,M., Lander,E.S., Getz,G. and Mesirov,J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26.

63. Brooks,A.F., Vélez-Martínez,C.S., Showalter,H.D.H. and Garcia,G.A. (2012) Investigating the prevalence of queuine in escherichia coli RNA by incorporation of the tritium-labeled precursor, preQ(1). *Biochem. Biophys. Res. Commun.*, **425**, 83–88.

64. Motorin,Y. and Helm,M. (2022) RNA nucleotide methylation: 2021 update, wiley interdisciplinary reviews. *RNA*. **15**, 2021.

65. Motorin,Y., Burhenne,J., Teimer,R., Koynov,K., Willnow,S., Weinhold,E. and Helm,M. (2011) Expanding the chemical scope of RNA:methyltransferases to site-specific alkynylation of RNA for click labeling. *Nucleic Acids Res.*, **39**, 1943–1952.

66. Fischer,T.R., Meidner,L., Schwickert,M., Weber,M., Zimmermann,R.A., Kersten,C., Schirmeister,T. and Helm,M. (2022) Chemical biology and medicinal chemistry of RNA methyltransferases. *Nucleic Acids Res.*, **50**, 4216–4245.

67. Schulz,D., Holstein,J.M. and Rentmeister,A. (2013) A chemo-enzymatic approach for site-specific modification of the RNA cap. *Angew. Chem. Int. Ed. Engl.*, **52**, 7874–7878.

68. Muttach,F. and Rentmeister,A. (2016) A biocatalytic cascade for versatile one-pot modification of mRNA starting from methionine analogues. *Angew. Chem. Int. Ed. Engl.*, **55**, 1917–1920.

69. Ovcharenko,A., Weissenboeck,F.P. and Rentmeister,A. (2021) Tag-Free internal RNA labeling and photocaging based on mRNA methyltransferases. *Angew. Chem. (Int. Ed. Engl.)*, **60**, 4098–4103.

70. Busby,K.N., Fulzele,A., Zhang,D., Bennett,E.J. and Devaraj,N.K. (2020) Enzymatic RNA biotinylation for affinity purification and identification of RNA-Protein interactions. *ACS Chem. Biol.*, **15**, 2247–2258.

71. Zhou,C.Y., Alexander,S.C. and Devaraj,N.K. (2017) Fluorescent turn-on probes for wash-free mRNA imaging via covalent site-specific enzymatic labeling. *Chem. Sci.*, **8**, 7169–7173.

72. Brooks,A.F., Garcia,G.A. and Showalter,H.D. (2021) Synthesis of azide congeners of preQ 1 as potential substrates for tRNA guanine transglycosylase. *J Heterocyclic Chem*, **58**, 1192–1198.

73. Xie,W., Liu,X. and Huang,R.H. (2003) Chemical trapping and crystal structure of a catalytic tRNA guanine transglycosylase covalent intermediate. *Nat. Struct. Biol.*, **10**, 781–788.

74. Croce,S., Serdjukow,S., Carell,T. and Frischmuth,T. (2020) Chemoenzymatic preparation of functional click-labeled messenger RNA. *Chembiochem*, **21**, 1641–1646.

75. Hartstock,K., Nilges,B.S., Ovcharenko,A., Cornelissen,N.V., Püllen,N., Lawrence-Dörner,A.-M., Leidel,S.A. and Rentmeister,A. (2018) Enzymatic or in vivo installation of propargyl groups in combination with click chemistry for the enrichment and detection of methyltransferase target sites in RNA. *Angew. Chem. (Int. Ed. Engl.)*, **57**, 6342–6346.

76. Hartstock,K., Ovcharenko,A., Kueck,N.A., Spacek,P., Cornelissen,N.V., Hüwel,S., Dieterich,C. and Rentmeister,A. (2022) MePMe-seq: Antibody-free simultaneous m6A and m5C mapping in mRNA by metabolic propargyl labeling and sequencing. bioRxiv doi: https://doi.org/10.1101/2022.03.16.484494, 16 March 2022, preprint: not peer reviewed.

77. Shu,X., Cao,J., Cheng,M., Xiang,S., Gao,M., Li,T., Ying,X., Wang,F., Yue,Y., Lu,Z. *et al.* (2020) A metabolic labeling method detects m6A transcriptome-wide at single base resolution. *Nat. Chem. Biol.*, **16**, 887–895.

78. Akimoto,H., Nomura,H., Yoshida,M., Shindo-Okada,N., Hoshi,A. and Nishimura,S. (1986) Queuine analogues. Their synthesis and

inhibition of growth of mouse L5178Y cells in vitro. *J. Med. Chem.*, **29**, 1749–1753.

79. Bednářová,A., Hanna,M., Durham,I., VanCleave,T., England,A., Chaudhuri,A. and Krishnan,N. (2017) Lost in translation: defects in transfer RNA modifications and neurological disorders. *Front. Mol. Neurosci.*, **10**, 135.

80. Skolnick,S.D. and Greig,N.H. (2019) Microbes and monoamines: potential neuropsychiatric consequences of dysbiosis. *Trends Neurosci.*, **42**, 151–163.

Supplementary Data

Functional integration of a semi-synthetic azido-queuosine derivative into translation and a tRNA modification circuit

Larissa Bessler[1,†], Navpreet Kaur[2,†], Lea-Marie Vogt[1,†], Laurin Flemmich[3], Carmen Siebenaller[4], Marie-Luise Winz[1], Francesca Tuorto[5], Ronald Micura[3], Ann E. Ehrenhofer-Murray[2,*] and Mark Helm[1,*]

[1]Institute of Pharmaceutical and Biomedical Sciences, Johannes Gutenberg-University Mainz, 55128 Mainz, Germany
[2]Institute of Biology, Humboldt-Universität zu Berlin, 10117 Berlin, Germany
[3]Department of Organic Chemistry, University of Innsbruck, 6020 Innsbruck, Austria
[4]Department of Chemistry – Biochemistry, Johannes Gutenberg-University Mainz, 55128 Mainz, Germany.
[5]Division of Biochemistry, Mannheim Institute for Innate Immunoscience (MI3), Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

* To whom correspondence should be addressed. Tel: +49 (0) 6131 39 25731; Fax: +49 (0) 6131 39 20373; Email: mhelm@uni-mainz.de
Correspondence can also be addressed to Ann E. Ehrenhofer-Murray. Tel: +49 (0) 30 2093 49630; Fax: +49 (0) 30 2093 49641; Email: ann.ehrenhofer-murray@hu-berlin.de

† Joint Authors

1

Supplementary Table S1: *S. pombe* strains used in this study.

Designation	Genotype	Source
AEP1	*h⁻ leu1-32 ura4-D18 his3-D3*	YGRC
AEP288	*h⁻ leu1-32 ura4-D18 his3-D3 qtr2Δ::NatMX*	(1)

Supplementary Table S2: Plasmids used in this study.

Designation	Genotype	Source
pAE1688	pJET1-tRNA^Asp (*S.pombe*)	(2)
pAE2975	pASK-IBA13Plus - *Z. mobilis* TGTStrep-tag® II N-terminal	(3)
pAE2963	pCDF-Duet - hQTRT1-6xHis & hQTRT2	(4)

Supplementary Table S3: Oligonucleotides used in this study.

Designation	Sequence	Purpose
5S rRNA	5'-biotin-ACCCCGGATTCCCATGTTGTCTCCAACCATAGTAC-3'	rRNA depletion
5.8S rRNA	5'-biotin-CGTTCTTCATCGATGCGAGAGCCAAGAGATCCGTT-3'	rRNA depletion
tRNA Asp RT-Primer	5'-CTCaactggattggctnnnnngataaatccagttgagtggCTCTCCCT-3' [a]	RT
tRNA Asp bisulfite.fwd	5'- TTAGTATAGGGGTAGTATAT-3'	Bisulfite sequencing
stemloop.rev	5'- CGATCANNNNCTCAACTGGATTGGCT -3' [b]	Bisulfite sequencing
tRNAAsp_probe	5'-biotin-GGGCTGCAAGCGTGACAGG-3'	Northern
snoR38_probe	5'-biotin-CTCAACACTATGCTTTAGACAGGG-3'	Northern
snoR69_probe	5'-biotin-GCGTACTCGTCAATGTAAATAC-3'	Northern
RT-Primer_tRNAAsp	5'-AATCACTCAACTGGATTGGCT nnnnnGATAAATCCAGTTGAGTGGCTCTCCCTC-3' [a]	RT
RT-Primer_sno38	5'-AATCACTCAACTGGATTGGCT nnnnnGATAAATCCAGTTGAGTGGCAGATTTAC-3' [a]	RT
RT-Primer_snoR69	5'-AATCACTCAACTGGATTGGCT nnnnnGATAAATCCAGTTGAGTGGGTTCAGATA-3' [a]	RT
qPCR_tRNAAsp_fwd	5'-ATAGGGGTAGTACACAAGCCTGT-3'	qPCR
qPCR_sno38_fwd	5'-ACAGTTATCCCTGTCTAAAGCATAG-3'	qPCR
qPCR_snoR69_fwd	5'-CTTCGTTAAACCCAGCTCAC-3'	qPCR
qPCR_rev	5'-CGATCAATCACTCAACTGGATTGGCT-3'	qPCR

[a] Nucleotides marked as "n" are random nucleotides used as indices
[b] Nucleotides marked as "N" are the barcode region

Supplementary Table S4: Software used in this study.

Software	Version
R	3.6.3
Skewer	0.2.2
Salmon	14.0
HISAT2	2.2.1
SAMtools	1.12
DESeq2	1.26.0
ExomePeak2	2.16.0
IGV	2.11.1
M6A viewer	1.6.1
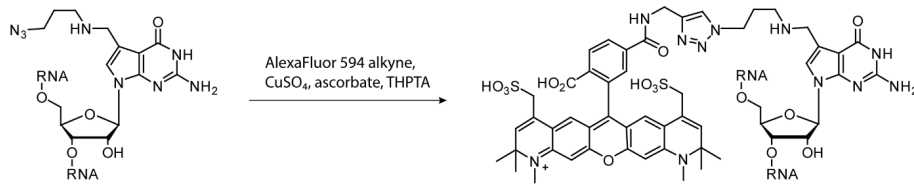
Supplementary Figure 1: Click derivatization of preQ$_1$-L1-containing RNA by a fluorescent alkyne. THPTA = Tris((1-hydroxy-propyl-1H-1,2,3-triazol-4-yl)methyl)amine.
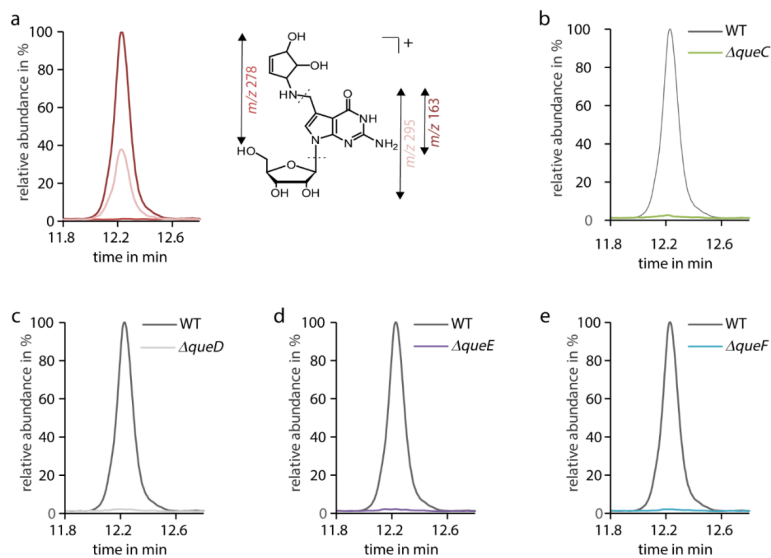
Supplementary Figure 2: LC-MS analysis of queuosine levels. a Extracted ion chromatograms displaying the fragmentation pattern of Q (m/z 410) in LC-MS/MS experiments, normalized to the highest peak area (m/z 163). Product ions are assigned in the structure of Q. **b** Relative LC-MS/MS quantification of Q levels in digested total tRNA from wild type and $\Delta queC$ mutant cells, normalized to the UV signal of adenosine and set in relation to the peak area of Q in the WT cells. **c, d** and **e** show similar analysis of total tRNA from wild type and $\Delta queD$, $\Delta queE$ and $\Delta queF$, respectively.
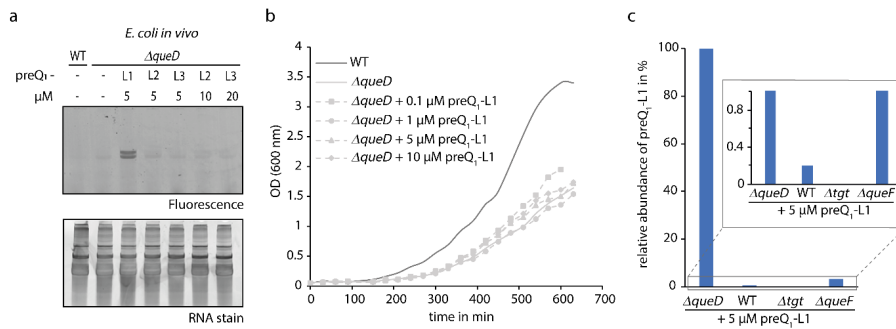
Supplementary Figure 3: Detection of incorporation of preQ₁-ligands L1-3 into total tRNA isolated from a *ΔqueD E. coli* strain via click derivatization. a Analysis of total tRNA from *ΔqueD* grown with the indicated concentrations of preQ₁-ligands L1-3 after click reaction by denaturing PAGE and subsequent scanning for fluorescence of AlexaFluor 594 (excitation: 532 nm, emission: 610 nm). **b** Growth of *E. coli* wild-type (WT) strain compared to the growth of the *ΔqueD* strain supplemented with indicated concentrations of preQ₁-L1. **c** Relative quantification of Q-L1 (*m/z* 395 → 163, blue) in total tRNA isolated from *E. coli* WT, *ΔqueD*, *Δtgt* and *ΔqueF* cells supplemented with 5 µM preQ₁-L1 via LC-MS/MS. Peak areas were normalized to the UV signal of adenosine and related the signal of Q-L1 in *ΔqueD* + 5 µM preQ₁-L1.
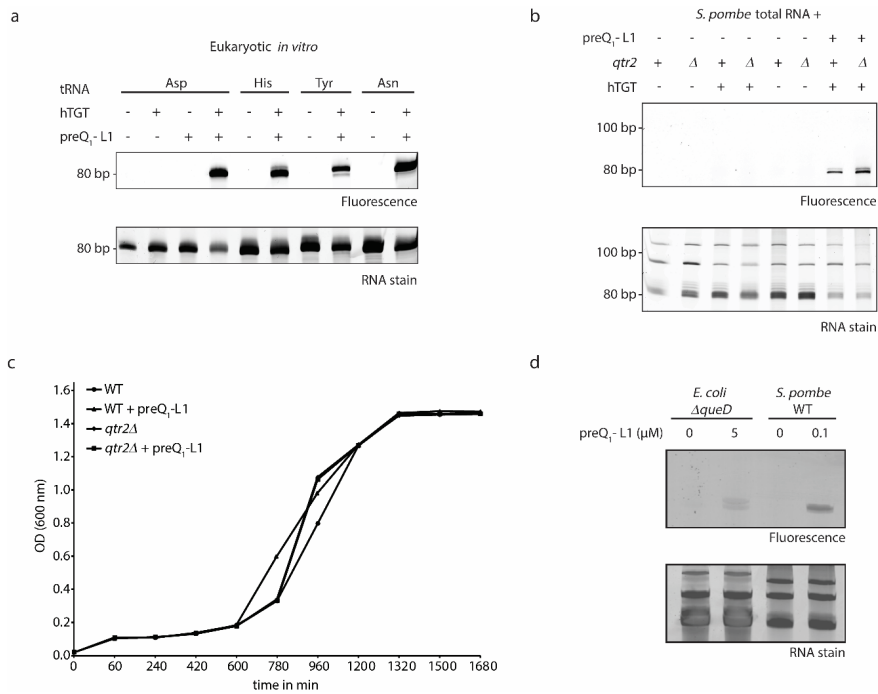
Supplementary Figure 4: Incorporation of preQ$_1$-L1 *in vitro* and *in vivo* in *S. pombe*. a Human tRNA guanine transglycosylase (hTGT) was incubated with respectively indicated *in vitro* transcribed *S. pombe* tRNA in presence of preQ$_1$-L1. After click reaction with AlexaFluor 594, the tRNAs were separated on a 10% polyacrylamide/ 8 M urea gel and visualized by scanning for fluorescence at 532 nm. Untreated tRNAs and tRNA^{Asp} incubated with hTGT or preQ$_1$-ligand L1, respectively, served as controls. A loading control was obtained by RNA staining with SybrGold. **b** Analysis of the total RNA click product after human tRNA guanine transglycosylase (hTGT)-catalysed incorporation of preQ$_1$-ligands L1 into RNA from *S. pombe* by denaturing PAGE and visualization by fluorescence scan for AlexaFluor 594 (excitation: 532 nm, emission: 610 nm). Total RNA was extracted from *S. pombe* WT cells containing functional TGT (+) and *qtr2Δ* cells that lack functional TGT (*Δ*), which were both cultured in the presence of queuine. The incubation of total RNA from WT and *qtr2Δ* cells without preQ$_1$-ligand or without hTGT, respectively, served as negative controls. A loading control was obtained by RNA staining with SybrGold. **c** Growth of *S. pombe* wild-type (WT) and *qtr2Δ* strains in the presence or absence of 0.1 µM preQ$_1$-L1. **d** Analysis of total tRNA from *E. coli ΔqueD* and *S. pombe* WT cells grown with the indicated concentrations of preQ$_1$-L1 after click reaction by denaturing PAGE and subsequent scanning for fluorescence of AlexaFluor 594. A loading control was obtained by RNA staining with GelRed.

5
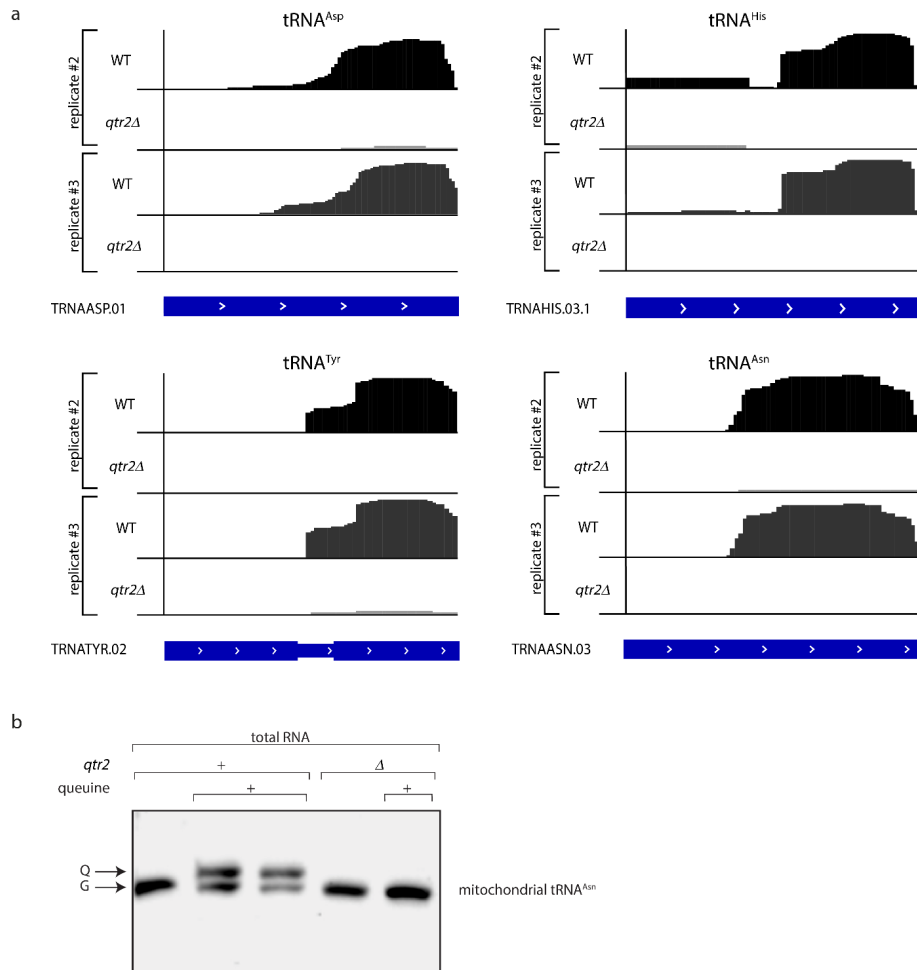
Supplementary Figure 5: *In vivo* **detection of Q-modified tRNAs in** *S. pombe* **using metabolic labelling with preQ$_1$-L1 combined with high-throughput sequencing (Q-RIP-Seq). a** Q-RIP-Seq of tRNA^{Asp}, tRNA^{His}, tRNA^{Tyr} and tRNA^{Asn} after metabolic labelling with preQ$_1$-L1 in *S. pombe* WT and *qtr2Δ* mutant cells and affinity purification. Coverage of the tRNA sequences from modified (WT, black) and unmodified (*qtr2Δ,* grey) samples is shown. The transcript architecture is shown below with thin and thick parts representing introns and mature tRNA sequences. Replicates #2 and #3 (c.f. Figure 5 for replicate #1) of three independent experiments are shown. Plots were generated using IGV. **b** Measurement of Q levels in mitochondrial tRNA^{Asn} using polyacrylamide gels covalently linked with N-acryloyl-3-aminophenylboronic acid (APB). Northern blotting with APB-gels was performed, and membranes were probed for mitochondrial tRNA^{Asn}. RNA samples from WT and *qtr2Δ* strains cultured with or without queuine are shown. The arrows indicate the migration distance of unmodified (G) and Q-modified (Q) tRNA, respectively.

6
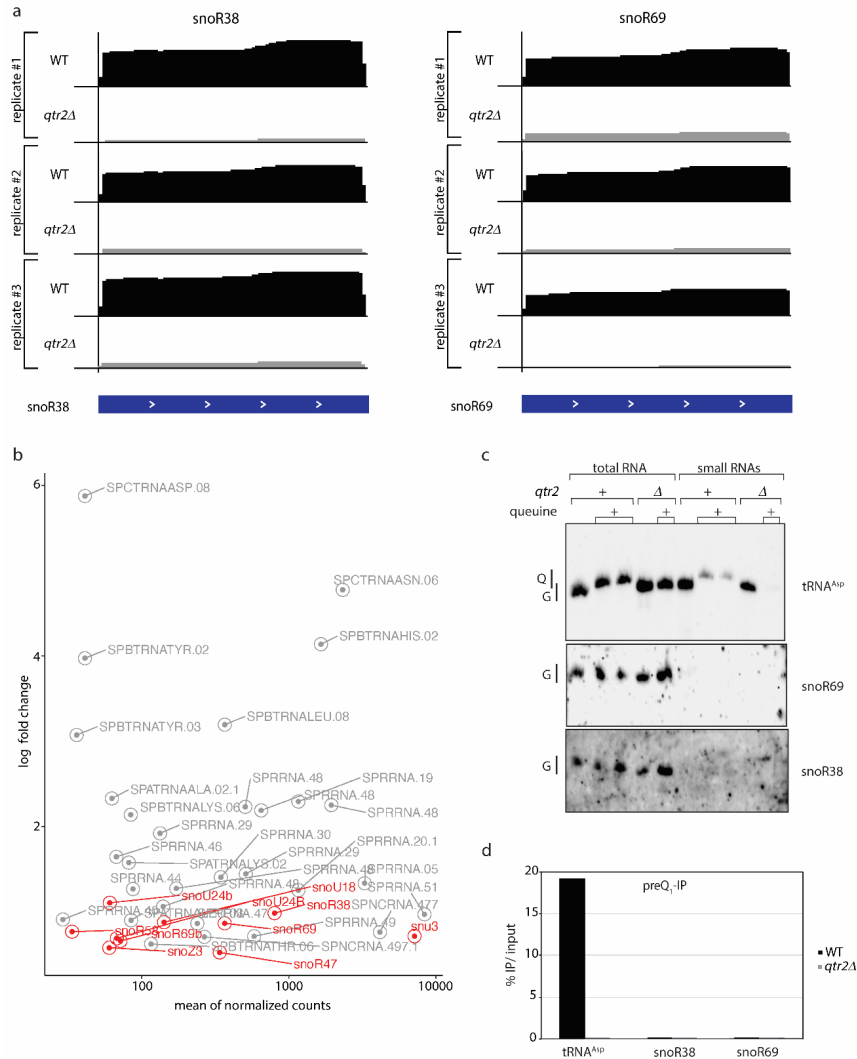
Supplementary Figure 6: Enrichment of snoRNA sequences of *S. pombe* in Q-RIP-Seq. RNAs that were metabolically labelled with azido-propyl-preQ$_1$ were bioconjugated with biotin-alkyne, enriched with streptavidin-coated magnetic beads and subjected to HTS as in Figure 5. **a** Coverage of the snoR38 and snoR69 sequences from WT ant *qtr2Δ S. pombe* RNAs are shown (three independent replicates). **b** Log2 fold change of normalized read counts of RNAs from WT compared to *qtr2Δ* determined by exomePeak2. Red: snoRNAs. **c** Measurement of Q levels in tRNA^{Asp}, snoR69 and snoR38 using polyacrylamide gels covalently linked with N-acryloyl-3-aminophenylboronic acid (APB). Northern blotting with APB-gels was performed, and membranes were probed with the indicated probes. RNA samples from WT and *qtr2Δ* strains cultured with or without queuine are shown. The label indicates the migration distance of unmodified (G) and Q-modified (Q) tRNA, respectively. **d** Measurement of

percentage of IP to input (%IP) for tRNA[Asp], snoR69 and snoR38 using qRT-PCR. RNAs from WT (black) and *qtr2Δ* (grey) were subjected to the Q-RIP method (IP) or only biotin-clicked (input).
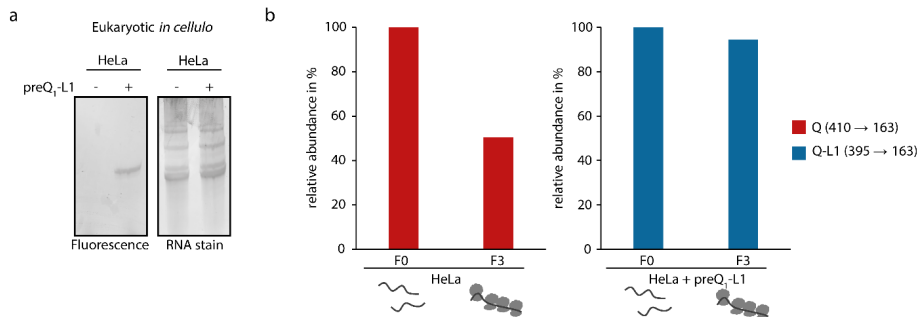
Supplementary Figure 7: *In cellulo* **incorporation of preQ₁-L1 in HeLa cells and analysis of tRNA purified from HeLa cell polysome preparations.** A single experiment is shown. **a** Analysis of total RNA from HeLa cells grown either in q-containing DMEM medium (-) or in q-free medium supplemented with 0.1 µM preQ$_1$-L1 (+) after click reaction by denaturing PAGE and subsequent scanning for fluorescence of AlexaFluor 594 (excitation: 532 nm, emission: 610 nm). b Relative quantification of Q (m/z 410 → 163, red) and Q-L1 (m/z 395 → 163, blue) in tRNA purified from fractions F0 and F3 of HeLa cells grown either in q-containing DMEM medium (HeLa) or in q-free medium supplemented with 0.1 µM preQ$_1$-L1 (HeLa + preQ$_1$-L1) via LC-MS/MS. Peak areas were normalized to the UV signal of adenosine and related to the respective F0 fraction.

Supplementary References

1. Müller, M., Hartmann, M., Schuster, I., Bender, S., Thüring, K.L., Helm, M., Katze, J.R., Nellen, W., Lyko, F. and Ehrenhofer-Murray, A.E. (2015) Dynamic modulation of Dnmt2-dependent tRNA methylation by the micronutrient queuine, *Nucleic acids research,* **43,** 10952–10962. First published on Sep 30, 2015.

2. Becker, M., Müller, S., Nellen, W., Jurkowski, T.P., Jeltsch, A. and Ehrenhofer-Murray, A.E. (2012) Pmt1, a Dnmt2 homolog in Schizosaccharomyces pombe, mediates tRNA methylation in response to nutrient signaling, *Nucleic acids research,* **40,** 11648–11658. First published on Oct 15, 2012.

3. Gerber, H.-D. and Klebe, G. (2012) Concise and efficient syntheses of preQ1 base, Q base, and (ent)-Q base, *Organic & biomolecular chemistry,* **10,** 8660–8668.

4. Johannsson, S., Neumann, P. and Ficner, R. (2018) Crystal Structure of the Human tRNA Guanine Transglycosylase Catalytic Subunit QTRT1, *Biomolecules,* **8.** First published on Aug 24, 2018.

8

5 Conclusion and Perspectives

The first part of this thesis follows up previous work by former members of the ▮▮▮▮ Group regarding the structural elucidation and characterization of hitherto unknown RNA modifications. For this purpose, two promising candidates were selected from the created list, namely candidate 338 and 470, both of which appeared to be hypermodified structures and thus might play important roles in proper tRNA function.

For the adenosine-derived candidate 338, an involvement of the modification enzymes MiaA and MiaB was confirmed, while the meanwhile proposed structure that contained a triple bond was discarded based on deuterium exchange experiments and the literature described instability of similar structures. Consideration of these findings as well as the remaining double bond equivalents resulted in the single alternative structure proposal of 2-methylthio-1,N6-ethenoadenosine. Finally, the organic synthesis of this hypermodified nucleoside and LC-MS/MS based alignment of both compounds confirmed their identity and thus the existence of this previously unknown modified nucleoside structure in *E. coli* tRNA. After successful structural elucidation and obvious structural as well as metabolic parallels to ms^2i^6A, the development of a method to label isoprene derived compounds in RNA allowed to demonstrate that $ms^2\varepsilon A$ originates from ms^2i^6A. Although this sophisticated concept was successfully applied for the purpose of this work, further optimization of the system is required before it can be applied as a routine tool in RNA analysis. An investigation of the biological effects of $ms^2\varepsilon A$ showed rare occurrence on tRNAs and predominant exclusion of accordingly modified tRNAs from translation, conferring $ms^2\varepsilon A$ dysfunctional character and warranting its classification as an RNA modification damage of ms^2i^6A. However, the mechanisms leading to the formation of $ms^2\varepsilon A$, i.e. formation by damaging agents or simply as a spurious side product of an enzymatic reaction, have not yet been elucidated and may be addressed by *in vitro* assays with the appropriate enzymes. Putting the results for candidate 338 in the context of discovery and structural elucidation of new RNA modifications and thus also the list of potentially new RNA modifications, not only the presence of possible artifacts and genuine RNA modifications with a biological function must be considered, but a distinction from RNA modification damages which might open up a new subfield of research should be drawn and critically questioned in further research in the field.

In the case of candidate 470, an investigation of the candidate's occurrence in RNA of various origins showed that the dimeric structure occurs natively in *E. coli* tRNA, whereas it is absent in human cell lines and *in vitro* transcribed RNA. However, in all three types of RNA (increased) formation of candidate 470 was induced upon UV irradiation *in vitro* as well as *in vivo*/*in cellulo* in *E. coli* and HEK cells. In the following experiments on the native formation of candidate 470, ThiI emerged as the modification enzyme involved in the *in vivo* formation and a spurious side reaction seemed plausible at first. Consequently, the corresponding structure proposal was organically synthesized, but subsequent LC-MS/MS analysis showed strongly different physicochemical properties compared to the native candidate 470, rejecting the proposed structure as well as the conceived mechanism. Alternatively, the involvement of ThiI suggested that s[4]U might play a role and thus also brought into play the s[4]U8-C13 crosslink, which would meet the elemental composition of candidate 470, but initially seemed unlikely to a certain extent, since the results with IVT RNA and HEK tRNA showed that an s[4]U modification was not necessarily needed. Pursuing the *in vitro* formation of candidate 470, the irradiation of the monomeric building blocks was not successful, but oligonucleotides with defined sequences allowed to further investigate the formation of the candidate from the main nucleosides and uncovered that the dimeric structure can be formed from two cytidines. Furthermore, a direct adjacency of cytidines for a potential intramolecular reaction was not crucial for formation of candidate 470 and, moreover, an intermolecular reaction proved possible as well. *In vitro* irradiation experiments with s[4]U-free HEK RNA showed a direct correlation between the UV energy dose and the abundance of candidate 470, which also reached saturation although at least the critical cytidine showed no signs of depletion. Despite the demonstrated formation of candidate 470 from unmodified RNA, a dethiolation experiment with subsequent irradiation showed that s[4]U is a promoting factor for the formation of the dimer. Although at first sight the alternative ways that lead to candidate 470 formation seemed very different, intensive literature research revealed that the structure postulated for a dimer resulting from the irradiation of polycytidylic acid is congruent with the structure of the s[4]U8-C13 crosslink, shaping a new perception of candidate 470. While known modified structures in tRNA were supposedly eliminated in the course of compiling the list of potentially new modifications, the findings obtained within this work provide strong evidence that candidate 470 might be this already known dimeric structure. In this case, the results of the present work would suggest that the UV-induced formation of the dimer from two cytidines can also take place in the biological context of an intact RNA structure both upon *in vitro* and upon *in vivo* irradiation and that these crosslinked structures arise not only in tRNA from *E. coli* but also in human cell lines, presumably in the context of UV damage to RNA. Still, the role of ThiI and s[4]U in the overall picture needs further clarification regarding the native occurrence of candidate 470 in *E. coli* because the currently available data does

not allow to completely rule out the possibility that the intramolecular crosslink might be formed from exposure to light during sample handling which may be addressed by performing bacterial culture and RNA extraction in the dark. In any case, the next step is to synthesize the s^4U8-C13 crosslink structure and to compare its LC-MS/MS behavior to that of candidate 470, in order to evaluate if both substances have one and the same structure.

Moving from the world of new RNA modification structures to the area of known RNA modifications, the second major goal of the present work deals with targeted manipulation of the hypermodified queuosine for a deeper understanding of its biological role on molecular level. In close collaboration with the groups of ██████████████████, ██████████ and ████████████, azide-containing preQ$_1$-derivatives were incorporated into tRNA of unicellular prokaryotes as well as eukaryotes. After successfully demonstrating the tolerance of TGT towards the supplemented congeners on *in vitro* level by subsequent click reaction, metabolic labeling experiments with one of the compounds were established *in vivo*. In *E. coli*, not only the incorporation of the supplied preQ$_1$-derivative by endogenous TGT was demonstrated by click reaction as well as LC-MS/MS analysis, but also the functional integration into the cellular process of translation by analysis of polysomal tRNA. In *S. pombe*, the azide moiety was used for affinity purification and sequencing experiments confirmed position 34 of the four known tRNA substrates as exclusive sites for TGT-catalyzed modification. Moreover, the semi-synthetic nucleoside proved to mimic the stimulating effect of Q in the modification circuit with m^{5}C38 in the anticodon stem loop of tRNA^{Asp}, representing the first manipulation of such an interaction *in vivo*. Even though queuosine was linked to a wide variety of diseases, a mechanism to explain these implications remains to be elucidated. Given the functional integration of the semi-synthetic nucleoside in the physiological processes mediated by queuosine *in vivo*, the workflow developed here represents a minimally invasive system to install a non-natural, clickable nucleobase within cellular tRNA and paves the way to scrutinize the purpose of queuosine in normal and pathogenic human cell lines on molecular level as well as potential therapeutic applications in this context.

6 Material and Methods

The material and methods described below refer to the unpublished results described in 4.2. Experimental protocols of the published results can be found in the materials and methods section of the respective publication or in the corresponding supporting information.

6.1 Material

6.1.1 Chemicals

Acetic acid, LC-MS grade	Sigma-Aldrich (Steinheim, Germany)
Acetic acid, glacial	Honeywell (Morris Plains, USA)
Acetonitrile, LC-MS grade (ACN)	Honeywell (Morris Plains, USA)
Ammonium acetate	Carl Roth (Karlsruhe, Germany)
Ammonium acetate, LC-MS grade	Sigma-Aldrich (Steinheim, Germany)
Ammonium chloride	Merck Millipore (Burlington, USA)
Ammonium chloride, ^{15}N-labeled	Sigma-Aldrich (Steinheim, Germany)
Ammonium hydroxide	Sigma-Aldrich (Steinheim, Germany)
Ammonium persulfate (APS)	Carl Roth (Karlsruhe, Germany)
Ampicillin	Carl Roth (Karlsruhe, Germany)
ATP $\geq$ 90%, lyophilized	Carl Roth (Karlsruhe, Germany)
Bacteriological agar	Sigma-Aldrich (Steinheim, Germany)
Bromphenol Blue	Merck (Darmstadt, Germany)
Bovine serum albumine (BSA)	Thermo Fisher Scientific (Waltham, USA)
Calcium chloride	Grüssing GmbH (Filsum, Germany)
Chloroform, HPLC grade	Sigma-Aldrich (Steinheim, Germany)
CTP $\geq$ 98%, lyophilized	Carl Roth (Karlsruhe, Germany)
Disodium hydrogen phosphate $\geq$ 99%	Carl Roth (Karlsruhe, Germany)
Dichlorodimethylsilane	Sigma-Aldrich (Steinheim, Germany)
Diethyl ether	Sigma-Aldrich (Steinheim, Germany)
Dithiothreitol (DTT)	Thermo Fisher Scientific (Waltham, USA)
Dulbecco's Balanced Salt Solution (DPBS)	Thermo Fisher Scientific (Waltham, USA)
Ethanol 99.9% (*V/V*, absolute)	Carl Roth (Karlsruhe, Germany)
Ethylenediaminetetraacetic acid (EDTA)	Carl Roth (Karlsruhe, Germany)

GelRed™ (3x)	Biotium (Hayward, USA)
Glucose (D-glucose)	Carl Roth (Karlsruhe, Germany)
Glucose (D-glucose), ^{13}C-labeled	Sigma-Aldrich (Steinheim, Germany)
Glycerol	Sigma-Aldrich (Steinheim, Germany)
Glycogen, RNA-grade	Thermo Fisher Scientific (Waltham, USA)
Guanosine 5′-monophosphate	Sigma-Aldrich (Steinheim, Germany)
GTP $\geq$ 90%, lyophilized	Carl Roth (Karlsruhe, Germany)
HEPES $\geq$ 99.5%	Carl Roth (Karlsruhe, Germany)
Hydrogen peroxide, 3% (*m/m*)	Sigma-Aldrich (Steinheim, Germany)
Isopropyl alcohol, HPLC grade	Carl Roth (Karlsruhe, Germany)
Isopropyl-β-D-thiogalactopyranosid (IPTG)	Carl Roth (Karlsruhe, Germany)
Kanamycin	Carl Roth (Karlsruhe, Germany)
LB broth with agar (Lennox)	Sigma-Aldrich (Steinheim, Germany)
LB broth (Lennox)	Carl Roth (Karlsruhe, Germany)
Magnesium chloride hexahydrate	Merck (Darmstadt, Germany)
Magnesium sulfate $\geq$ 99%	Carl Roth (Karlsruhe, Germany)
Magnesium sulfate heptahydrate	Carl Roth (Karlsruhe, Germany)
N,N,N',N'-Tetramethylethylenediamine (TEMED)	Carl Roth (Karlsruhe, Germany)
Pentostatine $\geq$ 95%	Sigma-Aldrich (Steinheim, Germany)
Potassium chloride $\geq$ 99.5%	Carl Roth (Karlsruhe, Germany)
Potassium dihydrogen phosphate $\geq$ 99%	Carl Roth (Karlsruhe, Germany)
Primocin	InvivoGen (San Diego, USA)
RNaseOUT™	Thermo Fisher Scientific (Waltham, USA)
Roti-Aqua Phenol, pH 4 (for RNA)	Carl Roth (Karlsruhe, Germany)
Roti Phenol (for DNA)	Carl Roth (Karlsruhe, Germany)
Rotiphorese™ Gel 40% acrylamide mix (19:1)	Carl Roth (Karlsruhe, Germany)
Sodium chloride $\geq$ 99%	Carl Roth (Karlsruhe, Germany)
Sodium sulfite	Carl Roth (Karlsruhe, Germany)
Spermidine trihydrochloride $\geq$ 99.5%	Sigma-Aldrich (Steinheim, Germany)
TBE buffer, 10x	Carl Roth (Karlsruhe, Germany)
Tetrahydrouridine, InSolution	Merck (Darmstadt, Germany)
Total tRNA, *E. coli*	Roche Diagnostics (Basel, Schweiz)
TRI Reagent®	Sigma-Aldrich (Steinheim, Germany)
Tris-HCl	Carl Roth (Karlsruhe, Germany)
Trisodium citrate	Sigma-Aldrich (Steinheim, Germany)
TritonX-100	Sigma-Aldrich (Steinheim, Germany)
Trypan Blue	Thermo Fisher Scientific (Waltham, USA)
Trypsin-EDTA, 0.05%	Thermo Fisher Scientific (Waltham, USA)
Tryptone ex casein	Carl Roth (Karlsruhe, Germany)
UTP $\geq$ 90 %, lyophilized	Carl Roth (Karlsruhe, Germany)

Yeast extract   Carl Roth (Karlsruhe, Germany)
Zinc chloride   Sigma-Aldrich (Steinheim, Germany)

6.1.2 Buffers, solutions and growth media

Unless stated otherwise, buffers and growth media were prepared with Milli-Q H_2O.

6.1.2.1 Buffers and solutions

APS solution	10% APS (m/V)
DNase I buffer, 10x	100 mM Tris-HCl (pH 7.5), 25 mM $MgCl_2$, 1 mM $CaCl_2$
Fast AP Buffer, 10x	500 mM NH_4OAc (pH 9.0, adjusted with NH_4OH)
HEPES buffer, 5x	125 mM HEPES, 750 mM NaCl
LC-MS/MS solvent A	5 mM NH_4OAc (pH 5.3—5.4, adjusted with acetic acid), 1% V/V ACN
Nuclease P1 Buffer, 10x	225 mM NH_4OAc (pH 5.0), 0.2 mM $ZnCl_2$
One pot digestion (OPD) buffer , 10x	250 mM NH_4OAc (pH 7.5)
Phosphate buffer, 1x (pH 8)	3.52 mM NaH_2PO_4, 46.48 mM Na_2HPO_4, 1 mM EDTA
Straßbourg (SB) buffer (transcription buffer)	40 mM Tris-HCl (pH 8.1), 1 mM spermidine, 5 mM DTT, 0.01% Triton X-100
TBE buffer, 10x (pH 8.3)	1 M tris boric acid, 20 mM EDTA

6.1.2.2 Growth media

eukaryotic cell culture medium	DMEM supplemented with 10% (V/V) FBS and 100 µg/mL primocin
LB agar (plates)	15 g/L bacteriological agar 20 g/L LB broth with agar
LB medium	20 g/L LB broth
M9-minimal medium (standard)	1.0 g/L NH_4Cl, 4.0 g/L D-glucose, 6.8 g/L Na_2HPO_4, 3.0 g/L KH_2PO_4, 0.5 g/L NaCl, 0.1 mM $CaCl_2$, 2.0 mM $MgSO_4$

M9-minimal medium (^{13}C-labeled)	1.0 g/L NH$_4$Cl, 4.0 g/L ^{13}C$_6$-D-glucose, 6.8 g/L Na$_2$HPO$_4$, 3.0 g/L KH$_2$PO$_4$, 0.5 g/L NaCl, 0.1 mM CaCl$_2$, 2.0 mM MgSO$_4$
M9-minimal medium (^{15}N-labeled)	1.0 g/L ^{15}NH$_4$Cl, 4.0 g/L D-glucose, 6.8 g/L Na$_2$HPO$_4$, 3.0 g/L KH$_2$PO$_4$, 0.5 g/L NaCl, 0.1 mM CaCl$_2$, 2.0 mM MgSO$_4$
SOC medium	3.6 g/L D-glucose, 2.03 g/L MgCl$_2 \cdot$ 6H$_2$O, 2.47 g/L MgSO$_4 \cdot$ 7H$_2$O, 0.2 g KCl, 0.6 g/L NaCl, 20 g/L tryptone, 5 g/L yeast extract

6.1.3 Enzymes and kits

Benzonase® Nuclease ($>$ 250 U/µL)	Sigma Aldrich (Steinheim, Germany)
Cfr42I (SacII) (10 U/µL)	Thermo Fisher Scientific (Waltham, USA)
DNase I (1 U/µL), RNase-free	Thermo Fisher Scientific (Waltham, USA)
Fast AP thermosensitive Alkaline Phosphatase (1 U/µL)	Thermo Fisher Scientific (Waltham, USA)
Invitrogen™ Library Efficiency™ DH5α Competent Cells kit	Thermo Fisher Scientific (Waltham, USA)
MEGAclear™ Kit	Thermo Fisher Scientific (Waltham, USA)
Nuclease P1 from *Penicillium citrinum* (lyophilized)	Sigma Aldrich (Steinheim, Germany)
Snake Venom Phosphodiesterase from *Crotalus adamanteus venom* (lyophilized)	Worthington Biochemical Corporation (Lakewood, USA)
T7 RNA polymerase (20 U/µL)	Thermo Fisher Scientific (Waltham, USA)

6.1.4 Oligonucleotides and plasmids

6.1.4.1 RNA oligonucleotides

ID	Sequence	Comment	Supplier
MH272	AGAAAUAUGUCUGAUAAA	only 1 C	in-lab synthesized (Salifu Seidu-Larry)
MH527	ACCAACAAACCACCACAACC	only C & A	Sigma Aldrich (Steinheim, Germany)
MH553	UUCUUUCUUUCCCUUCCUUU	only C & U	Biomers (Ulm, Germany)

| MH554 | UUAUAUUUAUAAAUAUUAAA | only U & A | Biomers (Ulm, Germany) |
| MH1213 | UCAUUCACAUCUAUGUUAAU UCUACUAUACUCAUAUUAUC UCUUAUCAUA | no adjacent Cs | IDT (Coralville, USA) |

6.1.4.2 DNA oligonucleotides

ID	Sequence	Comment	Supplier
MH572	CGCGCCGGCCGGCGGGCCGG	only C and G	Biomers (Ulm, Germany)
MH1250	/6-FAM/ AGACGTGTGCTCTTC CGATCTTGGCGTCCCCTAGGG GATTCGAACCCCTGTTACCGC CGTGGATCGTCGGACTGTAGA ACTCTGAAC	complementary to *E. coli* tRNA^{Glu}	IDT (Coralville, USA)
MH1254	/6-FAM/ AGACGTGTGCTCTTC CGATCTTGGTGCCCGGACTCG GAATCGAACCAAGGACACGGG GATTGATCGTCGGACTGTAGA ACTCTGAAC	complementary to *E. coli* tRNA^{Phe}	IDT (Coralville, USA)
MH1291	/6-FAM/ AGACGTGTGCTCTTCC GATCTTGGTGCCGATAATAGGA GTCGAACCTACGACCTTCGCAT TGATCGTCGGACTGTAGAACTC TGAAC	complementary to *E. coli* tRNA$^{Thr}_{(GGT)}$	IDT (Coralville, USA)

Abbreviations: 6-FAM = 6-carboxyfluorescein.

6.1.4.3 Plasmids

Vector	Encoding for
pUC 57	full length EMCV IRES eGFP mRNA: GGCGAAUUGGGUACCGGGCCCCCCCU CGAGGUCAUCGAAUUCCGCCCCUCUCCCUCCCCCCCCCCUAACGUUACUG GCCGAAGCCGCUUGGAAUAAGGCCGGUGUGCGUUUGUCUAUAUGUUAUU UUCCACCAUAUUGCCGUCUUUUGGCAAUGUGAGGGCCCGGAAACCUGGC CCUGUCUUCUUGACGAGCAUUCCUAGGGGUCUUUCCCCUCUCGCCAAAG GAAUGCAAGGUCUGUUGAAUGUCGUGAAGGAAGCAGUUCCUCUGGAAGC UUCUUGAAGACAAACAACGUCUGUAGCGACCCUUUGCAGGCAGCGGAAC

CCCCCACCUGGCGACAGGUGCCUCUGCGGCCAAAAGCCACGUGUAUAAG
AUACACCUGCAAAGGCGGCACAACCCCAGUGCCACGUUGUGAGUUGGAU
AGUUGUGGAAAGAGUCAAAUGGCUCUCCUCAAGCGUAUUCAACAAGGGG
CUGAAGGAUGCCCAGAAGGUACCCCAUUGUAUGGGAUCUGAUCUGGGGC
CUCGGUGCACAUGCUUUACAUGUGUUUAGUCGAGGUUAAAAAAACGUCU
AGGCCCCCCGAACCACGGGGACGUGGUUUUCCUUUGAAAAACACGAUGA
UAAUAUGGCCACAACCAUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGG
GUGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGUAAACGGCCACAAGU
UCAGCGUGUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGAC
CCUGAAGUUCAUCUGCACCACCGGCAAGCUGCCCGUGCCCUGGCCCACC
CUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCCGCUACCCCG
ACCACAUGAAGCAGCACGACUUCUUCAAGUCCGCCAUGCCCGAAGGCUA
CGUCCAGGAGCGCACCAUCUUCUUCAAGGACGACGGCAACUACAAGACC
CGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACCGCAUCGAGC
UGGUGAACCGCAUCGAGCUGAAGGAGGACGGCAACAUCCUGGGGCACAA
GCUGGAGUACAACUACAACAGCCACAACGUCUAUAUCAUGGCCGACAAG
CAGAAGAACGGCAUCAAGGUGAACUUCAAGAUCCGCCACAACAUCGAGG
ACGGCAGCGUGCAGCUCGCCGACCACUACCAGCAGAACACCCCCAUCGG
CGACGGCCCCGUGCUGCUGCCCGACAACCACUACCUGAGCACCCAGUCC
GCCCUGAGCAAAGACCCCAACGAGAAGCGCGAUCACAUGGUCCUGCUGG
AGUUCGUGACCGCCGCCGGGAUCACUCUCGGCAUGGACGAGCUGUACAA
GUAAAGCGGCCGCCACCGCGG

pGEX-
5X-1-H

thiI: ATGAAATTTATTATTAAACTGTTTCCGGAAATTACCATTAAAAGCCAG
AGCGTGCGCCTGCGCTTTATTAAAATTCTGACCGGCAACATTCGCAACGTG
CTGAAACATTATGATGAAACCCTGGCGGTGGTGCGCCATTGGGATAACATT
GAAGTGCGCGCGAAAGATGAAAACCAGCGCCTGGCGATTCGCGATGCGCT
GACCCGCATTCCGGGCATTCATCATATTCTGGAAGTGGAAGATGTGCCGTT
TACCGATATGCATGATATTTTTGAAAAAGCGCTGGTGCAGTATCGCGATCA
GCTGGAAGGCAAAACCTTTTGCGTGCGCGTGAAACGCCGCGGCAAACATG
ATTTTAGCAGCATTGATGTGGAACGCTATGTGGGCGGCGGCCTGAACCAGC
ATATTGAAAGCGCGCGCGTGAAACTGACCAACCCGGATGTGACCGTGCATC
TGGAAGTGGAAGATGATCGCCTGCTGCTGATTAAAGGCCGCTATGAAGGCA
TTGGCGGCTTTCCGATTGGCACCCAGGAAGATGTGCTGAGCCTGATTAGCG
GCGGCTTTGATAGCGGCGTGAGCAGCTATATGCTGATGCGCCGCGGCTGCC
GCGTGCATTATTGCTTTTTTAACCTGGGCGGCGCGGCGCATGAAATTGGCG
TGCGCCAGGTGGCGCATTATCTGTGGAACCGCTTTGGCAGCAGCCATCGCG
TGCGCTTTGTGGCGATTAACTTTGAACCGGTGGTGGGCGAAATTCTGGAAA
AAATTGATGATGGCCAGATGGGCGTGATTCTGAAACGCATGATGGTGCGCG
CGGCGAGCAAAGTGGCGGAACGCTATGGCGTGCAGGCGCTGGTGACCGGC
GAAGCGCTGGGCCAGGTGAGCAGCCAGACCCTGACCAACCTGCGCCTGATT
GATAACGTGAGCGATACCCTGATTCTGCGCCCGCTGATTAGCTATGATAAA

```
GAACATATTATTAACCTGGCGCGCCAGATTGGCACCGAAGATTTTGCGCGC
ACCATGCCGGAATATTGCGGCGTGATTAGCAAAAGCCCGACCGTGAAAGCG
GTGAAAGCAAAATTGAAGCGGAAGAAGAAAAATTTGATTTTAGCATTCT
GGATAAAGTGGTGGAAGAAGCGAACAACGTGGATATTCGCGAAATTGCGC
AGCAGACCGAACAGGAAGTGGTGGAAGTGGAAACCGTGAACGGCTTTGGC
CCGAACGATGTGATTCTGGATATTCGCAGCATTGATGAACAGGAAGATAAA
CCGCTGAAAGTGGAAGGCATTGATGTGGTGAGCCTGCCGTTTTATAAACTG
AGCACCAAATTTGGCGATCTGGATCAGAACAAAACCTGGCTGCTGTGGTGC
GAACGCGGCGTGATGAGCCGCCTGCAGGCGCTGTATCTGCGCGAACAGGGC
TTTAACAACGTGAAAGTGTATCGCCCGTAG
```

Plasmids including the oligo corresponding to the respective gene of interest were synthesized by GeneScript (Piscataway, USA) and purchased commercially.

6.1.5 Cell lines

6.1.5.1 Prokaryotic

The *E. coli* strains (K-12) used within this work were part of the commercially available Dharmacon™ Keio collection which was purchased from GE Healthcare (United Kingdom).[323] The selection of strains that was used for this thesis is listed below.

Name	ID	Affected enzyme
Keio parent	BW25113	none (wild type)
thiI KO	JW0413	tRNA sulfurtransferase
mnmA KO	JW1119	tRNA-specific 2-thiouridylase

6.1.5.2 Eukaryotic

The human embryonic kidney (HEK) cell line 293 (DSMZ no. ACC 305) was purchased from the DSMZ-German Collection of Microorganisms and Cell Cultures GmbH (Braunschweig, Germany)

6.1.6 Disposables

Countess™ Cell Counting Chamber Slides	Thermo Fisher Scientific (Waltham, USA)
Eppendorf tubes (1.5 mL)	Carl Roth (Karlsruhe, Germany)
Falcon tubes (15, 50 mL)	Sarstedt (Nümbrecht, Germany)
Filter top vacuum bottles, PES, 0.2 µm pore (250, 500 mL)	Sarstedt (Nümbrecht, Germany)
Inserts, conical, clear glass (0.1 mL)	neoLab Migge GmbH (Heidelberg, Germany)

Mμltiflex Round Tips	Sorenson Bioscience Inc. (Murray, USA)
Nanosep® MF Centrifugal Devices (0.45 μM)	Pall (New York, USA)
Needles (Sterican)	Braun (Melsungen, Germany)
Parafilm®	VWR (Darmstadt, Germany)
PCR Softtubes	Biozym (Hessisch Oldendorf, Germany)
Petri dishes	Sarstedt (Nümbrecht, Germany)
Pipette tips (with filter, sterile, RNase/DNase-free)	Greiner Bio-One (Frickenhausen, Germany)
Polypropylene centrifugation tubes (50 mL)	Sarstedt (Nümbrecht, Germany)
Reaction tubes with screw cap (1.5 mL)	Carl Roth (Karlsruhe, Germany)
Screw-cap ND9 (blue, septum red rubber/PTFE beige)	neoLab Migge GmbH (Heidelberg, Germany)
Serological pipettes (2, 5, 10, 25 mL)	Sarstedt (Nümbrecht, Germany)
Semi-micro cuvette, PS, transparent (3 mL)	Sarstedt (Nümbrecht, Germany)
Short thread vials ND9 (1.5 mL)	neoLab Migge GmbH (Heidelberg, Germany)

6.1.7 Instruments

Analytical balances

Mettler Toledo PM460	Mettler Toledo (Gießen, Germany)
Sartorius Cubis Analytical Balance	Sartorius (Goettingen, Germany)

Centrifuges

Beckman Avanti J25	Beckman Coulter (Krefeld, Germany)
Eppendorf Centrifuge 5810R	Eppendorf (Hamburg, Germany)
Eppendorf Centrifuge 5430R	Eppendorf (Hamburg, Germany)
Sprout mini centrifuge	Biozym (Hessisch Oldendorf, Germany)

***E. coli* culture**

Heraeus BB15 incubator	Thermo Fisher Scientific (Waltham, USA)
InFors Ecotron	InFors HT (Basel, Switzerland)
Biochrom Ultrospec™ cell density meter	Thermo Fisher Scientific (Waltham, USA)

Eukaryotic cell culture

Heraeus BB15 CO_2 incubator	Thermo Fisher Scientific (Waltham, USA)
Herasafe™ HS 12	Thermo Fisher Scientific (Waltham, USA)
Microscope DM IRB	Leica Microsystems (Wetzlar, Germany)
Invitrogen™ Countess™ 3FL automated cell counter	Thermo Fisher Scientific (Waltham, USA)

Gel electrophoresis

Electrophoresis Power Supply – EPS 3500XL	GE Healthcare (Buckinghamshire, UK)
LSG-400-20 NA vertical chamber	C.B.S. Scientific (San Diego, USA)
Model 250/2.5 power supply	BioRad (Munich, Germany)
Shaker (DOS-10L)	neoLab Migge GmbH (Heidelberg, Germany)
TapeStation 4200	Agilent Technologies (Waldbronn, Germany)
Typhoon TRIO+ variable mode imager	GE Healthcare (Chicago, USA)

High performance liquid chromatography & columns

Agilent 1260 Infinity (II) LC	Agilent Technologies (Waldbronn, Germany)
Synergy Fusion RP C18 column (4 µm particle size, 80 Å pore size, 250 mm length, 2 mm inner diameter)	Phenomenex (Aschaffenburg, Germany)

General equipment

NanoDrop™ ND-2000	PeqLab (Erlangen, Germany)
pH-meter FiveEasy™ FE20	Mettler Toledo (Gießen, Germany)
Pipette boy Integra	VWR (Darmstadt, Germany)
Ultrapure water purification system Milli-Q®	Millipore (Schwalbach, Germany)
Variable micropipettes Discovery Comfort (2, 10, 20, 100, 200, 1000 µL)	Abimed (Langenfeld, Germany)
Vortex Mixer 7-2020	neoLab Migge GmbH (Heidelberg, Germany)

Irradiation

CAMAG™ UV Lamp 4	CAMAG (Muttenz, Switzerland)
UV Stratalinker 1800 Crosslinker	Stratagene (San Diego, USA)

Mass Spectrometry

Agilent 6470B triple quadrupole	Agilent Technologies (Waldbronn, Germany)
Genius XE-70, nitrogen generator	Peak Scientific (Düren, Germany)

Thermoshaker & heatblocks

BIOER ThermoCell	BIOER (Hangzhou, China)
Eppendorf Thermomixer Comfort	Eppendorf (Hamburg, Germany)
VWR Digital Heatblock	VWR International (Radnor, USA)

6.1.8 Software

Adobe Illustrator 2022	Adobe Inc. (San José, USA)
ChemBioDraw Ultra 14.0	Cambridge Soft/PerkinElmer (Waltham, USA)

Citavi 6	Swiss Academic Software (Wädenswil, Switzerland)
IrfanView	created by Irfan Skiljan
MassHunter, version 10.0	Agilent Technologies (Waldbronn, Germany)
Microsoft Office 365	Microsoft (Redmont, USA)
PyMOL, version 2.3.4	Schrödinger LLC (Braunschweig, Germay)
SnapGene 6.0.2	GSL Biotech LLC (Chicago, USA)
TeXstudio 4.5.1	created by Benito van der Zander
Typhoon scanner software	GE Healthcare (Chicago, USA)

6.2 Methods

6.2.1 *E. coli* culture

Within this work, *E. coli* strains from the commercially available Keio collection (GE Healthcare (Dharmacon™), United Kingdom) containing the wild-type Keio parent strain as well as strains with single gene deletions for every non-essential gene in the *E. coli* K-12 strain were used.[323] Of note, the Keio parent has no antibiotic resistance while the KO strains are equipped with a kanamycin resistance (final concentration 25 µg/mL) The collection was stored in 96-well plates at $-80℃$.

6.2.1.1 General procedure

In a first step for *E. coli* culture, the strain of interest was plated on an LB-agar plate with (KO strains) or without kanamycin (Keio parent) and grown overnight in an incubator at 37 ℃. In case the colonies were not directly used for culture, a storage of the plates at 4 ℃ for 1-2 weeks was possible. In order to prepare an overnight culture, one colony was picked from the agar plate with a pipet tip, added to 10 mL of the respective growth medium and incubated overnight (190 rpm, 37 ℃). The next day, the optical density (at 600 nm, OD_{600}) of these pre-cultures was measured and used to calculate the required amount to inoculate the main culture to an optical density of 0.05. The main cultures were grown at 37 ℃ and 190 rpm, mainly in 300 mL Erlenmeyer flasks in a total volume of 60 mL of the appropriate culture medium which was supplemented with antibiotics if necessary. When an OD of approximately 0.8 was reached, the culture was transferred into 50 mL centrifugation tubes to harvest the cells by centrifugation (10000 rpm, 4 ℃, 20 min) and, depending on the volume of the culture, an appropriate amount of TRI Reagent® (usually 1-2 mL) was added and the mixture homogenized by vortexing. Subsequently, RNA or DNA was extracted from these samples as described in 6.2.3.2 and 6.2.4.1, respectively.

6.2.1.2 *E. coli* knockout experiments

For *E. coli* knockout experiments $\Delta thiI$ (JW0413) and $\Delta mnmA$ (JW1119) strains from the Keio collection were used as well as the corresponding parent strain (BW25113).[323] The cells were cultured in standard M9 medium and RNA was isolated as described below (6.2.3.2).

6.2.1.3 Preparation of unlabeled and isotopically labeled nucleic acid samples (unlabeled, ^{13}C, ^{15}N)

To obtain unlabeled nucleic acid samples, the strain of interest (Keio parent) was cultured in standard M9 medium, while isotopically labeled nutrients, namely ^{15}NH$_4$Cl for ^{15}N-labeled M9 medium or ^{13}C-glucose for ^{13}C-labeled M9 medium replaced their unlabeled counterparts in the recipe. Subsequent extraction (see 6.2.3.2 and 6.2.4.1) yielded the accordingly labeled nucleic acid.

6.2.1.4 Preparation of chemically competent cells

1 mL of an overnight culture of the $\Delta thiI$ strain in LB medium supplemented with kanamycin (25 µg/mL) was added to 99 mL LB medium (+ 25 µg/mL kanamycin) in a 500 mL baffled flask and cultured at 37 ℃, 190 rpm until an optical density of 0.3-0.5 was reached. A respectively equal volume of culture was then transferred into 3 centrifugation tubes and the bacteria were harvested (10 min, 7000 rpm, 4 ℃). The supernatant was discarded, the remaining pellet gently resuspended in 20 mL CaCl$_2$ (100 mM, sterile) and the solution incubated on ice for 15 min. After another centrifugation step (10 min, 7000 rpm, 4 ℃), the cells were resuspended gently in 5 mL of the ice-cold CaCl$_2$ solution supplemented with 15% (*V/V*) glycerol. For further experiments, the competent cells were stored in aliquots of 50 µL at −80 ℃. As a quality control, the transformation efficiency of the competent cells was quantified by transformation with a pUC-19 plasmid (see 6.2.1.5).

6.2.1.5 Transformation

The chemically competent cells to be transformed were thawed on ice. In a 15 mL polypropylene tube, the plasmid of interest (50 pg of pUC-19 control plasmid from the Invitrogen™ Library Efficiency DH5α™ Competent Cells kit or 100 ng of *thiI*-containing plasmid, both conferring ampicillin resistance) was gently mixed with 50 µL of these competent cells by carefully pipetting up and down and incubated on ice for 30 min. The cells were then heat-shocked in a pre-heated water bath (42 ℃) for 45 s and subsequently incubated with 950 µL SOC medium (170 rpm, 37 ℃, 60 min). Meanwhile, LB-agar plates selective for kanamycin (25 µg/mL) as well as ampicillin resistance (50 µg/mL) were

warmed in the incubator at 37 ℃ and after the incubation used to plate the transformed cells. For the *thiI*-expressing plasmid, 75 μL of the transformation mixture were plated while in case of the control plasmid a dilution of the mixture at a ration of 1:5 with SOC medium was performed before plating 150 μL of this diluted mixture on an agar plate. After the plates were incubated overnight at 37 ℃, the number of colonies was counted, in order to determine transformation efficiency as colony forming units per μg ($\frac{CFU}{\mu g}$):

$$\frac{CFU}{\mu g} = \frac{CFU\,(P)}{m\,(pUC19)} \cdot \frac{1 \cdot 10^6\,pg}{g} \cdot \frac{V(T)}{V(P)} \cdot f \tag{6.1}$$

CFU (P) = colonies counted on control LB agar plate
m (pUC19) = amount of pUC19 plasmid DNA used for transformation in pg (here: 50 pg)
V (T) = volume of transformation batch in mL (here: 1 mL)
V (P) = volume of plated transformation batch in mL (here: 0.15 mL)
f = dilution factor (here: 5)

The transformation efficiency of the DH5α™ cells from the Invitrogen ™ kit served as a reference.

6.2.1.6 Culture of transformed cells

Culture of the transformed *E. coli* cells followed the initially described general protocol until an OD of 0.8 was reached. At that time, gene expression was induced by addition of 0.5 mM isopropyl-β-ᴅ-thiogalactopyranosid (IPTG). After maintaining the culture conditions for 4 h, the cells were harvested and RNA extracted according to the described procedures (6.2.1.1 and 6.2.3.2)

6.2.1.7 UV irradiation of *E. coli* cultures

For the UV irradiation of *E. coli* cells *in vivo*, the Keio parent strain and the *thiI* strain were grown in M9 medium as previously described. Upon reaching an OD of 0.8, the culture was split in 2 petri dishes in equal volumes. One of these petri dishes was irradiated in the Stratalinker at 254 nm with a UV energy dose of 1 J/cm^2 by placing the open dish in the device while the other dish was placed in the stratalinker for the same time span without being irradiated (control). Afterwards, cells were harvested and RNA or DNA extracted as described below (6.2.3.2 and 6.2.4.1).

6.2.2 Eukaryotic cell culture

Eukaryotic cells were cultured by ▮▮▮▮▮▮▮▮.

6.2.2.1 General procedure

Aliquots of cell pellets were stored in liquid nitrogen and all media and solutions required for eukaryotic cell culture were sterile filtered and warmed to 37 ℃ prior to use. Eukaryotic cells were generally cultivated in a humidified incubator at 37 ℃ and 5% CO_2. Frozen cell aliquots of HEK293 were thawed by gently resuspending them in warm medium and cultivated in DMEM which was supplemented with FBS (10%, *V/V*) and a final concentration of 100 µg/mL primocin overnight and passaged in fresh medium the next day. Cultures were grown to a confluency of approximately 80% and routinely split in a ratio of 1:20.

6.2.2.2 UV irradiation of HEK293 cells

An amount of $6 \cdot 10^6$ cells in 12 mL DMEM (+ 10% (*V/V*) FBS and 100 µg/mL primocin) was seeded in a petri dish of 100 mm diameter. After 2 h of incubation at 37 ℃, the cells were adhered to the dish and irradiated in the Stratalinker at 254 nm with a UV energy dose of 1 J/cm^2 by placing the open dish in the device. As a control, another dish was placed in the stratalinker for the same amount of time without being irradiated. After the procedure, the medium was transferred in a 15 mL centrifugation tube and the cells were washed with 1 mL DPBS. The washing solution was also transferred to the centrifugation tube. Subseqeuntly, the cells were trypsinized by addition of 1 mL ready-to-use trypsin-EDTA (0.05%) and also transferred to the centrifugation tube. 20 µL of each cell suspension were used for a viability assay, i.e. staining with trypan blue and evaluation of viability using a cell counter and disposable slides. The cells within the remaining solution were harvested by centrifugation (400 g, 5 min) and the resulting pellet homogenized in 3 mL TRI Reagent® for subsequent RNA or DNA extraction (see 6.2.3.2 and 6.2.4.1).

6.2.3 RNA sample preparation

6.2.3.1 Synthesis of *in vitro* transcribed RNA

In a first step, a commercially purchased plasmid DNA (pDNA) encoding for full-length EMCV IRES-eGFP mRNA, which was kindly provided by ▮▮▮▮▮▮▮▮, was linearized by use of the restriction enzyme SacII according to the manufacturer's protocol in an overnight incubation at 37 ℃. Isolation of the linearized pDNA was achieved by extracting the mixture with chloroform and DNA phenol at a ratio of 2:1:1 twice. The joint aqueous phases were then extracted twice with the same volume of chloroform and residual phenol was removed by subsequent extraction with the same volume of diethyl ether. Each extraction step was performed by harsh vortexing which was followed by

centrifugation for clear phase separation, respectively. Finally, the linearized DNA was purified by ammonium acetate ethanol precipitation (see 6.2.3.3).

For *in vitro* transcription, a mixture of the substances displayed in Table 6.1 was incubated at 37 ℃ for 4 h.

Subsequently, the DNA template was hydrolyzed by incubation of the *in vitro* transcription mixture with 2 U DNase I per μg DNA in the appropriate buffer at 37 ℃ for 30 min. The *in vitro* transcribed RNA was purified using the MEGAclear™ kit (Thermo Fisher Scientific, Germany) following the manufacturer's instructions and ammonium acetate ethanol precipitation (see 6.2.3.3).

Table 6.1: Representative pipetting scheme for *in vitro* transcription.

compound	stock	final	V / μL
Plasmid (ng/μL)	1680	50	11.90
5x SB Buffer	5	1	80.00
DTT (mM)	1000	5	2.00
$MgCl_2$ (mM)	1000	30	12.00
BSA (μg/mL)	1000	2.5	1.00
GTP (mM)	100	2	8.00
GMP (mM)	125	16	51.20
CTP (mM)	100	5	20.00
UTP (mM)	100	5	20.00
ATP (mM)	100	5	20.00
RNase Out	40		1.70
T7 RNA polymerase (μg/μL)			20.00
Milli-Q H_2O			152.20
final volume			400.00

6.2.3.2 RNA extraction

RNA was extracted from harvested cell pellets (*E. coli* or eukaryotic cells) using TRI Reagent® according to the manufacturer's protocol. In a first step, the cell pellets were homogenized in an appropriate volume of TRI Reagent® by vortexing, split in portions of 1 mL homogenized mixture per Eppendorf tube and 200 μL chloroform were added to each tube. The samples were mixed by vortexing, incubated at room temperature for 5 min and then centrifuged (13000 rpm, 4 ℃, 10 min). The aqueous phase was transferred to a new tube and the procedure repeated with another 200 μL of chloroform. After centrifugation (13000 rpm, 4 ℃, 10 min), the aqueous phase was transferred to a new Eppendorf tube and 500 μL isopropanol were added for precipitation. After vortexing the mixture was incubated at room temperature for 15 min and then centrifuged (13000 rpm, −4 ℃, at least 45 min). The resulting RNA pellet was washed with 200 μL

icecold 75% (*V/V*) ethanol (13000 rpm, −4 ℃, at least 15 min) and then air-dried for a few minutes. Finally, the RNA was dissolved in Milli-Q water and its concentration and purity determined by NanoDrop™ measurement. In case of impurities, the samples were subjected to ammonium acetate ethanol precipitation.

Of note, for the validation of candidate 470 as RNA modification, the homogenized solutions of harvested cell pellets from two independently grown were mixed and extracted together.

6.2.3.3 Ammonium acetate ethanol precipitation

In order to recover RNA (or alternatively DNA) from a solution, respective samples were mixed with 1/10 volume of 5 M ammonium acetate, 1 μL glycogen and 2-3 volumes of ice cold 100% ethanol. The mixture was incubated at −80 ℃ for 2 h or at −20 ℃ overnight and subsequently centrifuged at 12000 g and 4 ℃ for at least 45 min. The RNA pellet was the washed with icecold 75% ethanol (12000 g, 4 ℃, 15 min), air-dried for a few minutes and finally dissolved in Milli-Q water.

Concentrations were determined by NanoDrop™ measurements.

6.2.3.4 Purification of single RNA species: tRNA^{Phe}, tRNA^{Thr}, tRNA^{Glu}

The basis for this method were DNA oligonucleotides with sequences that are reverse complementary to the sequence of the tRNA species of interest (see 6.1.4.2). 5 μg of total tRNA extracted from *E. coli* cells (Keio parent or *thiI* KO, respectively) were mixed with 3 μL of a 3.5 μM solution of the respective DNA oligonucleotide and 2 μL 5x HEPES buffer in a total volume of 10 μL and subjected to the following program in a PCR cycler:

Heat to 110 ℃→ 2 min at 95 ℃→ 1 min at 90 ℃→ 1 min at 85 ℃→ 1 min at 80 ℃→ 1 min at 75 ℃→ 1 min at 70 ℃→ 3 min at 60 ℃→ 3 min at 50 ℃→ 3 min at 40 ℃→ 5 min at 30 ℃→ 5 min at 25 ℃→ storage at 4 ℃ for infinite time

Separation of the hybridized tRNA of interest from other tRNA species was achieved by native polyacrylamide gel electrophoresis (PAGE). Therefore, a 10% native polyacrylamide gel was prepared by mixing 25 mL of the commercially available 40% acrylamide mix (19:1), 10 mL of 10x TBE, 65 mL of Milli-Q water, 250 μL of a 10% (*m/V*) APS solution and 40 μL of a TEMED solution. After polymerization the gel was pre-run in 1x TBE buffer at 10 W for 10 min. Meanwhile the RNA samples were mixed with a 50% glycerol solution and then loaded onto the gel which was run at 10 W for 45 min. Then the gel was stained with GelRed and the bands were visualized by scanning with a Typhoon Imager at an excitation wavelength of 532 nm. With the printed image as a template, the bands of interest were excised, mashed with a scalpel and shaken with 300 μL of 0.5 M ammonium acetate at 15 ℃ and 700 rpm overnight. Gel pieces were removed by filtration

though NanoSep® centrifugal filters and the hybrids precipitated following the ammonium acetate ethanol precipitation protocol. Subsequently, DNA oligonucleotides were removed by DNase I digestion by incubation of samples with 5 µL DNase I (1 U/µL) and 5 µL of the appropriate 10x buffer in a total volume of 50 µL at 37 ℃ for 1 h. Isolation of the RNA was achieved by phenol chloroform extraction. Therefore the sample mixture (diluted to 100 µL with Milli-Q water) was extracted with 50 µL chloroform and 50 µL Roti-Aqua Phenol by harsh vortexing and centrifugation for phase separation. This extraction step was repeated and then the combined aqueous phases were extracted twice with the same volume of chloroform before the RNA was precipitated according to the ammonium acetate ethanol protocol. Finally, the yield of isolated tRNA was determined by NanoDrop™ measurement and purity was analyzed using the Agilent 4200 TapeStation according to the manufacturer's protocol.

6.2.3.5 RNA dethiolation

The protocol used for dethiolation of RNA was based on the protocols published by Watanabe *et al.*.[322] 225 µg of RNA extracted from the Keio parent strain was diluted in dethiolation buffer to a final concentration of 450 ng/µL and hydrogen peroxide 3% (*m/m*) was added to a final concentration of 0.1 M. The mixture was incubated in the dark at 20 ℃ and an aliquot (50 µL) was taken every 30 min. In order to quench the H_2O_2 reaction, aliquots were immediately mixed with 10 µL of a 0.5 M sodium sulfite solution and incubated for 30 min at 20 ℃ and daylight. Samples were purified by ammonium acetate ethanol precipitation and either directly prepared for LC-MS/MS analysis or first irradiated as described in 6.2.5.2 and then prepared for LC-MS/MS analysis.

6.2.4 DNA sample preparation

6.2.4.1 DNA extraction

DNA was extracted by using TRI Reagent® according to the manufacturer's protocol. Harvested cell pellets (*E. coli* or eukaryotic cells) were homogenized in an appropriate volume of TRI Reagent® by vortexing and 1 mL of TRI Reagent® solution was mixed with 200 µL chloroform. The solution was vortexed again, incubated at room temperature for 2 min and centrifuged for phase separation (13000 rpm, 4 ℃, 10 min). The aqueous phase was removed and DNA precipitated from the residual interphase and organic phase. Therefore, 300 µL ethanol (99.9%, *V/V*) were added, sample tubes inverted for mixing and incubated at room temperature for 3 min. After centrifugation (2000 g, 4 ℃, 5 min) the supernatant was removed and the DNA pellet washed twice with 1 mL of an ethanolic (10%) trisodium citrate solution (0.1 M) and incubated for 30 min at room temperature under occasional shaking, respectively. Each washing step was followed by

centrifugation (2000 g, 4 ℃, 5 min) and removal of the washing solution. The DNA pellet obtained after the second washing step was air-dried at room temperature for 15 min and dissolved in 500 µL Milli-Q water. Insoluble residues were removed after another centrifugation step (12000 rpm, 4 ℃, 10 min) by transferring the supernatant into a new tube. In a final step, the DNA was purified by ammonium acetate ethanol precipitation and its concentration determined by NanoDrop™ measurement.

6.2.5 UV irradiation of different samples *in vitro*

6.2.5.1 Irradiation with a UV handlamp

Initial irradiation experiments, including irradiation of *in vitro* transcribed RNA, commercial total tRNA as well as oligonucleotides (MH527, MH553, MH554 and MH1213) and nucleosides/nucleotides, were performed with a UV handlamp (CAMAG™ UV Lamp 4) at 254 nm for 2 h.

In vitro transcribed RNA, commercial tRNA as well as oligonucleotides were prepared as solutions of 70 ng/µL in a volume of 100 µL, irradiated in cuvettes and precipitated after the irradiation procedure to facilitate subsequent preparation for LC-MS/MS analysis.

In the case of nucleoside irradiation, solutions of cytidine alone (1 µM or 10 µM), uridine alone (1 µM or 10 µM) and a combination of cytidine and uridine (1 µM or 10 µM, respectively) were treated according to the described procedure. Furthermore, a mixture of all four main nucleosides cytidine, uridine, guanosine and adenosine (4 mM, respectively) was irradiated in cuvettes (25 ℃) or unsealed Eppendorf tubes (45 ℃). Samples were then diluted and an amount of 10 pmol of theoretically formed dimer (assuming a quantitative reaction) was analyzed by LC-MS/MS.

For nucleotide irradiation, *in vitro* transcribed RNA was hydrolyzed to nucleotide level by using the two-step protocol described below to avoid digestion to nucleoside level in the first step. The nucleotide mixture was then irradiated and prepared for LC-MS/MS analysis (digestion to nucleosides).

6.2.5.2 Irradiation with a UV crosslinker

For better reproducibility and in order to guarantee similar treatment conditions, further irradiation experiments were performed in a UV Stratalinker 1800 crosslinker employing the energy mode, i.e. the samples were irradiated at a wavelength of 254 nm until the built-in energy-sensor had absorbed the programmed UV energy dose. To achieve maximal conformance with this energy dose, the samples were placed at approximately the same height in the device. Since the maximum UV energy dose to be programmed was 1 J/cm^2, multiple cycles had to be applied for higher doses. Upturned vial caps served

as sample vessels and contained a solution of 5 μg of the nucleic acid sample of interest (including RNA from diverse biological backgrounds as well as oligonucleotides (MH272 and MH572)) in 50 μL Milli-Q water. Due to the heat generated during irradiation, water evaporated from the sample solution was replaced after 5 irradiation cycles. Once the irradiation was finished, samples were transferred to Eppendorf tubes. If necessary, the initial volume was restored and the concentration of the samples determined by Nano-Drop™ measurement.

6.2.6 Nucleoside analysis by LC-MS/MS

6.2.6.1 Sample preparation

Prior to the LC-MS/MS analysis nucleic acid samples, including commercial tRNA as well as DNA and RNA oligonucleotides but also DNA and RNA extracted from *E. coli* or HEK cells with different biological backgrounds were hydrolyzed to nucleoside level. Untreated samples served as control samples. Usually, a one pot digestion protocol was applied, except for the experiment regarding formation of candidate 470 from IVT hydrolyzed to nucleotide level. Hydrolyzed samples were either directly subjected to LC-MS/MS measurement or stored at −20 ℃.

One pot digestion

Up to 10 μg of the nucleic acid sample of interest were hydrolyzed to nucleoside level by incubation with a mixture of 0.6 U nuclease P1, 0.2 U snake venom phosphodiesterase, 2 U fast alkaline phosphatase (Fast AP), 10 U benzonase, 200 ng pentostatin and 500 ng tetrahydrouridine in 25 mM ammonium acetate (pH 7.5) overnight at 37 ℃.

Two step digestion

7 μg of *in vitro* transcribed RNA were first hydrolyzed to nucleotide level by incubation with 0.3 U nuclease P1, 0.1 U snake venom phosphodiesterase, 200 ng Pentostatin and 500 ng tetrahydrouridine in 1/10 volume of 10x nuclease NP1 buffer (0.25 M NH_4OAc, pH 5.0, 0.2 mM $ZnCl_2$) for 4 h at 37 ℃. After irradiation (see 6.2.5.1), the nucleotide mixture was hydrolyzed to nucleoside level by incubation with 1/10 volume of 10x Fast AP buffer (0.5 M NH_4OAc, pH 9.0) and 2 U Fast AP for 5 h at 37 ℃.

6.2.6.2 LC-MS/MS analysis

After enzymatic hydrolysis, 1-10 μg of the nucleic acid sample (depending on the type of measurement, see below) were subjected to LC-MS/MS analysis. Measurements were performed an an Agilent 1260 infinity (II) series HPLC coupled to an Agilent 6470 QQQ mass spectrometer which was interfaced with an ESI source. The mixture was separated on a Synergi Fusion RP-C18 column (250 × 2.0 mm, 4 μM, 80 Å; Phenomenex, Germany)

at a flow of 0.35 mL/min and a temperature of 35 ℃ with 5 mM ammonium acetate buffer (pH 5.3; solvent A) and LC-MS grade acetonitrile (solvent B) as solvents. The LC gradient started with 100% solvent A and solvent B was linearly increased to 10% at 20 min. After another linear increase to 100% solvent B at 50 min (hold for 3 min), the initial conditions were restored within 7 min. Finally, the column was re-equilibrated for 5 min prior to the next injection. The mass spectrometer was consistently operated in the positive ion mode and ESI parameters were as follows: gas temperature 300 ℃, gas flow 7 L/min, nebulizer 60 psi, sheath gas temperature 400 ℃, sheath gas flow 12 L/min, capillary voltage 3000 V.

For relative quantification, the Agilet MassHunter Qualitative Analysis Software (version 10.0) was used to extract ion chromatograms. Peaks corresponding to the respective compound were integrated, yielding the area under the curve. While modified nucleosides were detected mass spectrometrically, main nucleosides were monitored with a diode array detector (DAD) at 254 nm (UV trace) and the UV signal of A (or dA, respectively) was used for normalization.

Neutral loss scan

Neutral loss scans were used for qualitative analysis and investigation of mass shifts as a consequence of stable isotope labeling. For this type of analysis, 10 µg of digested RNA or DNA sample were injected into the LC-MS system. The Agilent MassHunter software (version 10.0) was employed in the NLS mode, detecting either the loss of 132 Da in the case of ribonucleosides (scan range m/z = 300-500) or 116 Da (scan range m/z = 400-600) in the case of 2′-deoxyribonucleosides. For ^{13}C-labeled DNA, the neutral loss was adjusted to 121 Da. Importantly, since the ribose moiety is not affected by ^{15}N-labeling, in this case no adjustments were required. Method settings were as follows: fragmentor voltage 80 V. collision energy 12 eV, cell accelerator voltage 4 V, scan time 555 ms and mass range as indicated above.

Multiple reaction monitoring

Multiple reaction monitoring which does not require information on retention time of the compound of interest was the method of choice for the first analysis of compounds with a known mass transition, as it was the case when investigating the occurrence of a candidate 470 DNA analog or the LC-MS/MS behavior of the synthetic dimer. Therefore, 5 µg of nucleic acid sample were subjected to the LC-MS system and the mass spectrometer was employed in the multiple reaction monitoring mode with a selection of the following compounds (Table 6.2), depending on the sample. For co-injection of hydrolyzed ^{13}C-labeled tRNA with the synthetic NH-linked structure proposal for candidate 470 (kindly provided by ▇▇▇▇▇▇▇ from the ▇▇▇▇ group), 5 µg of hydrolyzed ^{13}C-labeled tRNA were spiked with 5000 fmol of the (unlabeled) synthetic substance.

Table 6.2: Settings for MRM analysis of candidate 470 and its derivatives.

Compound	Precursor ion (*m/z*)	Product ion (*m/z*)	Fragmentor voltage / V	Collision energy / eV	Cell accelerator voltage
470	470	338	80	15	4
^{13}C-470	488	351	80	15	4
d470	438	322	80	15	4

Dynamic multiple reaction monitoring

Once retention times were determined by NLS or MRM measurement, dMRM methods were applicable and usually used, due to their high sensitivity. Therefore, 1 µg of hydrolyzed RNA sample was analyzed with the mass spectrometer operating in the dMRM mode with a selection of the following compounds (Table 6.3), depending on the sample. Of note, the retention time of s^{4}U was determined by analysis of the appropriate reference substance.

Table 6.3: Settings for dMRM analysis of candidate 470 and s^{4}U.

Compound	Precursor ion (*m/z*)	Product ion (*m/z*)	RT / min	ΔRT / min	Fragmentor voltage / V	Collision energy / eV	Cell accelerator voltage
470	470	338	17.4	2	80	15	4
s^{4}U	261	129	12.1	4	90	9	4

Bibliography

[1] R. Dahm, *Human Genetics* **2008**, *122*, 565–581.

[2] F. H. Crick, *Symposia of the Society for Experimental Biology* **1958**, *12*, 138–163.

[3] F. Crick, *Nature* **1970**, *227*, 561–563.

[4] Z. Shabarova, A. Bogdanov, *Advanced Organic Chemistry of Nucleic Acids*, Wiley, **1994**.

[5] A. Rich, D. R. Davies, *Journal of the American Chemical Society* **1956**, *78*, 3548–3549.

[6] M. B. Hoagland, M. L. Stephenson, J. F. Scott, L. I. Hecht, P. C. Zamecnik, *The Journal of Biological Chemistry* **1958**, *231*, 241–257.

[7] S. Brenner, F. Jacob, M. Meselson, *Nature* **1961**, *190*, 576–581.

[8] T. A. Steitz, *Nature Reviews. Molecular Cell Biology* **2008**, *9*, 242–253.

[9] H. Dana, G. M. Chalbatani, H. Mahmoodzadeh, R. Karimloo, O. Rezaiean, A. Moradzadeh, N. Mehmandoost, F. Moazzen, A. Mazraeh, V. Marmari, M. Ebrahimi, M. M. Rashno, S. J. Abadi, E. Gharagouzlo, *International Journal of Biomedical Science : IJBS* **2017**, *13*, 48–57.

[10] L. Statello, C.-J. Guo, L.-L. Chen, M. Huarte, *Nature Reviews. Molecular Cell Biology* **2021**, *22*, 96–118.

[11] T. R. Cech, J. A. Steitz, *Cell* **2014**, *157*, 77–94.

[12] R. W. Holley, J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick, A. Zamir, *Science (New York N.Y.)* **1965**, *147*, 1462–1465.

[13] C. Lorenz, C. E. Lünse, M. Mörl, *Biomolecules* **2017**, *7*, 35.

[14] M. Helm, J. D. Alfonzo, *Chemistry & Biology* **2014**, *21*, 174–185.

[15] S. H. Kim, F. L. Suddath, G. J. Quigley, A. McPherson, J. L. Sussman, A. H. Wang, N. C. Seeman, A. Rich, *Science (New York N.Y.)* **1974**, *185*, 435–440.

[16] H. Shi, P. B. Moore, *RNA (New York N.Y.)* **2000**, *6*, 1091–105.

[17] T. Suzuki, *Nature Reviews. Molecular Cell Biology* **2021**, *22*, 375–392.

[18] E. M. Phizicky, A. K. Hopper, *Genes & Development* **2010**, *24*, 1832–1860.

[19]   M. Ibba, D. Soll, *Annual Review of Biochemistry* **2000**, *69*, 617–650.

[20]   F. H. Crick, L. Barnett, S. Brenner, R. J. Watts-Tobin, *Nature* **1961**, *192*, 1227–1232.

[21]   J. M. Goodenbour, T. Pan, *Nucleic Acids Research* **2006**, *34*, 6137–6146.

[22]   F. H. Crick, *Journal of Molecular Biology* **1966**, *19*, 548–555.

[23]   Y. Saletore, K. Meyer, J. Korlach, I. D. Vilfan, S. Jaffrey, C. E. Mason, *Genome Biology* **2012**, *13*, 175.

[24]   P. Boccaletto, F. Stefaniak, A. Ray, A. Cappannini, S. Mukherjee, E. Purta, M. Kurkowska, N. Shirvanizadeh, E. Destefanis, P. Groza, G. Avşar, A. Romitelli, P. Pir, E. Dassi, S. G. Conticello, F. Aguilo, J. M. Bujnicki, *Nucleic Acids Research* **2022**, *50*, D231–D235.

[25]   F. F. Davis, F. W. Allen, *The Journal of Biological Chemistry* **1957**, *227*, 907–915.

[26]   W. E. Cohn, *Biochimica et Biophysica Acta* **1959**, *32*, 569–571.

[27]   C. Dal Magro, P. Keller, A. Kotter, S. Werner, V. Duarte, V. Marchand, M. Ignarski, A. Freiwald, R.-U. Müller, C. Dieterich, Y. Motorin, F. Butter, M. Atta, M. Helm, *Angewandte Chemie International Edition* **2018**, *57*, 7893–7897.

[28]   B.-I. Kang, K. Miyauchi, M. Matuszewski, G. S. D'Almeida, M. A. T. Rubio, J. D. Alfonzo, K. Inoue, Y. Sakaguchi, T. Suzuki, E. Sochacka, T. Suzuki, *Nucleic Acids Research* **2017**, *45*, 2124–2136.

[29]   T. Carell, C. Brandmayr, A. Hienzsch, M. Müller, D. Pearson, V. Reiter, I. Thoma, P. Thumbs, M. Wagner, *Angewandte Chemie International Edition* **2012**, *51*, 7110–7131.

[30]   D. L. Lafontaine, D. Tollervey, *Nature Reviews. Molecular Cell Biology* **2001**, *2*, 514–520.

[31]   H. Khatter, A. G. Myasnikov, S. K. Natchiar, B. P. Klaholz, *Nature* **2015**, *520*, 640–645.

[32]   K. E. Sloan, A. S. Warda, S. Sharma, K.-D. Entian, D. L. J. Lafontaine, M. T. Bohnsack, *RNA Biology* **2017**, *14*, 1138–1152.

[33]   M. Taoka, Y. Nobe, Y. Yamaki, K. Sato, H. Ishikawa, K. Izumikawa, Y. Yamauchi, K. Hirota, H. Nakayama, N. Takahashi, T. Isobe, *Nucleic Acids Research* **2018**, *46*, 9289–9298.

[34]   R. Brimacombe, P. Mitchell, M. Osswald, K. Stade, D. Bochkariov, *FASEB journal : Official Publication of the Federation of American Societies for Experimental Biology* **1993**, *7*, 161–167.

[35]   A. Ben-Shem, N. Garreau de Loubresse, S. Melnikov, L. Jenner, G. Yusupova, M. Yusupov, *Science (New York N.Y.)* **2011**, *334*, 1524–1529.

[36] A. Bakin, B. G. Lane, J. Ofengand, *Biochemistry* **1994**, *33*, 13475–13483.

[37] L. Lestrade, M. J. Weber, *Nucleic Acids Research* **2006**, *34*, D158–62.

[38] N. J. Watkins, M. T. Bohnsack, *Wiley Interdisciplinary Reviews. RNA* **2012**, *3*, 397–414.

[39] W. A. Decatur, M. J. Fournier, *Trends in Biochemical Sciences* **2002**, *27*, 344–351.

[40] A. Baudin-Baillieu, C. Fabret, X.-h. Liang, D. Piekna-Przybylska, M. J. Fournier, J.-P. Rousset, *Nucleic Acids Research* **2009**, *37*, 7665–7677.

[41] Y. S. Polikanov, S. V. Melnikov, D. Söll, T. A. Steitz, *Nature Structural & Molecular Biology* **2015**, *22*, 342–344.

[42] X.-h. Liang, Q. Liu, M. J. Fournier, *Molecular Cell* **2007**, *28*, 965–977.

[43] H. Demirci, F. Murphy, R. Belardinelli, A. C. Kelley, V. Ramakrishnan, S. T. Gregory, A. E. Dahlberg, G. Jogl, *RNA (New York N.Y.)* **2010**, *16*, 2319–2324.

[44] F. E. Darnell, *Progress in Nucleic Acid Research and Molecular Biology* **1976**, *19*, 493–511.

[45] F. Mignone, C. Gissi, S. Liuni, G. Pesole, *Genome Biology* **2002**, *3*, reviews0004.1.

[46] S. Nachtergaele, C. He, *Annual Review of Genetics* **2018**, *52*, 349–372.

[47] G. Jia, Y. Fu, X. Zhao, Q. Dai, G. Zheng, Y. Yang, C. Yi, T. Lindahl, T. Pan, Y.-G. Yang, C. He, *Nature Chemical Biology* **2011**, *7*, 885–887.

[48] G. Zheng, J. A. Dahl, Y. Niu, P. Fedorcsak, C.-M. Huang, C. J. Li, C. B. Vågbø, Y. Shi, W.-L. Wang, S.-H. Song, Z. Lu, R. P. G. Bosmans, Q. Dai, Y.-J. Hao, X. Yang, W.-M. Zhao, W.-M. Tong, X.-J. Wang, F. Bogdan, K. Furu, Y. Fu, G. Jia, X. Zhao, J. Liu, H. E. Krokan, A. Klungland, Y.-G. Yang, C. He, *Molecular Cell* **2013**, *49*, 18–29.

[49] D. Dominissini, S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, S. Osenberg, K. Cesarkas, J. Jacob-Hirsch, N. Amariglio, M. Kupiec, R. Sorek, G. Rechavi, *Nature* **2012**, *485*, 201–206.

[50] K. D. Meyer, Y. Saletore, P. Zumbo, O. Elemento, C. E. Mason, S. R. Jaffrey, *Cell* **2012**, *149*, 1635–1646.

[51] R. P. Perry, D. E. Kelley, K. Friderici, F. Rottman, *Cell* **1975**, *4*, 387–394.

[52] B. Slobodin, R. Han, V. Calderone, J. A. F. O. Vrielink, F. Loayza-Puch, R. Elkon, R. Agami, *Cell* **2017**, *169*, 326–337.e12.

[53] H. Huang, H. Weng, K. Zhou, T. Wu, B. S. Zhao, M. Sun, Z. Chen, X. Deng, G. Xiao, F. Auer, L. Klemm, H. Wu, Z. Zuo, X. Qin, Y. Dong, Y. Zhou, H. Qin, S. Tao, J. Du, J. Liu, Z. Lu, H. Yin, A. Mesquita, C. L. Yuan, Y.-C. Hu, W. Sun, R. Su, L. Dong, C. Shen, C. Li, Y. Qing, X. Jiang, X. Wu, M. Sun, J.-L. Guan, L. Qu, M. Wei, M. Müschen, G. Huang, C. He, J. Yang, J. Chen, *Nature* **2019**, *567*, 414–419.

[54] S. Zaccara, R. J. Ries, S. R. Jaffrey, *Nature Reviews. Molecular Cell Biology* **2019**, *20*, 608–624.

[55] W. Xiao, S. Adhikari, U. Dahal, Y.-S. Chen, Y.-J. Hao, B.-F. Sun, H.-Y. Sun, A. Li, X.-L. Ping, W.-Y. Lai, X. Wang, H.-L. Ma, C.-M. Huang, Y. Yang, N. Huang, G.-B. Jiang, H.-L. Wang, Q. Zhou, X.-J. Wang, Y.-L. Zhao, Y.-G. Yang, *Molecular Cell* **2016**, *61*, 507–519.

[56] K. I. Zhou, H. Shi, R. Lyu, A. C. Wylder, Ż. Matuszek, J. N. Pan, C. He, M. Parisien, T. Pan, *Molecular Cell* **2019**, *76*, 70–81.e9.

[57] I. U. Haussmann, Z. Bodi, E. Sanchez-Moran, N. P. Mongan, N. Archer, R. G. Fray, M. Soller, *Nature* **2016**, *540*, 301–304.

[58] I. A. Roundtree, G.-Z. Luo, Z. Zhang, X. Wang, T. Zhou, Y. Cui, J. Sha, X. Huang, L. Guerrero, P. Xie, E. He, B. Shen, C. He, *eLife* **2017**, *6*, e31311.

[59] T. Lence, J. Akhtar, M. Bayer, K. Schmid, L. Spindler, C. H. Ho, N. Kreim, M. A. Andrade-Navarro, B. Poeck, M. Helm, J.-Y. Roignant, *Nature* **2016**, *540*, 242–247.

[60] S. Sommer, U. Lavi, J. E. Darnell, *Journal of Molecular Biology* **1978**, *124*, 487–499.

[61] X. Wang, Z. Lu, A. Gomez, G. C. Hon, Y. Yue, D. Han, Y. Fu, M. Parisien, Q. Dai, G. Jia, B. Ren, T. Pan, C. He, *Nature* **2014**, *505*, 117–120.

[62] X. Wang, B. S. Zhao, I. A. Roundtree, Z. Lu, D. Han, H. Ma, X. Weng, K. Chen, H. Shi, C. He, *Cell* **2015**, *161*, 1388–1399.

[63] R. J. Jackson, C. U. T. Hellen, T. V. Pestova, *Nature Reviews. Molecular Cell Biology* **2010**, *11*, 113–127.

[64] K. D. Meyer, D. P. Patil, J. Zhou, A. Zinoviev, M. A. Skabkin, O. Elemento, T. V. Pestova, S.-B. Qian, S. R. Jaffrey, *Cell* **2015**, *163*, 999–1010.

[65] S. Lin, J. Choe, P. Du, R. Triboulet, R. I. Gregory, *Molecular Cell* **2016**, *62*, 335–345.

[66] I. Barbieri, T. Kouzarides, *Nature Reviews. Cancer* **2020**, *20*, 303–322.

[67] A. Paramasivam, J. V. Priyadharsini, S. Raghunandhakumar, *Cellular & Molecular Immunology* **2020**, *17*, 550–551.

[68] M. Han, Z. Liu, Y. Xu, X. Liu, D. Wang, F. Li, Y. Wang, J. Bi, *Frontiers in Neuroscience* **2020**, *14*, 98.

[69]   P. J. Batista, *Genomics Proteomics & Bioinformatics* **2017**, *15*, 154–163.

[70]   C. Yang, Y. Hu, B. Zhou, Y. Bao, Z. Li, C. Gong, H. Yang, S. Wang, Y. Xiao, *Cell Death & Disease* **2020**, *11*, 960.

[71]   T. Wang, S. Kong, M. Tao, S. Ju, *Molecular Cancer* **2020**, *19*, 88.

[72]   C. M. Wei, A. Gershowitz, B. Moss, *Cell* **1975**, *4*, 379–386.

[73]   J. M. Adams, S. Cory, *Nature* **1975**, *255*, 28–33.

[74]   J. Mauer, X. Luo, A. Blanjoie, X. Jiao, A. V. Grozhik, D. P. Patil, B. Linder, B. F. Pickering, J.-J. Vasseur, Q. Chen, S. S. Gross, O. Elemento, F. Debart, M. Kiledjian, S. R. Jaffrey, *Nature* **2017**, *541*, 371–375.

[75]   N. Sonenberg, A. G. Hinnebusch, *Cell* **2009**, *136*, 731–745.

[76]   P. J. Sikorski, M. Warminski, D. Kubacka, T. Ratajczak, D. Nowis, J. Kowalska, J. Jemielity, *Nucleic Acids Research* **2020**, *48*, 1607–1626.

[77]   S. Schwartz, D. A. Bernstein, M. R. Mumbach, M. Jovanovic, R. H. Herbst, B. X. León-Ricardo, J. M. Engreitz, M. Guttman, R. Satija, E. S. Lander, G. Fink, A. Regev, *Cell* **2014**, *159*, 148–162.

[78]   D. Arango, D. Sturgill, N. Alhusaini, A. A. Dillman, T. J. Sweet, G. Hanson, M. Hosogane, W. R. Sinclair, K. K. Nanan, M. D. Mandler, S. D. Fox, T. T. Zengeya, T. Andresson, J. L. Meier, J. Coller, S. Oberdoerffer, *Cell* **2018**, *175*, 1872–1886.e24.

[79]   D. Dominissini, S. Nachtergaele, S. Moshitch-Moshkovitz, E. Peer, N. Kol, M. S. Ben-Haim, Q. Dai, A. Di Segni, M. Salmon-Divon, W. C. Clark, G. Zheng, T. Pan, O. Solomon, E. Eyal, V. Hershkovitz, D. Han, L. C. Doré, N. Amariglio, G. Rechavi, C. He, *Nature* **2016**, *530*, 441–446.

[80]   J. E. Squires, H. R. Patel, M. Nousch, T. Sibbritt, D. T. Humphreys, B. J. Parker, C. M. Suter, T. Preiss, *Nucleic Acids Research* **2012**, *40*, 5023–5033.

[81]   Q. Dai, S. Moshitch-Moshkovitz, D. Han, N. Kol, N. Amariglio, G. Rechavi, D. Dominissini, C. He, *Nature Methods* **2017**, *14*, 695–698.

[82]   A. V. Grozhik, S. R. Jaffrey, *Nature Chemical Biology* **2018**, *14*, 215–225.

[83]   D. Wiener, S. Schwartz, *Nature Reviews. Genetics* **2021**, *22*, 119–131.

[84]   K. Karikó, M. Buckstein, H. Ni, D. Weissman, *Immunity* **2005**, *23*, 165–175.

[85]   K. Karikó, D. Weissman, *Current Opinion in Drug Discovery & Development* **2007**, *10*, 523–532.

[86]   I. Freund, T. Eigenbrod, M. Helm, A. H. Dalpke, *Genes* **2019**, *10*, 92.

[87]   O. Andries, S. Mc Cafferty, S. C. de Smedt, R. Weiss, N. N. Sanders, T. Kitada, *Journal of Controlled Release* **2015**, *217*, 337–344.

[88] K. D. Nance, J. L. Meier, *ACS Central Science* **2021**, *7*, 748–756.

[89] T. R. Damase, R. Sukhovershin, C. Boada, F. Taraballi, R. I. Pettigrew, J. P. Cooke, *Frontiers in Bioengineering and Biotechnology* **2021**, *9*, 628137.

[90] S. Qin, X. Tang, Y. Chen, K. Chen, N. Fan, W. Xiao, Q. Zheng, G. Li, Y. Teng, M. Wu, X. Song, *Signal Transduction and Targeted Therapy* **2022**, *7*, 166.

[91] M. A. Machnicka, A. Olchowik, H. Grosjean, J. M. Bujnicki, *RNA Biology* **2014**, *11*, 1619–1629.

[92] E. M. Phizicky, J. D. Alfonzo, *FEBS Letters* **2010**, *584*, 265–271.

[93] M. Sprinzl, K. S. Vassilenko, *Nucleic Acids Research* **2005**, *33*, D139–40.

[94] P. F. Agris, *Biochimie* **1991**, *73*, 1345–1349.

[95] P. F. Agris, F. A. P. Vendeix, W. D. Graham, *Journal of Molecular Biology* **2007**, *366*, 1–13.

[96] A. Rozov, N. Demeshkina, I. Khusainov, E. Westhof, M. Yusupov, G. Yusupova, *Nature Communications* **2016**, *7*, 10457.

[97] R. Schaffrath, S. A. Leidel, *RNA Biology* **2017**, *14*, 1209–1222.

[98] P. F. Agris, *EMBO reports* **2008**, *9*, 629–635.

[99] J. Cabello-Villegas, M. E. Winkler, E. P. Nikonowicz, *Journal of Molecular Biology* **2002**, *319*, 1015–1034.

[100] J. W. Stuart, K. M. Koshlap, R. Guenther, P. F. Agris, *Journal of Molecular Biology* **2003**, *334*, 901–918.

[101] F. Tuorto, F. Lyko, *Open Biology* **2016**, *6*, 160287.

[102] G. R. Björk, J. U. Ericson, C. E. Gustafsson, T. G. Hagervall, Y. H. Jönsson, P. M. Wikström, *Annual Review of Biochemistry* **1987**, *56*, 263–287.

[103] J. Urbonavicius, G. Stahl, J. M. B. Durand, S. N. Ben Salem, Q. Qian, P. J. Farabaugh, G. R. Björk, *RNA (New York N.Y.)* **2003**, *9*, 760–768.

[104] T. N. Lamichhane, N. H. Blewett, R. J. Maraia, *RNA (New York N.Y.)* **2011**, *17*, 1846–1857.

[105] Y. Motorin, M. Helm, *Wiley Interdisciplinary Reviews. RNA* **2022**, *13*, e1691.

[106] L. Han, E. M. Phizicky, *RNA (New York N.Y.)* **2018**, *24*, 1277–1284.

[107] M. P. Guy, E. M. Phizicky, *RNA (New York N.Y.)* **2015**, *21*, 61–74.

[108] A. G. Arimbasseri, J. Iben, F.-Y. Wei, K. Rijal, K. Tomizawa, M. Hafner, R. J. Maraia, *RNA (New York N.Y.)* **2016**, *22*, 1400–1410.

[109] L. Han, E. Marcus, S. D'Silva, E. M. Phizicky, *RNA (New York N.Y.)* **2017**, *23*, 406–419.

[110] N. Kleiber, N. Lemus-Diaz, C. Stiller, M. Heinrichs, M. M.-Q. Mai, P. Hackert, R. Richter-Dennerlein, C. Höbartner, K. E. Bohnsack, M. T. Bohnsack, *Nature Communications* **2022**, *13*, 209.

[111] J. M. Lentini, R. Bargabos, C. Chen, D. Fu, *The Journal of Biological Chemistry* **2022**, *298*, 101788.

[112] E. Schöller, J. Marks, V. Marchand, A. Bruckmann, C. A. Powell, M. Reichold, C. D. Mutti, K. Dettmer, R. Feederle, S. Hüttelmaier, M. Helm, P. Oefner, M. Minczuk, Y. Motorin, M. Hafner, G. Meister, *Molecular Cell* **2021**, *81*, 4810–4825.e12.

[113] A. Rich, U. L. RajBhandary, *Annual Review of Biochemistry* **1976**, *45*, 805–860.

[114] J. J. Dalluge, T. Hashizume, A. E. Sopchik, J. A. McCloskey, D. R. Davis, *Nucleic Acids Research* **1996**, *24*, 1073–1079.

[115] Y. Motorin, M. Helm, *Biochemistry* **2010**, *49*, 4934–4944.

[116] R. Giegé, F. Jühling, J. Pütz, P. Stadler, C. Sauter, C. Florentz, *Wiley Interdisciplinary Reviews. RNA* **2012**, *3*, 37–61.

[117] M. Charette, M. W. Gray, *IUBMB Life* **2000**, *49*, 341–351.

[118] F. Spenkuch, Y. Motorin, M. Helm, *RNA Biology* **2014**, *11*, 1540–1554.

[119] P. C. Durant, D. R. Davis, *Journal of Molecular Biology* **1999**, *285*, 115–131.

[120] F. Lecointe, O. Namy, I. Hatin, G. Simos, J.-P. Rousset, H. Grosjean, *The Journal of Biological Chemistry* **2002**, *277*, 30445–30453.

[121] Y. Motorin, M. Helm, *Wiley Interdisciplinary Reviews. RNA* **2011**, *2*, 611–631.

[122] H. Hori, *Frontiers in Genetics* **2014**, *5*, 144.

[123] A. Alexandrov, I. Chernyakov, W. Gu, S. L. Hiley, T. R. Hughes, E. J. Grayhack, E. M. Phizicky, *Molecular Cell* **2006**, *21*, 87–96.

[124] J. M. Dewe, J. M. Whipple, I. Chernyakov, L. N. Jaramillo, E. M. Phizicky, *RNA (New York N.Y.)* **2012**, *18*, 1886–1896.

[125] P. Barraud, C. Tisné, *IUBMB Life* **2019**, *71*, 1126–1140.

[126] M. Pereira, S. Francisco, A. S. Varanda, M. Santos, M. A. S. Santos, A. R. Soares, *International Journal of Molecular Sciences* **2018**, *19*, 3738.

[127] C. Fergus, D. Barnes, M. A. Alqasem, V. P. Kelly, *Nutrients* **2015**, *7*, 2897–2929.

[128] H. Kasai, Z. Oashi, F. Harada, S. Nishimura, N. J. Oppenheimer, P. F. Crain, J. G. Liehr, D. L. von Minden, J. A. McCloskey, *Biochemistry* **1975**, *14*, 4198–4208.

[129] H. M. Goodman, J. Abelson, A. Landy, S. Brenner, J. D. Smith, *Nature* **1968**, *217*, 1019–1024.

[130] F. Harada, S. Nishimura, *Biochemistry* **1972**, *11*, 301–308.

[131] J. R. Katze, B. Basile, J. A. McCloskey, *Science (New York N.Y.)* **1982**, *216*, 55–56.

[132] J. P. Reyniers, J. R. Pleasants, B. S. Wostmann, J. R. Katze, W. R. Farkas, *The Journal of Biological Chemistry* **1981**, *256*, 11591–11594.

[133] J. S. Reader, D. Metzgar, P. Schimmel, V. de Crécy-Lagard, *The Journal of Biological Chemistry* **2004**, *279*, 6280–6285.

[134] S. G. van Lanen, J. S. Reader, M. A. Swairjo, V. de Crécy-Lagard, B. Lee, D. Iwata-Reuyl, *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 4264–4269.

[135] G. Phillips, B. El Yacoubi, B. Lyons, S. Alvarez, D. Iwata-Reuyl, V. de Crécy-Lagard, *Journal of Bacteriology* **2008**, *190*, 7876–7884.

[136] R. M. McCarty, A. Somogyi, G. Lin, N. E. Jacobsen, V. Bandarian, *Biochemistry* **2009**, *48*, 3847–3852.

[137] R. M. McCarty, A. Somogyi, V. Bandarian, *Biochemistry* **2009**, *48*, 2301–2303.

[138] N. Okada, S. Nishimura, *The Journal of Biological Chemistry* **1979**, *254*, 3061–3066.

[139] N. Okada, S. Noguchi, H. Kasai, N. Shindo-Okada, T. Ohgi, T. Goto, S. Nishimura, *The Journal of Biological Chemistry* **1979**, *254*, 3067–3073.

[140] R. K. Slany, M. Bösl, H. Kersten, *Biochimie* **1994**, *76*, 389–393.

[141] Z. D. Miles, R. M. McCarty, G. Molnar, V. Bandarian, *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 7368–7372.

[142] I. Biela, N. Tidten-Luksch, F. Immekus, S. Glinca, T. X. P. Nguyen, H.-D. Gerber, A. Heine, G. Klebe, K. Reuter, *PloS One* **2013**, *8*, e64240.

[143] C. Boland, P. Hayes, I. Santa-Maria, S. Nishimura, V. P. Kelly, *The Journal of Biological Chemistry* **2009**, *284*, 18218–18227.

[144] Y.-C. Chen, V. P. Kelly, S. V. Stachura, G. A. Garcia, *RNA (New York N.Y.)* **2010**, *16*, 958–968.

[145] B. I. Patel, M. Heiss, A. Samel-Pommerencke, T. Carell, A. E. Ehrenhofer-Murray, *Biochemical and Biophysical Research Communications* **2022**, *624*, 146–150.

[146] F. Meier, B. Suter, H. Grosjean, G. Keith, E. Kubli, *The EMBO Journal* **1985**, *4*, 823–827.

[147] M. Müller, C. Legrand, F. Tuorto, V. P. Kelly, Y. Atlasi, F. Lyko, A. E. Ehrenhofer-Murray, *Nucleic Acids Research* **2019**, *47*, 3711–3727.

[148] R. C. Morris, K. G. Brown, M. S. Elliott, *Journal of Biomolecular Structure & Dynamics* **1999**, *16*, 757–774.

[149] H. Grosjean, E. Westhof, *Nucleic Acids Research* **2016**, *44*, 8020–8040.

[150] N. Manickam, K. Joshi, M. J. Bhatt, P. J. Farabaugh, *Nucleic Acids Research* **2016**, *44*, 1871–1881.

[151] S. Noguchi, Y. Nishimura, Y. Hirota, S. Nishimura, *The Journal of Biological Chemistry* **1982**, *257*, 6544–6550.

[152] F. Tuorto, C. Legrand, C. Cirzi, G. Federico, R. Liebers, M. Müller, A. E. Ehrenhofer-Murray, G. Dittmar, H.-J. Gröne, F. Lyko, *The EMBO Journal* **2018**, *37*, e99777.

[153] J. Zhang, R. Lu, Y. Zhang, Ż. Matuszek, W. Zhang, Y. Xia, T. Pan, J. Sun, *Cancers* **2020**, *12*, 628.

[154] G. Dirheimer, W. Baranowski, G. Keith, *Biochimie* **1995**, *77*, 99–103.

[155] S. Varghese, M. Cotter, F. Chevot, C. Fergus, C. Cunningham, K. H. Mills, S. J. Connon, J. M. Southern, V. P. Kelly, *Nucleic Acids Research* **2017**, *45*, 2029–2039.

[156] P. Richard, L. Kozlowski, H. Guillorit, P. Garnier, N. C. McKnight, A. Danchin, X. Manière, *PloS One* **2021**, *16*, e0253216.

[157] J. K. Hurt, S. Olgen, G. A. Garcia, *Nucleic Acids Research* **2007**, *35*, 4905–4913.

[158] M. Müller, M. Hartmann, I. Schuster, S. Bender, K. L. Thüring, M. Helm, J. R. Katze, W. Nellen, F. Lyko, A. E. Ehrenhofer-Murray, *Nucleic Acids Research* **2015**, *43*, 10952–10962.

[159] A. E. Ehrenhofer-Murray, *Biomolecules* **2017**, *7*, 14.

[160] M. Schaefer, T. Pollex, K. Hanna, F. Tuorto, M. Meusburger, M. Helm, F. Lyko, *Genes & Development* **2010**, *24*, 1590–1595.

[161] F. Tuorto, F. Herbst, N. Alerasool, S. Bender, O. Popp, G. Federico, S. Reitter, R. Liebers, G. Stoecklin, H.-J. Gröne, G. Dittmar, H. Glimm, F. Lyko, *The EMBO Journal* **2015**, *34*, 2350–2362.

[162] S. Johannsson, P. Neumann, A. Wulf, L. M. Welp, H.-D. Gerber, M. Krull, U. Diederichsen, H. Urlaub, R. Ficner, *Scientific Reports* **2018**, *8*, 8880.

[163] T. P. Jurkowski, R. Shanmugam, M. Helm, A. Jeltsch, *Biochemistry* **2012**, *51*, 4438–4444.

[164] J. E. Klaunig, *Current Pharmaceutical Design* **2018**, *24*, 4771–4778.

[165] T. Lindahl, *Nature* **1993**, *362*, 709–715.

[166] D. Fu, J. A. Calvo, L. D. Samson, *Nature Reviews. Cancer* **2012**, *12*, 104–120.

[167] J. Cadet, E. Sage, T. Douki, *Mutation Research* **2005**, *571*, 3–17.

[168] E. J. Wurtmann, S. L. Wolin, *Critical Reviews in Biochemistry and Molecular Biology* **2009**, *44*, 34–49.

[169] L. L. Yan, H. S. Zaher, *The Journal of Biological Chemistry* **2019**, *294*, 15158–15171.

[170] A. Nunomura, P. I. Moreira, R. J. Castellani, H.-G. Lee, X. Zhu, M. A. Smith, G. Perry, *Neurotoxicity Research* **2012**, *22*, 231–248.

[171] S. Blanco, S. Dietmann, J. V. Flores, S. Hussain, C. Kutter, P. Humphreys, M. Lukk, P. Lombard, L. Treps, M. Popis, S. Kellner, S. M. Hölter, L. Garrett, W. Wurst, L. Becker, T. Klopstock, H. Fuchs, V. Gailus-Durner, M. Hrabě de Angelis, R. T. Káradóttir, M. Helm, J. Ule, J. G. Gleeson, D. T. Odom, M. Frye, *The EMBO Journal* **2014**, *33*, 2020–2039.

[172] M. C. Popis, S. Blanco, M. Frye, *Current Opinion in Oncology* **2016**, *28*, 65–71.

[173] T. Hofer, C. Badouard, E. Bajak, J.-L. Ravanat, A. Mattsson, I. A. Cotgreave, *Biological Chemistry* **2005**, *386*, 333–337.

[174] Z. Huang, Y. Chen, Y. Zhang, *Journal of Biosciences* **2020**, *45*, 84.

[175] M. Pearson, H. E. Johns, *Journal of Molecular Biology* **1966**, *20*, 215–229.

[176] J. Eigner, H. Boedtker, G. Michaels, *Biochimica et Biophysica Acta* **1961**, 165–168.

[177] J. Tenhunen, *Molecular and Cellular Probes* **1989**, *3*, 391–396.

[178] B. Singer, *Virology* **1971**, *45*, 101–107.

[179] N. Miller, P. Cerutti, *Proceedings of the National Academy of Sciences of the United States of America* **1968**, *59*, 34–38.

[180] C. Sun, M. Jora, B. Solivio, P. A. Limbach, B. Addepalli, *ACS Chemical Biology* **2018**, *13*, 567–572.

[181] C. Sun, P. A. Limbach, B. Addepalli, *Biomolecules* **2020**, *10*, 1527.

[182] A. Favre, E. Hajnsdorf, K. Thiam, A. Caldeira de Araujo, *Biochimie* **1985**, *67*, 335–342.

[183] A. Favre, M. Yaniv, A. M. Michelson, *Biochemical and Biophysical Research Communications* **1969**, *37*, 266–271.

[184] X. Zou, X. Dai, K. Liu, H. Zhao, Di Song, H. Su, *The Journal of Physical Chemistry. B* **2014**, *118*, 5864–5872.

[185] P. Bommisetti, V. Bandarian, *ACS Omega* **2022**, *7*, 4011–4025.

[186] G. R. Björk, T. G. Hagervall, *EcoSal Plus* **2014**, *6*, DOI `10.1128/ecosalplus.ESP-0007-2013`.

[187] T. L. de Jager, A. E. Cockrell, S. S. Du Plessis, *Advances in Experimental Medicine and Biology* **2017**, *996*, 15–23.

[188] T. Finkel, N. J. Holbrook, *Nature* **2000**, *408*, 239–247.

[189] M. Schieber, N. S. Chandel, *Current Biology* **2014**, *24*, R453–62.

[190] N. Chondrogianni, I. Petropoulos, S. Grimm, K. Georgila, B. Catalgol, B. Friguet, T. Grune, E. S. Gonos, *Molecular Aspects of Medicine* **2014**, *35*, 1–71.

[191] C. Mylonas, D. Kouretas, *In Vivo (Athens Greece)* **1999**, *13*, 295–309.

[192] J. Barciszewski, M. Z. Barciszewska, G. Siboska, S. I. Rattan, B. F. Clark, *Molecular Biology Reports* **1999**, *26*, 231–238.

[193] X. Shan, Y. Chang, C.-l. G. Lin, *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology* **2007**, *21*, 2753–2764.

[194] M. Tanaka, P. B. Chock, E. R. Stadtman, *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 66–71.

[195] Z. Li, J. Wu, C. J. Deleo, *IUBMB Life* **2006**, *58*, 581–588.

[196] C. L. Simms, H. S. Zaher, *Cellular and Molecular Life Sciences* **2016**, *73*, 3639–3653.

[197] C. Fimognari, *Oxidative Medicine and Cellular Longevity* **2015**, *2015*, 358713.

[198] J. Liu, E. Head, A. M. Gharib, W. Yuan, R. T. Ingersoll, T. M. Hagen, C. W. Cotman, B. N. Ames, *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 2356–2361.

[199] A. Nunomura, G. Perry, M. A. Pappolla, R. Wade, K. Hirai, S. Chiba, M. A. Smith, *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience* **1999**, *19*, 1959–1964.

[200] Z. Xu, J. Huang, M. Gao, G. Guo, S. Zeng, X. Chen, X. Wang, Z. Gong, Y. Yan, *GeroScience* **2021**, *43*, 487–505.

[201] F. el Ghissassi, A. Barbin, J. Nair, H. Bartsch, *Chemical Research in Toxicology* **1995**, *8*, 278–283.

[202] F. P. Guengerich, P. P. Ghodke, *Genes and Environment : The Official Journal of the Japanese Environmental Mutagen Society* **2021**, *43*, 24.

[203] R. J. Laib, H. M. Bolt, *Toxicology* **1977**, *8*, 185–195.

[204] W. J. Krzyzosiak, J. Biernat, J. Ciesiołka, K. Gulewicz, M. Wiewiórowski, *Nucleic Acids Research* **1981**, *9*, 2841–2851.

[205] D. R. Hoffman, D. W. Marion, W. E. Cornatzer, J. A. Duerre, *The Journal of Biological Chemistry* **1980**, *255*, 10822–10827.

[206] B. Rydberg, T. Lindahl, *The EMBO Journal* **1982**, *1*, 211–216.

[207] R. Ougland, C.-M. Zhang, A. Liiv, R. F. Johansen, E. Seeberg, Y.-M. Hou, J. Remme, P. Ø. Falnes, *Molecular Cell* **2004**, *16*, 107–116.

[208] F. Drabløs, E. Feyzi, P. A. Aas, C. B. Vaagbø, B. Kavli, M. S. Bratlie, J. Peña-Diaz, M. Otterlei, G. Slupphaug, H. E. Krokan, *DNA Repair* **2004**, *3*, 1389–1407.

[209] V. F. Reichle, V. Weber, S. Kellner, *ChemBioChem : a European Journal of Chemical Biology* **2018**, *19*, 2575–2583.

[210] V. F. Reichle, D. P. Petrov, V. Weber, K. Jung, S. Kellner, *Nature Communications* **2019**, *10*, 5600.

[211] C. Borek, V. F. Reichle, S. Kellner, *ChemBioChem : a European journal of chemical biology* **2020**, *21*, 2768–2771.

[212] Y. Zhang, L. Lu, X. Li, *Experimental & Molecular Medicine* **2022**, *54*, 1601–1616.

[213] M. Helm, Y. Motorin, *Nature Reviews. Genetics* **2017**, *18*, 275–291.

[214] Y. Motorin, M. Helm, *Genes* **2019**, *10*, 35.

[215] Y. Motorin, V. Marchand, *Genes* **2021**, *12*, 278.

[216] O. Begik, J. S. Mattick, E. M. Novoa, *RNA (New York N.Y.)* **2022**, *28*, 1430–1439.

[217] S. C. Pomerantz, J. A. McCloskey, *Methods in Enzymology* **1990**, *193*, 796–824.

[218] S. Kellner, A. Ochel, K. Thüring, F. Spenkuch, J. Neumann, S. Sharma, K.-D. Entian, D. Schneider, M. Helm, *Nucleic Acids Research* **2014**, *42*, e142.

[219] D. Su, C. T. Y. Chan, C. Gu, K. S. Lim, Y. H. Chionh, M. E. McBee, B. S. Russell, I. R. Babu, T. J. Begley, P. C. Dedon, *Nature Protocols* **2014**, *9*, 828–841.

[220] K. Thüring, K. Schmid, P. Keller, M. Helm, *Methods (San Diego Calif.)* **2016**, *107*, 48–56.

[221] N. Klöcker, F. P. Weissenboeck, A. Rentmeister, *Chemical Society Reviews* **2020**, *49*, 8749–8773.

[222] P. Asadi-Atoi, P. Barraud, C. Tisne, S. Kellner, *Biological Chemistry* **2019**, *400*, 847–865.

[223] C. W. Gehrke, K. C. Kuo, R. A. McCune, K. O. Gerhardt, P. F. Agris, *Journal of Chromatography* **1982**, *230*, 297–308.

[224] C. W. Gehrke, K. C. Kuo, R. W. Zumwalt, *Journal of Chromatography* **1980**, *188*, 129–147.

[225]   C. W. Gehrke, K. C. Kuo, *Journal of Chromatography* **1989**, *471*, 3–36.

[226]   J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, *Science (New York N.Y.)* **1989**, *246*, 64–71.

[227]   P. F. Crain, *Methods in Enzymology* **1990**, *193*, 782–790.

[228]   M. Jora, K. Borland, S. Abernathy, R. Zhao, M. Kelley, S. Kellner, B. Addepalli, P. A. Limbach, *Angewandte Chemie International Edition* **2021**, *60*, 3961–3966.

[229]   K. Miyauchi, S. Kimura, T. Suzuki, *Nature Chemical Biology* **2013**, *9*, 105–111.

[230]   S. Kaiser, S. R. Byrne, G. Ammann, P. Asadi Atoi, K. Borland, R. Brecheisen, M. S. DeMott, T. Gehrke, F. Hagelskamp, M. Heiss, Y. Yoluç, L. Liu, Q. Zhang, P. C. Dedon, B. Cao, S. Kellner, *Angewandte Chemie International Edition* **2021**, *60*, 23885–23893.

[231]   Y. Yoluç, G. Ammann, P. Barraud, M. Jora, P. A. Limbach, Y. Motorin, V. Marchand, C. Tisné, K. Borland, S. Kellner, *Critical Reviews in Biochemistry and Molecular Biology* **2021**, *56*, 178–204.

[232]   Agilent Technologies, *Agilent 6400 Series Triple Quadrupole LC/MS Ssystem. Concepts Guide. The Big Picture.* **2012**.

[233]   Agilent Technologies, *Ion optics innovations for increased sensitivity in hybrid MS systems*, **2015**.

[234]   S. Kellner, J. Neumann, D. Rosenkranz, S. Lebedeva, R. F. Ketting, H. Zischler, D. Schneider, M. Helm, *Chemical Communications (Cambridge England)* **2014**, *50*, 3516–3518.

[235]   Agilent Technologies, *Agilent Ultivo Triple Quadrupole LC/MS System. Concepts Guide. The Big Picture.* **2017**.

[236]   C. Buré, W. Gobert, D. Lelièvre, A. Delmas, *Journal of Mass Spectrometry* **2001**, *36*, 1149–1155.

[237]   A. S. Abdelhameed, M. W. Attwa, H. A. Abdel-Aziz, A. A. Kadi, *Chemistry Central Journal* **2013**, *7*, 16.

[238]   M. Jora, A. P. Burns, R. L. Ross, P. A. Lobue, R. Zhao, C. M. Palumbo, P. A. Beal, B. Addepalli, P. A. Limbach, *Journal of the American Society for Mass Spectrometry* **2018**, *29*, 1745–1756.

[239]   M. Jora, D. Corcoran, G. G. Parungao, P. A. Lobue, L. F. L. Oliveira, G. Stan, B. Addepalli, P. A. Limbach, *Analytical Chemistry* **2022**, *94*, 13958–13967.

[240]   M. Meselon, F. W. Stahl, *Proceedings of the National Academy of Sciences of the United States of America* **1958**, *44*, 671–682.

[241] J. J. Dalluge, T. Hashizume, J. A. McCloskey, *Nucleic Acids Research* **1996**, *24*, 3242–3245.

[242] T. Brückl, D. Globisch, M. Wagner, M. Müller, T. Carell, *Angewandte Chemie International Edition* **2009**, *48*, 7932–7934.

[243] C. Brandmayr, M. Wagner, T. Brückl, D. Globisch, D. Pearson, A. C. Kneuttinger, V. Reiter, A. Hienzsch, S. Koch, I. Thoma, P. Thumbs, S. Michalakis, M. Müller, M. Biel, T. Carell, *Angewandte Chemie International Edition* **2012**, *51*, 11162–11165.

[244] K. Borland, J. Diesend, T. Ito-Kureha, V. Heissmeyer, C. Hammann, A. H. Buck, S. Michalakis, S. Kellner, *Genes* **2019**, *10*, 26.

[245] E. P. Nikonowicz, A. Sirr, P. Legault, F. M. Jucker, L. M. Baer, A. Pardi, *Nucleic Acids Research* **1992**, *20*, 4507–4513.

[246] C. E. Dumelin, Y. Chen, A. M. Leconte, Y. G. Chen, D. R. Liu, *Nature Chemical Biology* **2012**, *8*, 913–919.

[247] B. J. Landgraf, A. J. Arcinas, K.-H. Lee, S. J. Booker, *Journal of the American Chemical Society* **2013**, *135*, 15404–15416.

[248] S. M. Huber, P. van Delft, A. Tanpure, E. A. Miska, S. Balasubramanian, *Journal of the American Chemical Society* **2017**, *139*, 1766–1769.

[249] M. Heiss, K. Borland, Y. Yoluç, S. Kellner, *Methods in Molecular Biology (Clifton N.J.)* **2021**, *2298*, 279–306.

[250] V. F. Reichle, S. Kaiser, M. Heiss, F. Hagelskamp, K. Borland, S. Kellner, *Methods (San Diego Calif.)* **2019**, *156*, 91–101.

[251] M. Heiss, V. F. Reichle, S. Kellner, *RNA Biology* **2017**, *14*, 1260–1268.

[252] M. Heiss, F. Hagelskamp, V. Marchand, Y. Motorin, S. Kellner, *Nature Communications* **2021**, *12*, 389.

[253] Y. Yoluç, E. van de Logt, S. Kellner-Kaiser, *Genes* **2021**, *12*, 1344.

[254] Press Release. NobelPrize.org. NobelPrizeOutreach AB2023, `https://www.nobelprize.org/prizes/chemistry/2022/press-release/` (visited on 01/11/2023).

[255] J. E. Moses, A. D. Moorhouse, *Chemical Society Reviews* **2007**, *36*, 1249–1262.

[256] W. H. Binder, R. Sachsenhofer, *Macromolecular Rapid Communications* **2007**, *28*, 15–54.

[257] C. D. Hein, X.-M. Liu, D. Wang, *Pharmaceutical Research* **2008**, *25*, 2216–2230.

[258] P. Thirumurugan, D. Matosiuk, K. Jozwiak, *Chemical Reviews* **2013**, *113*, 4905–4979.

[259] N. K. Devaraj, *ACS Central Science* **2018**, *4*, 952–959.

[260] R. E. Bird, S. A. Lemmel, X. Yu, Q. A. Zhou, *Bioconjugate Chemistry* **2021**, *32*, 2457–2479.

[261] N. K. Devaraj, M. G. Finn, *Chemical Reviews* **2021**, *121*, 6697–6698.

[262] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angewandte Chemie International Edition* **2001**, *40*, 2004–2021.

[263] H. C. Kolb, K. B. Sharpless, *Drug Discovery Today* **2003**, *8*, 1128–1137.

[264] C. W. Tornøe, C. Christensen, M. Meldal, *The Journal of Organic Chemistry* **2002**, *67*, 3057–3064.

[265] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angewandte Chemie International Edition* **2002**, *41*, 2596–2599.

[266] V. Hong, S. I. Presolski, C. Ma, M. G. Finn, *Angewandte Chemie International Edition* **2009**, *48*, 9879–9883.

[267] N. Z. Fantoni, A. H. El-Sagheer, T. Brown, *Chemical Reviews* **2021**, *121*, 7122–7154.

[268] Y. Motorin, J. Burhenne, R. Teimer, K. Koynov, S. Willnow, E. Weinhold, M. Helm, *Nucleic Acids Research* **2011**, *39*, 1943–1952.

[269] D. Schulz, J. M. Holstein, A. Rentmeister, *Angewandte Chemie International Edition* **2013**, *52*, 7874–7878.

[270] J. M. Holstein, D. Schulz, A. Rentmeister, *Chemical Communications (Cambridge England)* **2014**, *50*, 4478–4481.

[271] C. Y. Jao, A. Salic, *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 15779–15784.

[272] K. Hartstock, B. S. Nilges, A. Ovcharenko, N. V. Cornelissen, N. Püllen, A.-M. Lawrence-Dörner, S. A. Leidel, A. Rentmeister, *Angewandte Chemie International Edition* **2018**, *57*, 6342–6346.

[273] S. Nainar, S. Beasley, M. Fazio, M. Kubota, N. Dai, I. R. Corrêa, R. C. Spitale, *ChemBioChem : a European Journal of Chemical Biology* **2016**, *17*, 2149–2152.

[274] H. Rao, A. A. Tanpure, A. A. Sawant, S. G. Srivatsan, *Nature Protocols* **2012**, *7*, 1097–1112.

[275] J. M. Holstein, D. Stummer, A. Rentmeister, *Protein Engineering Design & Selection : PEDS* **2015**, *28*, 179–186.

[276] F. Muttach, A. Rentmeister, *Angewandte Chemie International Edition* **2016**, *55*, 1917–1920.

[277] F. Ehret, C. Y. Zhou, S. C. Alexander, D. Zhang, N. K. Devaraj, *Molecular Pharmaceutics* **2018**, *15*, 737–742.

[278] E. Paredes, S. R. Das, *ChemBioChem : a European Journal of Chemical Biology* **2011**, *12*, 125–131.

[279] N. A. Kueck, A. Ovcharenko, K. Hartstock, N. V. Cornelissen, A. Rentmeister, *Methods in Enzymology* **2021**, *658*, 161–190.

[280] K. Hartstock, A. Ovcharenko, N. A. Kueck, P. Spacek, N. V. Cornelissen, S. Hüwel, C. Dieterich, A. Rentmeister, *bioRxiv* **2022**, DOI `10.1101/2022.03.16.484494`.

[281] N. J. Agard, J. A. Prescher, C. R. Bertozzi, *Journal of the American Chemical Society* **2004**, *126*, 15046–15047.

[282] M. L. Blackman, M. Royzen, J. M. Fox, *Journal of the American Chemical Society* **2008**, *130*, 13518–13519.

[283] N. K. Devaraj, R. Weissleder, S. A. Hilderbrand, *Bioconjugate Chemistry* **2008**, *19*, 2297–2299.

[284] J. M. Holstein, D. Stummer, A. Rentmeister, *Chemical Science* **2015**, *6*, 1362–1369.

[285] F. Muttach, N. Muthmann, D. Reichert, L. Anhäuser, A. Rentmeister, *Chemical Science* **2017**, *8*, 7947–7953.

[286] K. N. Busby, N. K. Devaraj, *Methods in Enzymology* **2020**, *641*, 373–399.

[287] S. C. Alexander, K. N. Busby, C. M. Cole, C. Y. Zhou, N. K. Devaraj, *Journal of the American Chemical Society* **2015**, *137*, 12756–12759.

[288] K. N. Busby, A. Fulzele, D. Zhang, E. J. Bennett, N. K. Devaraj, *ACS Chemical Biology* **2020**, *15*, 2247–2258.

[289] D. Zhang, C. Y. Zhou, K. N. Busby, S. C. Alexander, N. K. Devaraj, *Angewandte Chemie International Edition* **2018**, *57*, 2822–2826.

[290] D. Zhang, S. Jin, X. Piao, N. K. Devaraj, *ACS Chemical Biology* **2020**, *15*, 1773–1779.

[291] M.-L. Winz, A. Samanta, D. Benzinger, A. Jäschke, *Nucleic Acids Research* **2012**, *40*, e78.

[292] J. M. Holstein, L. Anhäuser, A. Rentmeister, *Angewandte Chemie International Edition* **2016**, *55*, 10899–10903.

[293] L. Anhäuser, A. Rentmeister, *Current Opinion in Biotechnology* **2017**, *48*, 69–76.

[294] L. Anhäuser, S. Hüwel, T. Zobel, A. Rentmeister, *Nucleic Acids Research* **2019**, *47*, e42.

[295] A. Bollu, A. Peters, A. Rentmeister, *Accounts of Chemical Research* **2022**, *55*, 1249–1261.

[296] N. Klöcker, F. P. Weissenboeck, M. van Dülmen, P. Špaček, S. Hüwel, A. Rentmeister, *Nature Chemistry* **2022**, *14*, 905–913.

[297] C. Dal Magro, PhD thesis, Helm Group, Johannes Gutenberg-Universität Mainz, **2018**.

[298] P. Keller, PhD thesis, Helm Group, Johannes Gutenberg-Universität Mainz, **2018**.

[299] L. Bessler, MA thesis, Helm Group, Johannes Gutenberg-Universität Mainz, **2018**.

[300] J. A. Moore, C. D. Poulter, *Biochemistry* **1997**, *36*, 604–614.

[301] H. C. Leung, Y. Chen, M. E. Winkler, *The Journal of Biological Chemistry* **1997**, *272*, 13073–13083.

[302] B. Esberg, H. C. Leung, H. C. Tsui, G. R. Björk, M. E. Winkler, *Journal of Bacteriology* **1999**, *181*, 7256–7265.

[303] F. Pierrel, G. R. Björk, M. Fontecave, M. Atta, *The Journal of Biological Chemistry* **2002**, *277*, 13367–13370.

[304] I. Lieberman, *The Journal of Biological Chemistry* **1956**, *222*, 765–775.

[305] G. A. O'Donovan, J. Neuhard, *Bacteriological Reviews* **1970**, *34*, 278–343.

[306] R. I. Christopherson, L. R. Finch, *European Journal of Biochemistry* **1978**, *90*, 347–358.

[307] M. Čavužić, Y. Liu, *Biomolecules* **2017**, *7*, 27.

[308] T. Ohira, K. Minowa, K. Sugiyama, S. Yamashita, Y. Sakaguchi, K. Miyauchi, R. Noguchi, A. Kaneko, I. Orita, T. Fukui, K. Tomita, T. Suzuki, *Nature* **2022**, *605*, 372–379.

[309] H.-W. Winter, C. Wentrup, *Angewandte Chemie International Edition* **1980**, *19*, 720–721.

[310] S. Lund, R. Hall, G. J. Williams, *ACS Synthetic Biology* **2019**, *8*, 232–238.

[311] L. S. Behlen, J. R. Sampson, O. C. Uhlenbeck, *Nucleic Acids Research* **1992**, *20*, 4055–4059.

[312] B. Singer, *Virology* **1971**, *45*, 101–107.

[313] A. Favre, M. Yaniv, A. M. Michelson, *Biochemical and Biophysical Research Communications* **1969**, *37*, 266–271.

[314] N. J. Leonard, D. E. Bergstrom, G. L. Tolman, *Biochemical and Biophysical Research Communications* **1971**, *44*, 1524–1530.

[315] T. V. Ramabhadran, T. Fossum, J. Jagger, *Photochemistry and Photobiology* **1976**, *23*, 315–321.

[316] V. Emilsson, A. K. Näslund, C. G. Kurland, *Nucleic Acids Research* **1992**, *20*, 4499–4505.

[317] C. T. Lauhon, W. M. Erwin, G. N. Ton, *The Journal of Biological Chemistry* **2004**, *279*, 23022–23029.

[318] S. Y. Wang, D. F. Rhoades, *Journal of the American Chemical Society* **1971**, *93*, 3779–3781.

[319] D. F. Rhoades, S. Y. Wang, *Biochemistry* **1971**, *10*, 4603–4611.

[320] G. Deboer, O. Klinghoffer, H. E. Johns, *Biochimica et Biophysica Acta* **1970**, *213*, 253–268.

[321] M. Tao, M. P. Gordon, E. W. Nester, *Biochemistry* **1966**, *5*, 4146–4152.

[322] K. Watanabe, *Biochemistry* **1980**, *19*, 5542–5549.

[323] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, H. Mori, *Molecular Systems Biology* **2006**, *2*, 2006.0008.
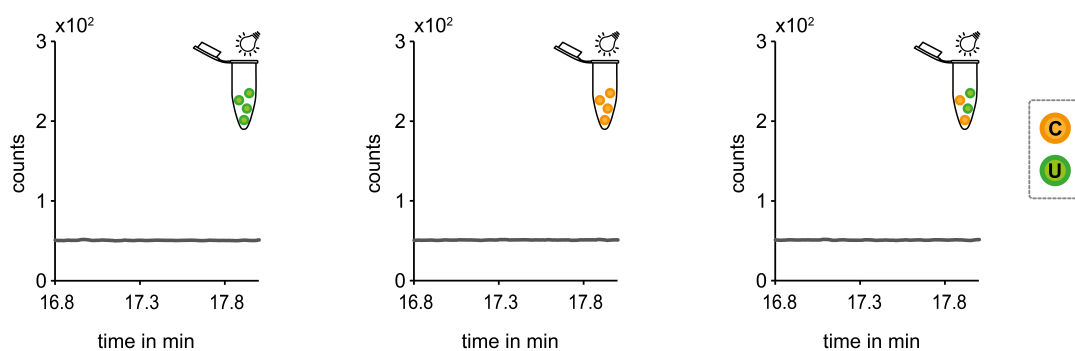
Appendix

Figure A1: Qualitative LC-MS/MS analysis of nucleoside irradiation experiments. Representative extracted ion chromatograms for the mass transition corresponding to candidate 470 (m/z 470 → 338) in LC-MS/MS analysis subsequent to irradiation of the indicated nucleoside mixtures (conditions: 10 µM, 25 ℃) with a UV handlamp at 254 nm for 2 h. The nucleosides are illustrated by different colors according to the color scheme on the right.
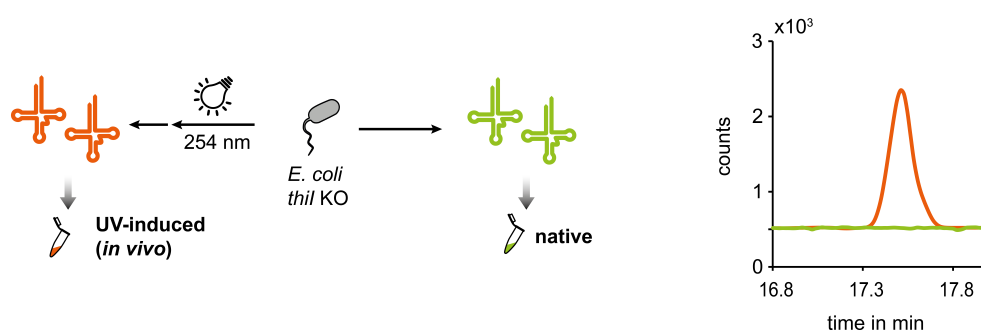

Figure A2: UV-induced formation of candidate 470 in the *thiI* KO *in vivo*. tRNA was extracted from the *thiI* KO strain that was either untreated as a control (native) or irradiated with a UV energy dose of 1 J/cm^2. Smoothed extracted ion chromatograms for the mass transition corresponding to candidate 470 (m/z 470 → 338) are shown.
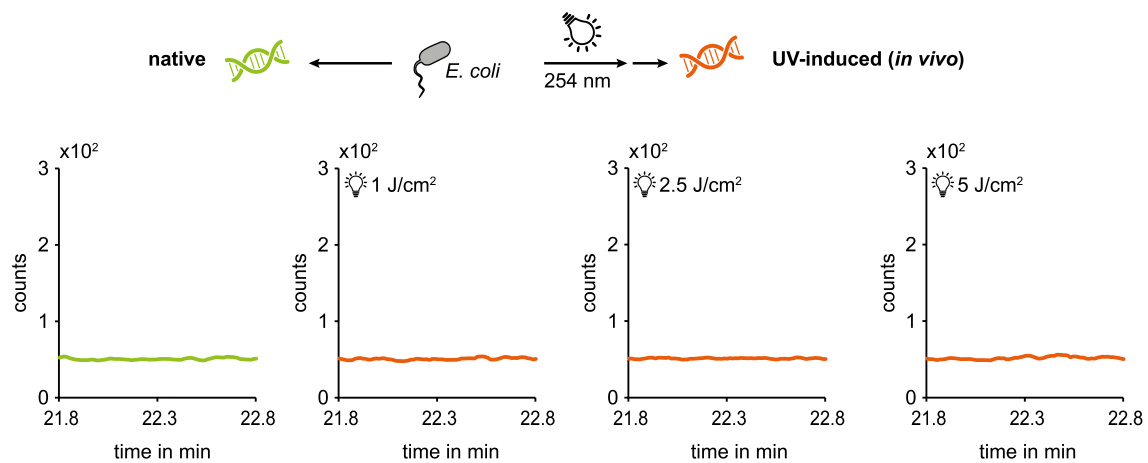
Figure A3: UV-induced formation of the candidate 470 DNA analog d470 *in vivo*. DNA was extracted from *E. coli* that were either untreated as a control (native) or irradiated with a UV energy doses of 1 J/cm^2, 2.5 J/cm^2 or 5 J/cm^2. Extracted ion chromatograms for the mass transition corresponding to the candidate 470 DNA analog d470 (*m/z* 438 → 322) are shown.

Curriculum Vitae