Method Development for the Detection and Quantification of Biological and Synthetic RNA Cap Structures by LC-MS/MS

Dissertation

zur Erlangung des Grades

“Doktor der Naturwissenschaften”

im Promotionsfach Pharmazie

am Fachbereich Chemie, Pharmazie, Geographie und Geowissenschaften

der Johannes Gutenberg-Universität Mainz

Martina Christina Krämer, geb. Schmidt-Dengler

geboren in [redacted]

Mainz, Januar 2025

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

JG|U

Dekanin: Prof. Dr. Eva Rentschler

1. Berichterstatter: [redacted]

2. Berichterstatter: [redacted]

Datum der mündlichen Prüfung: 04. Februar 2025

D77 (Dissertation Mainz)

Die vorgelegte Dissertation 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 [redacted] angefertigt.

1. Berichterstatter:

[redacted]

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

2. Berichterstatterin:

[redacted]

Fachbereich Chemie, Pharmazie 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

[redacted]

Contents

Abstract

5’ caps on messenger RNA (mRNA) play a crucial role in regulating mRNA function and processing. Apart from the canonical cap 0 structure, comprising a 7-methylguanosine (m^7G) linked to the first transcribed nucleotide by a reverse 5’-5’-triphosphate bridge, additional methylations, such as 2’*O*-methylations of the penultimate nucleotide and additional *N*6-methylation of adenosine contribute to the complex cap epitranscriptome. Based on the critical role of 5’ caps in mRNA function, this study developed a sensitive and reliable liquid chromatography tandem mass spectrometry (LC-MS/MS) approach to detect and quantify these structures, while also providing the capability to simultaneously analyze modified nucleosides. The comprehensive method development of chromatographic and mass spectrometric parameters considerably impacted signal intensity and peak separation, pushing detection limits into the attomole to femtomole range. Examination of eukaryotic RNA revealed distinct cap methylation patterns for cell and tissue samples. Investigating the knockout of the cap-modifying enzyme CMTR1 provided insights into its substrate specificity and contributed to the understanding of RNA cap methylation processes. To enhance mass spectrometry-based quantification, an accessible approach for isotope labeling of cap dinucleotides was developed. The method aimed to achieve deuteration at the C-8 position of purines through incubation in deuterium oxide at elevated temperatures. Studies with *N*7-unmethylated cap dinucleotides revealed that temperatures of at least 85 °C and extended incubation times were necessary for near-complete deuteration. However, m^7G exhibited rapid exchange at the C-8 position in contact with water, precluding stable isotope labeling of m^7G-containing cap structures. In view of the increasing relevance of capped synthetic mRNA for therapeutic applications, another part of this study assessed a novel LC-MS-based method for determining capping efficiencies. Application to co- and post-transcriptionally capped *in vitro* transcribed (IVT) RNA revealed challenges in IVT purification and quantification accuracy, highlighting the need to optimize IVT processes and purification techniques for reliable evaluation of capping reactions.

Zusammenfassung

5'-Caps von mRNA spielen eine entscheidende Rolle bei der Regulierung der Funktion und der Prozessierung von mRNA. Neben der kanonischen Cap 0-Struktur, die aus 7-Methylguanosin (m^7G) besteht, welches über eine umgekehrte 5'-5'-Triphosphatbrücke mit dem ersten transkribierten Nukleotid verbunden ist, tragen zusätzliche Methylierungen wie die 2'*O*-Methylierungen des ersten transkribierten Nukleotids und zusätzliche *N*6-Methylierung von Adenosin zum komplexen Cap-Epitranskriptom bei. Aufgrund der entscheidenden Rolle der 5'-Caps für die Funktion der mRNA wurde in dieser Arbeit eine empfindliche und zuverlässige Flüssigchromatographie gekoppelte Tandem-Massenspektrometrie (LC-MS/MS) -Methode zur Detektion und Quantifizierung dieser Strukturen entwickelt, die zudem die gleichzeitige Analyse von modifizierten Nukleosiden ermöglicht. Die umfangreiche Methoden-entwicklung der chromatographischen und massenspektrometrischen Parameter zeigte erheblichen Einfluss auf die Signalintensität und die Peaktrennung, wodurch die Nachweisgrenzen in den attomolaren bis femtomolaren Bereich verschoben wurden. Die Untersuchung von eukaryotischer RNA ergab unterschiedliche Cap-Methylierungsmuster für Zell- und Gewebeproben. Die weitere Untersuchung des Knockouts des Cap-modifizierenden Enzyms CMTR1 lieferte Einblicke in seine Substratspezifität und trug zum Verständnis der RNA-Cap-Methylierungsprozesse bei. Um die massenspektrometrische Quantifizierung zu verbessern, wurde eine leicht zugängliche Herangehensweise für die Isotopenmarkierung von Cap-Dinukleotiden entwickelt. Die Methode zielt darauf ab, eine Deuterierung an der C-8-Position von Purinen durch Inkubation in Deuteriumoxid bei erhöhten Temperaturen zu erreichen. Untersuchungen von *N*7-unmethylierten Cap-Dinukleotiden ergaben, dass Temperaturen von mindestens 85 °C und längere Inkubationszeiten für eine nahezu vollständige Deuterierung erforderlich waren. Allerdings zeigte m^7G im Kontakt mit Wasser einen schnellen Austausch an der C-8-Position, was eine stabile Isotopenmarkierung von m^7G-haltigen Cap-Strukturen verhinderte. Angesichts der zunehmenden Bedeutung von gecappter synthetischer mRNA für therapeutische Anwendungen wurde in einem weiteren Teil dieser Arbeit eine neuartige LC-MS-basierte Methode zur Bestimmung der Capping-Effizienz untersucht. Die Anwendung auf co- und posttranskriptionell gecappte *in vitro* transkribierte (IVT) RNA zeigte Herausforderungen bei der Aufreinigung und der Quantifizierungs-genauigkeit auf. Dies unterstreicht die Notwendigkeit, IVT-Prozesse und Aufreinigungs-techniken für eine zuverlässige Bewertung von Capping-Reaktionen zu optimieren.

Abbreviations and symbols

2'*O*-MTase	2'*O*-methyltransferase
3'*O*-Me-m^{7}G	3'*O*-methylated 7-methylguanosine
A	adenosine
ACN	acetonitrile
ADPr	adenosine diphosphate-ribose
ADPRC	adenosine diphosphate-ribosylcyclase
Am	2'*O*-methyladenosine
Ap$_4$A	diadenosine tetraphosphate
APS	ammonium persulfate
AQbD	analytical quality by design
ARCA	anti-reverse cap analog
ARE	AU-rich element
A_S	symmetry factor
ATP	adenosine triphosphate
B. subtilis	*Bacillus subtilis*
C	cytidine
C. elegans	*Caenorhabditis elegans*
CBC	cap-binding complex
CBP20	cap-binding protein 20
CBP80	cap-binding protein 80
cDNA	complementary DNA
CID	collision-induced dissociation
Cm	2'*O*-methylcytidine
CMTR	cap methyltransferase
CoA	coenzyme A
COVID-19	coronavirus disease 19
CTIF	CBP80-CBP20-dependent translation initiation factor
Ctrl	control
CuAAC	copper-catalyzed azide-alkyne cycloaddition
CTP	cytidine triphosphate
d	column diameter

$d_{0.05}$	distance between the leading edge of the curve and the peak maximum, measured at 5% of the peak height
D_2O	deuterium oxide
DMEM	Dulbecco's Modified Eagle Medium
dMRM	dynamic multiple reaction monitoring
DNA	deoxyribonucleic acid
DoE	design of experiments
DPBS	Dulbecco's Balanced Salt Solution
dpCoA	3'-dephospho-CoA
dsRNA	double-stranded RNA
DXO	decapping exoribonuclease
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
eGFP	enhanced green fluorescent protein
eIF	eukaryotic translation initiation factor
ESI	electrospray ionization
F	flow rate
FAD	flavin adenine dinucleotide
FastAP	fast alkaline phosphatase
FBS	fetal bovine serum
FCE	faustovirus capping enzyme
FDA	Food and Drug Administration
FTO	fat mass and obesity-associated protein
G	guanosine
GDP	guanosine diphosphate
GMP	guanosine monophosphate
GTase	guanylyltransferase
GTP	guanosine triphosphate
HEK	human embryonic kidney 293
HILIC	hydrophilic interaction liquid chromatography
His	histidine
HPLC	high performance liquid chromatography
ICH	International Council for Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use

IDL	instrument detection limit
IFIT	interferon-induced proteins with tetratricopeptide repeats
IFN	interferon
IMP	Importin
IVT	*in vitro* transcription
k.o.	knockout
l	column length
LC-MS	liquid chromatography coupled mass spectrometry
LC-MS/MS	liquid chromatography coupled tandem mass spectrometry
LEGO	ligation-enabled mRNA-oligonucleotide assembly
LNA	locked nucleic acid
lncRNA	long non-coding RNA
LOD	limit of detection
LOQ	limit of quantification
m/z	mass-to-charge ratio
m^2A	2-methyladenosine
m^5C	5-methylcytidine
$m^{6,6}A$	*N*6,*N*6-dimethyladenosine
m^6A	*N*6-methyladenosine
m^6Am	*N*6,2'*O*-dimethyladenosine
m^7G	7-methylguanosine
METTL4	methyltransferase-like 4
miRNA	micro RNA
MODR	method operable design region
MRM	multiple reaction monitoring
mRNA	messenger RNA
mRNP	messenger RNP
MS	mass spectrometry
MS/MS	tandem-mass spectrometry
N	noise
n	number of replicate injections
*N*7-MTase	guanine-*N*7-methyltransferase
NAD	nicotinamide dinucleotide
NCIN	non-canonical initiating nucleotide
NELF	negative elongation factor

Nm	2'*O*-methylated nucleoside
NMR	nuclear magnetic resonance
NP1	nuclease P1
Np_nN	dinucleoside polyphosphate
Nudix	nucleoside diphosphate linked to another moiety X
nt	nucleotides
NTP	nucleoside triphosphate
OFAT	one-factor-at-a-time
PABP	poly(A)-binding protein
PAGE	polyacrylamide gel electrophoresis
PCIF1	phosphorylated CTD interacting factor 1
PCR	polymerase chain reaction
pDNA	plasmid DNA
Ph. Eur.	European Pharmacopoeia
PHAX	phosphorylated adaptor for RNA export protein
piRNA	Piwi-interacting RNA
Pol II	RNA polymerase II
Pol III	RNA polymerase III
PPAT	phosphopantetheine adenylyltransferase
PPH	pyrophosphohydrolase
PRNTase	GDP polyribonucleotidyl transferase
PRR	pattern recognition receptor
Ψ	pseudouridine
RdRp	RNA-dependent RNA polymerase
REF	RNA export factor
RIG-I	retinoic acid-inducible gene I
RNA	ribonucleic acid
RNGTT	RNA guanylyltransferase and 5'-phosphatase
RNMT	RNA guanine-7 methyltransferase
RNPs	ribonucleoproteins
RP	reversed-phase
RppH	RNA 5'-pyrophosphohydrolase
rRNA	ribosomal RNA
R_S	resolution
RSD	relative standard deviation

RTPase	RNA triphosphatase
S	analyte signal
S. cerevisiae	*Saccharomyces cerevisiae*
S. venezuelae	*Streptomyces venezuelae*
S/N	signal-to-noise ratio
SAM	*S*-adenosyl methionine
SARS-CoV-2	severe acute respiratory syndrome coronavirus type 2
SILIS	stable isotope labeled internal standard
SIM	single ion monitoring
SMN	survival motor neuron
snRNA	small nuclear RNA
snRNP	small nuclear RNP
snoRNA	small nucleolar RNA
snoRNP	small nucleolar RNP
SPAAC	strain-promoted azide-alkyne cycloaddition
SRM	single reaction monitoring
$t_{(n-1,1-\alpha=0.99)}$	Student's t value for a 99% confidence interval and n-1 degrees of freedom
TBE	tris-borate-ethylenediaminetetraacetic acid
TEMED	*N,N,N',N'*-tetramethylethylenediamine
TGS	trimethylguanosine synthase
TLC	thin layer chromatography
TMG	2,2,7-trimethylguanosine
TOF	time-of-flight
t_0	dead time
t_R	retention time
TREX	transcription export complex
tRNA	transfer RNA
TTI	total transmission ion
U	uridine
UDP	uridine diphosphate
UDP-Glc	uridine diphosphate glucose
UDP-GlcNAc	uridine diphosphate *N*-acetylglucosamine
UTP	uridine triphosphate
UTR	untranslated region

UV	ultraviolet
VCE	vaccinia capping enzyme
$w_{0.05}$	peak width at 5% of the peak height
$w_{0.5}$	peak width at half maximum
w_h	peak width at the baseline

1. Introduction

1.1. Caps in RNA

1.1.1. The mRNA cap

Messenger RNA (mRNA) plays a pivotal role in gene expression. It is generated during transcription in the nucleus, where RNA polymerase II creates a copy of a coding DNA sequence. After maturation and transport to the cytoplasm, the mRNA serves as a template for protein synthesis during translation.[1,2] The core of a eukaryotic mRNA is its coding sequence, which begins with a start codon and ends with a stop codon, and is flanked by untranslated regions (UTR) at both sides (Figure 1.1 A). A poly(A) tail is located at the 3'-end, whereas the 5'-end is equipped with a cap structure, which is involved in several important biological processes, including RNA export, translation initiation and the protection of RNA from degradation.[3]

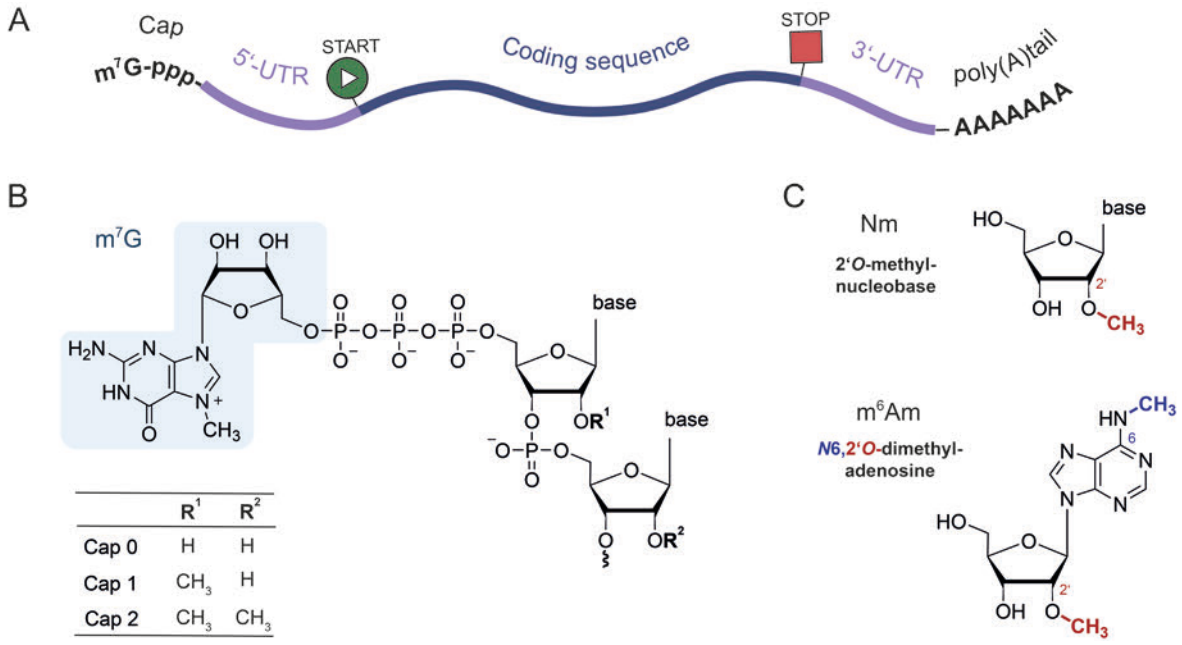

Figure 1.1. A.) Structure of a eukaryotic mRNA transcript with functional regions highlighted. B.) Depiction of a canonical cap structure with respective nomenclature regarding 2'*O*-methylation pattern at the first and second transcribed nucleotide. C.) Structures of relevant modified nucleosides, which are part of the cap structure, including general 2'*O*-methylated nucleosides (Nm) and specific *N*6,2'*O*-dimethyladenosine (m^6Am).

The canonical cap on most eukaryotic mRNAs consists of a 7-methylguanosine (m^7G) that is connected to the transcript by a 5'-5'-triphosphate linkage. While lower eukaryotes, such as

yeast, predominantly contain the so-called cap 0 structure (m^7GpppN), the adjacent nucleotides can undergo 2'*O*-methylation in higher eukaryotes, resulting in the formation of cap 1 structures, when only the first transcribed nucleotide is ribose-methylated (m^7GpppNm) or cap 2 structures, when the first and the second transcribed nucleotide are ribose-methylated (m^7GpppNmpNm).[4–6] If the first transcribed nucleotide consists of a 2'*O*-methyladenosine (Am), additional methylation at the *N*6-position can occur, leading to *N*6,2'*O*-dimethyl-adenosine (m^6Am) formation (Figure 1.1 B,C).[7]

However, not all cap structures follow this standard pattern. For instance, deviating cap structures have been identified in insect oocyte mRNA, where non-methylated guanosines (GpppN caps) are described.[8] Additional cap variants, including hypermethylation at the cap guanosine and single methylations of the triphosphate, which are observed in other RNA species, and metabolite caps, will be detailed in the following sections.

1.1.2. Caps in various RNA species

In addition to mRNA, the presence of cap structures has been observed in a variety of RNA species, exhibiting diverse forms of caps. These structures are predominantly found in RNA polymerase II (Pol II) transcripts, including pre-mRNAs, pre-micro RNAs (miRNA), pre-long non-coding RNAs (lncRNA), small nucleolar RNAs (snoRNAs) and small nuclear RNAs (snRNA).[9] Additionally, some RNA polymerase III (Pol III) transcripts have also been shown to possess cap structures.

Long non-coding RNAs are mainly transcribed by Pol II. They often possess m^7G caps and poly(A) tails and are spliced similarly to mRNA, although they display distinctive characteristics with regard to expression and extent of nuclear localization according to their functional roles.[10–12] Also, several miRNA precursors have been described to be m^7G-capped.[13,14] Xie and colleagues presented evidence for a non-canonical, microprocessor-independent biogenesis pathway, in which m^7G-capped pre-miRNAs are directly transported to the cytoplasm, where Dicer processing occurs to yield uncapped mature miRNAs.[15,16] Similarly, precursor transcripts of Piwi-interacting RNA (piRNA) carry an m^7G cap that is endonucleolytically cleaved to generate monophosphorylated ends required for PIWI protein binding.[17] In *Caenorhabditis elegans* (*C. elegans*), a trimer of Schlafen-like-domain proteins called PUCH has recently been identified to perform piRNA precursor cleavage, thereby strictly requiring the m^7G cap for activity.[18]

Another cap variant is the hypermethylated 2,2,7-trimethylguanosine (TMG) cap. Most snRNAs are transcribed by Pol II and acquire an m^7G-cap in the nucleus.[19] After nuclear export in the cytoplasm, assembly of the snRNA and Sm proteins is mediated by the survival motor neuron (SMN) protein complex to form core ribonucleoproteins (RNPs).[20] Following maturation events comprise hypermethylation of the 5' cap by the trimethylguanosine synthase 1 (TGS) protein to yield a TMG cap.[21] Import receptors that recognize the TMG cap re-import snRNPs into the nucleus, where spliceosome assembly and pre mRNA splicing events can take place.[22–24] snoRNAs play crucial roles in ribosomal RNA (rRNA) processing and modification in the nucleolus, including 2′*O*-methylation and pseudouridylation.[25–27] snoRNAs generated by Pol II transcription possess a TMG cap. In contrast to snRNAs, which undergo cap hypermethylation in the cytoplasm, hypermethylation of snoRNAs occurs normally in the nucleoplasm, before transport to the nucleolus.[28,29] It is currently thought that the export of snoRNAs is disabled by the combination of assembly with core snoRNP proteins and the hypermethylation of the cap.[30] In a recent study, Ohira and Suzuki demonstrated that certain precursors of transfer RNA (tRNA) contain methylated cap structures, despite being Pol III transcripts, which typically does not produce capped transcripts. Pre-tRNA capping occurs subsequent to transcription during tRNA maturation, resulting in m^7G-caps and subsequent hypermethylated variants and plays a role in protecting pre-tRNAs from 5'-exonucleolytic degradation.[31]

The variety of cap structures is also exemplified by the minimalistic cap-like structure that encompasses a 5' γ-monomethyl phosphate. This structure has been described in small nuclear snRNAs that are transcribed by Pol III, such as mammalian U6 and plant U3,[32,33] as well as in small non-coding RNAs, including B2 and 7SK.[34] While the cap structure was demonstrated to enhance the stability of RNAs in *Xenopus* oocytes, it was not identified as a requirement for transport across the nuclear membrane.[35] Furthermore, the monomethyl cap has been observed to reduce the affinity of Pol III transcripts for the La protein, which may facilitate the release of these RNA transcripts.[36]

1.1.3. Metabolite caps

A variety of non-canonical caps have recently been described. These originate from coenzymes such as nicotinamide dinucleotide (NAD), coenzyme A (CoA), and flavin adenine dinucleotide (FAD), from metabolite alarmones including dinucleoside polyphosphates, and from nucleotide sugars like uridine diphosphate (UDP) -glucose. The chemical structures are depicted in Figure 1.2.

In 2009, NAD-capped RNA was detected in *Escherichia coli* (*E. coli*) and *Streptomyces venezuelae* (*S. venezuelae*) using liquid chromatography coupled mass spectrometry (LC-MS).[37] Six years later, the development of NAD captureSeq allowed the sequencing-based identification of NAD-linked RNA.[38] The approach is based on transglycosylation of NAD with a clickable alkynyl group using adenosine diphosphate-ribosylcyclase (ADPRC), followed by biotinylation, streptavidin pull down for enrichment and subsequent reverse transcription and next generation sequencing.[39] NAD captureSeq was the basis for several other sequencing methods for the analysis of NAD-RNA, incorporating modified protocols such as a different tagging strategy to also enable the use of Oxford Nanopore technology,[40] exchanging copper-catalyzed azide-alkyne cycloaddition (CuAAC) by strain-promoted azide-alkyne cycloaddition (SPAAC) click chemistry to avoid RNA degradation by copper ions,[41,42] or additional depletion of m^7G-capped RNA to avoid false positive hits due to low activity of ADPRC towards m^7G-RNA.[42,43] An alternative sequencing method, called CapZyme-seq uses decapping enzymes and ligates the resulting 5'-monophosphate RNA to a barcoded RNA adaptor.[44] Further detection and quantification methods are based on mass spectrometry (MS),[37,45] boronate affinity electrophoresis,[46] or colorimetry.[47] These detection methods revealed the occurrence of NAD-capped RNAs in a number of different organisms, including bacteria,[37,38,48,49] archaea,[50] and eukaryotes.[45,51–53] The bacterial RNA polymerase as well as the eukaryotic Pol II can use NAD as non-canonical initiating nucleotide (NCIN) at promoters that bear an adenosine at the +1 position.[48,54,55] Furthermore, the presence of NAD-caps on mammalian snoRNAs, which are predominantly processed from intronic regions, indicates the potential for an alternative post-transcriptional NAD-capping mechanism.[52] A subset of decapping enzymes has also been identified. The nucleoside diphosphate linked to another moiety X (Nudix) hydrolase termed NudC hydrolyzes the pyrophosphate bond of NAD-RNA in *E. coli*, leaving nicotinamide mononucleotide and 5'-adenosine monophosphate RNA,[56] while the RNA 5'-pyrophosphohydrolase (RppH) was found to decap NAD-RNAs in *Bacillus subtilis* (*B. subtilis*).[48] In eukaryotes, the decapping enzymes DXO/Rai1 remove the NAD cap (deNADding), which is followed by RNA degradation.[52] Recently, Rat1 and Xrn1 from yeast have also been shown to exhibit deNADding activity.[57] Additionally, an alternative pathway in eukaryotes involves the CD38 glycoprotein, which converts NAD-capped RNA to adenosine diphosphate-ribose (ADPr)-capped RNA by removal of nicotinamide.[58] NAD-RNA was shown to influence several cellular processes and gene regulation. In prokaryotes, the NAD cap protects bacterial transcripts from 5'-degradation.[38,48] Additional roles include regulation of toxin expression in *Staphylococcus aureus*[49] and growth phase-dependent regulation in *E. coli* and *B. subtilis*.[41,48] Also, NAD-RNA is involved in the infection process of *E. coli* by the T4

bacteriophage, where the phage-encoded ADP-ribosyltransferase ModB uses NAD-RNA as a substrate to RNAylate proteins.[24,59] In contrast to the stabilizing role of NAD-RNA in prokaryotes, the cap promotes RNA decay by the deNADding enzyme DXO in eukaryotes.[52] With regard to the translation process, it was shown that NAD-capped mRNAs are not translated in mammalian cells[52] and budding yeast,[60] whereas studies in *Arabidopsis* showed association of NAD-RNAs with translating ribosomes.[53] In addition, transcripts in mitochondria are NAD-capped. Remarkably, mitochondrial RNA polymerases demonstrate high efficient NAD capping with efficiencies of up to 60% in yeast.[54] The deNADding enzyme Xrn1 was recently shown to modulate mitochondrial NAD-RNA levels, thereby additionally displaying an interconnection between NAD and NAD-RNA levels.[57,61] Overall, NAD-RNA is involved in a multitude of biological functions across different organisms, and ongoing research aims to uncover its full range of roles and interactions.[24,61]

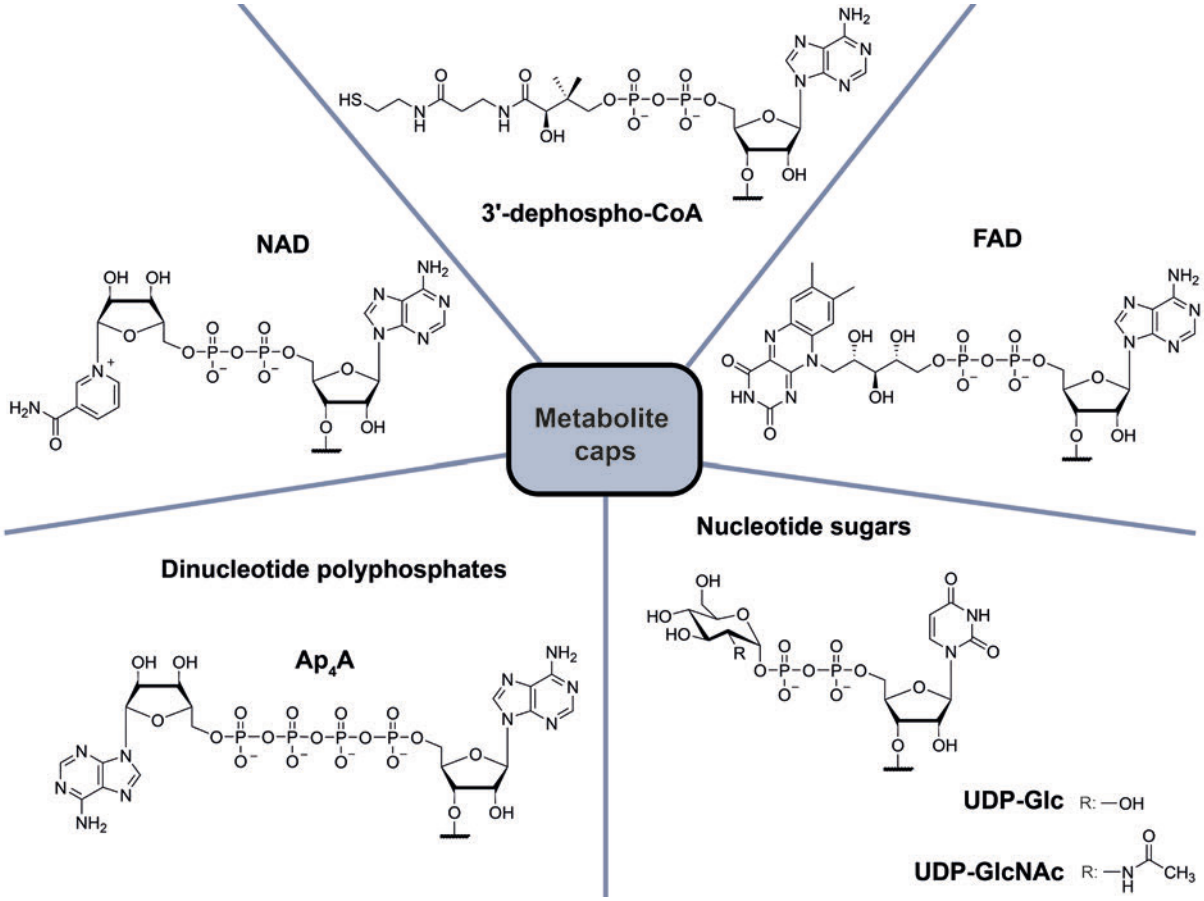

Figure 1.2. Overview of several metabolite caps, including 3'-dephospho-coenzyme A- (dpCoA), flavin adenine dinucleotide- (FAD), uridine diphosphate glucose- (UDP-Glc) and uridine diphosphate *N*-acetylglucosamine- (UDP-GlcNAc), diadenosine tetraphosphate- (Ap_4A), and nicotinamide dinucleotide- (NAD) capped RNA.

Kowtuniok and coworkers identified 3'-dephospho-CoA (dpCoA) and its succinyl- and acetyl-thioester derivates linked to the 5'-terminus of *E. coli* and *S. venezuelae* RNA.[62] Later, dpCoA

caps were also identified and quantified in the small RNA fraction of *S. aureus* with quantities of up to 14 fmol/µg RNA.[63] While other metabolite caps are only introduced as NCIN to initiate transcription, the *E. coli* enzyme phosphopantetheine adenylyltransferase (PPAT), an enzyme, which is involved in the CoA biosynthetic pathway,[64] has been demonstrated to accept RNA transcripts as a substrate, resulting in the formation of CoA-RNA. This provides a potential post-transcriptional mechanism for the generation of CoA-RNAs in bacteria.[65] However, reliable sequencing methods to further investigate and validate CoA-RNA transcripts are still required to explore the occurrence and functions of this metabolite cap structure.[66]

Recently, Julius and Yuzenkova demonstrated that bacterial RNA polymerase can utilize various nucleotide metabolites as non-canonical initiating substrates for transcription *in vitro*, including FAD, UDP-glucose, and UDP-*N*-acetylglucosamine.[67] Liquid chromatography coupled tandem mass spectrometry (LC-MS/MS) analysis demonstrated the existence of these caps in bacteria, eukaryotic species, as well as in viral RNA.[45] The colorimetric quantification method FAD capQ indicated the appearance of FAD caps on short RNAs in human cells, but the biological role remains elusive.[68] Most recently, hepatitis C virus RNA has been demonstrated to be 5'-FAD capped with a high capping frequency of approximately 75%, which protects the RNA from retinoic acid-inducible gene I (RIG-I) mediated innate immune response.[24] The latter demonstrates a new viral RNA-capping strategy utilizing cellular metabolites.[69]

Dinucleoside polyphosphates (Np_nN) are present in all domains of life and have been shown to increase under stress conditions, leading to their designation as “alarmones”.[70] They were first discovered as 5’ RNA caps in bacteria in two different studies using LC-MS and boronate gel electrophoresis, respectively.[71,72] The structural variation is demonstrated through the variety of polyphosphate bridge lengths, ranging from 3 to 5 phosphates, and additional partial methylation of the nucleosides. These methylations lead to protection from cleavage by the decapping enzyme RppH and were more abundant under stress conditions, such as starvation.[71] Additionally, capped transcripts exhibited elevated levels during exposure to disulfide stress,[72] suggesting a potential role in stabilizing capped transcripts under stress conditions.[24] In mammalian cells, diadenosine tetraphosphate (Ap_4A)-capped RNA was discovered. The precise biological function of this structure remains uncertain, as the capped RNA was not translated. However, the absence of an innate immune response following Ap_4A-RNA transfection suggests that it is a naturally occurring element within the transcriptome.[66] Concludingly, the exact mechanisms and specific functions of Np_nN-capped RNAs have not yet been entirely elucidated.

1.2. Functions of the mRNA cap

The cap structure in mRNAs fulfils several important functions during the life cycle of mRNAs. Many biological processes are mediated by cap-binding proteins. Of particular note are the cap-binding complex (CBC) primary located in the nucleus, and the eukaryotic translation initiation factor (eIF) 4E in the cytoplasm.[3]

1.2.1. RNA processing and export

In the nucleus, the m^7G cap co-transcriptionally binds to CBC, a heterodimer, consisting of cap-binding protein 80 (CBP80) and 20 (CBP20).[73] This complex plays a role in pre-mRNA splicing, 3'-end processing, and mRNA export. The m^7G cap structure on pre-mRNA enhances splicing efficiency as demonstrated by *in vitro* experiments.[74,75] The essential role of CBC in mediating these effects on splicing has been shown using immunodepletion and antibody blocking techniques.[73,76] CBC recruits snRNPs, thereby promoting the spliceosome assembly to initiate splicing.[77–79] A further RNA processing step is the 3'-end formation, which consists of two major steps: endonucleolytic cleavage which is followed by the addition of a poly(A) tail. Depletion of CBC in HeLa cells resulted in diminished poly(A) site cleavage. This effect was restored upon the addition of recombinant CBC. Furthermore, the interaction of the CBC/cap complex with processing factors at the poly(A) site was demonstrated by co-immuno-precipitation.[80] Additionally, it was shown that CBC interacts with the negative elongation factor (NELF) to promote 3'-end processing in replication-dependent histone mRNAs.[81] Cap structures bound to CBC are also involved in mediating mRNA export, as the CBP 80 subunit recruits the transcription export complex (TREX) through interaction with RNA export factor (REF)/Aly.[82] Following interaction with the nuclear export receptor NXF1, bulk mRNA is exported to the cytoplasm.[79] Notably, CBC also participates in the nuclear export of snRNAs during biogenesis by interacting with the phosphorylated adaptor for RNA export protein (PHAX), which promotes the assembly of the export complex.[83]

1.2.2. Translation initiation

The majority of mRNAs are translated in a cap-dependent manner. The canonical pathway of eukaryotic translation initiation (Figure 1.3) depends on the cap structure to bind eIF4E, which is part of the complex eIF4F, along with the RNA helicase eIF4A and the scaffold protein eIF4G. The 43S pre-initiation complex, comprising the ternary complex eIF2-GTP-$tRNA^{Met}$, the 40S ribosome subunit, and additional initiation factors (e.g. eIF3), interacts with the activated mRNA

through the binding of eIF4G with eIF3. The resulting large complex, designated the 48S pre-initiation complex, is capable of scanning the 5' UTR in 5'- to 3'-direction. Upon recognition of the start codon, some initiation factors dissociate, and the 60S ribosomal subunit joins. Subsequent to codon-anticodon base pairing, displacement of eIF1 results in a conformational change of the ribosome, marking the starting point of the elongation phase (reviewed by Topisirovic and colleagues).[79]

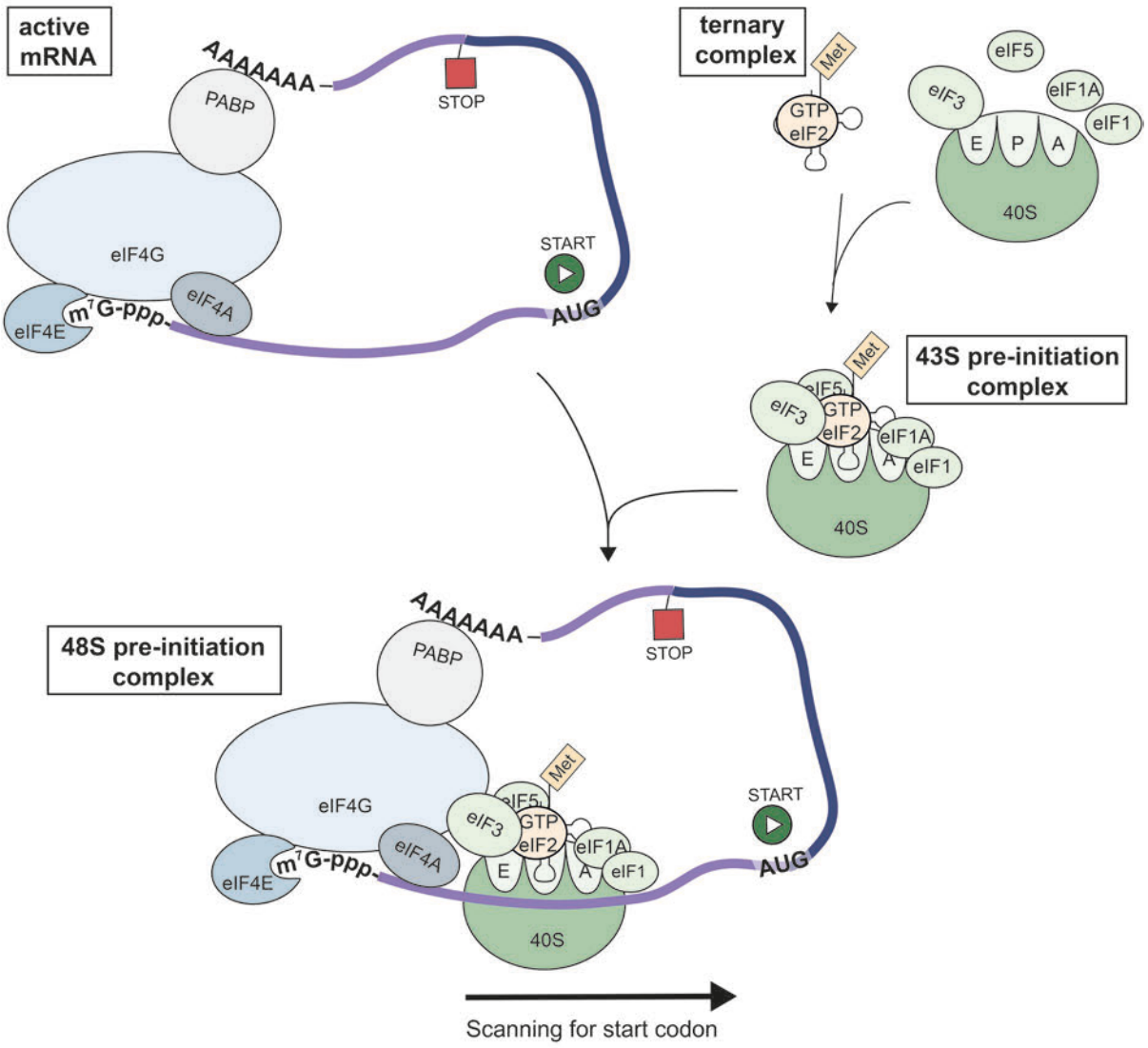

Figure 1.3. Cap-dependent eukaryotic translation initiation. mRNA is activated by binding to eIF4E and recruitment of the eukaryotic translation initiation factors (eIF) 4G and 4A, depicted in blue. The closed loop conformation is established by binding of the poly(A)-binding protein (PABP, grey) to eIF4G (upper left). The 43S pre-initiation complex is formed by recruitment of the ternary complex (eIF2-GTP-tRNA^{Met}, orange) to the 40S ribosomal subunit, which is associated with several translation initiation factors, depicted in green (upper right). The active mRNA and the 43S pre-initiation complex form the 48S pre-initiation complex, which subsequently scans the 5' UTR for a start codon. This figure was adapted and substantially modified from Hanson *et al*.[84]

While eIF4F is bound to the cap, eIF4G binds the poly(A)-binding protein (PABP), leading to circularization of the mRNA.[85] This configuration is presumed to ensure the translation of properly processed mRNA and to enhance recruitment of the ribosomes by stabilizing the interaction between the translation initiation factors and the mRNA.[86,87] The purpose of the translation of eIF4E-bound mRNA is to generate the majority of cellular proteins. However, newly synthesized mRNAs undergo a pioneer round of translation, serving as a control mechanism for the quality of gene expression by supporting non-mediated decay to eliminate transcripts with premature termination codons.[88] The composition of the translation complex differs between the pioneer round and steady-state translation. After export into the cytoplasm, CBC stays bound to the capped mRNA and recruits eIF4G and eIF4A, while in subsequent rounds eIF4E is used as cap-binding protein.[3,89] Additionally, the pioneer round involves primarily nuclear PABP in addition to cytoplasmic PABP, the CBP80-CBP20-dependent translation initiation factor CTIF, as well as exon junction complexes associated with spliced mRNAs.[88,90,91] Despite these differences, similar mechanisms of translation initiation take place, as indicated by the utilization of analogous translation initiation factors by CBC- and eIF4E-bound mRNAs, including eIF4G, eIF3, and eIF4A.[88,92] After the first round of translation, the transition to steady-state translation involves Importin (IMP)-β, which interacts with the CBC-bound IMP-α, leading to the replacement of CBC by eIF4E.[93]

In conclusion, the m^7G cap acts as specific binding site for protein factors, that regulate and mediate essential functions for the mRNP.[3]

1.2.3. RNA stability

The enhanced stability of capped mRNA towards uncapped transcripts is attributed to the protection from 5'-exonucleolytic degradation.[94,95] The removal of cap structures by decapping enzymes represents a regulatory step in mRNA stability, which in turn affects mRNA function.[96] Dcp2, the first discovered eukaryotic decapping enzyme, is a member of the Nudix hydrolase superfamily and cleaves the cap triphosphate bridge between the α- and β-phosphate, leading to m^7GDP and 5’-monophosphate RNA. Dcp2 binds to mRNAs with a specific stem-loop structure near the 5' cap, which is crucial for its recognition and decapping activity. In addition to this direct RNA recruitment, cis-elements such as AU-rich elements in 3' UTRs or deadenylated transcripts containing a U-tract, can recruit proteins to enhance Dcp2-mediated decapping, demonstrating the complexity of the process regulation.[96,97] Further reported decapping enzymes are Nudt16 and Nudt3, which also have different target preferences and only function on a subset of mRNAs.[96] While Nudt16 is ubiquitous expressed in various tissues

and has been shown to regulate among others the decay of some AU-rich element- (ARE)-containing mRNAs,[98] Nudt3 has been reported to modulate cell migration by decapping of integrin β6 and lipocalin-2 mRNAs.[99] Additional Nudix proteins, such as Nudt2, Nudt12, Nudt17, and Nudt19, are capable of decapping RNA *in vitro*.[100] The DXO (decapping exonuclease) family of enzymes exhibits distinct decapping mechanisms.[101] In yeast, Rai1 was found to possess 5'-end pyrophosphohydrolase (PPH) activity, which results in the release of pyrophosphate from 5'-triphosphorylated RNA.[102] Furthermore, it displays decapping activity, whereby the entire cap structure (m^7GpppN/GpppN) is removed. Studies showed stronger activity towards GpppRNA in comparison to m^7GpppRNA.[103] Rat1, a 5'-3'-exonuclease which is in a complex with Rai1, subsequently degrades the 5'-monophosphate RNA, generated by PPH and decapping activities of Rai1.[102] An additional member of the enzyme family is Dxo1, which has weak sequence homology with Rai1 in yeast. While it lacks PPH activity, it exhibits decapping activity and additional 5'-3'-exoribonuclease activity.[104] The mammalian homolog of Rai1 and Dxo1, termed DXO, combines the activities of both, which encompass PPH, decapping, and exoribonuclease activity.[105] *In vitro*, decapping activity occurs towards methylated and unmethylated caps, whereas *in vivo*, cap binding proteins protect methylated caps from degradation.[105] Consequently, DXO family enzymes function in capping quality control, by degrading uncapped or unmethylated capped transcripts (reviewed by Jurado and colleagues).[101] Following the removal of the cap, the 5'-monophosphate RNA is degraded exonucleolytically by Dxo1 and DXO, or by the heterodimer Rai1-Rat1.

1.2.4. Antiviral response

Innate immune response involves the activation of pattern recognition receptors (PRRs), such as RIG-I, that recognize non-self RNA and trigger the production of type I interferons (IFNs) and pro-inflammatory cytokines.[106,107] RIG-I exhibits preferential binding to double-stranded RNA (dsRNA) with 5'-diphosphate or 5'-triphosphate moieties, or cap 0 structures.[108] Notably, the presence of 2'*O*-methylation, particularly at the first transcribed nucleotide but also at the second one, significantly attenuates RIG-I binding and subsequent IFN signaling.[9,109] RIG-I mutation of His830, which is proximal to the 2'OH of the first transcribed nucleotide, resulted in almost same binding affinities for 2'*O*-methylated and 2'*O*-unmethylated dsRNAs, indicating that His830 is crucial for the discrimination between cap 1 and cap 0 RNA.[108] Interferon-induced proteins with tetratricopeptide repeats (IFIT) are induced by type I interferons and bind to single stranded RNA that lacks 2'*O*-methylation, inhibiting its translation.[110,111] IFIT1 binds unmethylated cap 0 RNA, thereby blocking translation factors like eIF4E and eIF4F, and as a

consequence inhibits viral RNA replication.[3,112,113] The capacity of 2'*O*-methylation to circumvent the innate immune response was exemplified by diminishing 2'*O*-methylation of endogenous RNA, which resulted in a considerable increase in type I IFN response.[109] The examination of a West Nile virus mutant that lacks 2'*O*-methyltransferase activity resulted in the attenuation of virus infections in primary cells. However, the mutant demonstrated virulence in cells with deficient IFN signaling, indicating that 2'*O*-methylation of the viral mRNA cap represents a strategy to evade the antiviral response.[114] The mechanisms employed by viruses to produce capped RNA are detailed in chapter 1.3.2.

1.2.5. Role of cap methylations

The cap structure is a highly methylated modification. The specific roles of these methylations are summarized in the following section.

The *N*7-methylation of guanosine is important for binding CBC and eIF4E, which mediate splicing, export, and translation initiation. It was shown that the binding affinity of eIF4E for m^7GpppG is ~10^3-folds higher than that for GpppG.[115] Similarly, CBC exhibits a higher affinity for the *N*7-methylated cap.[116] The involvement of the m^7G cap moiety in protein biosynthesis was demonstrated by translation of vaccinia virus mRNAs in wheat germ extract, where methylated *in vitro* mRNA was translated more efficiently than unmethylated mRNA.[117]

The previous chapter described the function of 2'*O*-methylation at the first transcribed nucleotide as a signature for self-RNA in the context of antiviral defense. In addition, a role in translation has been reported in the literature by demonstrating a positive influence of this methylation on the ribosome binding.[117–119] Furthermore, translation and oocyte maturation are enhanced by cap ribose methylation of c-mos mRNA in *Xenopus laevis*.[120]

Additional *N*6-methylation of Am as the first transcribed nucleotide leads to the formation of m^6Am, a frequent cap modification that is part of the major cap species in HeLa mRNA (m^7Gpppm^6Am), accounting for up to 30% of the cap pool.[7] The m^6Am modification has been reported to regulate gene expression, affecting processes such as mRNA stability and translation.[121] However, the reported roles of m^6Am in these processes are contradictory. It has been observed that transcripts initiated by m^6Am exhibit longer half-lives.[122,123] *In vitro* decapping experiments have demonstrated that the enhanced stability of m^6Am transcripts is due to a reduced susceptibility to decapping by DCP2.[122] In contrast, some other reports indicate that the *N*6-methylation does not affect transcript stability, as depletion of phosphorylated CTD interacting factor 1 (PCIF1), the methyltransferase responsible for *N*6-

methylation of Am at the first transcribed nucleotide position in mRNA, did not affect the abundance of m^{6}Am-modified mRNAs.[124,125] Similarly conflicting, some findings show that m^{6}Am promotes translation, while others report the suppression or no substantially effect on translation.[123–125] These controversial findings, presumably due to different mRNA targets and cell types as well as analytical approaches and their limitations, necessitate further investigation to elucidate the biological function of m^{6}Am in the mRNA cap.[121]

Several groups showed that the 2'*O*-methylation of the second transcribed nucleotide, leading to the so-called cap 2 structure, also contributes to evasion of the innate immune system. Abbas and colleagues demonstrated that transcripts bearing an additional 2'*O*-methylation at the second transcribed nucleotide completely disrupted binding to IFIT1, while 2'*O*-methylation exclusively at the first transcribed nucleotide only partially reduced the interaction.[126] Similarly, Despic and Jaffrey showed that the cap 2 structure further reduced the binding and activation of RIG-I.[127] Moreover, their development of a transcriptome-wide mapping method to quantify cap 2 structures, designated CLAM-Cap-Seq, revealed an enrichment of these caps on long-lived mRNAs. This established a correlation between mRNA ageing and the occurrence of additional methylation of the second transcribed nucleotide, resulting from the prolonged persistence of these mRNAs within the cytoplasm, where the cap methyltransferase (CMTR)2, responsible for cap 2 methylation, is also located.[127] Furthermore, small effects on mRNA translation or stability where shown.[127,128] Transfection of mRNA bearing cap 2 methylations, synthesized by *in vitro* transcription using tetranucleotide cap analogs, revealed a cell line dependent impact on protein production levels.[128] Additionally, resistance towards decapping by DXO was observed. Recently, CMTR1 and 2 were found to be essential for mouse embryonic development, with analysis of CMTR mutant embryos showing misregulation of genes, demonstrating an additional role for cap 2 modifications in gene regulation.[129]

1.3. Capping process

1.3.1. Eukaryotic mRNA capping

Capping of nascent Pol II transcripts occurs co-transcriptionally once the transcripts have reached 20-25 nucleotides in length, representing the initial stage of pre-mRNA processing in the nucleus.[130] This involves the recruitment of capping enzymes to the pre-mRNA through the carboxy-terminal domain of Pol II.[131,132] The generation of the m^{7}G-cap comprise three enzymatical activities: first, RNA triphosphatase (RTPase) activity removes the terminal phosphate of the 5'-triphosphate RNA creating 5'-diphosphate RNA. Next, guanylyltransferase

(GTase) activity transfers a GMP, which resulted from GTP hydrolysis, to the 5'-diphosphate RNA, forming the cap-typical 5'-5'-triphosphate bridge that connects guanosine to the first transcribed nucleotide. Finally, guanine-*N*7-methyltransferase (*N*7-MTase) activity leads to methylation of the *N*7-position of guanosine, generating the so-called cap 0 structure (Figure 1.4).[3]

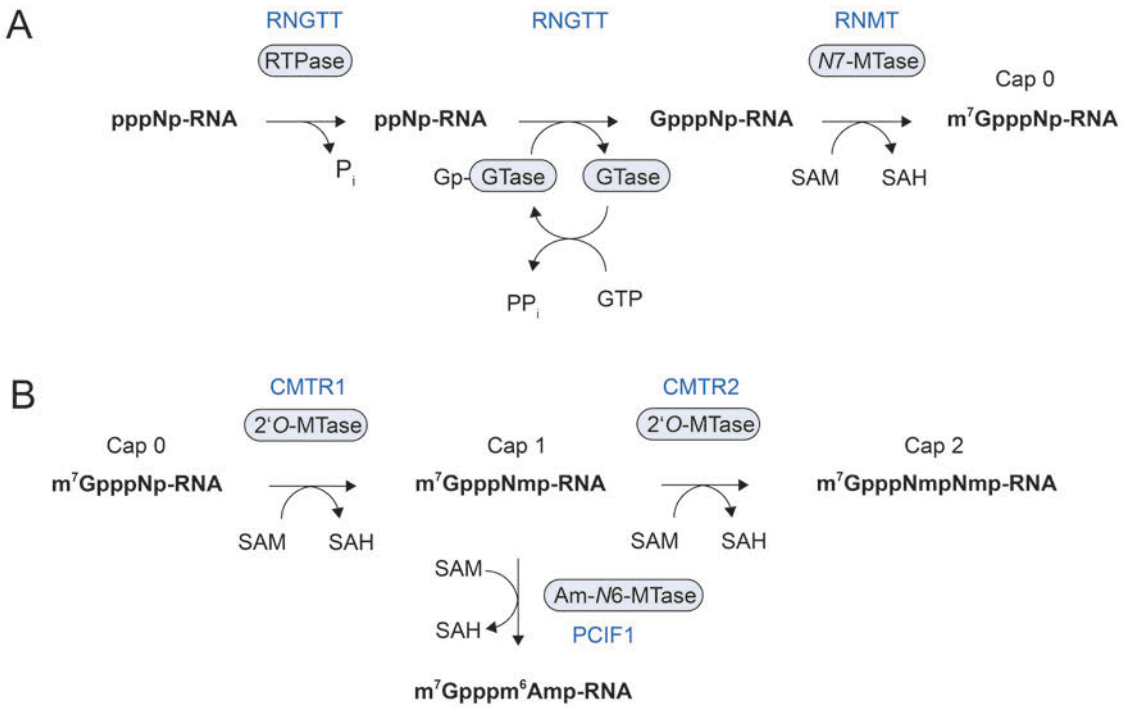

Figure 1.4. Canonical capping pathway, showing capping enzyme activities (framed). The respective mammalian capping enzymes are depicted in blue. A.) Pathway to yield cap 0 structures. This involves the activity of an RNA triphosphatase (RTPase), which removes the γ-phosphate of 5'-triphosphate RNA, resulting in 5'-diphosphate RNA. A guanylyltransferase (GTase) forms an intermediate with GMP, which results from the hydrolysis of GTP, and transfers this molecule onto the 5'-diphosphate RNA. In mammalians, RNA guanylyltransferase and 5'-phosphatase (RNGTT) carries both functions. Finally, an *N*7-methyltransferase (*N*7-MTase), in mammalians RNA guanine-7 methyltransferase (RNMT), transfers a methyl group from *S*-adenosyl methionine (SAM) to the *N*7-position of guanosine, forming the cap 0 structure. B.) Possible methylations of a cap 0 structure. The cap methyltransferases (CMTR)1 and 2 are 2'*O*-methyltransferases (2'*O*-MTases) that lead to formation of cap 1 and cap 2 structures by methylating the 2'*O*-positions of the triphosphate-adjacent nucleotides, respectively. When the first transcribed nucleotide is an adenosine, phosphorylated CTD interacting factor 1 (PCIF1), a 2'*O*-methyladenosine-*N*6-methyltransferase (Am-*N*6-MTase), can transfer a methyl group to the *N*6-position of adenosine.

In *Saccharomyces cerevisiae* (*S. cerevisiae*), three distinct enzymes are responsible for the enzymatic functions involved in mRNA capping.[133,134] In contrast, mammals use the bifunctional enzyme RNGTT (RNA guanylyltransferase and 5'-phosphatase) to catalyze both RTPase and GTase activities, while a separate protein, the RNA guanine-7 methyltransferase (RNMT), carries the *N*7-MTase activity.[3] In mammalian cells, CMTR1 is responsible for 2'*O*-

methylation of the cap-adjacent nucleotide, while additional *N*6-methylation of adenosine can be introduced by PCIF1 in vertebrates.[118,123,124,135,136] *In vitro* biochemical studies showed that PCIF1 can also act on ribose-unmethylated m^{7}GpppA-capped RNA to form an m^{7}Gpppm^{6}A cap moiety, even though kinetic experiments demonstrated an approximately 8-fold preference for the 2'*O*-methylated adenosine substrate. The absence of m^{7}Gpppm^{6}A detection in biological samples led to the assumption that 2'*O*-methylation occurs prior to m^{6}A formation.[124] CMTR2 methylates the second transcribed nucleotide at the 2'*O*-position.[9] Werner and colleagues demonstrated that CMTR2 is present in the nucleus and cytosol, suggesting that while 2'*O*-methylation of the first transcribed nucleotide predominantly takes place in the nucleus, CMTR2-mediated capping occurs in either the nucleus or after export to the cytoplasm.[137]

Of note, cytoplasmic recapping has also been reported. In the early 1990s, the Maquat laboratory discovered that cytoplasmic β-globin RNA could undergo recapping, indicating that capping occurs not only during transcription but also post-transcriptionally in mammalian cells. This finding was supported by the identification of capping enzymes, such as RNGTT and RNMT, in the cytoplasm, although at lower levels than in the nucleus. RNAs that have lost their caps through decapping can serve as substrate for recapping. Recently, a kinase was identified as cofactor for RNGTT, which can convert the 5'-monophosphate end of a decapped RNA substrate (e.g., resulting from cleavage by DCP2) to a 5'-diphosphate, which is necessary for RNGTT activity. This cytoplasmic recapping may serve as a mechanism to regulate protein synthesis with uncapped RNA to be stored until further use (reviewed by Borden *et al.*)[138]

1.3.2. Viral RNA capping

As described in chapter 1.2.4., viruses have developed strategies to cap RNA, thereby facilitating efficient protein synthesis and enabling them to circumvent the innate immune response. Viruses employ various mechanisms to acquire capped RNA. Most DNA viruses and some types of RNA viruses utilize the host's cellular capping machinery to add caps to their mRNA. Other viruses perform the so-called "cap snatching", stealing caps from host mRNA or encoding their own capping enzymes.[139] Different capping machineries are summarized in the following section.[3,139]

An example of a virus that utilizes the canonical capping pathway is the vaccinia virus, which encodes a multifunctional protein with RTPase, GTase, and *N*7-MTase activity operating in the canonical sequential order (see Figure 1.4).[140] Additionally, the bifunctional protein VP39

catalyzes the 2'*O*-methylation of the triphosphate-adjacent nucleotide, forming cap 1 RNA, in addition to its role as poly(A) polymerase processivity factor.[141] However, there are also noncanonical capping mechanisms. In alphaviruses, the sequential order of the enzymatic steps is rearranged. Here, GTP is already *N*7-methylated, before the GTase activity transfers m^7GMP to the γ-dephosphorylated 5'-diphosphate RNA.[3,142] In contrast, mononegavirales, such as vesicular stomatitis virus and rabies virus, encode a GDP polyribonucleotidyl transferase (PRNTase), which forms a covalent complex with the viral RNA through a monophosphate group. Transfer of GDP leads to formation of the triphosphate bridge, followed by *N*7-methylation and 2'*O*-methylation of the triphosphate bridge-adjacent nucleotide.[139] Additionally, different mechanisms of cap snatching have been described. In influenza virus, the RNA-dependent RNA polymerase (RdRp) also exhibits endonuclease activity. It binds to host RNA and cleaves the first 10-14 nucleotides of the capped RNA, which are subsequently used as a primer for viral RNA synthesis by the RdRp.[143,144] In contrast, in yeast totiviruses, m^7GMP is cleaved from host RNA and co-transcriptionally transferred to the 5'-diphosphate of the viral transcript.[3,145]

1.4. Caps in synthetic mRNA

Synthetic mRNA is a crucial instrument in laboratory research, enabling the assessment of mRNA functions within a cellular context as well as the utilization for therapeutic applications. The idea of using synthetic mRNA for direct gene transfer started in 1990 with the injection of unmodified RNA in mouse muscles[146] and recently had its preliminary highlight, with the development and regulatory improvement of mRNA vaccines from Biontech/Pfizer and Moderna against COVID-19.[147–149] In between, there have been years of effort and development to improve mRNA translation and stability, modulate immunostimulatory activity, and optimize mRNA delivery. Structural modifications to tune mRNA properties include the use of modified nucleosides, codon optimization, and manipulation of the cap structure and poly(A) tail.[150] The pivotal roles of caps in enhancing stability, translational efficacy, and circumventing innate immune responses underscore the necessity to cap synthetic RNAs.[149]

1.4.1. Co- and posttranscriptional capping

Due to the size limitations of approximately 100 nucleotides, solid-phase synthesis is not a feasible option for the preparation of biologically relevant mRNAs with lengths of several hundreds to thousands of nucleotides.[151,152] Contrary, i*n vitro* transcription (IVT) of a DNA

template is the method of choice resulting in an uncapped 5'-triphosphorylated transcript. The capping process can be carried out post-transcriptionally or co-transcriptionally.

Post-transcriptional capping can be performed by downstream enzymatic modification of the 5'-end, with the IVT transcript being subjected to three sequential enzymatic steps. These steps reflect the canonical capping pathway found in higher eukaryotes as well as in some eukaryotic viruses, which enables the use of recombinant viral capping enzymes for *in vitro* capping.[3] In the first step, the 5'-terminal γ-phosphate is removed by an RNA triphosphatase. Subsequently, a GMP moiety derived from GTP is transferred to the 5'-diphosphate terminus by RNA guanylyl transferase activity, thereby forming the 5'-5'-triphosphate bridge. Ultimately, the guanosine cap is methylated at the *N*7-position by a guanine-*N*7-methyltransferase using *S*-adenosyl methionine (SAM) as methyl donor, schematically depicted in Figure 1.5 A. As early as the 1970s, the protein complex responsible for the capping reaction in the vaccinia virus was isolated[153,154] and later successfully utilized to cap *in vitro* transcribed RNA.[155] Since that time, the capping enzymes of the vaccinia virus have become commercially available and are therefore most commonly used. The vaccinia capping enzyme is composed of two subunits bearing the three enzymatic functions necessary to produce cap 0 structures. The 2'*O*-methyltransferase VP39 from vaccinia is also commercially available to produce IVT RNA with a cap 1 structure. These vaccinia capping enzymes have been used in the pre-clinical production of mRNA vaccines against SARS-CoV-2.[148]

The process of co-transcriptional capping employs the ability of RNA polymerases with flexible substrate requirements, such as *E. coli*, T3, T7, or SP6 RNA polymerases, to be primed by cap analogs, like m^7GpppG, which directly leads to the synthesis of 5'-capped IVT RNA.[156] However, co-transcriptional capping faces problems regarding the capping efficiency, since the cap analog competes with GTP as initiation starter, resulting in a fraction of uncapped transcript. This effect can be decreased by using GTP at a limited concentration, as the capping efficiency is determined by the ratio of cap analog to GTP. An additional limitation of these cap analogs is their susceptibility to incorporation in reverse orientation, potentially generating Gppp(m^7G)-capped RNA (Figure 1.5 B).[157]

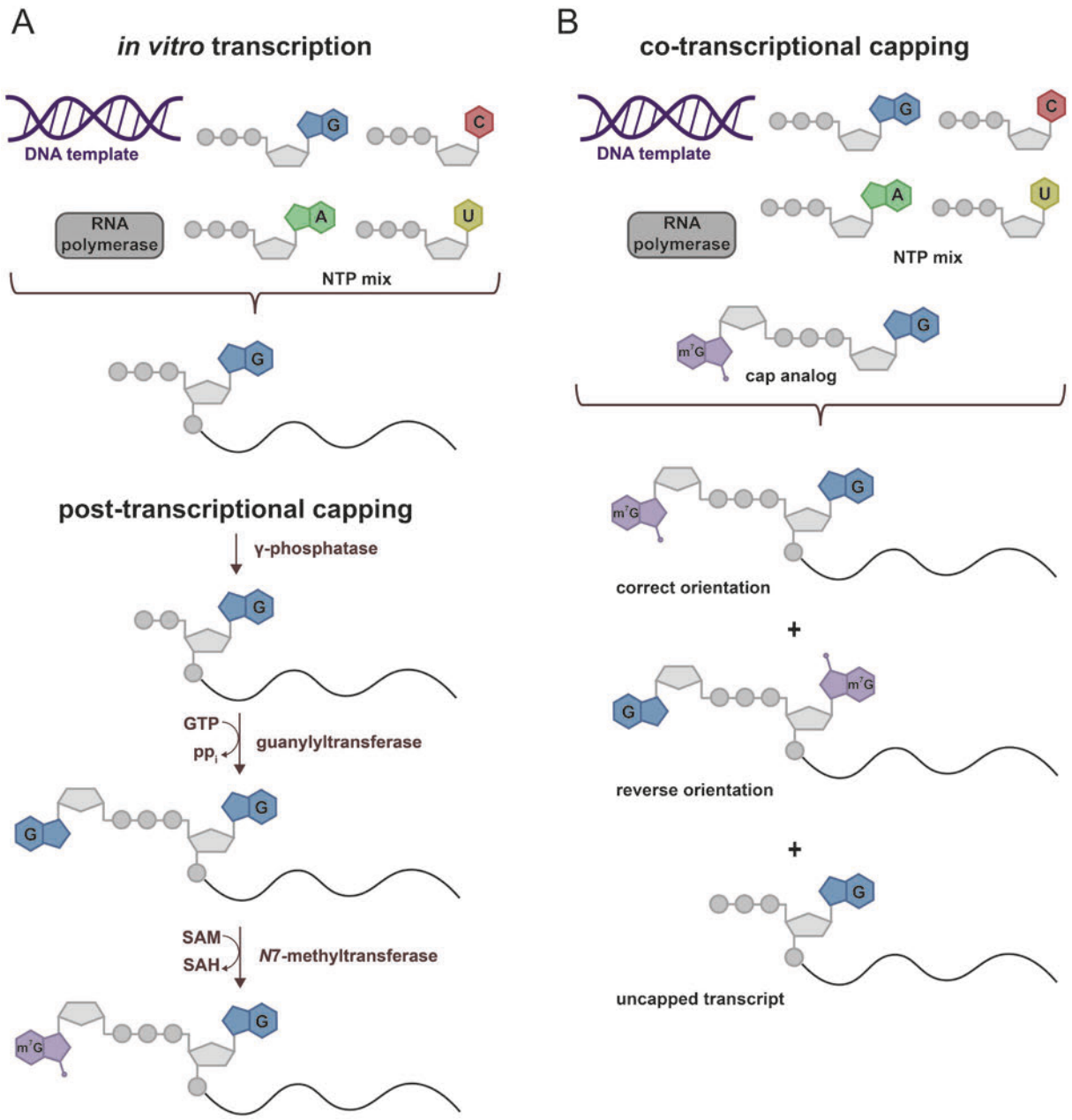

Figure 1.5. A.) Schematic illustration of *in vitro* transcription involving a DNA template, an RNA polymerase, and nucleoside triphosphates (NTP), followed by post-transcriptional enzymatic capping. B.) In the co-transcriptional capping process, a cap analog is added to the IVT reaction yielding different outcomes: uncapped transcript, capped transcript with the cap analog in the correct orientation, and capped transcript, with the cap analog in the reverse orientation.

This issue was subsequently circumvented by the development of anti-reverse cap analogs (ARCA), which are modified at the 3'OH group of m^7G through methylation (Figure 1.6 A) or deoxygenation, thereby preventing elongation at the wrong nucleotide.[158] Remarkably, the reverse incorporation of the cap analog was also prevented by methylating the 2′*O*-position of

m^{7}G.[159] Another notable advancement was the demonstration that trinucleotide cap analogs are also accepted by T7 RNA polymerases.[160] These trinucleotides possess distinct advantages, including the inability to be incorporated in an inverted orientation due to an additional base pair of the cap analog with the template, and the capacity for direct incorporation of cap 1 structures, which is enabled by lacking a direct involvement of the ribose-methylated nucleotide at the transcription initiation stage.[160]

The analysis of various trinucleotide analogs (m^{7}GpppNpG) demonstrated that high incorporation efficiencies of approximately 90% were achieved when N = A or Am were used as triphosphate-adjacent nucleotides, while analogs with N = G, Gm, m^{6}A, and m^{6}Am exhibited lower efficiencies (80-85%) and pyrimidine-containing analogs only reached 55-60%.[161] Recently, some cap trinucleotides have emerged as commercially available products under the label CleanCap® reagents (Figure 1.6 A).[162] The range of cap analogs suitable for co-transcriptional capping has also been expanded to cap tetranucleotides, which allow the direct incorporation of cap 2 structures into synthetic RNA.[128] Additionally, the incorporation efficiency of metabolite caps can be enhanced by employing cap analogs such as NADpG (Figure 1.6 B) or FADpG, which demonstrated significantly elevated incorporation rates approaching 95% in comparison to NAD and FAD as transcription initiator.[163]

1.4.2. Incorporation of non-natural caps

The previous part focused on capping options to produce natural caps, except for 3'*O*-methylation of m^{7}G (3'*O*-Me-m^{7}G) to prevent inverse incorporation of the cap analog. However, these capping methods also enable the production of RNA with chemically modified caps to modulate biological characteristics or to introduce new functionalities. The aim here is to affect the translational activity, which is linked to the affinity for the translation initiation factor 4E, or the stability as expressed by relative susceptibility to decapping enzymes.[149] Co-transcriptional capping using cap analogs provides diverse opportunities for structural modifications, which have been comprehensively reviewed in recent literature.[149,156] An overview of different cap analog structures is depicted in Figure 1.6 A.

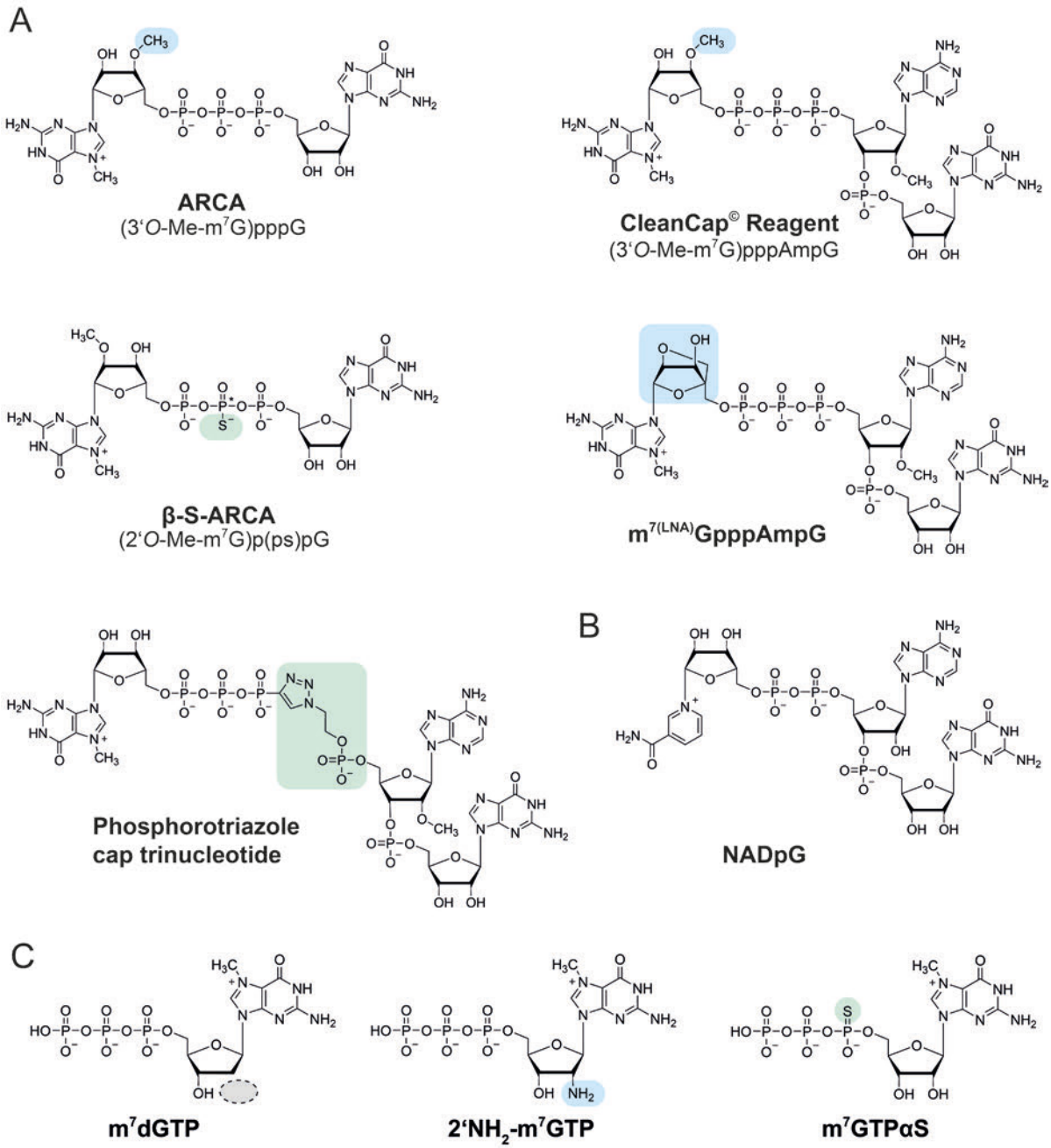

Figure 1.6. A.) Depiction of several non-natural cap analogs, comprising di- and trinucleotides. The locked nucleic acid moiety is abbreviated as LNA. B.) Structure of a cap trinucleotide for synthesis of NAD-capped RNA. C.) GTP analogs to post-transcriptionally introduce cap variants using the vaccinia virus capping enzyme. Modifications are highlighted in blue (ribose modifications) and green (triphosphate modifications).

The ARCA-capped RNA demonstrated a greater than twofold increase in translation efficiency when compared to the m^{7}GpppG-capped RNA in a rabbit reticulocyte lysate.[158] Intending to modify the cap triphosphate bridge, the *O*-to-*S* substitution within the β-phosphate of 2'*O*-ARCA resulted in the development of a new compound, termed β-S-ARCA (Figure 1.6 A). Of the isomers produced, particularly the D1 diastereomer demonstrated enhanced stability and

translational activity.[164] Moreover, the use of this analog in antigen-encoding RNA led to an enhanced induced immune response, and the compound was employed for mRNA capping in clinical trials for RNA vaccines.[165–167] Other cap analogs that are chemically modified at the triphosphate bridge include triazole moieties (e.g. phosphorotriazole caps, see Figure 1.6 A), which enables the introduction of the cap via click chemistry.[168] The incorporation of benzyl moieties at the *N*7- or *N*2-position of guanosine caps revealed promising candidates exhibiting enhanced translational properties when incorporated into mRNA in comparison to m^7GpppG- and ARCA-capped RNA.[169,170] Also, another new cap trinucleotide analog bearing a locked nucleic acid (LNA) moiety (see Figure 1.6 A) was translated three times more efficient compared to m^7G-capped RNA.[171] The introduction of chemically modified cap analogs can also be achieved through enzymatic reactions, as some capping enzymes were shown to exhibit weak substrate specificity, thereby allowing preparation of non-natural caps. Vaccinia virus and Paramecium bursaria Chlorella virus-1 capping enzymes showed the ability to use several GTP analogs (Figure 1.6 C) as substrates, with some analogs showing enhanced (m^7dG, 3'*O*-Me-m^7G, (m^7)GαS) or comparable protein expression (2'*O*-Me-m^7G, 2'NH_2-m^7G) relative to m^7G.[172,173] The advantage of these techniques is the accessibility of commercially available GTP analogs, thereby avoiding time-consuming chemical synthesis of the corresponding cap dinucleotides. A recently published chemo-enzymatic approach employs propargyl-SAM analogs to modify the *N*6-position of cap-adjacent adenosine utilizing PCIF1. The resulting capped RNA demonstrated maintained translational efficiency in combination with a moderate increase in immune response.[174]

In addition to the aforementioned variety of chemical modifications used to modulate RNA functions, which is far from complete, further features can be implemented in the cap structure. The introduction of photocleavable cap analogs by several research groups has opened up a wide range of functions, including the study of cap-protein interaction through RNA-protein crosslinking,[175] the light-mediated activation of mRNAs,[176,177] and the facilitation of high performance liquid chromatography (HPLC)-based purification of capped RNA by exploiting the hydrophobicity of the labeled cap.[178] Additionally, the introduction of fluorescent- and biotin-labeled cap analogs has been demonstrated,[179–181] as well as the integration of amino groups, capable of covalently binding molecules conjugated with *N*-hydroxysuccinimide-ester, and the integration of azide or terminal alkyne groups, enabling the application of click chemistry for bioconjugation.[173,182,183]

As mentioned previously, the chemical synthesis of capped RNA based on solid-phase synthesis is size-restricted, which hinders the direct preparation of biologically relevant

mRNAs.[156] However, combining chemical and enzymatic methods can overcome these limitations. For instance, specific mutations in the replicative DNA polymerase from *Thermococcus gorgonarius* have altered its substrate specificity, enabling the synthesis of RNA from a single-stranded DNA template. This modified polymerase showed the ability to accept capped oligoribonucleotides as primer.[184] In addition to primer extension, click chemistry provides a versatile approach for preparing capped RNAs and cap analogs. Using CuAAC, an azido-modified cap analog can be covalently linked to an alkyne-modified RNA molecule.[185] As an alternative, alkyne-modified triphosphorylated RNA can be clicked with a 5'-azido-modified m^7G.[156,186] Recently, Chen and colleagues introduced a technique called LEGO (ligation-enabled mRNA-oligonucleotide assembly), which allows chemically synthesized oligonucleotides containing various cap variants to be ligated to IVT-mRNA.[187] Furthermore, the capability of introducing an additional cap at the 5'-end of mRNA using click-chemistry was demonstrated, resulting in dual-capped mRNA with enhanced translation capacity.

The variety of options to cap synthetic RNA, including the development of novel cap analog strategies, offers the possibility of generating differently capped RNA. This allows for a highly efficient introduction of advanced cap analogs into mRNA for therapeutic applications, as well as the production of caps with customized functionalities to further explore cap functions and associated biological processes.[149,156]

1.5. Analytical methods to investigate the cap epitranscriptome

1.5.1. Overview of detection methods

The identification and investigation of the cap structure and its modifications were accomplished by using radioisotope labeling, enzymatic hydrolysis and different types of chromatography.[6,7,188] These principles are analogous to the techniques currently used to investigate cap structures in RNA derived from tissues and cell lines, which include thin layer chromatography (TLC) and gel electrophoresis, in combination with prior ^{32}P-labelling, HPLC coupled with MS following RNA digestion, as well as sequencing methods.[189]

Kruse and colleagues described a TLC-based method, in which poly(A) RNA is enzymatically decapped, dephosphorylated and subsequently 5'-end radiolabeled.[190] Following digestion to nucleoside monophosphates by nuclease P1 (NP1), TLC separation is performed, which allows to further investigate the occurrence of cap-associated adenosine modifications in mouse mRNA, in particular the relative comparison of m^6Am and Am as first transcribed

nucleotide. In a similar approach, poly(A) RNA is subjected to enzymatic decapping and subsequent recapping by vaccinia capping enzyme using ^{32}P-labeled GTP. Following RNase I digestion, which leaves 2'*O*-methylated sites uncleaved, the resulting fragments are separated using denaturing polyacrylamide gel electrophoresis (PAGE). This approach determines the number of 2'*O*-methylated nucleotides adjacent to the cap while not revealing their specific identity or sequence.[191,192] However, only small amounts of input RNA are required, which is an advantage of these methods.

Several high-throughput sequencing approaches are used to determine the cap-typical m^{6}Am modification (reviewed by Benak and colleagues).[193] The utilization of *N*6-methyladenosine (m^{6}A) antibodies in detection methods often lacks specificity in distinguishing between m^{6}A and m^{6}Am. To address this issue, techniques such as MeRIPseq, miCLIP, and m6ACE-seq employ special computational analyses based on the distinct localization of m^{6}Am or the use of PCIF1-knockout cell lines enabling the differentiation of m^{6}Am profiles.[123,194,195] Further sequencing approaches comprise m6Am-exo-seq, which enriches 5'-end RNA fragments containing m^{6}Am by removing uncapped RNA and subsequently decapping for antibody recognition.[125] Similarly, m6Am-seq first isolates 5'-terminal RNA fragments using cap-m^{7}G immunoprecipitation and then applies recombinant FTO treatment to selectively remove m^{6}Am, given its higher affinity for m^{6}Am compared to m^{6}A, prior to sequencing.[196,197] Recently, Despic and Jaffrey developed CLAM-cap-seq, a novel transcriptome-wide sequencing method for detecting cap 1 and cap 2 structures in mRNA.[127] This method starts with reverse transcription of previously decapped RNA. Following CircLigase treatment to connect the mRNA 5'-end with the complementary DNA (cDNA) 3'-end, RNase T2 degrades the RNA, while preserving 2'*O*-modified nucleosides. This process results in a two-nucleotide-long cap tag for cap 1 structures and a three-nucleotide-long cap tag for cap 2 structures attached to the cDNA. After adaptor ligation and PCR (polymerase chain reaction) amplification, sequencing is conducted. Finally, data analysis focuses on identifying the resulting lengths of palindrome sequences that are specific to cap 1 or cap 2 structures. However, this approach cannot detect m^{6}Am modifications, presumably as a consequence of limitations in the reverse transcription process.[189]

1.5.2. LC-MS/MS analysis of cap structures

LC-MS/MS analysis of cap structures is usually preceded by a digestion step of RNA into smaller fragments ranging from nucleosides to oligonucleotides of variable length. Total digestion of RNA is commonly achieved by using endonucleases such as nuclease P1 and a

phosphodiesterase to produce 5’-nucleoside monophosphates, followed by dephosphorylation using alkaline phosphatase, which yields RNA nucleosides.[198] A disadvantage is the complete loss of sequence information. Therefore, in the context of examining cap structures, this method is only reliable for detecting cap nucleoside modifications that are unique to the cap and do not occur naturally elsewhere within the analyzed RNA.[199] Consequently, such an approach is more frequently employed in the analysis of synthetic RNA.

An alternative digestion protocol involves the use of nuclease P1, optionally in combination with alkaline phosphatase, while omitting phosphodiesterase. The integrity of the 5’-5’-triphosphate bridge is preserved, as neither nuclease P1 nor alkaline phosphatase target this linkage, resulting in the release of cap dinucleotides.[45,199] This approach enables reliable detection and identification of the cap triphosphate-adjacent nucleotides and their methylation status, thus facilitating the analysis of cap 0 and cap 1 structures. However, cap 2 structures cannot be investigated using the digestion to cap dinucleotides. Recently, two approaches, CAP-MAP and CapQuant, have been developed to detect and quantify global levels of cap structures based on the analysis of cap dinucleotides.[200,201] CAP-MAP directly analyzes nuclease P1-digested poly(A) RNA, while CapQuant includes an additional step of HPLC enrichment for cap dinucleotides. The synthesis and use of cap dinucleotide standards in these approaches allow accurate absolute quantification of detected cap modifications, providing insights into the cap epitranscriptome.

Alternative LC-MS approaches utilize site-specific ribonucleases to analyze fragments of capped RNA. One such approach, developed by Akichika and colleagues, employs RNase T1 digestion, which specifically cleaves the phosphodiester bonds on the 3’-side of guanosines, yielding 3’-monophosphate ends through formation of a 2’,3’-cyclic phosphate bond.[202] This approach enables the relative comparison of fragments containing different cap structures, such as m^7GpppAmpGp and m^7Gpppm^{6}AmpGp.[124] In addition, a new LC-MS approach to determine the capping efficiencies of synthetic mRNAs has recently been reported. The method involves RNase H-based cleavage after annealing of a complementary biotin-tagged oligonucleotide. The resulting digestion generates short 5'-cleavage products, which are subsequently resolved by LC-MS, with the amounts of capped and uncapped fragments being compared.[203]

The analysis of oligonucleotides is commonly performed using reversed-phase (RP)-HPLC while often requiring ion-pairing reagents in the mobile phase to achieve chromatographic retention of these hydrophilic and negatively charged molecules. However, the use of ion-pairing reagents can lead to ion suppression, leading to the investigation of alternative

techniques such as hydrophilic interaction liquid chromatography (HILIC).[204] Mass spectrometric analysis of oligonucleotides frequently employs tandem-mass spectrometry (MS/MS), which facilitates identification through molecule-specific fragmentation patterns. The use of high-resolution MS instruments, including time-of-flight (TOF) and orbitrap analyzers, enables the detection of exact masses, which are important for the decoding of complex mass spectra. In contrast, the analysis of nucleosides and cap dinucleotides typically relies on RP-HPLC without the need for ion pairing reagents, although alternative separation methods such as HILIC or hypercarb columns have also been reported.[200] MS detection is often performed using triple quadrupole mass spectrometers, which are of low-resolution but offer advantages in terms of sensitivity and specificity (reviewed by Yoluç *et al.*)[205] Considering its significance to the present thesis, the following sections outline the basic principles of RP-LC-MS/MS using a triple quadrupole mass spectrometer.

1.5.2.1. RP-HPLC

As a sophisticated analytical technique for separating components in liquid mixtures, HPLC allows the detection and quantification of these components. Following injection of the analyte, pumps carry the analyte-containing liquid mobile phase through a column packed with stationary phase material. The sample components interact differently with the stationary phase, which results in them travelling through the column at different speeds and ultimately leading to their separation.[206] In RP-HPLC, hydrophobic stationary phases are employed, with C18 columns being the most commonly used. These columns are composed of silica particles that are bonded to octadecyl chains.[207] The mobile phase typically consists of an aqueous buffer and a water-miscible organic solvent, such as acetonitrile or methanol.[208] The analytes are separated based on their hydrophobic interactions, with more hydrophobic compounds exhibiting stronger retention due to increased affinity for the non-polar stationary phase, while hydrophilic analytes elute faster.[208] After leaving the column, the analytes can be detected using various methods, such as UV-Vis spectroscopy or mass spectrometry, with the latter enabling the detection of analytes based on their mass-to-charge (*m/z*) ratios.

An example of a resulting chromatogram plotting the analyte-triggered signal against time is shown in Figure 1.7. Several chromatographical parameters can be determined directly or indirectly from the chromatogram to assess peak retention and peak separation. One parameter determined directly from the resulting chromatogram is the retention time (t_R), which is defined as the time required for an analyte to pass through the column. Additionally, peak widths can be measured at different points on the chromatographic peak. Common

measurements include the peak width at the baseline (w_h), the peak width at 5% of the peak height ($w_{0.05}$), the peak width at half maximum ($w_{0.5}$), or the distance between the leading edge of the curve and the peak maximum, measured at 5% of the peak height ($d_{0.05}$). These parameters are used to calculate characteristic values describing peak shape and separation quality, such as resolution and symmetry factor.

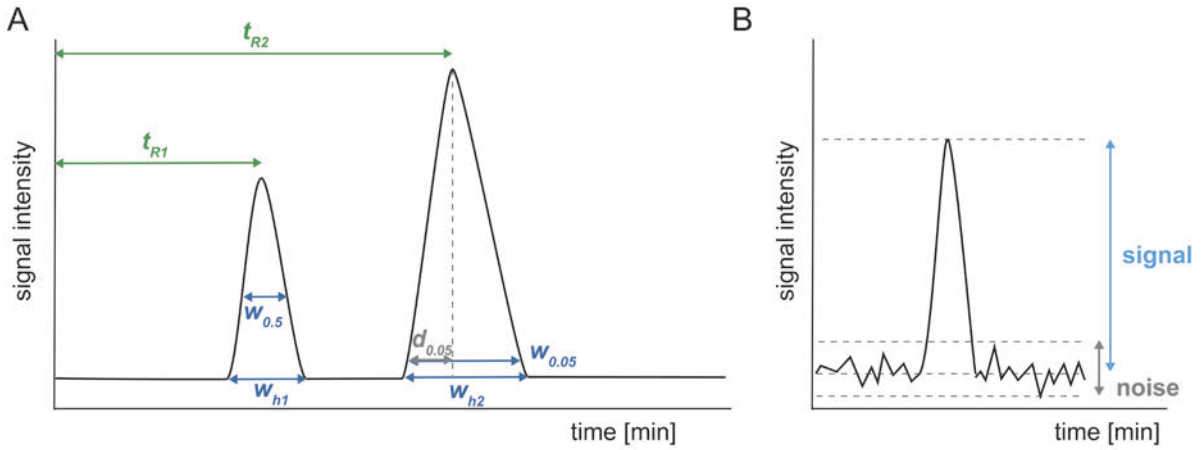

Figure 1.7. A.) Exemplary chromatogram to illustrate the chromatographical parameters retention time (t_R), peak width at the baseline (w_h), peak width at 5% of the peak height ($w_{0.05}$), peak width at half maximum ($w_{0.5}$), and distance between the leading edge of the curve and the peak maximum, measured at 5% of the peak height ($d_{0.05}$). B.) Depiction of signal and noise level in a chromatogram.

The resolution (R_s) indicates the degree of separation between two peaks in a chromatogram and is calculated according to equation 1. It is a function of the difference between the retention times of two peaks in combination with their peak widths at the baseline. The threshold for achieving complete baseline separation between adjacent peaks is widely accepted to be a resolution of 1.5.[209,210]

(1) $$R_s = \frac{1.18 \cdot (t_{R2} - t_{R1})}{w_{h2} + w_{h1}}$$

The symmetry factor (A_s) is a tool to assess peak shapes and is calculated according to equation 2. It is defined as the ratio of the peak width at 5% of the maximum peak height to twice the distance between the leading edge of the curve and the peak maximum, measured at 5% of the peak height.

(2) $$A_s = \frac{w_{0.05}}{2d_{0.05}}$$

A value of $A_s = 1$ indicates an ideally symmetrical peak, $A_s < 1$ denotes peak fronting, while $A_s > 1$ signifies peak tailing. These deviations from ideal peak symmetry can be attributed to multiple factors, including column overloading, column aging, or secondary interactions with

residual silanol groups inside the column.[211] Adjustment of injection volumes, temperature, or mobile phase compositions (e.g. proportion of organic solvent, ion strength, pH) can help to improve the peak shape.

Furthermore, the signal-to-noise ratio (*S/N*) ratio can be determined, which is a measure that compares the height of an analyte signal (*S*) to the level of background noise (*N*) and is particularly used as a tool to assess the reliability of detection and quantification at low analyte concentrations. The limit of detection (LOD) is defined as the lowest amount of analyte that can be reliably distinguished from the background noise and is established at an *S/N* ratio of 3 (equation 3). In contrast, the limit of quantification (LOQ) refers to the lowest amount of an analyte that can be used for quantification and requires a signal height 10 times greater than the background level (equation 4).[201]

(3) $LOD: \ S/N = 3$

(4) $LOQ: \ S/N = 10$

These thresholds are crucial for evaluating a method's ability to detect low concentrations of analytes, as well as for assessing whether observed peaks in chromatograms are suitable for reliable quantification.

1.5.2.1.1. HPLC method development and optimization

The optimization of HPLC methods is a crucial step in ensuring the reliability and efficiency of analytical results. During the development of a new method, it is essential to consider several key factors. This includes the separation system, which can involve different types of columns such as reverse-phase, normal phase or ion exchange. Furthermore, the functional groups of the column, as well as the particle size, and column length, have a significant impact on the chromatographic outcome.[212,213]

The interaction of analytes in the mobile phase with the stationary phase is influenced by several parameters. The main optimization parameters include mobile phase composition, buffer selection, pH, column temperature, and flow rate. The composition of the mobile phase, particularly in terms of ion strength and pH, impacts the retention and separation of analytes. In RP-HPLC, hydrophobic compounds exhibit stronger retention on the stationary phase. As a result, ionized forms of analytes typically elute faster than their neutral counterparts due to their increased polarity and reduced interaction with the hydrophobic stationary phase.[214] Consequently, given that the pH of the mobile phase can significantly impact the retention of

ionizable analytes, it is essential that ion strength provides adequate buffering capacity to prevent potential retention time shifts.[215] Additionally, high ion strength impacts peak shapes and retention by masking silanols inside the column, thereby minimizing interactions between ionized analytes and the stationary phase.[216] Column temperature is another critical parameter influencing the chromatographic behavior in various ways. For ionized compounds, the temperature impacts the pH value of mobile phases as well as the pK_a value of the analytes, which in turn influences their degree of ionization.[217–219] The extent of analyte interaction between the mobile and the stationary phases is also temperature-dependent. According to the principles of Le Chatelier, elevated temperature shifts the equilibrium towards the mobile phase,[220] leading to reduced retention times. Recent studies have demonstrated the effects of different column temperatures on retention and separation selectivity,[219,221] particularly demonstrating that higher temperatures can reduce run times of large molecules.[222]

The detection method also plays a critical role in method development and optimization. For instance, when coupling HPLC with MS, it is important to select buffers and salts that are compatible with MS.[213,223] As outlined in the next chapter, the flow rate is another crucial factor, which is typically used at lower rates in MS applications compared to traditional HPLC methods.[213] Additionally, mass spectrometric detection allows for the distinction of analytes based on their mass, reducing the need for complete chromatographic separation, which leads to lower separation requirements compared to UV-based detection.[224]

Before starting the optimization process, it is beneficial to collect information about the analytes of interest, including their physicochemical properties such as solubility and pK_a values.[225] This also includes information about any existing method previously employed for similar analyses.[226] Additionally, it is advisable to identify the objective for method optimization, such as minimizing analysis time or separating multiple analytes. This background knowledge enables the identification of parameters with the most impact, allowing for a prioritization of these parameters during the optimization process.[227]

However, given the complexity of HPLC method development, it is important to consider that many parameters are dependent on each other, resulting in interaction effects. To address this complexity, more laboratories are adopting multivariate approaches, instead of the previously common one-factor-at-a-time (OFAT) approaches.[227] Statistical planning techniques, such as Design of Experiments (DoE), help minimize the number of experiments needed by systematically examining multiple factors simultaneously.[228] Several reports that utilize DoE methods to optimize critical factors in parallel, including the volume of organic solvent, the pH of the mobile phase, temperature, and flow rates, have been reviewed by Latrous.[229] In LC-

MS method development, the factors for simultaneous examination additionally comprise MS parameters such as collision energy, ion mode or capillary voltage.[230] These DoEs are part of the analytical quality by design (AQbD) concept, which provides a systematic approach to the development of analytical methods, as emphasized by FDA (Food and Drug Administration) and ICH (International Council for Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use) guidelines.[227,231] AQbD principles involve defining analytical method objectives and identifying critical method parameters, followed by optimization through DoE and data interpretation to establish a method operable design region (MODR), where all performance criteria are achieved. The process concludes with validation for quality control based on parameters such as specificity and sensitivity, and planning of control strategies.[227,232] In brief, the optimization of HPLC methods by careful consideration of interconnected parameters and the application of statistical strategies enables the development of robust methods that increase the analysis efficiency.

1.5.2.2. Triple quadrupole mass spectrometry

Triple quadrupole mass spectrometers are characterized by their excellent sensitivity. The detection of several analytes via their distinctive fragmentation patterns enables enhanced selectivity and reduced interference from coeluting compounds and matrix components.[205,233]

After chromatographic separation, the eluate enters the coupled triple quadrupole mass spectrometer. In a first step, the analytes are ionized and dissolved in an upstream electrospray ion source. Electrospray ionization (ESI) begins with the formation of charged droplets at the electrospray capillary tip containing the analyte solution. The application of high voltage to the capillary results in the generation of an electric field that is responsible for the formation of these droplets. As the droplets move toward the counter electrode, solvent evaporation occurs, causing a reduction in size and an increase in the charge-to-volume ratio. Upon reaching the Rayleigh limit, droplets undergo a phenomenon known as Coulomb fission, resulting in the splitting of the droplets into smaller, highly charged droplets. This process continues, leading to the production of gas-phase ions. Two main mechanisms have been proposed for the final ion formation process. The ion evaporation model postulates that ions are directly emitted from the surface of small, highly charged droplets. In contrast, the charged residue model suggests that droplets are formed containing only a single analyte ion, and that complete solvent evaporation results in a single gas-phase ion (reviewed by Kebarle and Verkerk).[234] The ionization process is influenced by several chromatographic parameters. The solvent composition has an impact on the formation and transfer of ions, with higher percentages of

organic solvent resulting in improved ionization efficiencies.[235] Additionally, the use of additives, such as salts and buffers, as well as the pH of the mobile phase, can enhance or suppress the ionization efficiency.[236–239] Furthermore, the flow rate influences the sensitivity of ESI-MS by affecting droplet size and spray stability.[240–242] The device-specific ion source parameters such as capillary voltage, nebulizer pressure, sheath gas flow and temperature, drying gas flow and temperature, and nozzle voltage influence key processes in the ion formation such as effective desolvation, droplet charging and formation, and the focusing and concentration of ions.[243–246]

The formed gas-phase ions are accelerated and transferred from the electrospray chamber at atmospheric pressure through an ion capillary and an octopole into the first quadrupole, which is driven by applying a fragmentor voltage. The first quadrupole serves as a mass filter, enabling the selective transmission of ions with predefined *m/z* characteristics. The filtered, so-called precursor ions then enter the collision cell, which is often referred to as the second quadrupole. However, contemporary instruments employ a hexapole in lieu of the traditional quadrupole.[247] In a process known as collision-induced dissociation (CID), the ions collide with collision gas, producing characteristic product ions. The latter enter the third quadrupole, which serves as a mass filter. Only ions with pre-defined *m/z* characteristics can pass through to the subsequent detector. The application of diverse scan modes enables the identification of molecule-specific mass transitions, which can then be employed to sensitively detect and quantify RNA modifications. Such examination of molecule-specific mass spectrometric behavior is facilitated by the use of reference standards. The scan modes relevant in this thesis are depicted in Figure 1.8 and explained in the following section with regards to nucleoside and cap dinucleotide analysis. In the analysis of nucleosides, the mass detector is typically employed in positive ion mode.[248] However, for the analysis of cap dinucleotides, the use of positive and/or negative ion mode has been described.[45,199,200]

The full scan mode enables the identification of precursor ions during method development. In this configuration, the first quadrupole is set to total transmission ion (TTI) mode, which allows all ions to enter the system and pass through to the collision cell. In the subsequent stage, no collision energy is applied, thereby preventing any fragmentation. The third quadrupole is operated in scan mode, which reveals the *m/z* ratios of the ions entering the system. For structurally known molecules, the scan facilitates the identification of potential ion adducts and the determination of the charge status of the ionized molecule.

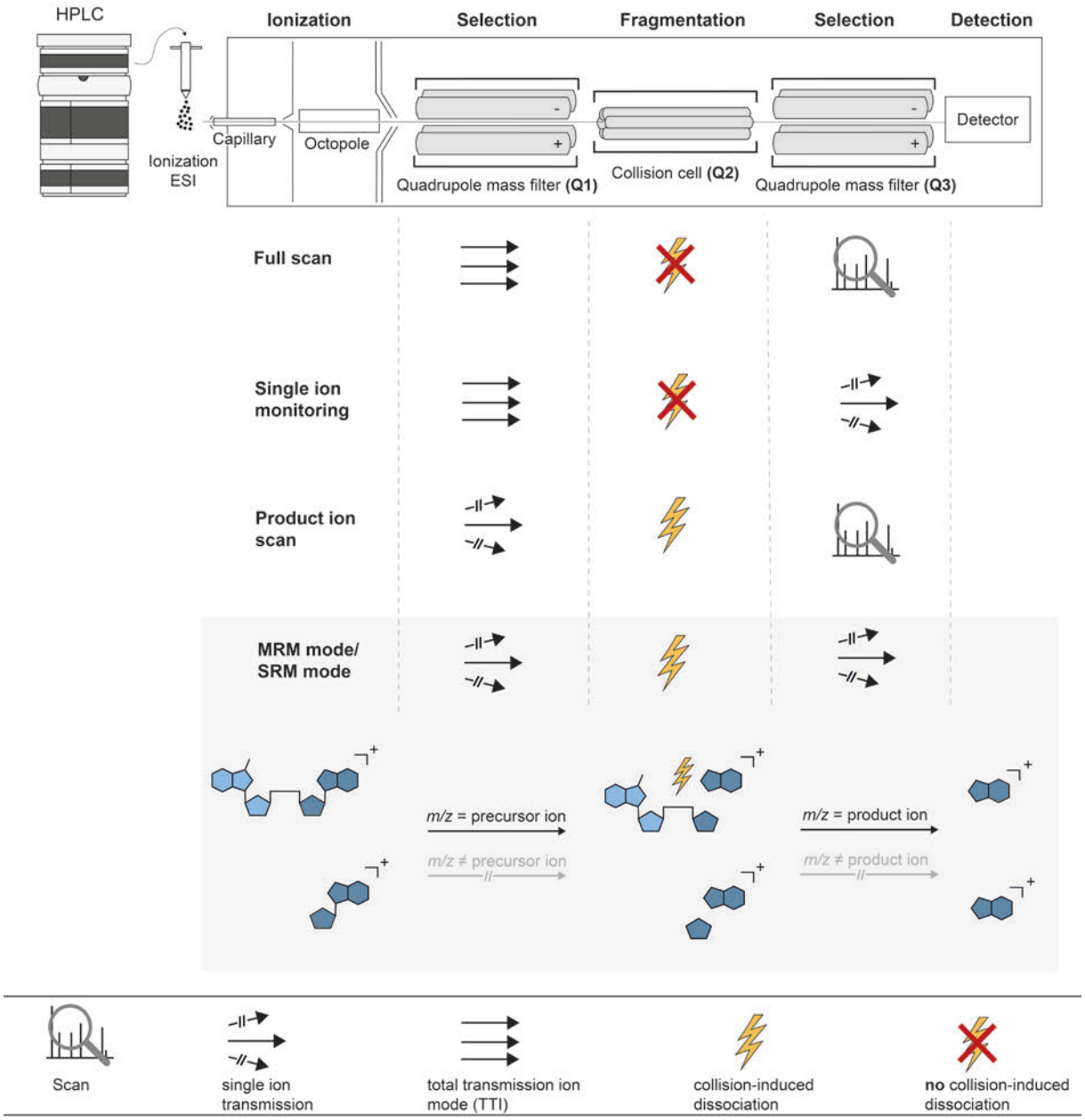

Figure 1.8. The upper part illustrates the setup of LC-MS/MS analysis using a triple quadrupole mass spectrometer. After HPLC separation, the eluent is ionized in the electrospray (ESI) ion source and enters the first quadrupole, which acts as a mass filter. In the collision cell, fragmentation is induced. The charged fragments then enter the third quadrupole, which again acts as a mass filter, and the selected ions are directed to the detector. The lower part shows different scan modes. The legend lists icons symbolizing the action modes of the quadrupoles, including total ion transmission, scan mode, or single ion transmission for pre-defined *m/z* ratios for quadrupole 1 and 3. In the collision cell, collision-induced dissociation may or may not be applied, depending on the selected mode. For MRM/SRM mode, the respective structures of cap dinucleotides and nucleosides are shown as examples.

The single ion monitoring (SIM) mode is used for method development and confirmation. Analogous to the full scan mode, the first quadrupole operates in TTI mode, and no collision is induced in the collision cell. In contrast to the full scan mode, the third quadrupole allows only

molecules with a pre-defined *m/z* ratio to pass through. Consequently, the results obtained from the full scan mode can be confirmed. Furthermore, this mode is utilized when optimizing molecule-specific mass parameters to test different fragmentor voltages.

The majority of nucleosides display a characteristic fragmentation pattern at the *N*-glycosidic bond.[249] In contrast, cap dinucleotides display a range of reported mass transitions.[45,199,200] This is partly attributed to the presence of two potential fragmentation sites at both *N*-glycosidic bonds, which challenges the prediction of the predominant fragmentation pathway. For examination of unknown fragmentation or verification of assumed fragmentation, a product ion scan can be applied. The identification of resulting product ions that are produced following fragmentation of known precursor ions is achieved by setting the first quadrupole to permit only those precursor ions with a pre-defined specific *m/z* ratio to pass. In the collision cell, collision energy is applied, leading to collision-induced dissociation. The third quadrupole operates in scan mode to identify resulting product ions.

The multiple reaction monitoring (MRM) mode is employed for analytes whose specific mass spectrometer characteristics have already been established. Only pre-defined precursor ions with a specific *m/z* pass through the first quadrupole and enter the collision cell, where CID occurs. The resulting characteristic product ions are then directed to the third quadrupole, where only ions with a specified *m/z* ratio can pass to the subsequent detector, ensuring the detection of only molecules with a pre-defined mass transition. The ability to monitor multiple mas transitions simultaneously gives the MRM mode its name. A variant is the dynamic MRM (dMRM) mode, which is characterized by the selection of additional retention time windows, enabling the detection of mass transitions exclusively within pre-defined time spans. This mode is particularly suitable when monitoring a high number of mass transitions, as this selection reduces the number of simultaneous analyzed mass transitions, which may otherwise limit sensitivity. The single reaction monitoring (SRM) mode operates similarly to the MRM mode, with the exception that only a single mass transition can be monitored in a sample run. The (d)MRM mode provides highest sensitivity, enabling the simultaneous detection of a plethora of RNA modifications, down to the sub-femtomolar range.[199,250] The use of external standards enables the measurement of calibration curves, facilitating absolute quantification. As several factors in addition to the quantity of analyte influence the MS signal intensity, including the efficiency of analyte ionization, salt loads, matrix effects, and instrument parameters, the addition of stable isotope labeled internal standards (SILIS) enables the compensation of occurring fluctuations and is therefore indispensable within the context of reliable absolute quantification.[250,251]

2. Motivation and Objectives

The 5' cap structure is an essential component of mRNA, fulfilling several crucial roles including protecting RNA from degradation by 5'-exonucleases and enhancing translation efficiency by recruiting the eukaryotic translation initiation factor 4E. Furthermore, it acts as a marker for self mRNA, helping to evade the innate immune response. The cap exhibits a range of structural variations in its methylation patterns, which are regulated during various cellular processes and are pivotal in modulating these functions.[109,122,196,252,253]

These aspects illustrate the significance of valid and reliable analytical tools for the detection and quantification of cap structures to provide deeper insights into the cap epitranscriptome. The low abundance of cap structures, limited to a maximum of one cap per mRNA molecule, underscores the necessity for a highly sensitive analytical approach capable of analyzing these structures in appropriate quantities of isolated mRNA. A significant objective of this thesis involved the development of an LC-MS/MS-based, highly sensitive method for the absolute quantification of various cap structures, with a particular emphasis on enabling the simultaneous detection of cap structures as well as other modified nucleosides, thus allowing for comprehensive analysis in a single run. To accomplish this, detailed chromatographic and mass spectrometric method optimization was aimed at in this work, as well as its application to biological samples of diverse origins.

Another objective of this work focused on isotope labeled cap dinucleotide standards. The strength of signals detected by mass spectrometry is not solely determined by the amount of the substance being analyzed, but also by various other elements, such as how effectively the analyte is ionized, the presence of salts, interactions with the sample matrix, and the specific instrument settings.[250,251] The incorporation of stable isotope labeled internal standards allows adjustment for any variations that may occur during the analysis process, making them essential for achieving accurate absolute quantification in MS experiments. Most current methods for the synthesis of isotope labeled and unlabeled cap dinucleotides involve intensive synthetic organic chemistry methods, with some potential for incorporating enzymatic strategies.[45,158,254,255] However, these approaches are frequently unfeasible for RNA biochemistry laboratories. Therefore, the development of a straightforward and widely implementable technique to produce isotope labeled cap dinucleotide standards was sought to fill this gap. This task ought to be addressed by employing deuterium exchange studies of

cap dinucleotides containing two purine nucleosides to evaluate the exchange behavior of C-8 hydrogens, with the aim of generating deuterated structures exhibiting a mass shift of +2. This included the evaluation of the suitability of this method in terms of completeness, stability, and applicability to differently methylated cap structures.

Most recently, the use of synthetic mRNA for therapeutic purposes was successfully demonstrated with the development and application of vaccines against COVID-19.[147,148] The presence of cap structures is also of great importance in this context, as they fulfil several crucial functions including improved RNA stability and translation efficiency as well as reducing the immunogenicity of the mRNA.[149] Capping of synthetic *in vitro* transcribed RNA can be performed either post-transcriptionally, by including enzymatic steps following IVT reaction, or co-transcriptionally by the addition of cap analogs into the IVT reaction mixture, but complete capping is not guaranteed. Uncapped mRNA is immunogenic and induces interferon response, thereby emphasizing the need for analytical tools to facilitate the quality control of mRNA therapeutics.[109,256] The aim of this work included the development of an LC-MS/MS approach to determine capping efficiencies of synthetic mRNAs by digesting the RNA to nucleoside level and subsequently quantifying the absolute amount of cap modification per RNA molecule. This method may prove to be a valuable tool in the study of suitable capping reaction parameters, in the comparison and evaluation of different cap analogs, as well as in the quality control of optimized standard capping procedures.

3. Results and Discussion

3.1. LC-MS/MS method development for absolute quantification of RNA caps

Tools to detect and quantify cap structures are essential for advancing knowledge of the cap epitranscriptome and its role in RNA regulation. This study focused on the development of a sensitive and robust LC-MS/MS method for the detection and quantification of cap structures in eukaryotic RNA, with the intention of detecting cap structures as cap dinucleotides. In the context of biological sample preparation, the key aspect to obtain these cap dinucleotides is the digestion of the RNA of interest to nucleosides while the triphosphate bridge of the cap structure is left intact. This is achieved through the enzymatic hydrolysis of RNA by NP1 and fast alkaline phosphatase (FastAP). A schematic workflow is illustrated in Figure 3.1.

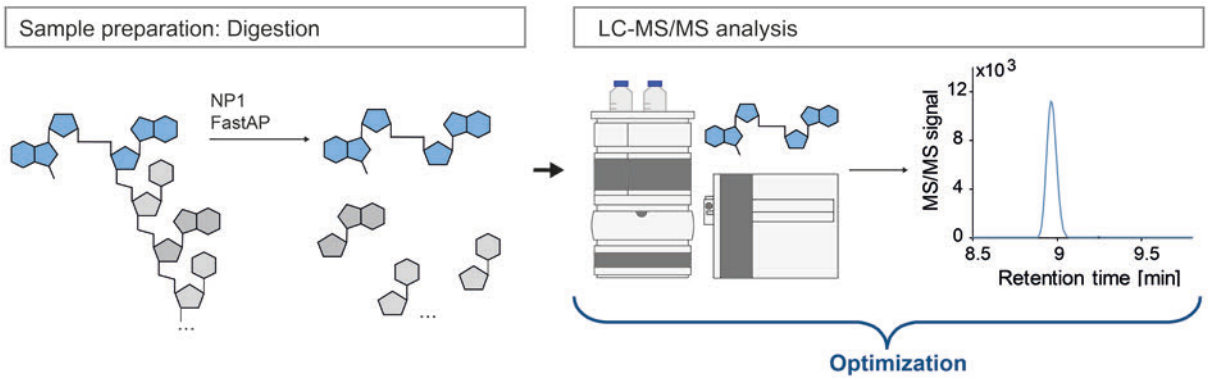

Figure 3.1. Schematic illustration of the workflow to analyze cap dinucleotides, starting with the digestion of capped RNA into nucleosides (gray) and cap dinucleotides (blue) using nuclease P1 (NP1) and fast alkaline phosphatase (FastAP). The subsequent LC-MS/MS analysis of cap dinucleotides is subject of a comprehensive method optimization.

The comprehensive LC-MS/MS method development detailed in the following chapters was conducted to optimize the detection of cap dinucleotides and focused on canonical caps with guanosine or 7-methylguanosine connected to adenosine or its methylation variants by a reverse 5'-5'-triphosphate bridge. Considering the individuality of the differently modified cap dinucleotides, the use of reference standards was inevitable to evaluate the mass spectrometric and chromatographic behavior of the corresponding substances, to ensure clear allocation within the parallel analysis of different species of cap dinucleotides. [redacted] synthesized and

kindly provided synthetic standards as well as stable isotope labeled variants containing a deuterated methyl group at the *N*7-position of guanosine. An overview of the utilized cap dinucleotide standards is depicted in Figure 3.2.

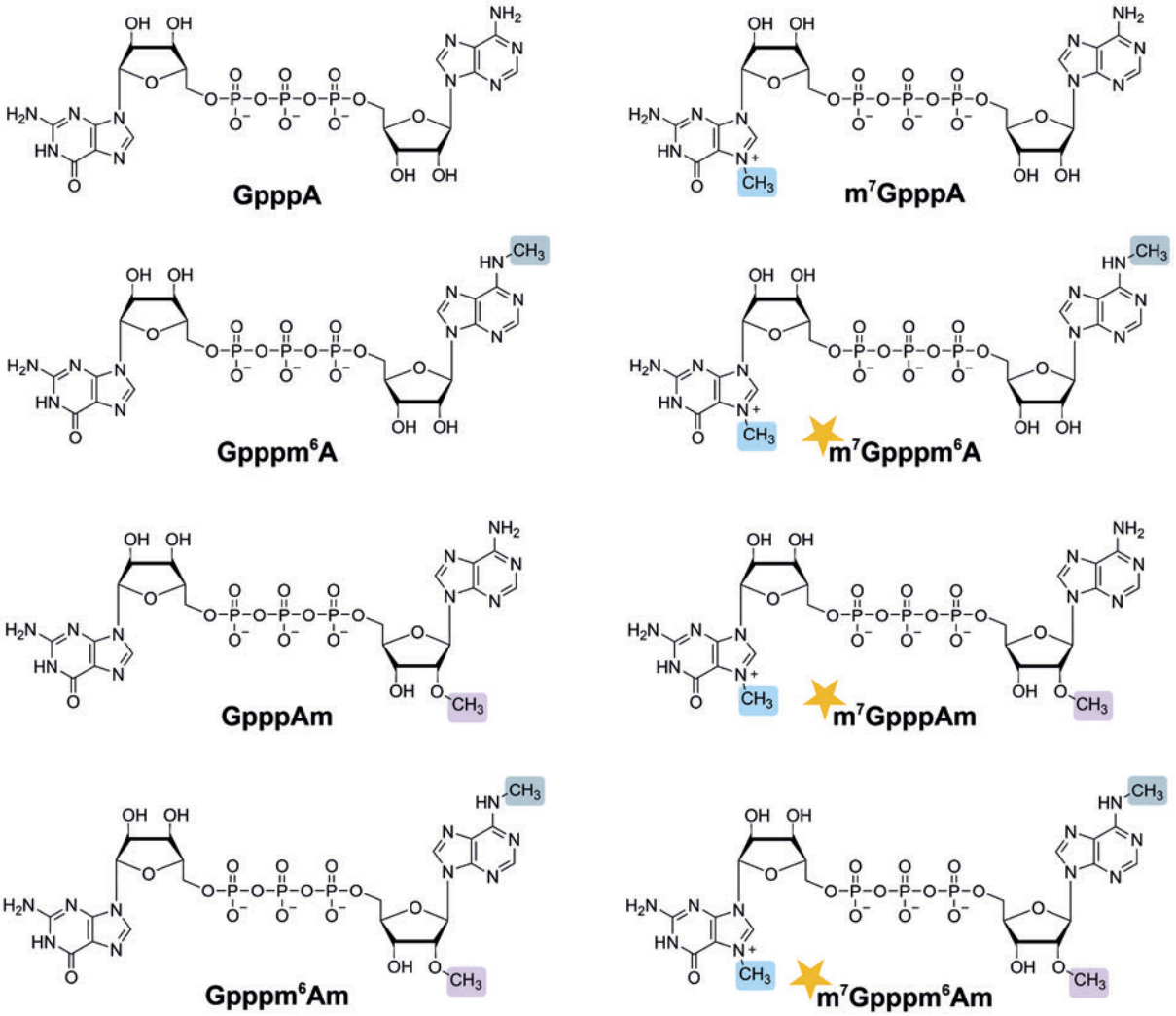

Figure 3.2. Overview of utilized cap dinucleotide standards. The methyl groups are highlighted in blue (*N*7-methylation of G), gray (*N*6-methylation of A), and purple (2'*O*-methylation of A). The cap dinucleotide standards for which a $(D_3C)^7$G-labeled variant exists, are highlighted with a yellow star.

3.1.1. Optimization process

The optimization procedure comprised mass spectrometric and chromatographic parameters. The mass spectrometric parameters can be classified into two categories: analyte-specific parameters, selected individually for each compound, and ion source parameters, which are instrument parameters applied to all analytes investigated.

The first parameters to be optimized were the analyte-specific parameters, including precursor and product ion *m/z*, fragmentor voltage, and collision energy, since detailed knowledge about the mass transitions is indispensable for detection of the cap dinucleotides by LC-MS/MS. This was followed by chromatographic optimization with respect to column temperature, ion strength, and pH of the mobile phase. In line with literature, ion source parameters were investigated in the last step of the optimization procedure since the process of electrospray ionization is affected by the physicochemical properties of the analytes and solvents, which in turn are influenced by chromatographic parameters, such as solvent composition, including pH of the mobile phase and the volume ratio of aqueous and organic solvent.[235,237,257]

3.1.1.1. Analyte specific parameters

The used LC-MS/MS system consists of an HPLC connected to a triple-quadrupole mass spectrometer equipped with an electrospray ionization source. To achieve high sensitivity, a quantification method operating in MRM mode was developed, which is characterized by enabling the simultaneous detection of several analytes based on their specific mass transitions. Given the extensive experience with the positive ion mode in nucleoside analytics and the successful implementation of this mode for detecting cap dinucleotides, as demonstrated in the literature, the mass spectrometer was operated in positive ion mode.[199,201]

Within this optimization process, pure standards were injected into the LC-MS system. To ensure unequivocal identification of individual cap dinucleotides and investigation of their behavior in mass spectrometric analysis, the cap dinucleotide standards were optimized one by one which, moreover, allowed to use a capillary loop instead of an HPLC column to reduce analysis time since chromatographic separation was not required. The optimization process is representatively shown for m^7GpppA. First, the MS was operated in full scan mode to detect precursor ions. The most abundant peak with an *m/z* of 787 was identified as precursor ion and corresponds to the positively charged m^7GpppA with a charge of $z = 1$. Minor signals were detected for *m/z* 166, which corresponds to a positively charged 7-methylguanine and resulted from in-source fragmentation and for *m/z* 394, which is a doubly protonated m^7GpppA (Figure 3.3 A).

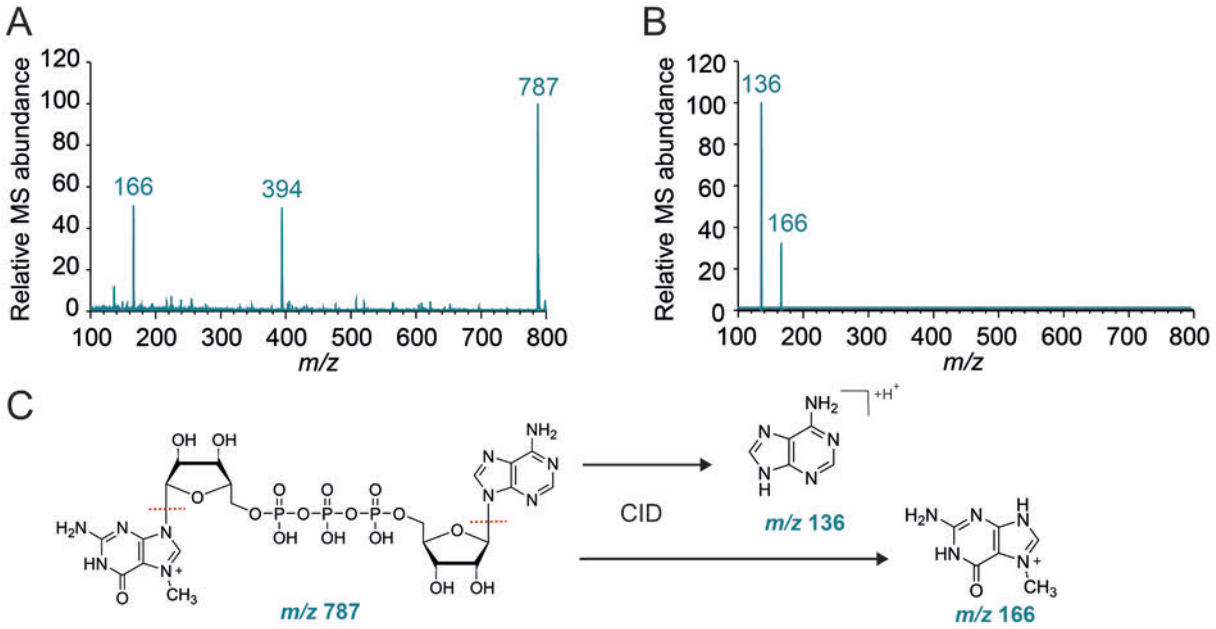

Figure 3.3. A.) Full scan mass spectrum of m^7GpppA standard. B.) Mass spectrum of a product ion scan with an *m/z* of 787 as precursor ion. C.) Collision-induced dissociation (CID) of m^7GpppA (*m/z* 787) resulting in the corresponding nucleobase product ions of adenine (*m/z* 136) and 7-methylguanine (*m/z* 166).

The next step of the optimization focused on the fragmentor voltage to find a balance between rapid transfer from the electrospray chamber into the mass spectrometer, to generate a strong MS signal, and minimizing unwanted fragmentation caused by collisions with other molecules at high speeds. Therefore, the MS was operated in SIM mode with different fragmentor voltages applied, ranging from 50 to 250 V. The one yielding maximum abundance was selected. Subsequently, a product ion scan was conducted to detect resulting product ions. Precursor *m/z* and fragmentor voltage were set to the already optimized parameters and a collision energy of 70 V was applied while the third quadrupole was operated in scan mode. Two major signals were detected corresponding to the nucleobases adenine (*m/z* 136) and 7-methylguanine (*m/z* 166) resulting from fragmentation at the *N*-glycosidic bonds (Figure 3.3 B,C), with *m/z* 136 being the most abundant signal which was therefore selected as product ion for further experiments. In the final step, the collision energy was optimized as it regulates the molecule-specific CID within the collision cell by operating in SRM mode with the previously selected mass transition and fragmentor voltage. Different collision energies in a range of 10 to 90 V were applied and the collision energy resulting in the maximum abundance was chosen. An overview of all optimized parameters is summarized in Supplement, Table S1.

3.1.1.2. Chromatographic parameters

After ensuring detection of the cap dinucleotides by the mass spectrometer, the chromatographic part was investigated further. Recently, Muthmann *et al.* published an LC-MS/MS approach to relatively quantify GpppG and m^7GpppG in synthetic mRNA.[199] The chromatographic parameters, including the use of a Poroshell C18 column as well as aqueous ammonium acetate solution and acetonitrile as mobile phases, were used as a starting point with minor changes in the form of a lowered flow rate and gradient adjustments to also tackle the more lipophilic and therefore later eluting adenosine containing cap dinucleotides. For detection, the previously optimized analyte specific mass parameters were used together with ion source parameters optimized for nucleoside analysis.[258] Based on these starting conditions, a series of experiments was conducted to determine the optimal composition of the mobile phases (ion strength of ammonium acetate and pH value) and column temperature for the analysis of a mixture of cap dinucleotide standards. The experiments involved the step-by-step testing of different parameters and the evaluation of the resulting data in terms of abundance, peak separation, and peak shape. A major focus was placed on the chromatographic separation of m^7GpppA-GpppAm and m^7Gpppm^{6}A-Gpppm^{6}Am, given that they have the same mass transition, respectively.

3.1.1.2.1. Ion strength

As previously stated, ammonium acetate salt in the aqueous mobile phase and acetonitrile as the organic eluent was utilized, following the method described by Muthmann *et al*. The first parameter investigated in this study was ion strength, focusing on its impact on analyte retention and signal abundance in mass spectrometry. Sufficient buffering capacity is essential to prevent retention time shifts, as the ionization degree of analytes affects their retention in the column.[214] While high ion strength can improve peak shapes, it also leads to an inverse relationship between buffer concentration and signal abundance, as increased salt concentrations can reduce ionization efficiency.[216,238,239] To explore these effects, the starting parameter of 20 mM ammonium acetate, as well as lower concentrations of 5 and 10 mM, were examined regarding their influence on signal abundance and retention by injecting a mixture of cap dinucleotide standards. The resulting chromatograms are depicted in Figure 3.4.

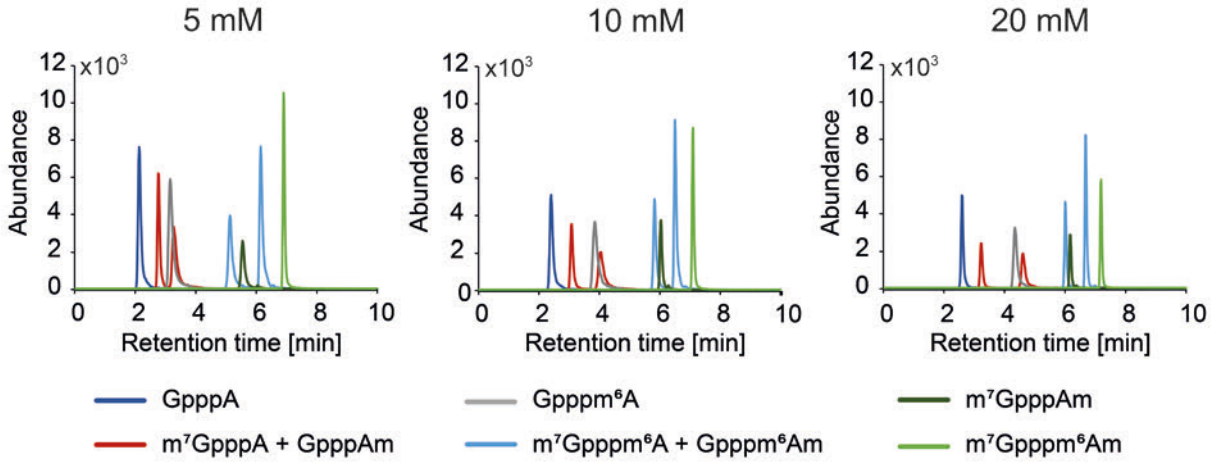

Figure 3.4. Overlay of extracted ion chromatograms of cap dinucleotides resulting from the analysis of a mixture of cap dinucleotide standards with a gradient of acetonitrile and aqueous mobile phases containing various amounts of ammonium acetate.

The resolution is used as a tool to assess the separation of two peaks and calculated according to equation 1, where R_s is the resolution, t_R is the retention time in min, and w_h is the peak width at the baseline in min. The aim was to achieve a value > 1.5 for the resolution since this indicates baseline separation.

(1) $$R_s = \frac{1.18 \cdot (t_{R2} - t_{R1})}{w_{h2} + w_{h1}}$$

A special focus was on the clear separation of the cap dinucleotides with identical mass transitions (m^{7}GpppA and GpppAm, m^{7}Gpppm^{6}A and Gpppm^{6}Am). Additionally, the adjacent peaks that show overlapping in the chromatograms (Gpppm^{6}A and GpppAm, m^{7}Gpppm^{6}A and m^{7}GpppAm) were examined in more detail. In Figure 3.5 A, it is apparent that the cap pairs with the same mass transition were baseline separated in each condition, although the resolution between m^{7}GpppA and GpppAm was reduced with decreasing salt concentration. The peak pairs Gpppm^{6}A-GpppAm and m^{7}Gpppm^{6}A-m^{7}GpppAm showed resolutions of less than 1.5 at an ammonium acetate concentration of 20 mM and thereby did not meet the requirements for baseline separation. While the peak pair Gpppm^{6}A-GpppAm showed even lower resolution with decreasing salt concentration, the opposite effect was observed for the peak pair m^{7}Gpppm^{6}A-m^{7}GpppAm. Here, the peaks were baseline separated at a salt concentration of 5 mM NH_4OAc. In a next step, the effect of decreasing salt concentrations on the signal abundance was further investigated and compared for each cap dinucleotide. Figure 3.5 B details the correlation of lower salt concentration with enhanced signals, showing gain factors ranging from 1.6 to 2.3 compared to the initial value (20 mM) for all cap dinucleotides.

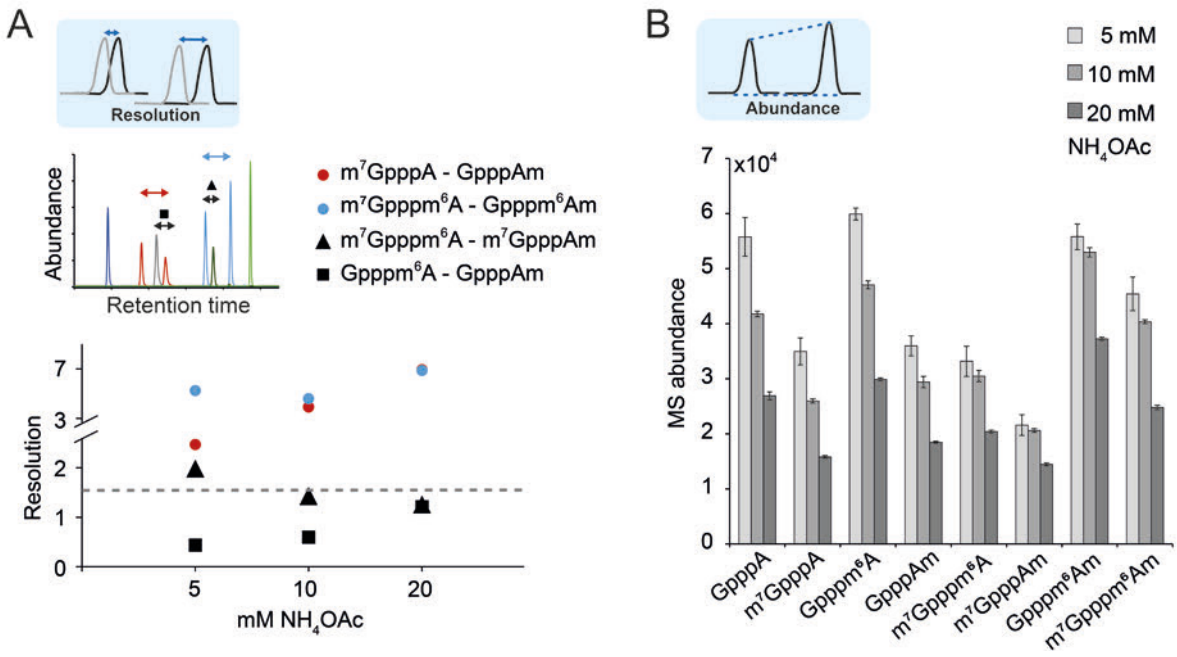

Figure 3.5. A.) Resolution of the peak pairs: m^7GpppA + GpppAm (red), m^7Gpppm^{6}A + Gpppm^{6}Am (light blue), m^7Gpppm^{6}A + m^7GpppAm (black triangle), and Gpppm^{6}A + GpppAm (black square), analyzed using aqueous mobile phases containing various amounts of ammonium acetate. The threshold of 1.5 is indicated by a gray dashed line. The resolutions were calculated from one representative technical replicate out of the three used to investigate MS abundances. B.) Peak areas (expressed as MS abundance) of cap dinucleotides analyzed using aqueous mobile phases containing various amounts of ammonium acetate. The average values from three technical replicates are shown, with error bars indicating standard deviation.

Taken together, the investigation of the change in resolution due to different salt concentrations in the aqueous mobile phase yielded contrasting results for different peak pairs. The primary objective of achieving chromatographic baseline separation, essential for cap dinucleotides with identical mass transitions to ensure method specificity, has been accomplished for all considered ammonium acetate concentrations. However, the detailed examination of the adjacent peak pairs revealed that no baseline separation was achieved for Gpppm^{6}A and GpppAm at the tested salt concentrations. Nevertheless, unambiguous analysis remains possible due to the specific mass transitions. A considerable advantage in signal intensity was observed for low salt concentrations, in alignment with literature, that indicates an inverse correlation between salt concentration and signal abundance in ESI-MS.[216,238,239] Since the aim of the optimization was to develop a highly sensitive method for the absolute quantification of cap dinucleotides, a concentration of 5 mM ammonium acetate in the aqueous mobile phase was chosen for further experiments.

3.1.1.2.2. pH

The subsequent step involved a more detailed examination of the pH value of the aqueous mobile phase, as this has a significant influence on the chromatographic separation process. In RP-HPLC, analytes with higher hydrophobicity are more strongly retained. Consequently, the retention of ionized analytes is lower than that of their neutral forms, which can result in significant peak broadening when partial ionization of the analyte occurs.[214] Furthermore, it has been demonstrated that the pH value also affects the ionization process in ESI-MS,[237,259,260] thereby highlighting the importance of conducting a more detailed investigation of the impact of the pH of the mobile phase. Based on the LC-MS/MS method of Muthmann *et al.* which uses a pH of 6 for the aqueous mobile phase and considering the pH limit of the used column (pH < 8), additional pH values within the range of 4-7 were considered.[199] The aqueous mobile phases were prepared with the optimized concentration of 5 mM NH_4OAc and the pH was adjusted with acetic acid or ammonium hydroxide. The analysis of a mixture of the cap dinucleotide standards revealed a significant impact of pH on the retention of the analytes (Figure 3.6).

At pH 4 and 5, extensive peak broadening was observed, with peak widths spanning several minutes. Consequently, these two pH values were excluded from further examination. The chromatograms obtained from the analysis of cap dinucleotide standards with the aqueous mobile phase at pH 6 and pH 7 were then subjected to further analysis with regard to peak shape, abundance and resolution. The latter was calculated using equation 1, as described in the prior chapter 3.1.1.2.1. As outlined before, the analysis was focused on the cap dinucleotides with identical mass transitions and the adjacent peaks that showed overlapping in the chromatograms (Figure 3.7 A). The resolution was elevated at pH 7 compared to pH 6 for all analyzed peak pairs. Nevertheless, the adjacent signals of Gpppm^{6}A and GpppAm still lacked baseline separation, whereas the peak pair m^{7}Gpppm^{6}A-m^{7}GpppAm and the cap dinucleotides m^{7}Gpppm^{6}A-Gpppm^{6}Am as well as m^{7}GpppA-GpppAm were separated at both pH values. In a next step, the peak shapes were assessed by calculating the symmetry factors according to equation 2, where A_s is the symmetry factor, $w_{0.05}$ is the peak width at 5% of the peak height, and $d_{0.05}$ is the distance between the leading edge of the curve and the peak maximum, measured at 5% of the peak height.

(2) $$A_s = \frac{w_{0.05}}{2d_{0.05}}$$

A symmetry factor of $A_s = 1$ indicates peak symmetry, $A_s < 1$ shows peak fronting, while $A_s > 1$ represents peak tailing. The European Pharmacopoeia (Ph. Eur.) recommends symmetry

factors between 0.8 and 1.8 for peaks used in quantification.[261] This specified range served as a guideline for target symmetry factors.

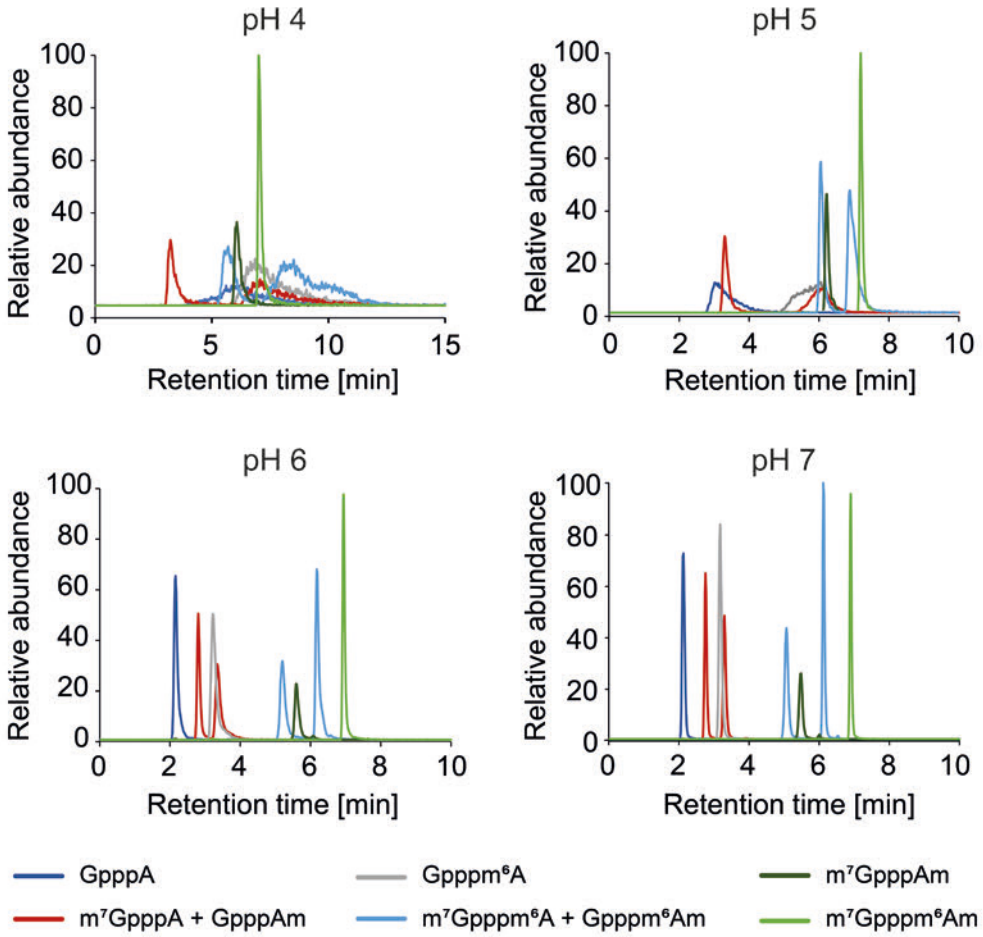

Figure 3.6. Overlay of extracted ion chromatograms of cap dinucleotides resulting from the analysis of a mixture of cap dinucleotide standards with a gradient of acetonitrile and aqueous mobile phases with various pH values.

As illustrated in Figure 3.7 B, tailing was observed at both pH values, albeit more pronounced at pH 6. At this pH, three distinct cap dinucleotides (GpppAm, Gpppm^{6}A, and Gpppm^{6}Am) displayed symmetry factors exceeding 1.8. In contrast, at pH 7, all cap dinucleotides exhibited symmetry factors within the targeted range, with the maximum value reaching only 1.4. In a next step, the signal abundances at pH 6 and 7 were further investigated by injecting different amounts of a mixture of cap dinucleotide standards into the LC-MS. The slope of the linear fit resulting from the plotting of peak areas against the amount injected was employed as a measure of signal abundances, with a higher slope representing increased abundances. A comparison of the slopes of the linear fit for each cap dinucleotide is illustrated in Figure 3.7 C, demonstrating comparable results at both pH levels.

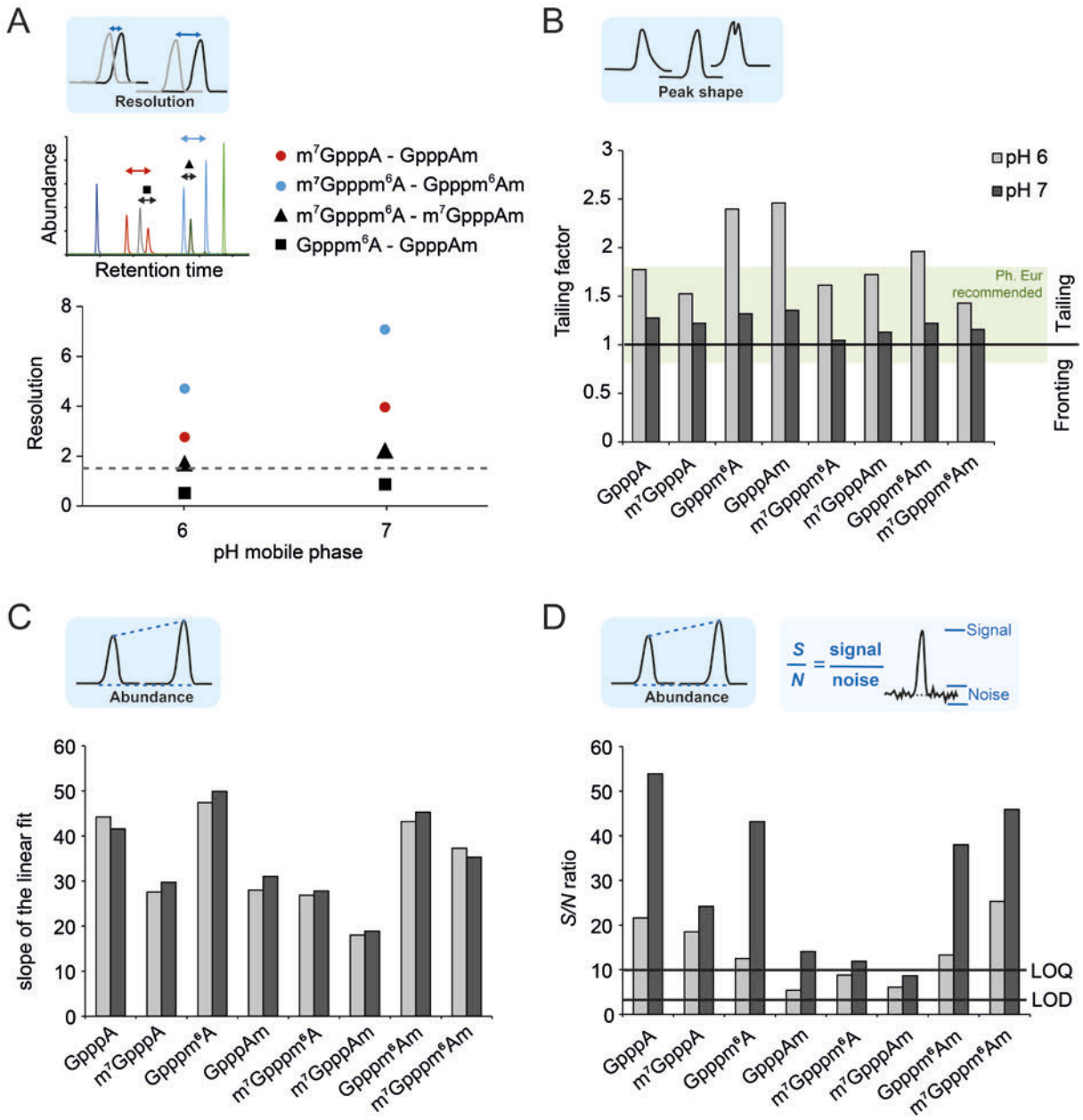

Figure 3.7. A.) Resolution of the peak pairs: m^{7}GpppA + GpppAm (red), m^{7}Gpppm^{6}A + Gpppm^{6}Am (light blue), m^{7}Gpppm^{6}A + m^{7}GpppAm (black triangle), and Gpppm^{6}A + GpppAm (black square), analyzed using aqueous mobile phases at pH 6 and 7. The threshold of 1.5 is indicated by a gray dashed line. B.) Investigation of peak shapes calculated as tailing factors at aqueous mobile phase pH values of 6 and 7. The range of $0.8 \leq A_s \leq 1.8$ recommended in the Ph. Eur., is highlighted in green. C.) Comparison of signal abundances obtained from analyses with mobile phases at pH 6 and 7, represented by the slopes of linear regression lines derived from plotting peak areas against injected sample amounts. D.) Signal-to-noise ratios for a mixture of cap dinucleotide standards (10 fmol each) analyzed using mobile phases at pH 6 and 7. Black lines indicate the limit of quantification (LOQ, *S/N* = 10) and the limit of quantification (LOQ, *S/N* = 3).

In the final stage of the comparison of signal abundance between pH 6 and 7, the *S/N* ratios of the cap dinucleotides at a low injected quantity of 10 fmol were examined to compare the

height of the signals to the background noise. The *S/N* ratios obtained were evaluated based on whether they exceeded or fell below the limit of quantification (*S/N* = 10) or the limit of detection (*S/N* = 3). As detailed in Figure 3.7 D, the *S/N* ratios were considerably higher at pH 7 with only one cap dinucleotide falling below the LOQ, in comparison to pH 6, where three cap dinucleotides were below the LOQ at 10 fmol.

In conclusion, the analysis of various aqueous mobile phases containing 5 mM ammonium acetate at pH values of 4, 5, 6, and 7 demonstrated that pH 4 and 5 were unsuitable due to extensive peak broadening of certain cap dinucleotides. This observed peak broadening is likely attributable to the presence of partly ionized analyte molecules. However, the estimation of their overall charge is complicated by the presence of multiple ionizable groups, which exhibit a range of pk_a values between 3 and 8.[262,263] Further investigation of pH 6 and 7 showed similar results for both pH values in terms of resolution and signal abundance. Visual inspection of the chromatograms indicated that the peaks at pH 7 were significantly sharper and more symmetrical. This observation was confirmed by calculating the symmetry factors, which showed that the peaks at pH 7 had symmetry factors closer to the ideal value of 1. The benefit of this peak shape became clear when determining the *S/N* ratios. Despite comparable peak areas at pH 6 and 7, the sharper and thus higher peaks at pH 7 resulted in notably elevated *S/N* ratios, demonstrating enhanced sensitivity as LODs and LOQs are reached at lower concentrations. In accordance with the column's pH limits (pH < 8), the investigation of more alkaline mobile phases was restricted to pH 7.5. A comparison of abundance, peak shape, and resolution between pH 7 and 7.5 showed no reproducible improvements, prompting a return to using the mobile phase at pH 7.0 (see Figure S1).

It is noteworthy that the pK_a values of ammonium (9.25) and acetate (4.75) indicate a reduction in the buffering capacity of ammonium acetate at neutral pH values.[223] Given the rarity of alternative salts with buffering capabilities in the neutral range or their incompatibility with ESI-MS due to non-volatility (e.g. phosphates),[223,264] the aqueous mobile phase with 5 mM ammonium acetate at pH 7 was rated as the best option. This was based on the observation that the analytes demonstrated consistent chromatographic and mass spectrometric behavior, and the aforementioned benefits outweighed the potential limitations of reduced buffer capacity.

3.1.1.2.3. Influence of injection volumes

Injection volumes influence chromatographic separation. Smaller volumes injection volumes can produce Gaussian-shaped peaks by ensuring simultaneous analyte migration, while volume overloading may cause peak broadening. Accordingly, the impact of diverse injection volumes on the developed LC-MS method was investigated by injecting the same amount of a mixture of cap dinucleotide standards in different volumes. Figure 3.8 illustrates that the four latest eluting cap dinucleotides m^7Gpppm^6A, $m^7GpppAm$, $Gpppm^6Am$, and m^7Gpppm^6Am had consistent retention times across a range of injection volumes (1-50 µL), while in terms of peak shapes, the signals showed sharper forms at smaller injection volumes. The four earliest eluting cap dinucleotides demonstrated strong peak splitting, particularly at an injection volume of 50 µL, resulting in an additional peak at 1.8 min.

The theoretical dead time for the chromatographic system was calculated using equation 5, where t_0 is the dead time, d the column diameter, l the column length, and F the flow rate. Additionally, a correction factor of 0.66 was applied, which is typically used for columns packed with fully porous material.[265]

$$t_0 = \pi \cdot \left(\frac{d}{2}\right)^2 \cdot l \cdot 0.66 \cdot F \quad (5)$$

A theoretical dead time of about 2 min was calculated, which is close to the detected value of 1.8 min. Of note, practical determination by injecting an analyte that does not interact with the stationary phase, such as uracil in reverse-phase chromatography,[266,267] would provide a more precise measurement by accounting for the entire chromatographic system. The performed calculation led to the assumption that the observed retention time indeed represented the dead time and therefore a proportion of these cap dinucleotides did not interact with the column at all at high injection volumes. To address the issue of lacking interaction between the more polar cap dinucleotides and the column, the gradient was adjusted by reducing the proportion of acetonitrile at the initial stage, resulting in enhanced interaction with the column and consequently stronger retention. Of note, the adjusted gradient (Figure 3.8) was also modified in terms of higher maximum acetonitrile proportions resulting in an elevated overall gradient time. This was implemented regardless of addressing the challenge of early eluting polar cap dinucleotides, but to ensure the gradient was suitable for the analysis of biological samples, where rinsing with high concentrations of acetonitrile is necessary to avoid column carry over. The low rinsing of the previous gradient was designed to minimize analysis time while measuring only clean reference standards.

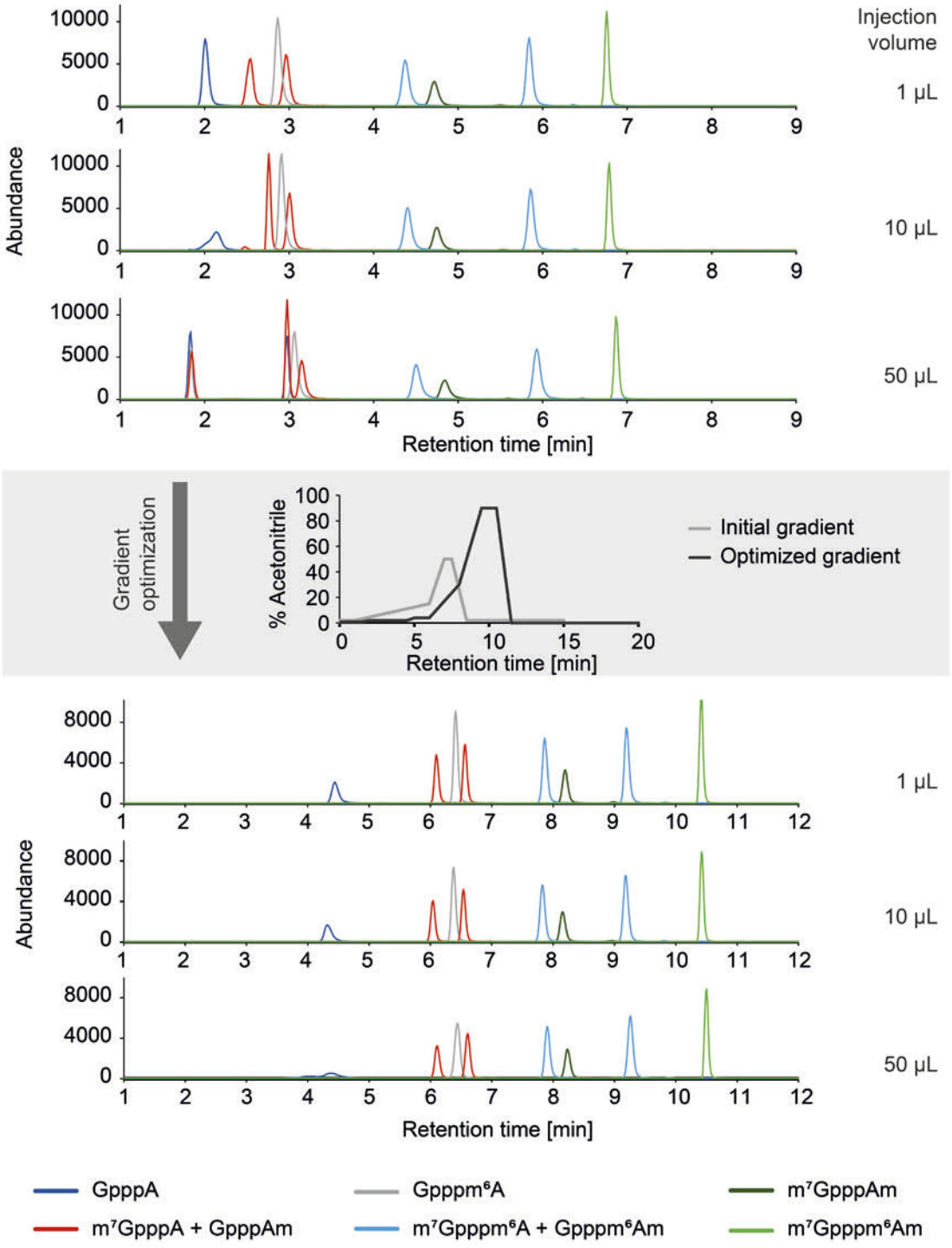

Figure 3.8. Overlay of extracted ion chromatograms of cap dinucleotides resulting from the injection of same amounts of a mixture of cap dinucleotide standards in different injection volumes. The upper three chromatograms were measured using the initial gradient, the lower three using the optimized gradient.

In Figure 3.8, it is apparent that the modified gradient resulted in higher retention times for all cap dinucleotides compared to the earlier gradient. The retention times remained consistent

over a range of 1-50 µL. Only the first eluting cap dinucleotide GpppA exhibited peak broadening and slight peak splitting at high injection volumes. The investigation of the impact of high injection volumes revealed that the initial gradient was unsuitable for more polar cap dinucleotides at high injection volumes due to the resulting extensive peak splitting. This issue was resolved by optimizing the gradient in terms of higher hydrophilicity in the initial stage, thereby ensuring a stronger interaction with the column. Ultimately, only GpppA exhibited peak broadening and slight peak splitting at high injection volumes, justifying the decision to proceed with the optimized gradient and to avoid high injection volumes when possible.

3.1.1.2.4. Column temperature

Column temperature is a crucial parameter in chromatography, influencing analyte behavior by affecting mobile phase pH, analyte pK_a, and the equilibrium between mobile and stationary phases. This impacts retention times and separation selectivity, highlighting the relevance of further investigating the impact of column temperature on the described LC-MS/MS method for cap dinucleotides. Column temperatures of 15 °C, 20 °C, 25 °C, and 35 °C were applied, and the chromatograms resulting from the injection of a mixture of cap dinucleotide standards were examined in terms of resolution, peak shape, and abundance.

As illustrated in Figure 3.9 A, the MS abundances were found to be within a similar range across all examined temperatures. However, specific temperature-dependent differences were observed for certain cap dinucleotides which exhibited slight preferences for lower temperatures compared to higher ones, as demonstrated by GpppA, m^7GpppA, Gpppm^{6}A, and Gpppm^{6}Am. Nevertheless, the results did not indicate a definitive preference for any single temperature. A similar outcome was observed in the analysis of peak shapes, utilizing peak widths measured at half maximum, instead of the previously used approach of calculating tailing factors, as this determination is significantly more time-efficient (Figure 3.9 B). Although some cap dinucleotides exhibited significantly smaller peak widths at specific temperatures (e.g., GpppA at 15 °C compared to 20 °C), it was not possible to identify a single temperature that was consistently superior for all analyzed cap dinucleotides. In the following step, the resolution of peak pairs was determined as detailed in the chapter 3.1.1.2.1., once more with a focus on the cap dinucleotides with identical mass transitions (m^7GpppA-GpppAm, m^7Gpppm^{6}A-Gpppm^{6}Am), as well as the adjacent peaks that show overlapping in the chromatograms (Gpppm^{6}A-GpppAm, m^7Gpppm^{6}A-m^7GpppAm).

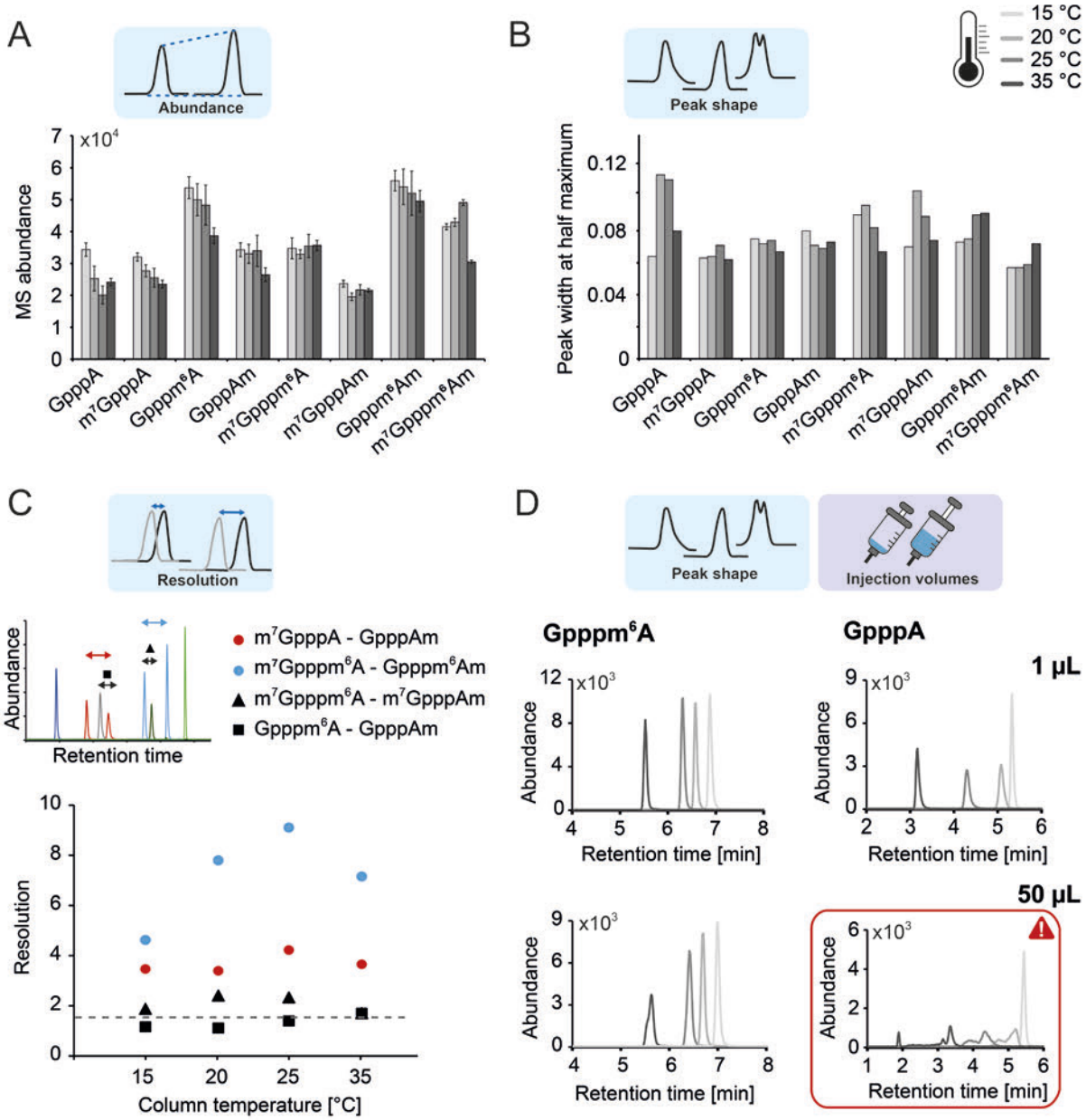

Figure 3.9. Examined temperatures are depicted in different shades of gray. A.) Peak areas (expressed as MS abundance) of cap dinucleotides analyzed at different column temperatures. The average values from three technical replicates are shown, with error bars indicating standard deviation. B.) Investigation of peak shapes as determined by peak width at half maximum at different column temperatures. The peak widths at half maximum were determined from one representative technical replicate out of the three used to investigate MS abundances. C.) Resolution of the peak pairs: m^7GpppA + GpppAm (red), m^7Gpppm^{6}A + Gpppm^{6}Am (light blue), m^7Gpppm^{6}A + m^7GpppAm (black triangle), and Gpppm^{6}A + GpppAm (black square), analyzed at different column temperatures. The threshold of 1.5 is indicated by a gray dashed line. D.) Overlay of extracted ion chromatograms of the cap dinucleotides GpppA and Gpppm^{6}A resulting from analysis at different column temperatures. The overlays are depicted for injection volumes of 1 µL and 50 µL for both cap dinucleotides, respectively.

Figure 3.9 C illustrates that the cap dinucleotide pairs with same mass transitions exhibited baseline separation with resolutions clearly above 1.5. Additionally, the peak pair m^7Gpppm^{6}A-m^7GpppAm was baseline separated at all investigated temperatures, in contrast to the peak pair Gpppm^{6}A-GpppAm which exhibited baseline separation only at 35 °C. At lower temperatures, the resolution was slightly below the targeted value of ≥ 1.5.

In a last step, the impact of temperature was investigated for high injection volumes. The graphical illustration of the chromatograms, obtained from the injection of a mixture of the cap dinucleotide standards in a volume of 50 µL at varying temperatures, revealed a distinct impact of temperature on the peak shape of GpppA (Figure 3.9 D). At 15 °C, GpppA eluted as a sharp peak, whereas at higher temperatures, a strong peak broadening and splitting was observed. The remaining cap dinucleotide standards showed pronounced peaks at all temperatures investigated, although slight peak broadening was observed for some cap dinucleotides at a column temperature of 35 °C in combination with the high injection volume of 50 µL (see Figure 3.9 D and Supplement, Figure S2).

In conclusion, the observed decrease in retention times with increasing temperatures aligns with the anticipated outcome, as the principle of Le Chatelier indicates that at higher temperatures, the equilibrium shifts towards the mobile phase, leading the analyte to remain in the mobile phase instead of interacting with the stationary phase.[220] Notably, the latest eluting cap dinucleotide, m^7Gpppm^{6}Am, demonstrated less sensitivity to temperature variations in terms of retention. It exhibited a shift of approximately half a minute between 15 °C and 35 °C, in comparison to GpppA, which shifted by approximately two minutes (Supplement, Figure S2). Accordingly, there was no significant reduction in analysis time achieved by increasing the temperature. However, the application of elevated temperatures combined with an adapted gradient could potentially result in a slight reduction in analysis time. By examining the peak retention at different temperatures with high injection volumes, a temperature of 15 °C yielded the most favorable results for the peak shape of GpppA, in contrast to higher temperatures, where pronounced peak splitting was observed with peak widths spanning 1-2 minutes. As the evaluated parameters abundance, peak shape, and resolution did not demonstrate a definite preference for one temperature, 15 °C was selected to ensure precise peak integration, even when injecting high volumes.

3.1.1.3. Ion source optimization

After optimizing the chromatographic parameters, the ion source parameters, including capillary voltage, nebulizer pressure, sheath gas flow and temperature, drying gas flow and temperature, and nozzle voltage, were also optimized to enhance effective ionization. The optimization of the parameters was conducted according to a one-factor-at-a-time strategy, whereby a single parameter was tested at varying levels while the remaining parameters were maintained at constant values. This optimization process was applied to a mixture of cap dinucleotide standards. The tested ranges were selected in accordance with the manufacturer's recommendations and the optimal parameters regarding signal abundance were selected in agreement with all analyzed cap dinucleotide standards. Capillary voltages were applied in steps of 500 V, ranging from 1500 to 5000 V. The highest signal intensity for all cap dinucleotides was observed at a capillary voltage of 2000 V (Figure 3.10 A). Drying gas temperatures were monitored in the range of 200-350 °C with steps of 30 °C and drying gas flows from 4-12 L/min with steps of 2 L/min. The highest abundances were observed at a temperature of 350 °C and a flow rate of 4 L/min (Figure 3.10 B,C). The examination of nozzle voltages between 0 and 2000 V, with steps of 500 V, revealed a negative impact of the applied nozzle voltage on signal intensities, with highest signal intensities obtained at 0 V (Figure 3.10 D). Nebulizer pressures were applied in the range of 25-60 V, with steps of 5 V and the highest signals were detected at 60 V (Figure 3.10 E). The impact of sheath gas temperature and sheath gas flow was tested for 200-400 °C, with steps of 50 °C and for 9-12 L/min with steps of 1 L/min, respectively. A sheath gas flow of 11 L/min yielded the highest signals (Figure 3.10 F). The different cap dinucleotides displayed distinct behavior when exposed to elevated sheath gas temperatures. While some exhibited the highest signal abundances at 400 °C, others reached their optimal abundances at 300 °C. Therefore, a compromise temperature of 350 °C was selected (Figure 3.10 G). Following the selection of the optimal parameters, a comparative measurement between the initial and the optimized source parameters was conducted to evaluate the impact of the optimization procedure and revealed increases of respective abundances by factors ranging from 1.3 to 2.1 (Figure 3.10 H).

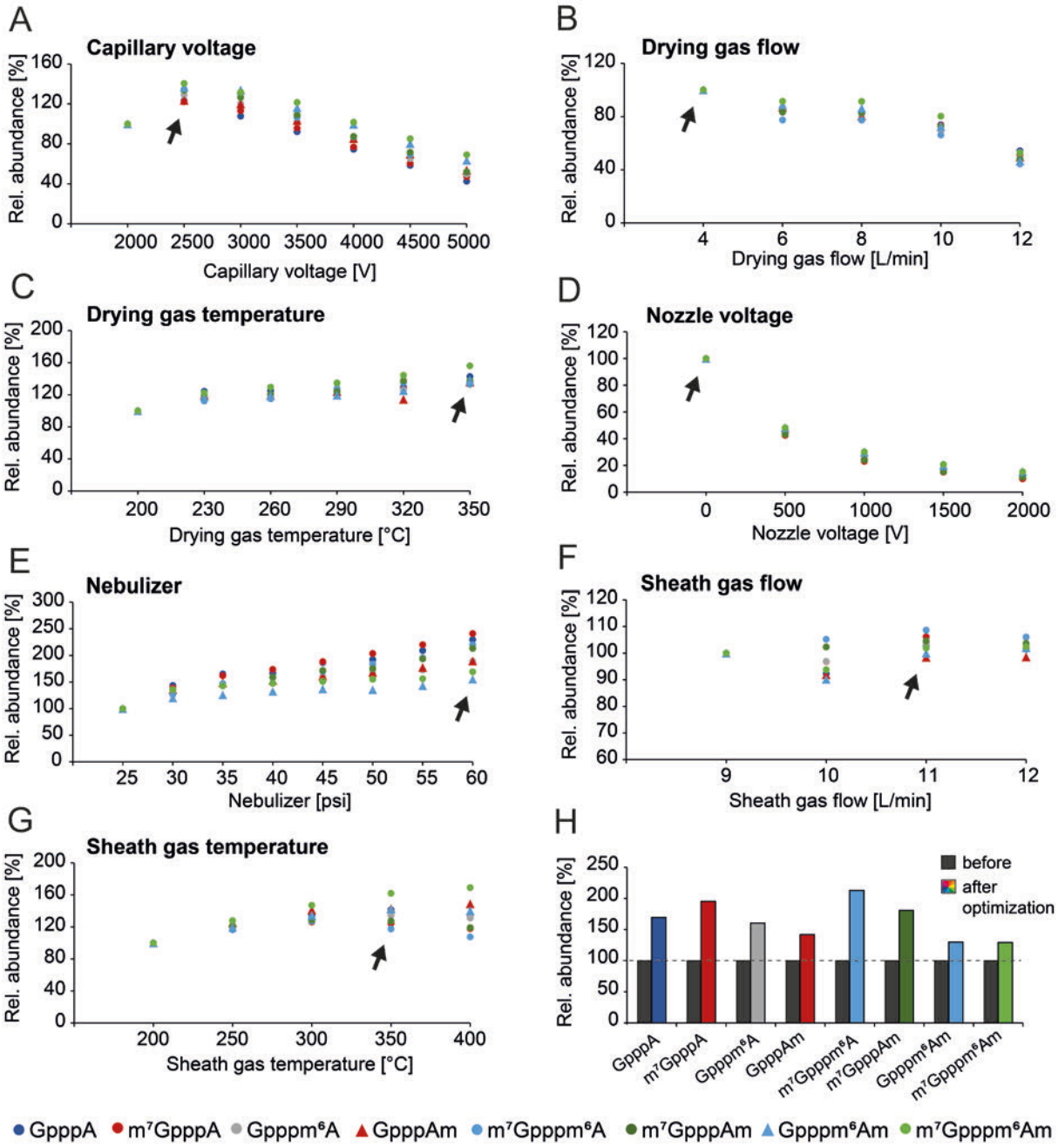

Figure 3.10. Relative abundances of cap dinucleotides towards different ion source parameters, showing A.) capillary voltages, B.) drying gas flows, C.) drying gas temperatures, D.) nozzle voltages, E.) nebulizer pressures, F.) sheath gas flows, and G.) sheath gas tempeatures. The lowest value for each parameter is set to 100%, respectively. The selected parameters are marked with a black arrow. H.) Relative comparison of abundances before (dark gray, set to 100%) and after (colored) ion source optimization.

3.1.1.4. Simultaneous detection of cap dinucleotides and (modified) nucleosides

The objective of this work was to develop a method for not only detecting but also quantifying cap structures. This involves determining the amount of cap dinucleotides in relation to the amount of RNA injected, which is calculated based on the quantity of main nucleosides detected. In addition, the simultaneous quantification of cap dinucleotides and other modified nucleosides was aimed for. To this end, it was tested whether the LC-MS method developed for cap dinucleotides is also capable of detecting and separating both main nucleosides and modified nucleosides. It is noteworthy that at this stage, a slight adjustment was implemented to the gradient to simplify and shorten the latter without affecting the retention of the cap dinucleotides. The resulting MS chromatogram of the simultaneous injection of cap dinucleotide and nucleoside modification standards is depicted in Figure 3.11 A, and a UV chromatogram resulting from the injection of the main nucleosides cytidine, uridine, guanosine, and adenosine is illustrated in Figure 3.11 B. Together, they demonstrate the capability of the developed method to simultaneously detect cap dinucleotides and modified nucleosides, as well as to separate the main nucleosides.

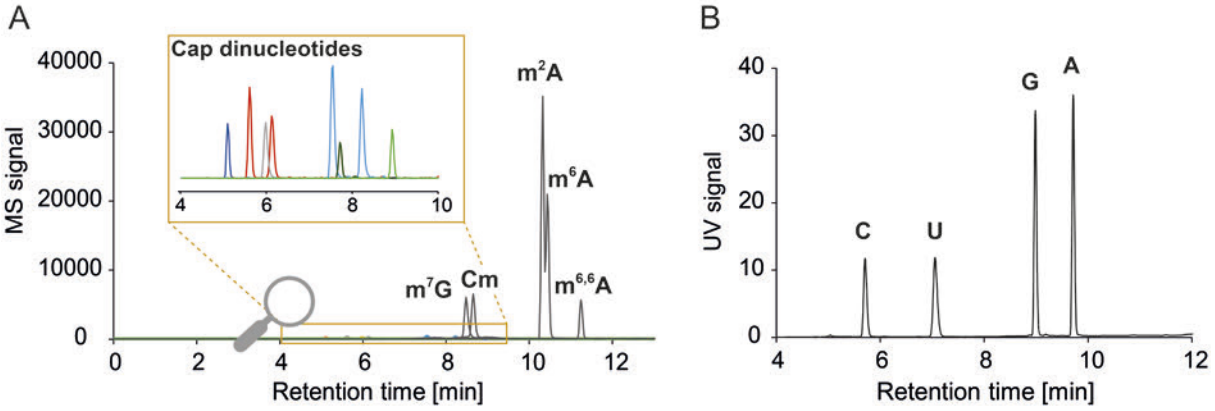

Figure 3.11. A.) Overlay of extracted ion chromatograms resulting from the simultaneous injection of 10 fmol of cap dinucleotides and modified nucleosides. The peaks of the modified nucleosides are assigned and abbreviated as m^7G – 7-methylguanosine, Cm – 2'*O*-methylcytidine, m^2A – 2-methyladenosine, m^6A – *N*6-methyladenosine, and $m^{6,6}A$ – *N*6,*N*6-dimethyladenosine. The cap dinucleotides are depicted in different colors: GpppA (dark blue), m^7GpppA and GpppAm (red), Gpppm^{6}A (gray), m^7Gpppm^{6}A and Gpppm^{6}Am (light blue), GpppAm (dark green), and m^7Gpppm^{6}Am (light green). B.) UV chromatogram of the main nucleosides cytidine (C), uridine (U), guanosine (G), and adenosine (A).

3.1.1.5. Instrument detection limits

After completing the LC-MS/MS optimization process and ensuring separation and detection of nucleosides, the instrument detection limits (IDLs) were determined for all analyzed cap dinucleotides and compared between the initial method and the optimized method. This was done to assess the analytical sensitivity of the optimized method and to evaluate the improvement resulting from the optimization process. The instrument detection limit is defined as the lowest amount of sample required to distinguish a measured signal from the background noise with a confidence interval of 99% and is calculated with the following equation 6, where IDL is the instrument detection limit, RSD is the relative standard deviation, $t_{(n-1,1-\alpha=0.99)}$ is the Student's t value for a 99% confidence interval and n-1 degrees of freedom, and n is the number of replicate injections:

$$\text{IDL} = t_{(n-1,\ 1-\alpha=0.99)} \cdot (\text{RSD}/100\%) \cdot \text{amount measured} \quad (6)$$

The IDLs were calculated according to Sheehan *et al.* and technical guidelines from Agilent Technologies.[268,269] The optimized method yielded IDLs between 0.9 and 1.2 fmol, representing a significant improvement over the initial method, which showed IDLs between 1.9 and 3.7 fmol (see Supplement, Table S2). The results demonstrate that the whole optimization process resulted in a notable increase in sensitivity, with a factor of approximately 2-3.

3.1.2. Method validation

The process of quantitative analysis of 5' cap structures in eukaryotic mRNA is schematically illustrated in Figure 3.12 A and commences with the extraction of total RNA and the subsequent isolation of polyadenylated RNA using oligo (dT) magnetic beads. Of note, the use of oligo (dT) beads to capture mRNA based on its typical 3'-end modification, the poly(A) tail, is a common method for isolating mRNA from total RNA. However, since other RNA species, such as lncRNAs and some rRNAs, can also contain poly(A) tails, the RNA treated with oligo (dT) beads is termed poly(A) RNA in the following.[12,270,271] Before proceeding with the poly(A) RNA digestion to nucleosides and cap dinucleotides, a staple isotope labeled cap dinucleotide internal standard is spiked into the samples. The sample is then subjected to LC-MS/MS analysis. This encompasses the MS detection of the cap dinucleotides, and their isotope labeled counterparts as well as the UV detection of the main nucleosides, which allows for the accounting of differences in the injected RNA amount.

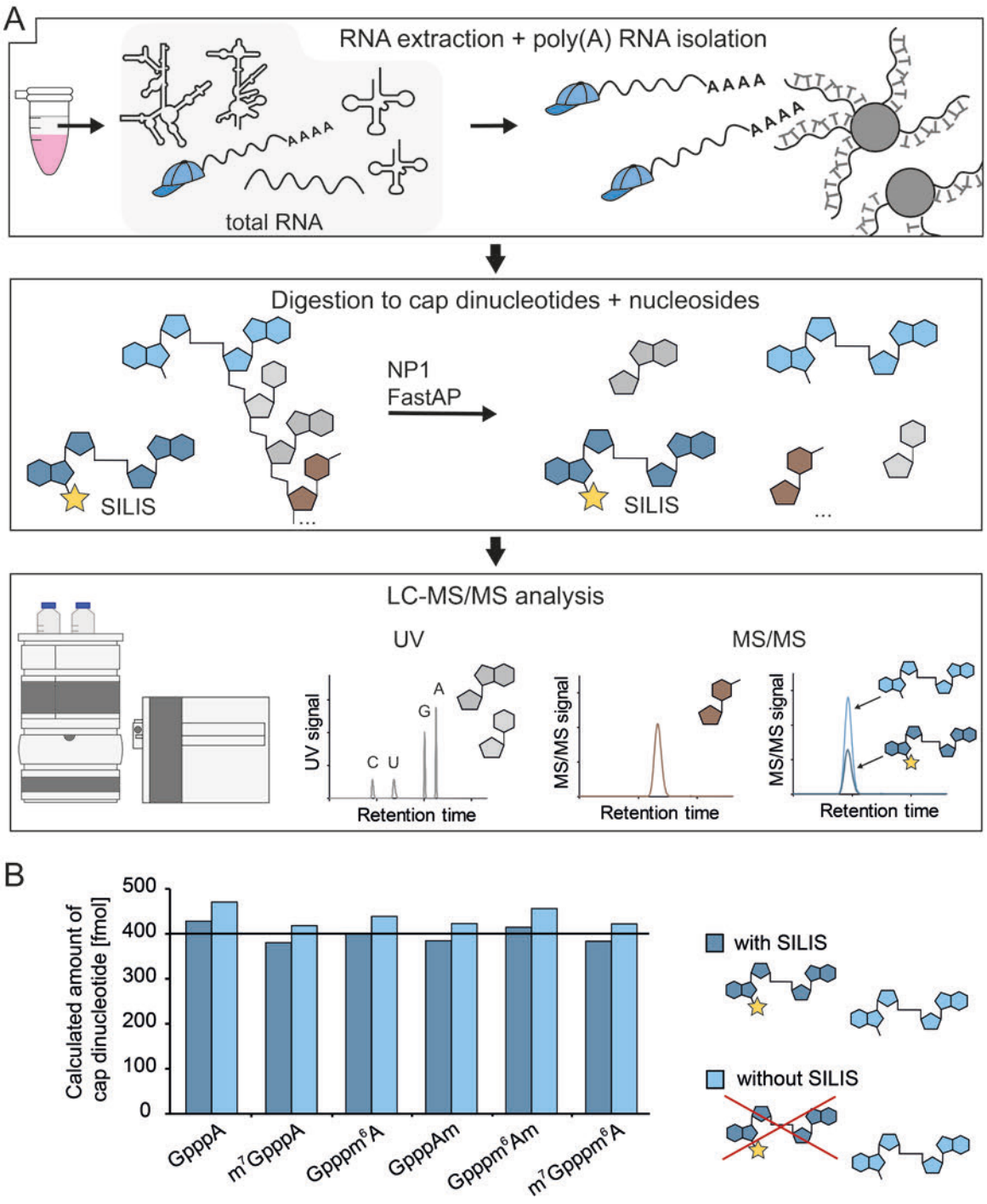

Figure 3.12. A.) Process of the quantitative analysis of 5' cap structures starting with organic extraction of total RNA followed by poly(A) RNA isolation using magnetic oligo (dT) beads. The cap structure is symbolized with a blue base cap. Stable isotope labeled cap dinucleotides, serving as stable isotope labeled internal standards (SILIS), are spiked into the sample prior to its digestion to nucleoside level and cap dinucleotides. The labeled cap dinucleotide is marked with a yellow star, cap dinucleotide structures are depicted in blue and an exemplary modified nucleoside in brown. The sample is subjected to LC-MS/MS, the main nucleosides cytidine (C), uridine (U), guanosine (G), and adenosine (A) depicted in gray are detected UV spectroscopically while cap dinucleotides (blue) as well as modified nucleosides (brown) are analyzed by MS. B.) Detected cap composition in RNA spiked with a

known amount of cap dinucleotides. Analysis was performed with (dark blue) and without (light blue) taking stable isotope labeled internal standards (SILIS) into account. The actual value of 400 fmol is illustrated by a black line.

Additionally, modified nucleosides can be detected simultaneously via MS. Absolute quantification is achieved by measuring against external calibration samples and including SILIS corrections. To validate the accuracy of this process of quantitative analysis of 5' cap structures, a biological sample with known cap dinucleotide content was mimicked. For this purpose, human embryonic kidney 293 (HEK) total RNA, in which the cap dinucleotides of interest were not detectable, was spiked with a known amount of cap dinucleotides. The resulting sample was subjected to the described process but skipping the first step of RNA isolation. The absolute quantification yielded deviations of 0.3-7% from the true value when SILIS corrections were included. In contrast, the absence of SILIS considerations resulted in a notable decrease in accuracy with deviations of up to 18% (see Figure 3.12 B). These results demonstrate the suitability of this approach for the accurate quantification of cap dinucleotides and allow the subsequent application of the method to biological samples.

3.1.3. Method application to biological samples

In the next sections, the possibilities and limitations of applying the developed workflow to biological samples were evaluated.

3.1.3.1. HEK cell poly(A) RNA

The study was initiated with an analysis of the poly(A) fraction of HEK cell RNA, as cells can be cultured in large numbers, thereby facilitating the extraction of high quantities of RNA. To gain insight into the poly(A) RNA quantities required for cap dinucleotide detection, a dilution series spanning from 5 ng to 25 µg was evaluated (Figure 3.13 A). This revealed that the two most prevalent cap dinucleotides, m^7Gpppm^6Am and $m^7GpppAm$ (cap 1 structures), were discernible by LC-MS/MS in only 5 ng and 100 ng of isolated poly(A) RNA, respectively. However, additional cap dinucleotides were only detected from an injection of at least 15 µg, which represents a large amount of extracted and isolated poly(A) RNA and thus indicates low occurrence of these cap species. In the same sample runs, the common mRNA modifications m^6A, pseudouridine (Ψ), and 5-methylcytidine (m^5C) were detected, as illustrated in Figure 3.13 B. Consequently, the capability to simultaneously quantify cap dinucleotides and modified nucleosides in biological samples was confirmed.

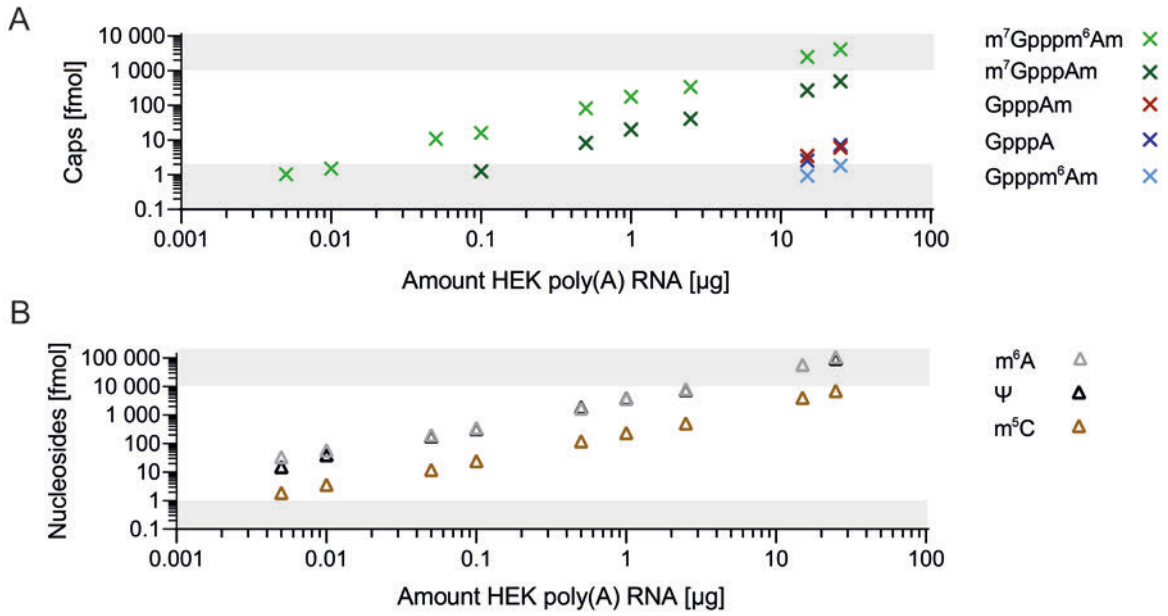

Figure 3.13. Quantification of A.) five cap species and B.) exemplary modified nucleosides (*N*6-methyladenosine (m^6A), pseudouridine (Ψ), and 5-methylcytidine (m^5C)) in a dilution series from HEK poly(A) RNA. The axes are represented in a logarithmic scale. The areas outside the calibrated range are shaded gray.

3.1.3.2. Mouse brain tissues

To illustrate the applicability of the method to the examination of biological tissue, poly(A) RNA extracted from disparate regions of the mouse brain, comprising the cerebellum, cortex, and hippocampus, was used. Subsequent analysis revealed the occurrence of around 0.08 cap dinucleotides per 1000 nucleotides (nt), with the exception of cortex tissue, which exhibited lower levels (Figure 3.14 A). These levels were calculated by relating the amount of quantified cap dinucleotide to the sum of quantified main nucleosides. Only the two cap 1 dinucleotides m^7Gpppm^6Am and $m^7GpppAm$ were detectable. Of note, the quantity of poly(A) RNA extracted from the brain area of one mouse (1 µg from the hippocampus, >2 µg from the other brain regions) was sufficient for cap dinucleotide analysis, demonstrating the sensitivity of this method.

A relative comparison of m^7Gpppm^6Am to $m^7GpppAm$ amounts was conducted to evaluate the extent of *N*6-methylation. Figure 3.14 B illustrates that the proportion of *N*6-methylated cap is markedly increased in the mouse brain tissues (96-98%) in comparison to its levels in HEK cells (90%).

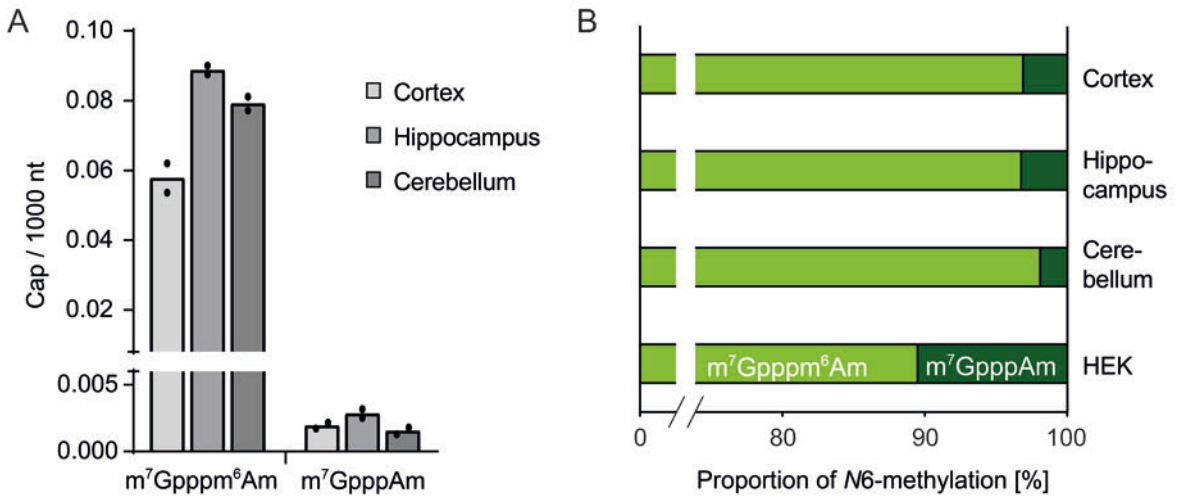

Figure 3.14. A.) Quantification of cap dinucleotides per 1000 nt in poly(A) RNA from different mouse brain sections, represented as averages from two biological replicates. B.) Proportion of m^7Gpppm^6Am to the sum of m^7Gpppm^6Am and $m^7GpppAm$ detected in HEK cells and mouse brain sections.

3.1.3.3. CMTR1 knock out cells

In mammalian cells, the methyltransferase CMTR1 performs 2'*O*-methylation at the cap-adjacent nucleotide, while PCIF1 can introduce additional *N*6-methylation of adenosine.[123,124,135,136] The developed LC-MS/MS method, in combination with CMTR1 knockout cells, offers the possibility to investigate whether PCIF1 can independently modify m^7GpppA-capped RNA without prior 2'*O*-methylation. [redacted] generated a CMTR1 deficient HEK293T cell line by CRISPR/Cas9 knock-out (k.o.) and a CMTR1 k.o. clone stably reconstituted by CMTR1 expressing retroviral vector as a control and kindly provided extracted total RNA.[272,273] The samples were subjected to the described 5' cap quantification process comprising poly(A) isolation, digestion and LC-MS/MS analysis. The control (Ctrl) sample demonstrated the presence of the cap 1 modifications m^7Gpppm^6Am and $m^7GpppAm$, whereas the CMTR1 k.o. sample predominantly exhibited the occurrence of the cap 0 modifications m^7GpppA and m^7Gpppm^6A, the latter demonstrating the ability of *N*6-methylation of cap-adjacent adenosines without previous 2'*O*-methylation (Figure 3.15 A). Of note, a small amount of 2'*O*-methylated m^7Gpppm^6Am was also detectable in CMTR1 k.o.

Finally, the extent of *N*6-methylation was determined for both the knockout and control samples. In CMTR1 knockout samples, the comparison of detected m^7Gpppm^6A and m^7GpppA

revealed a proportion of *N*6-methylation of 76%. In contrast, the control showed a slightly higher proportion of 81% among the 2'*O*-methylated cap dinucleotides (Figure 3.15 B).

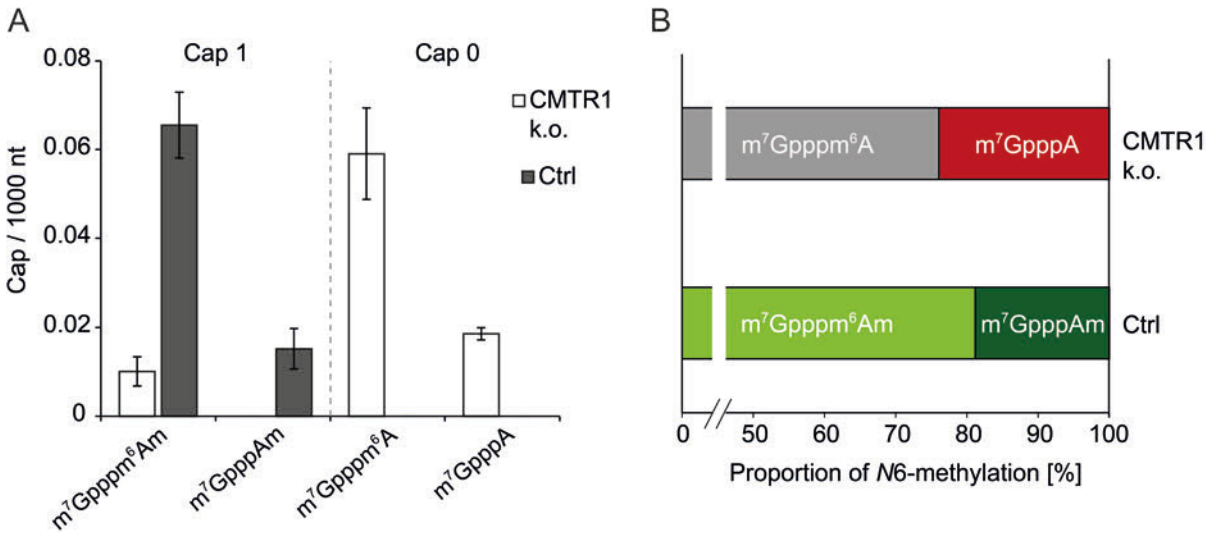

Figure 3.15. A.) Quantification of cap dinucleotides per 1000 nt in poly(A) RNA from CMTR1 k.o. and control (Ctrl), represented as average from three biological replicates, with error bars indicating standard deviation. B.) Proportion of *N*6-methylation comparing m^7Gpppm^6Am and $m^7GpppAm$ in Ctrl samples as well as m^7Gpppm^6A and m^7GpppA in CMTR1 k.o. samples.

3.1.4. Discussion

This work reports the development of a reliable and sensitive LC-MS/MS approach to investigate the cap epitranscriptome, with a focus on cap structures with adenosine and its methylation variants as first nucleotides along with modified nucleosides occurring within RNA species. Extensive HPLC parameter investigation and optimization demonstrated significant impact of mobile phase ion strength on signal amplitude, resulting in gain factors of 1.6-2.3 in abundance. The examination of the effect of mobile phase pH, injection volume, and column temperature revealed impact on peak width and symmetry, as well as retention, resulting in robust separation and increased sensitivity for the detection of cap dinucleotides. The optimization of the ion source parameters, previously optimized for nucleoside analytics, yielded a 1.3- to 2.1-fold increase in abundance, demonstrating the importance of adapting these parameters for chemically diverse structures. Finally, the comparison of the method before and after optimization by instrument detection limits displayed improvement factors ranging from 2.1 to 3.4. The straightforward approach of cap analysis involved the enrichment of poly(A)-tailed RNA, followed by the addition of an isotope labeled standard and subsequent

digestion. The resulting material was then subjected to LC-MS/MS analysis. Validation of this method highlighted the significance of employing isotope labeled standards, as it strongly reduced the discrepancy between the actual and the measured values, thereby enhancing the accuracy.

Recently, alternative approaches that also focus on the LC-MS/MS detection of cap structures as cap dinucleotides have been published.[199–201] The approach of Muthmann *et al.* focuses on the analysis of synthetically capped RNAs in the form of GpppG and m^7GpppG. However, the methods CAP-MAP from Galloway *et al.* and CapQuant from Wang *et al.* also aim to analyze the cap epitranscriptome. Notably, CapQuant includes non-canonical metabolite caps in addition to all canonical cap variants. The methods demonstrate considerable differences in terms of sample preparation. All approaches aiming to analyze biological samples involve the isolation of poly(A) RNA and subsequent digestion with nuclease P1, with the objective of obtaining cap dinucleotides. CapQuant employs an additional HPLC purification step to separate the cap dinucleotides from the nucleoside monophosphates in order to reduce matrix interference and to facilitate the discovery of new cap species by scanning collected fractions for new MS signals.[45] This procedure significantly extends the protocol time. Another disadvantage of this purification step is the use of ion pair reagents, which are known to be very difficult to remove from the HPLC system, requiring long rinsing times. Moreover, ion pair reagents can lead to ion suppression and reduce MS sensitivity.[274,275] Consequently, a dedicated system for ion pair chromatography is recommended, which presents a disadvantage as it requires additional instrumentation.[45] In contrast, CAP-MAP utilizes a significantly faster protocol, subjecting digested samples directly to LC-MS analysis. This practice aligns with the Helm group's long-established standard procedure for nucleoside analytics and was also applied in the newly developed method presented in this work.[248,250] While both CapQuant and the approach developed herein utilize an RP C18 column and measure in positive ion mode, CAP-MAP differs primarily by employing a hypercarb column at a pH of 9 and measuring in negative ion mode. A disadvantage of a hypercarb column is that it requires regular, extensive column regeneration to prevent peak distortion and retention time shifts.[200,276–278] Furthermore, the methodology developed within this thesis differs from those previously described in terms of the way the results are presented. The CAP-MAP and CapQuant approaches indicate the amount of detected cap dinucleotide per µg of injected RNA. In contrast, the approach developed within this thesis, which employs a phosphatase to digest the RNA to the nucleoside level, allows for the additional indication of detected cap dinucleotides per nucleotides. This enables the compensation for differences in the injected RNA amount or inaccuracies in the previous determination of the RNA amount. Moreover, this

approach offers the potential for the additional analysis of modified nucleosides that naturally occur in poly(A) RNA in a single experimental run. Regarding sensitivity, the method developed in this thesis demonstrated detection limits for all investigated cap dinucleotides between 0.9 and 1.2 fmol. The method of Muthmann *et al.* yielded significantly higher IDLs for GpppG and m^7GpppG, at 6.8 fmol and 34.1 fmol, respectively. However, the results are not strictly comparable, as the approach used in this study involved different cap dinucleotides.[199] In contrast to the determination of the IDL, which is based on statistical analysis, within the CapQuant approach the detection limits are determined via the *S/N* ratio as LOD.[201] Nevertheless, a comparison of the same analyzed cap dinucleotides reveals a wider range of quantification limits in the CapQuant approach, with LODs between 0.6 and 13 fmol, whereas in the developed approach presented in this work, 1.2 fmol is not exceeded by any cap dinucleotide. Upon application to biological samples, the major cap dinucleotides m^7GpppAm and m^7Gpppm^{6}Am were detected at a minimum amount of 100 ng poly(A) RNA, with m^7Gpppm^{6}Am even detected in only 5 ng of poly(A) RNA. Both CAP-MAP and CapQuant also showed the capability to successfully analyze biological samples using poly(A) RNA quantities significantly below 1 µg.[45,200] In conclusion, the method developed in this thesis combines the advantages of the previously described methods and was further expanded. It employs a rapid sample protocol that avoids the time-consuming HPLC purification step and the utilization of ion pair reagents. Furthermore, a common RP C18 column is used, which does not require special maintenance and is therefore easy to handle. The method is additionally characterized by very low detection limits for all investigated cap dinucleotides. Of particular note is the distinctive feature of this approach, which allows for the detection of nucleosides alongside cap dinucleotides, representing a novel contribution to the field.

The method developed within this work was applied to cell and tissue samples of higher eukaryotes, including HEK cell preparations as well as dissected mouse brain regions. The cap 1 dinucleotides, consisting of *N*7-methylated guanosine and 2'*O*-methylated adenosine (m^7GpppAm and m^7Gpppm^{6}Am), were the predominant species detected, in accordance with previous literature.[5–7,124] In high amounts of HEK cell poly(A) RNA, traces of *N*7-unmethylated variants, namely GpppA, GpppAm, and Gpppm^{6}Am were detectable. These were also detected in the publication using the CAP-MAP method, but not in the CapQuant analysis.[45,200] This discrepancy is presumably due to the fact that the latter only analyzed quantities up to 7 µg of biological samples. The presence of only traces of caps lacking m^7G indicates an efficient cap methylation process and/or efficient removal by decapping enzymes such as DXO as part of a quality control mechanism.[45,104,105,200] m^7Gpppm^{6}Am was detected as the most common cap dinucleotide and the comparison of HEK cells and brain tissue revealed

differences in the extent of *N*7-methylation of 90% and 97%. This is in line with recent investigations of the abundance of m^7Gpppm^6Am compared to $m^7GpppAm$, which demonstrate cell type and tissue-specific differences.[45,190,200] In a previous study, Akichika and colleagues reported the occurrence of 92% m^6Am in HEK293T mRNA (also isolated using poly(A) sections plus additional enrichment using m^7G-antibodies) which aligns with the findings in this work of 90%.[124] Conversely, earlier studies estimated the proportion of *N*6-methylation to be less than 75% in HEK and HeLa cells.[122,279] Further studies conducted comparative analyses of different mouse organs and revealed tissue-specific variations, with mouse brain tissue having the highest proportion of m^6Am at approximately 94%.[190,200] This complies with the resulting data in this thesis, which showed high proportions of *N*6-methylation of approximately 97% for all analyzed brain regions. These findings support the idea of cap methylations being a differently regulated process in diverse cell types and tissues. Recently, the methyltransferase PCIF1 was discovered to be responsible for *N*6-methylation at the triphosphate adjacent adenosine to form m^6Am. Biochemical studies showed that PCIF can also act on unmethylated adenosine to form m^7Gpppm^6A, but with around 8-fold lower preference as for the 2'*O*-methylated adenosine. Within this thesis, similar to previous studies by Akichika et al. and investigations using the CAP-MAP protocol, m^7Gpppm^6A was not detected in biological samples, supporting the proposal that CMTR1-mediated 2'*O*-methylation primarily occurs prior to m^6A formation.[124,200] In contrast, the study using the CapQuant method detected m^7Gpppm^6A in mouse liver but not kidney, suggesting tissue-specific differences in possible yet-to-be-determined cap demethylation processes or in the preferences of PCIF1 to act on adenosine without 2'*O*-methylation.[45] The cap analysis protocol developed in this thesis, when applied to CMTR1 k.o. cells, revealed the occurrence of m^7Gpppm^6A, which demonstrates the ability of PCIF1 to methylate unmodified adenosine in a biological context. The extent of *N*6-methylation in the k.o. was determined to be 76%. In contrast, the rescued knock out, which served as a control, demonstrated a higher proportion of 81% between the 2'*O*-methylated cap dinucleotides, presumably as a result of the higher preference of PCIF1 for 2'*O*-methylated adenosines, as determined in *in vitro* studies.[123,124] To conclude, the application of the developed LC-MS based cap analysis protocol to biological samples successfully demonstrated its potential to investigate the cap epitranscriptome and to gain insights into the regulation process of RNA cap methylations.

Of note, parts of this work including a summary of the LC-MS/MS optimization process and the analysis of HEK cells and mouse brain tissues was published recently.[255]

3.2. Isotope labeling of cap dinucleotides via deuterium exchange

As demonstrated in the previous chapter 3.1.2., isotope labeled cap dinucleotides effectively enhance method accuracy. This part of the work focused on developing a straightforward and widely applicable technique for producing isotope labeled cap dinucleotide standards, which are crucial for accurate quantification in mass spectrometry, while avoiding complex organic chemistry that can be challenging for RNA biochemistry laboratories. The research groups of Schweizer and Bullock demonstrated that during incubation in D_2O at elevated temperatures, the C-8 protons in purine nucleobases are replaced by deuterium, a property later applied in nucleoside structure elucidation via LC-MS to investigate C-8 substitutions.[280–284] Based on these findings, it can be concluded that the incubation of cap dinucleotides containing two purines in D_2O results in a mass shift of +2, thereby demonstrating the potential for their use as an isotope labeled internal standard in LC-MS, which will be explored in more detail in the following.

Parts of the experimental work were performed by [redacted]

3.2.1. Investigation of deuterium exchange parameters

The requirement for the internal standard is that it be labeled in the most complete manner possible. Consequently, the effectiveness of the deuterium exchange was tested in the next experiments. When incubating a cap dinucleotide comprising a guanosine and an adenosine moiety in D_2O, there are four different possible outcomes: no C-8 proton is deuterated, only guanosine, only adenosine, or both are C-8 deuterated, which is the favorable outcome. Due to different mass transitions, these four options are distinguishable by MS/MS (see Figure 3.16). Therefore, the mass transitions, determined in chapter 3.1.1.1., were used. A series of experiments was started with the cap dinucleotide GpppAm to investigate the incubation parameters temperature and duration and their impact on deuterium exchange.

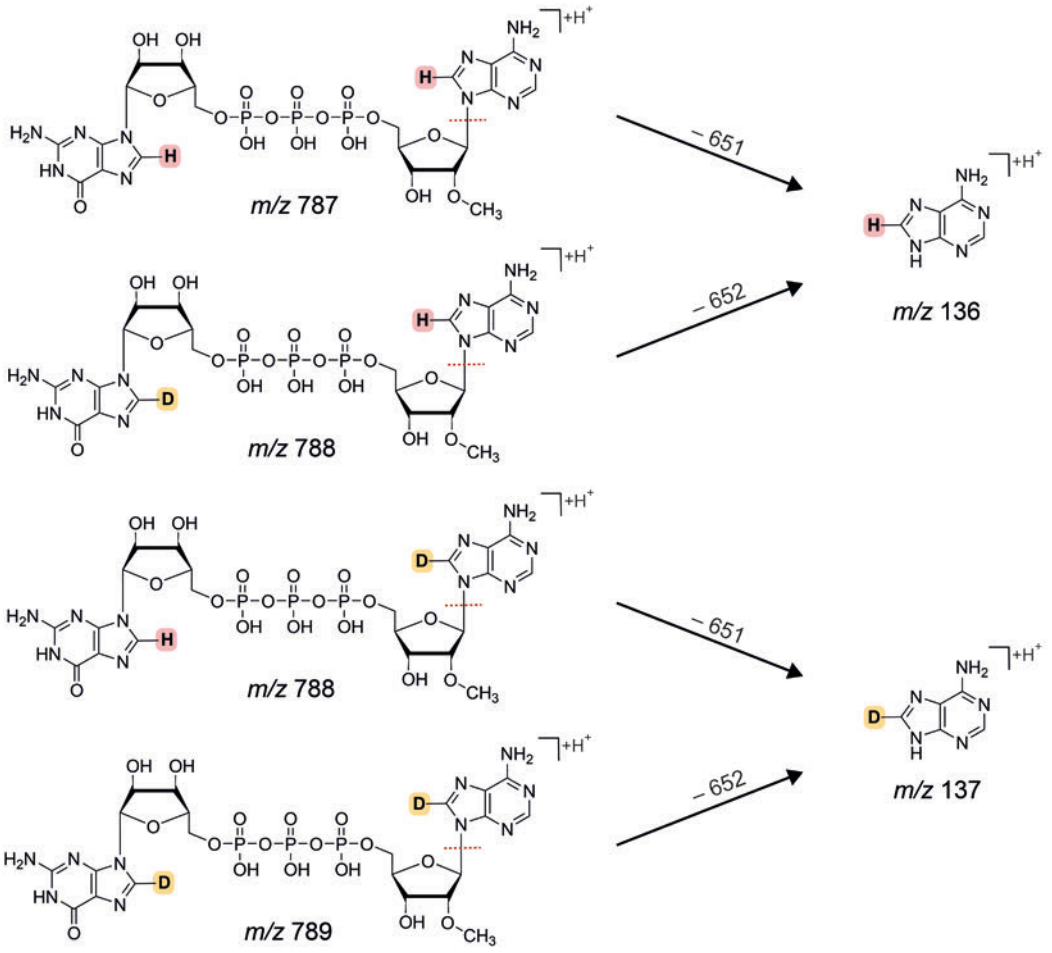

Figure 3.16. C-8 deuteration options of GpppAm with depicted mass transitions. C-8 hydrogen is marked in light red, C-8 deuterium is marked in yellow.

In a first step, GpppAm was incubated in D_2O at various temperatures between 50 °C and 95 °C for 0-8 hours. The relative proportion of doubly-deuterated cap is illustrated in Figure 3.17 A, which demonstrates a pronounced temperature dependency. At temperatures below 80 °C, the conversion to the doubly-deuterated cap was minimal, and temperatures above 85 °C were required to achieve quantitative conversion. The distribution pattern of deuteration was further investigated for an incubation temperature of 90 °C. It became evident, that after 3 h of incubation less than 2% of GpppAm were non-deuterated, whereas after 5 h of incubation, the proportion of cap with both C-8 positions being deuterated was above 90% (Figure 3.17 B). In addition, these experiments provided evidence that the deuteration of G occurs much faster, as can be seen from the fact that the proportion of GpppAm in which only G is deuterated is significantly higher than that of GpppAm that is only Am-deuterated.

Given the harsh incubation conditions of 90 °C for several hours, the effect on the stability of the cap structures was assessed using absolute signal strengths. Based on the same

ionization efficiency of isotopes, the sum of the peak areas for the four deuteration outcomes was calculated and compared across varying temperatures and incubation times. Figure 3.18 A illustrates a notable decline in signal strength for incubation temperatures exceeding 85 °C, with a 50% reduction observed after 3-4 hours. After 7-8 hours, the signal intensity decreased to 20-30% of the initial level observed without incubation, demonstrating degradation of the cap dinucleotides. This finding provides a rationale for further investigation into the deuteration process to ascertain ways of reducing incubation time and temperature while maintaining consistent deuteration efficiencies. For instance, this could involve working within specific pH ranges.

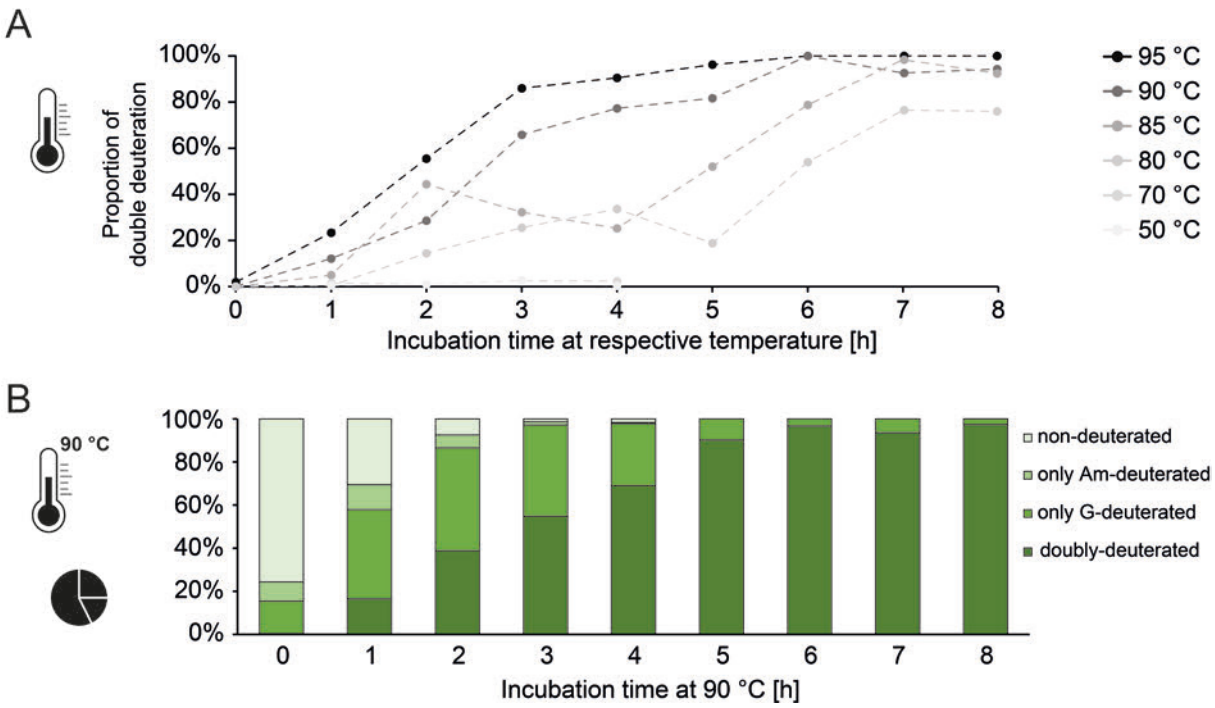

Figure 3.17. A.) Proportion of doubly-deuterated GpppAm resulting from varying incubation times at temperatures between 50 °C-95 °C. B.) Deuteration pattern of GpppAm after incubation in D_2O at 90 °C for varying time periods.

The objective of this subtopic is to establish an internal standard that can be employed routinely in the analysis of cap structures. Given the time-efficient benefits of storing larger quantities, an assessment of the stability of (deuterated) cap dinucleotides upon extended storage at −20 °C was carried out. Measurements of GpppAm, conducted immediately following incubation in D_2O at 90 °C and again after one week of storage at −20 °C, indicated no significant alterations in the proportions of the deuteration pattern (Figure 3.18 B), thereby demonstrating the suitability of this approach for the preparation of larger quantities of internal standard and subsequent storage until further use.

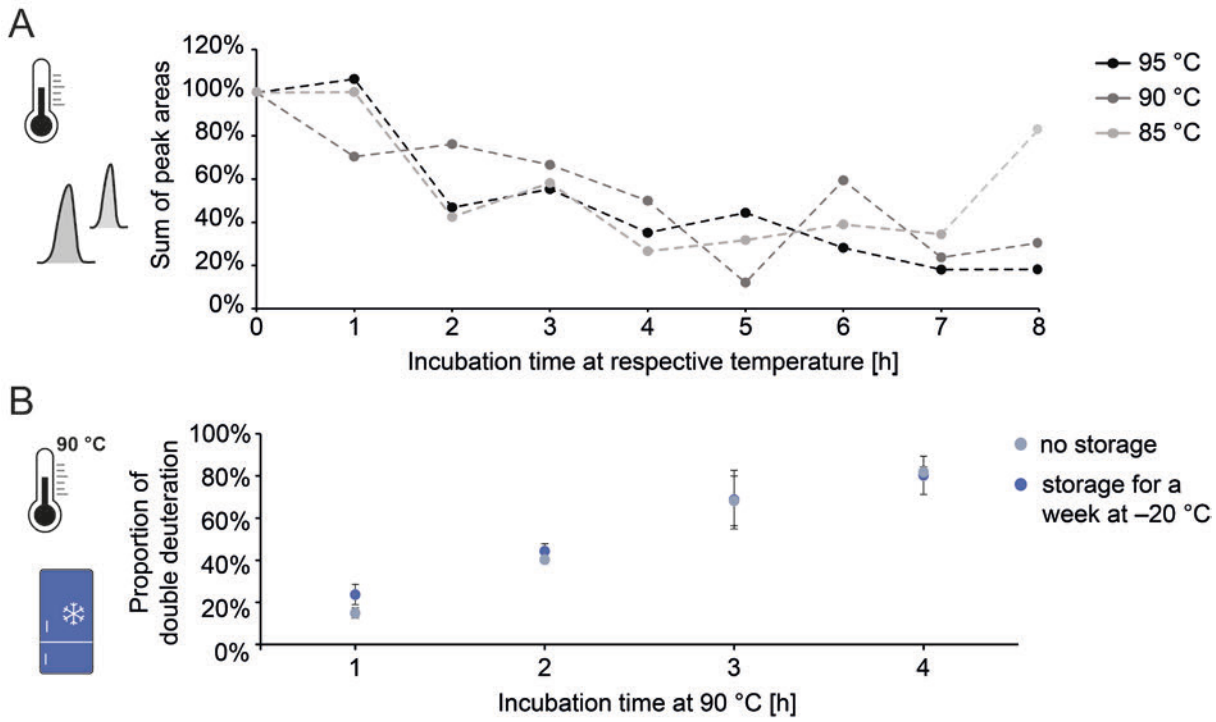

Figure 3.18. A.) Sum of the peak areas of the C-8 deuteration options for different incubation time periods at temperatures of 85 °C-95 °C. The sum of peak areas of the non-incubated sample (incubation time: 0 h) was set to 100%. B.) Proportion of doubly-deuterated GpppAm measured directly after incubation at 90 °C and again after storage for one week at −20 °C. The average values from three technical replicates are shown, with error bars indicating standard deviation.

3.2.2. Assessment of nucleoside modifications on deuterium exchange

Following the promising results of GpppAm's nearly complete deuteration at both C-8 positions upon incubation in D_2O at elevated temperatures, additional cap dinucleotides were investigated to assess the impact of nucleoside modifications. The illustration of the deuteration patterns of GpppAm compared to GpppA and Gpppm^{6}Am revealed similar outcomes for the non-methylated guanosine cap variants with different adenosine variants, demonstrating that the C-8 deuteration of adenosine did not depend on its methylation status (Figure 3.19). In contrast, m^{7}GpppA displayed a distinctive behavior in a way that the portion of doubly-deuterated m^{7}GpppA did not exceed 50%. Moreover, the proportion of deuterated m^{7}G remained consistent over incubation periods of 0-8 hours, indicating deuteration behavior independent of incubation at high temperature, which requires further investigation.

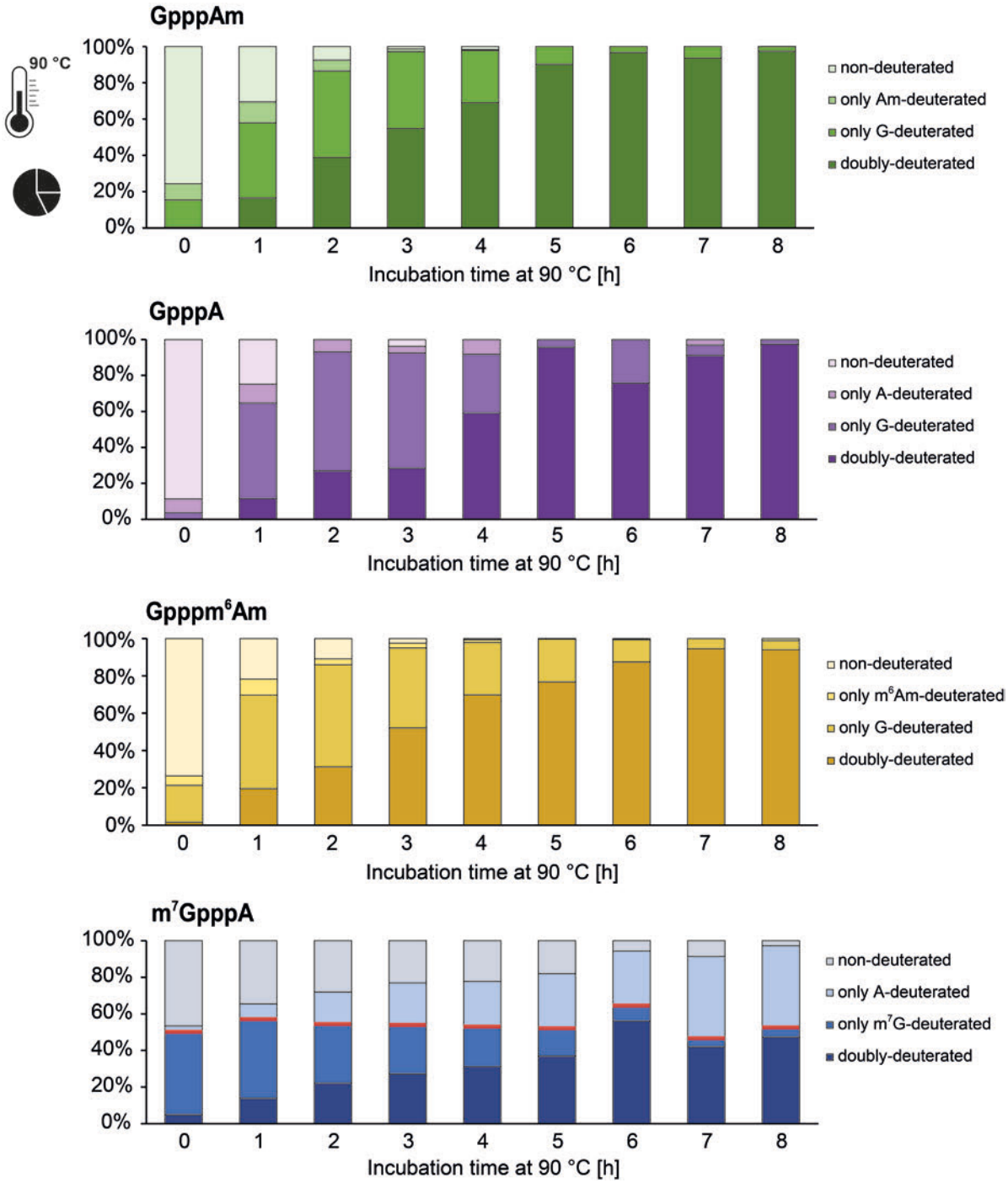

Figure 3.19. Deuteration patterns of GpppAm, GpppA, Gpppm^{6}Am, and m^{7}GpppA after incubation in D_2O at 90 °C for varying time periods. The part of deuterated m^{7}G, consisting of doubly-deuterated and only m^{7}G-deuterated m^{7}GpppA, is depicted with a red line.

Of note, the experiments were repeated in various D_2O ammonium acetate solutions covering a pH range of 5-9 to examine the impact of solvent pH on deuteration exchange reactions, particularly in the case of m^{7}GpppA. Additionally, the experiments aimed to assess whether

the harsh incubation parameters leading to degraded cap dinucleotides can be relaxed. However, pH dependency was not discernible (Supplement, Figure S3).

Due to its atypical behavior, the next step was to further investigate the behavior of m^7G in m^7GpppA upon deuterium exchange. The extensive reduction in the retention time, from several minutes to less than 10 seconds, by replacing the HPLC column with a capillary loop revealed a strong impact on the deuteration state of m^7G (Figure 3.20 A).

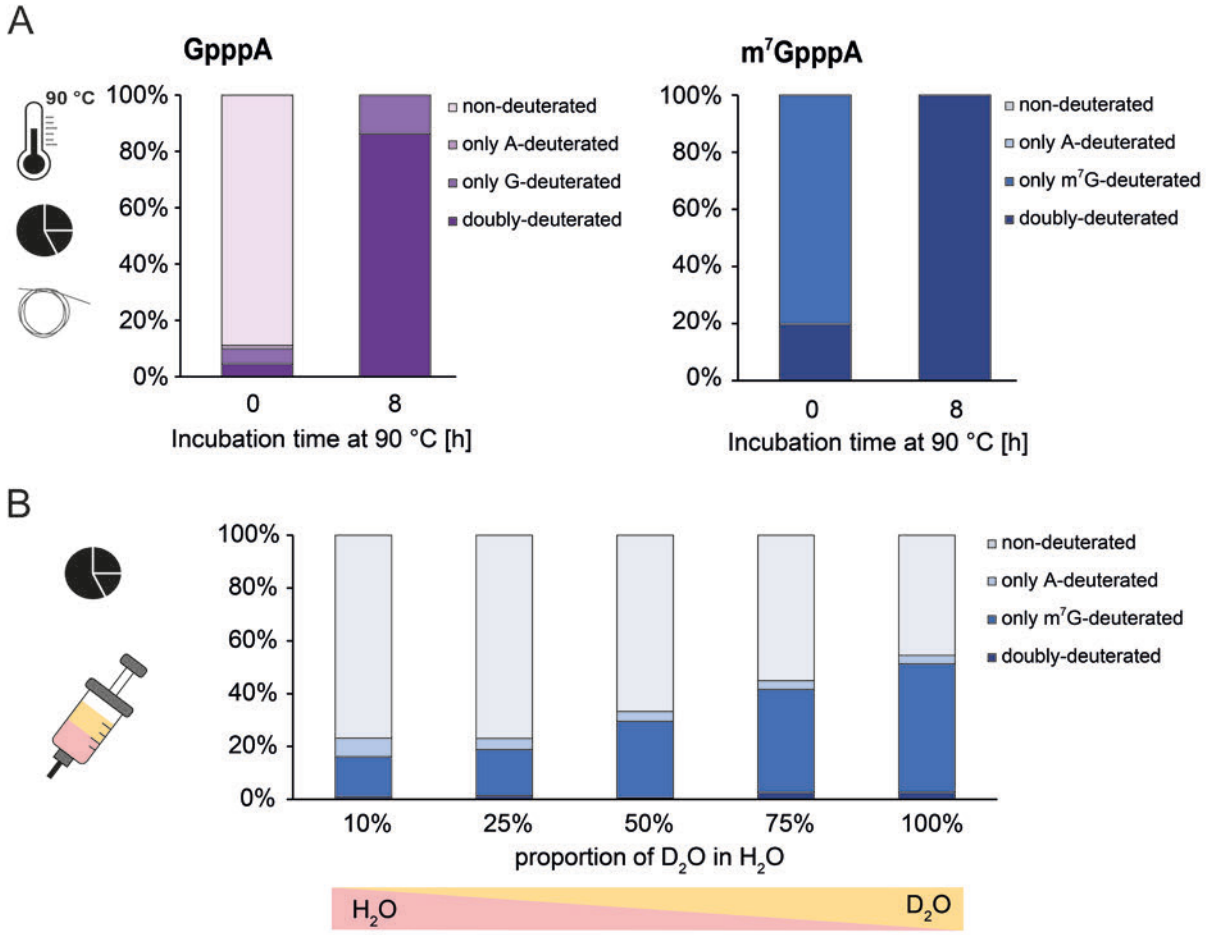

Figure 3.20. A.) Deuteration patterns of GpppA and m^7GpppA after 0 h and 8 h incubation at 90 °C. The HPLC column was exchanged by a capillary loop. B.) Deuteration pattern of m^7GpppA diluted in different ratios of D_2O/H_2O. No incubation at elevated temperatures was conducted.

Without incubation at 90 °C, m^7G showed 100% deuteration, predominantly with A remaining non-deuterated, while after incubation for 8 h, 100% of the doubly-deuterated form was detectable. In contrast, GpppA showed comparable deuteration patterns irrespective of the installed column or capillary loop, indicating that the C-8 proton exchange of purines was not reversible when exposed to H_2O from the HPLC buffer, with the exception of m^7G. The latter did not show stable deuterium exchange; however, it exhibited fast exchanging behavior even

when it was not incubated at 90 °C. This indicated that the extent of m^7G deuteration was dependent on the degree of contact with H_2O (including the HPLC buffer) and D_2O. Investigation of this assumption was important in the context of internal standard application. Here, the standard is spiked into the H_2O-containing sample solution to be analyzed, meaning that contact of the deuterated standard with high amounts of H_2O cannot be prevented. Injecting the cap dinucleotide, diluted in different ratios of H_2O and D_2O, did indeed show a diminished proportion of deuterated m^7G in conjunction with increased levels of H_2O in the solution (Figure 3.20 B).

The experimental work in this chapter was conducted by [redacted]

3.2.3. Discussion

Deuterium exchange experiments were successfully conducted on the cap dinucleotide GpppAm, thereby demonstrating the potential for the generation of an isotope labeled variant with a mass shift of +2 through the simple incubation of the latter in D_2O at an elevated temperature for several hours. It was demonstrated that temperatures exceeding 85 °C were essential to achieve nearly quantitative conversion. The proportion of doubly-deuterated cap dinucleotides was observed to comprise more than 90% of the total, following an incubation period of 5 hours at a temperature of 90 °C. Consistent with the literature, a faster exchange rate of guanosine relative to adenosine was observed.[285,286] After 3 hours of incubation, less than 2% of GpppAm remained non-deuterated. This finding is of particular importance because non-deuterated standards must be avoided, as this can result in inaccurate quantification due to the addition to the analysis sample. A potential limitation of this method is that the occurrence of non-deuterated standards cannot be unequivocally excluded.

The assessment of the stability of the cap structures incubated at 90 °C for up to 8 hours revealed a significant decline in signal intensity to only 20-30% of the initial level without incubation. This is consistent with the existing literature, which reports the half-life of the nucleoside adenosine to be 3.5 hours at 80 °C (pH 7).[287] Similarly, structurally related compounds of cap dinucleotides such as ATP demonstrate hydrolysis rates with half-lives of 0.5-6 hours at temperatures between 80 °C and 100 °C.[288–290] For the method developed within this thesis, this signifies that a larger quantity of starting material is required for the production of the labeled standard. Since triple quadrupole-MS was used for quantification, which was set

to examine only defined mass transitions, the presence of hydrolysis products or fragments had no discernible impact on the detection of the cap dinucleotides.

In a next step, further cap dinucleotides were subjected to deuterium exchange reactions to assess the influence of nucleoside modifications. While methylation of adenosine did not impact the isotopic exchange on the C-8 position, m^7G was observed to display strikingly disparate behavior, exhibiting rapid exchange rates in contact with D_2O or H_2O, independent of preceding heating steps. In literature, the fast exchange rates of m^7G are mechanistically explained by the stabilization of the ylid type intermediate formed after acidic dissociation by the carbene type resonance form, which is analogous to the behavior observed in thiamine-related compounds.[285,291] The half-life was found to be 5.5 min at a pH level of 4.1, while at pH 7 the exchange was too fast to obtain kinetic data.[291] Own pH studies with D_2O-solutions ranging from pH 5-9 did not show significant changes in the deuteration process. Further experiments at lower pH values to slow down the exchange to a greater extent were not carried out, as in method application to biological samples the internal standard, dissolved in D_2O, is spiked to the aqueous samples during sample preparation. Consequently, it is not feasible to reduce the interaction time of the standard with H_2O to a minimal level.

In summary, the deuterium exchange reactions to obtain isotope labeled standards with a mass shift of +2 were only successful for cap dinucleotides not containing an m^7G moiety, as m^7G exhibits a fast exchanging behavior that does not allow the generation of stable isotope labeling. Since this *N*7-methylation is prevalent in all major cap species, this method was found to be unsuitable for the generation of labeled canonical cap dinucleotides.

3.3. Determination of capping efficiencies via LC-MS/MS

The 5' cap structure of mRNA is critical for its stability and function, making effective capping essential for synthetic mRNA, especially in therapeutic applications.[149,292] The synthesis of mRNA predominantly relies on IVT. In this context, capping strategies can be classified into two main categories: post-transcriptional methods, where the cap structure is enzymatically generated after transcription, and co-transcriptional methods, where cap analogs are incorporated during the IVT reaction.[149,156] Despite these approaches, incomplete capping may result in variable levels of uncapped transcripts. Consequently, accurate determination of capping efficiencies is crucial for optimizing these methods and ensuring the quality of IVT RNA.

In this chapter, a novel method was employed to determine capping efficiencies via LC-MS, which is based on the complete digestion of RNA to nucleoside level, followed by absolute quantification of a cap-related modification, dependent on the capping reaction type (e.g., m^7G, 3'*O*-Me-m^7G). Finally, the modification is normalized to the amount of injected RNA, which is calculated according to equation 7.

$$(7) \qquad \text{Capping efficiency [\%]} = \frac{\text{cap-related modification [mol]}}{\text{RNA [mol]}} \cdot 100\%$$

$$\text{with} \quad \text{RNA [mol]} = \frac{\text{Main nucleoside N [mol]}}{\text{Number of N in RNA sequence}}$$

The concept of relating one modification to the amount of injected RNA molecules with known sequence has been previously described by Thüring et al. and has been successfully implemented in the assessments of the degree of modification of tRNA by the Helm group.[248,293] In contrast to the LC-MS/MS approaches used to detect cap dinucleotides, this method uses additionally a phosphodiesterase to digest the RNA, which results in the cleavage of the cap triphosphate bridge and the release of the aforementioned cap nucleoside modifications. The intended concept can also be used to relate cap dinucleotides to the amount of RNA. However, nucleosides were chosen as the preferred option due to their higher ionization efficiencies, which allow for the analysis of smaller sample amounts.

3.3.1. Proof of concept

To test this concept, a capped RNA was required for which the exact capping efficiency is known. For this, a synthetic oligonucleotide with a 5'-triphosphate and a length of 38 nt was ordered and subsequently capped with Faustovirus capping enzyme. Denaturing PAGE was

performed to separate the uncapped oligonucleotide from the capped fraction (Figure 3.21 A). The latter was excised and purified to serve as a control with a theoretical capping efficiency of 100%. LC-MS/MS analysis indeed revealed a number of 1.02 m^7G molecules per RNA molecule, expressed as a capping efficiency of 102%, while the measurement of the capped oligonucleotide without gel purification showed a capping efficiency of 63% (Figure 3.21 B). Notably, densitometric analysis of the gel image yielded a capping efficiency of 67%, which reflects a slight discrepancy with the LC-MS measurement. These results demonstrate the potential of LC-MS/MS analysis of digested RNA on nucleoside level for determining the proportion of capped RNA.

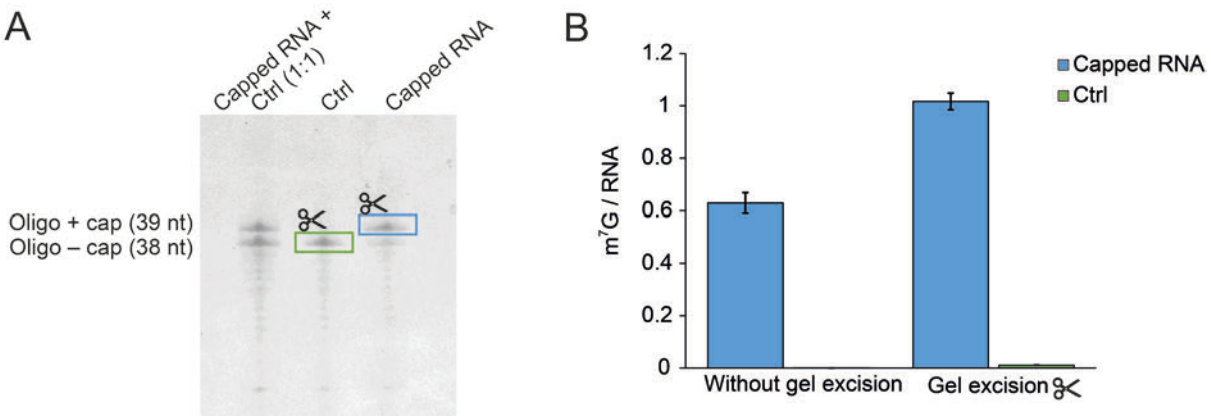

Figure 3.21. A.) Denaturing PAGE (20%, GelRed staining) of the capped RNA, the untreated control (Ctrl) and a 1:1 mixture of both. The marked areas were excised, purified, and subjected to LC-MS/MS analysis. The complete gel image is available in Supplement, Figure S4. B.) Determined number of m^7G per RNA molecule for capped RNA with and without gel excision. Error bars represent the standard deviation resulting from the calculation of the injected RNA amount based on the main nucleosides: cytidine, uridine, guanosine, and adenosine.

3.3.2. Application to *in vitro* transcribed RNA

In a first step, RNA was capped co-transcriptionally by using two different cap analogs: ARCA and CleanCap® to compare the resulting capping efficiencies. While ARCA consists of a cap dinucleotide with 3'*O*-Me-m^7G linked to guanosine by a 5'-5'-triphosphate bridge, the used CleanCap® analog consists of a cap trinucleotide with 3'*O*-Me-m^7G linked by the triphosphate bridge to Am and following guanosine (Figure 3.22 A). The presence of 3'*O*-Me-m^7G in both cap analogs enabled the quantification of the identical cap-related modification per RNA molecule. These structures mean that different template designs are required for IVT: while the initiator sequence must start with a G after the promoter sequence to use ARCA, an initiator sequence starting with AG is required for CleanCap®. IVT reactions with the respective cap analogs were performed to result in capped eGFP mRNA with a transcript length of 828 nt,

followed by DNase digestion and purification using a transcription clean-up kit. Negative controls were prepared by performing the IVT reaction without cap analog and adding the respective cap analog after template digestion but before purification, in order to account for the effectiveness of removal of free cap analog during purification. Subsequent LC-MS/MS analysis revealed capping efficiencies between 200-350% and up to 64% in the negative control (Figure 3.22 B). As these capping efficiencies were far too high, an additional step of gel purification by denaturing PAGE was implemented to remove RNA molecules not displaying the expected full length of 828 nt as well as free cap analogs. Subsequent LC-MS/MS analysis resulted in significantly reduced capping efficiencies of 147% for CleanCap®-capped RNA and 76% for ARCA-capped RNA (Figure 3.22 B).

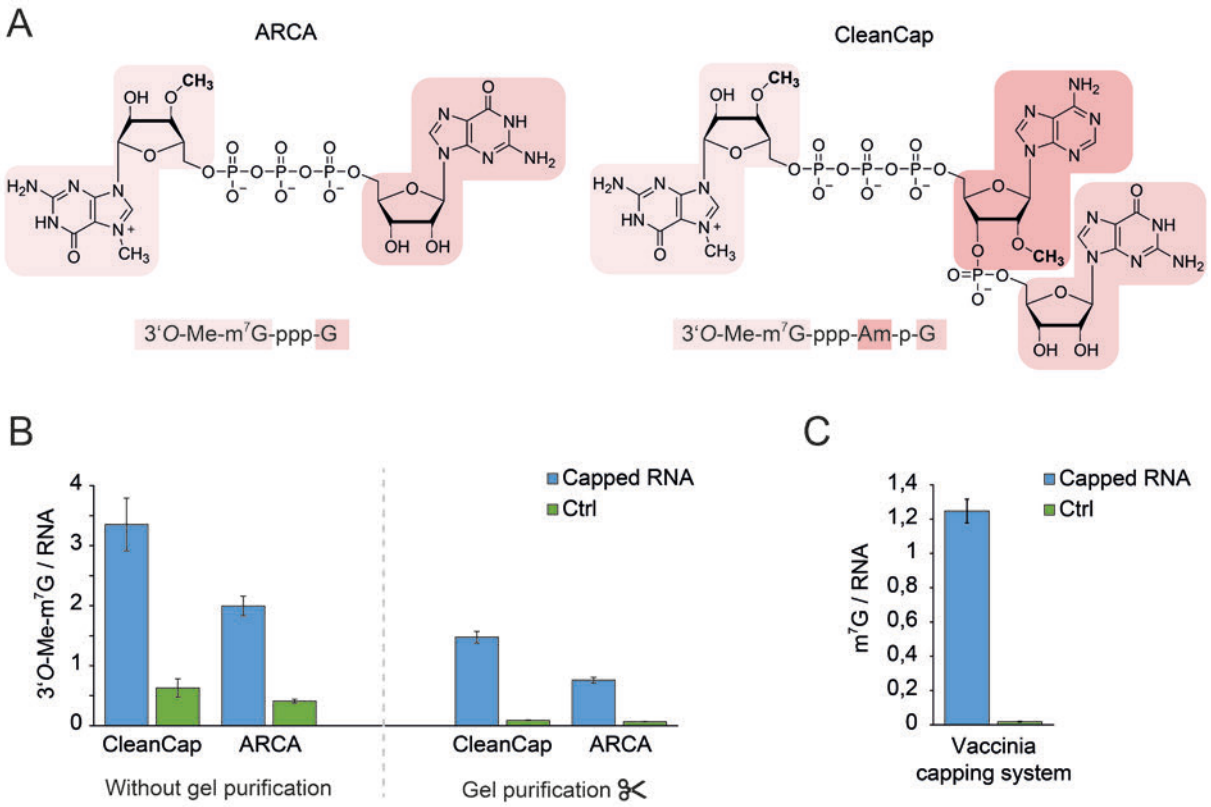

Figure 3.22. A.) Chemical structures of ARCA cap dinucleotide and the used CleanCap® reagent. B.) Determined number of 3'*O*-Me-m^{7}G per RNA molecule when using CleanCap® or ARCA cap analogs for co-transcriptional capping. Quantification was performed with and without an additional gel purification step. Error bars represent the standard deviation resulting from the calculation of the injected RNA amount based on the main nucleosides: cytidine, uridine, guanosine, and adenosine. The negative control was prepared by performing the IVT reaction in the absence of a cap analog but spiking the latter directly prior to RNA purification. C.) Determined number of m^{7}G per RNA molecule after post-transcriptional capping using the vaccinia capping system. The negative control reaction was performed in the absence of the capping enzyme. Error bars represent the standard deviation resulting from the calculation of the injected RNA amount based on the main nucleosides: cytidine, uridine, guanosine, and adenosine.

Next, the method was applied to post-transcriptionally capped RNA. Therefore, gel-purified eGFP mRNA was enzymatically capped using the vaccinia capping system and then purified with a clean-up kit. Subsequent LC-MS/MS analysis revealed a capping efficiency of 125%, whereas the negative control, obtained by repeating the procedure without addition of capping enzyme yielded negligible signals (Figure 3.22 C).

3.3.3. Discussion

The concept of the described method to determine capping efficiencies relies on the digestion of RNA to nucleoside level and subsequent absolute quantification of a cap modification related to the amount of injected RNA molecules. The analysis of a fully capped oligonucleotide, ensured by gel excision, revealed an appropriate precision with only 2% deviation from the assumed value. In contrast, the application of this concept on IVT RNA with a length of more than 800 nucleotides did not show promising results.

One potential source of error was revealed by negative control samples containing significant quantities of cap modifications. For co-transcriptional capping, the negative control samples were prepared by performing the IVT reaction without cap analog and adding the respective cap analog after template digestion but before purification. This served as a control for evaluating the efficiency of free cap analog removal during purification and is particular important since free cap analog cannot be distinguished from inserted cap analog after total digestion, which may result in inaccurate quantification.

The necessity of the described negative controls also applies to post-transcriptionally capped RNA. During the capping reaction, GTP is added in excess, thus, any insufficient purification may interfere with the quantification of guanosines when calculating the injected amount of RNA. In the experiments, non-negligible amounts of free analogs, reaching up to 0.62 cap/RNA, were detected and showed insufficient purification. After gel excision of the desired products, the negative control samples showed significantly reduced signals. Nevertheless, removal of the free cap analogs is recommended even in the absence of gel purification. Even though the used MEGAClear™ Kit is designed for purification after transcription, the results demonstrated that the kit did not achieve the required level of purification when used according to the manufacturer's instructions. To address this challenge, additional washing steps or repetition of the purification may be required. An alternative approach is the use of conventional precipitation methods for purification. In particular, lithium

chloride is a commonly used option following *in vitro* transcription, as it does not precipitate DNA, small RNA fragments, or nucleotides.[294,295]

The analysis of co-transcriptionally capped RNA resulted in capping efficiencies of 335% and 200%. However, after gel excision of the desired product, significantly lower capping efficiencies of 148% and 76% were obtained. This indicates that IVT transcripts not displaying the intended length represent an additional source of error that has a significant impact, apart from the inadequate removal of capping reaction components. The concept of the method described in this thesis relies on the calculation of injected RNA by dividing the obtained amount of a main nucleoside by its number in the RNA sequence. Therefore, only the presence of transcripts of the precise length is assumed. Transcripts of different lengths thus influence the outcome of calculated RNA amount and especially capped short transcripts would lead to a distinct overestimation of cap per full length RNA.

Potential by-products of IVTs are degraded RNA, abortive fragments, double stranded RNA, and RNA:DNA hybrids.[296] Especially in early stages of transcription short oligonucleotides of 9-12 nt length are produced, termed abortive initiation.[297,298] Also, the use of one nucleoside triphosphate (NTP) at a limited concentration can lead to truncated full-length products.[297] This is especially relevant in the context of co-transcriptional capping with ARCA, given that GTP, which competes with ARCA as an initiation starter, is typically limited to allow for enhanced capping efficiency.[162] Optimization of an IVT protocol concerning the amount of template DNA, the ratio of NTPs, the Mg^{2+} concentration, or the incubation temperature can lead to enhanced output and less by-products.[296,299–302] Such optimization may also be appropriate for the IVT reactions that have been presented in this chapter. Methods to purify IVT mRNA include chromatography purification methods, enzymatic approaches, tangential flow filtration, and gel electrophoresis, with the latter being performed within this study.[295,303]

The gel-purified CleanCap® RNA exhibited a capping efficiency of 148%. However, the previously described sources of error should have been eliminated by gel excision of the respective product length. Since cap analogs do not provide a free 5'-triphosphate, the possibility of internal incorporation leading to transcripts containing more than one cap was excluded. One possible explanation for the observed discrepancies is a systematic error in the absolute quantification. Sources of inaccuracy may relate to the preparation of external standards, including weighing, dissolving, and diluting, or to the analysis itself. Since no 3'*O*-Me-m^7G standard was available, the external standard was prepared by digesting the ARCA cap dinucleotide. The ARCA cap analog is provided as a dried salt with a specified molar amount. However, the accuracy and purity of this reagent are not guaranteed, as it is typically

utilized within a concentration range that is deemed sufficient for its intended applications, and therefore, it is a source of inaccuracy. Nevertheless, the measurement of m^7G in post-transcriptionally capped, gel-purified eGFP mRNA resulted in a capping efficiency of 125%, which is significantly higher than 100%, despite the availability of a solid standard. In this case, inaccuracies may occur due to the weighing process and further dilutions. Based on own experience in weighing nucleoside quantities in the low milligram range, variations of about 10% were observed, which does not only concern the cap-related modifications but also the main nucleosides used for calculation of the injected RNA amount. One option to determine the exact concentration of a standard solution may be the use of quantitative NMR, a common method for determining concentration and purity of small molecules.[304–306] Another option is the use of UV absorbance-based quantification using known extinction coefficients.

The LC-MS/MS analysis itself was performed according to established protocols, while avoiding known pitfalls in RNA modification quantification.[248,307] However, it is important to note that an isotope labeled internal standard was only available for m^7G and not for 3'*O*-Me-m^7G. This introduces an additional source of inaccuracy, since the use of isotopically labeled standards is essential for reliable quantification by compensating for fluctuations in signal intensity, e.g. due to matrix effects of the sample.

Several methods to determine capping efficiencies of IVT-mRNAs have been described, including radiolabeling approaches using [γ-^{32}P]GTP for enzymatic post-transcriptional capping and subsequent monitoring of the elimination of mRNA radioactivity.[308] Another approach involves the assessment of the incorporation of several cap analogs in the context of co-transcriptional capping using [α-^{32}P]GTP or [α-^{32}P]ATP. Following digestion by RNase T2, cap structures containing [^{32}P]-labeled 3'-monophosphate are resolved by anion exchange HPLC.[158,309] Recently, further approaches were developed that avoid the use of radioactivity. These involve the use of specifically designed ribozymes, RNase A- or RNase H-based cleavage after annealing of a complementary biotin-tagged oligonucleotide, which results in the generation of short 5'-cleavage products. These products are then subjected to analysis by PAGE or LC-MS to compare capped and uncapped fragments.[203,310,311]

In particular, the methods that lead to short 5'-cleavage products (using ribozymes or RNases) have the advantage that uncapped fragments can also be assigned. This benefit allows a direct relative comparison between capped and uncapped transcripts, building the main difference to the approach investigated in this thesis. Since in this investigated method no resulting fragment can be assigned to the uncapped part, an absolute quantification of the cap modification and the RNA amount is necessary. This allows for the potential sources of error

previously identified, including standard and calibration inaccuracies, insufficient removal of capping reaction components, and the presence of IVT by-products that do not align with the desired product length.

In summary, the method presented is therefore not sufficiently precise and is error-prone, making it inadequate to ascertain an exact capping efficiency. However, the majority of identified potential errors were associated with the performance of IVT reactions and especially with the subsequent purification. Solving these issues is independent of the method presented for determining capping efficiencies of high relevance, to avoid undesired immune responses and to improve mRNA translation levels.[295] Especially considering the increased relevance of IVT RNA for therapeutic applications, suitable purification methods such as chromatography-based approaches are recommended.[312] Once the sources of error associated with the implementation of IVT reactions have been eliminated, the described LC-MS/MS method can be employed for the comparative evaluation of capping reactions.

4. Conclusion and Perspectives

This thesis focused on the detection and quantification of biological and synthetic cap structures using LC-MS/MS-based approaches. One major objective was to develop an LC-MS/MS method for the absolute quantification of RNA cap modifications, detecting cap dinucleotides after the digestion of RNA to nucleoside level while leaving the cap triphosphate bridge intact. Cap dinucleotides, which were chemically synthesized by collaboration partners, were subjected to comprehensive method development and optimization. The examination of chromatographic parameters, comprising buffer ion strength, pH, and column temperature, revealed a significant impact on signal strength, peak shape, and separation. The subsequent optimization of ion source parameters led to an additional increase in signal abundance, resulting in final instrument detection limits in the attomole to femtomole range. The enhanced sensitivity of this LC-MS/MS approach was further demonstrated in the context of biological samples, with the successful detection of cap structures in amounts exceeding 100 ng of poly(A) RNA. Method validation revealed that the utilization of isotope labeled cap dinucleotide standards considerably increased quantification accuracy. Furthermore, the feasibility of simultaneously detecting cap dinucleotides and modified nucleosides in a single run was successfully demonstrated. The application to biological samples, including HEK cells and mouse brain tissues, confirmed previous studies on the appearances of major cap structures and tissue-specific differences in *N*6-methylation extents of adenosines. Additionally, analyzing the knockout of the cap-modifying enzyme CMTR1 provided insights into substrate preferences in the cap methylation process.

In conclusion, this analytical method offers precise and reliable quantification of cap structures in biological samples while enabling the simultaneous detection of modified nucleosides. The straightforward and sensitive method is characterized by low detection limits and low-input mRNA requirements, thus offering broad applicability for further investigation of cap abundances, functions, and regulation processes in cells and tissues.

A further objective of this study was to develop an accessible and widely applicable method for the preparation of isotope labeled cap dinucleotide standards, since internal standards contribute to increasing LC-MS/MS accuracy, as shown in the previous part of the thesis. To this end, the property of purines to exchange C-8 hydrogens by deuterium during incubation in D_2O at elevated temperatures was exploited to obtain doubly-deuterated cap dinucleotides,

resulting in a mass shift of +2. Accordingly, incubation parameters, which include temperature and duration, were examined and the deuteration pattern was monitored via LC-MS/MS. Cap dinucleotides not containing an m^7G moiety showed nearly complete C-8 deuteration at incubation temperatures exceeding 85 °C for at least 6 hours. A drawback was the limited stability of these cap dinucleotides under the described harsh incubation conditions, resulting in a decline of signal intensity to 20-30% of the initial level. In contrast, m^7G exhibited distinct behavior compared to the other investigated purines, showing rapid exchange in contact with D_2O or H_2O, which prevented stable isotope labeling. Studies testing the pH range of 5-9 did not reveal significant changes in the deuteration process.

These findings indicate that the deuterium exchange approach is not suitable for preparing isotope labeled cap dinucleotide standards containing m^7G, which is the predominant cap structure in eukaryotic and viral mRNAs. Thus, this work underscores the need for alternative strategies for producing such standards. Future research could focus on the development of other accessible and widely applicable methods for the generation of isotope labeled cap dinucleotide standards. Exploring different labeling techniques, such as enzymatic incorporation of labeled nucleotides or optimized chemical synthesis strategies, could provide valuable solutions in this area.

The third part of the thesis dealt with synthetic mRNA, more specifically the application of an LC-MS/MS method to determine capping efficiencies of IVT RNAs. The concept is based on the complete enzymatic digestion of RNA to the nucleoside level, including the hydrolysis of the cap triphosphate bridge, resulting in the release of cap-related modifications, such as m^7G or 3'*O*-Me-m^7G, dependent on the capping reaction performed. This cap modification is quantified and related to the amount of injected RNA, yielding the number of cap modifications per RNA. While the application to a fully capped oligonucleotide revealed promising accuracy, the analysis of IVT RNAs with lengths of over 800 nucleotides encountered several challenges. Analysis of co-transcriptionally capped RNA revealed capping efficiencies of more than 300%, suggesting potential difficulties in the purification of the resulting IVT product, including inadequate removal of capping reaction components and the presence of IVT by-products which did not correspond to the expected product length. These findings were substantiated by subsequent gel-purified IVT product measurements, which revealed a notable decline in calculated capping efficiencies. However, the outcome of co- and post-transcriptionally capped gel-purified IVT RNA remained above 100%, indicating the presence of additional factors likely attributed to standard and calibration inaccuracies. In contrast to alternative methodologies for determining capping efficiency, which compare the proportion of uncapped transcript with that

of capped transcript, this is not feasible with the described method, as uncapped transcript cannot be identified. This necessitates the precise absolute quantification of cap modification and RNA, thereby having to deal with the previously mentioned sources of error, including inadequate removal of capping reaction components, the presence of IVT by-products that do not align with the correct product length, and standard and calibration inaccuracies.

In conclusion, while the method presented lacks the precision required to determine exact capping efficiencies, it offers potential for future applications. If pure IVT RNA products were available, measuring cap-related modifications after total digestion could facilitate the relative comparison of different cap analogs and capping reaction parameters, as well as support quality control of standard capping procedures. Future research could focus on improving the method's precision and assessing its applicability to purified RNA samples. Although currently with limitations, this study suggests promising directions for further research and development in RNA capping analysis.

5. Material and Methods

5.1. Material

5.1.1. Chemicals

Acetic acid, LC-MS grade	Sigma-Aldrich (Steinheim, Germany)
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 hydroxide solution, 25%, LC-MS grade	Sigma-Aldrich (Steinheim, Germany)
Ammonium persulfate (APS)	Carl Roth (Karlsruhe, Germany)
ATP, 100 mM (HiScribe T7 Kit)	New England Biolabs (Ipswich, USA)
Chloroform, HPLC grade	Sigma-Aldrich (Steinheim, Germany)
CTP, 100 mM (HiScribe T7 Kit)	New England Biolabs (Ipswich, USA)
Deuterium oxide >99.9%	Deutero GmbH (Kastellaun, Germany)
Diethyl ether	Sigma-Aldrich (Steinheim, Germany)
Dulbecco's Balanced Salt Solution (DPBS)	Thermo Fisher Scientific (Waltham, USA)
Dulbecco's Modified Eagle Medium (DMEM)	Thermo Fisher Scientific (Waltham, USA)
Ethanol 99.9% (*V/V*, absolute)	Carl Roth (Karlsruhe, Germany)
Ethylenediaminetetraacetic acid (EDTA)	Carl Roth (Karlsruhe, Germany)
Fetal bovine serum (FBS)	Thermo Fisher Scientific (Waltham, USA)
Gel Loading Dye, Purple (6X)	New England Biolabs (Ipswich, USA)
GelRed™ (3x)	Biotium (Hayward, USA)
Glycerol	Sigma-Aldrich (Steinheim, Germany)
Glycogen, RNA-grade	Thermo Fisher Scientific (Waltham, USA)
GpppA	New England Biolabs (Ipswich, USA)
GTP, 10 mM (Vaccinia Capping Kit)	New England Biolabs (Ipswich, USA)
GTP, 100 mM (HiScribe T7 Kit)	New England Biolabs (Ipswich, USA)

Isopropyl alcohol, HPLC grade	Carl Roth (Karlsruhe, Germany)
Lithium chloride	Carl Roth (Karlsruhe, Germany)
m^7GpppA	New England Biolabs (Ipswich, USA)
Magnesium chloride solution, 1 M	Sigma-Aldrich (Steinheim, Germany)
2-Methyladenosine	BOC Sciences (Shirley, USA)
5-Methylcytidine	Berry&Associates (Dexter, USA)
7-Methylguanosine	Sigma-Aldrich (Steinheim, Germany)
2'*O*-Methylcytidine	Berry&Associates (Dexter, USA)
N,N,N',N'-Tetramethylethylenediamine (TEMED)	Carl Roth (Karlsruhe, Germany)
Oligo d(T)25 Magnetic Beads	New England Biolabs (Ipswich, USA)
Penicillin-Streptomycin (10.000 U/mL)	Thermo Fisher Scientific (Waltham, USA)
Pentostatin ≥ 95%	Sigma-Aldrich (Steinheim, Germany)
Pseudouridine	Sigma-Aldrich (Steinheim, Germany)
RNaseOUT™	Thermo Fisher Scientific (Waltham, USA)
Roti® Phenol (for DNA)	Carl Roth (Karlsruhe, Germany)
Rotiphorese (10x) TBE buffer	Carl Roth (Karlsruhe, Germany)
Rotiphorese Sequencing gel buffer concentrate	Carl Roth (Karlsruhe, Germany)
Rotiphorese Sequencing gel concentrate	Carl Roth (Karlsruhe, Germany)
Rotiphorese Sequencing gel diluent	Carl Roth (Karlsruhe, Germany)
SAM, 32 mM (Vaccinia Capping Kit)	New England Biolabs (Ipswich, USA)
Tetrahydrouridine, InSolution	Merck (Darmstadt, Germany)
TRI Reagent®	Sigma-Aldrich (Steinheim, Germany)
Tris, 1 M, pH 8.0, RNase-frei	Thermo Fisher Scientific (Waltham, USA)
Tris-HCl	Carl Roth (Karlsruhe, Germany)
Trypsin-EDTA, 0.05%	Thermo Fisher Scientific (Waltham, USA)
UTP, 100 mM (HiScribe T7 Kit)	New England Biolabs (Ipswich, USA)
Zinc chloride	Sigma-Aldrich (Steinheim, Germany)

5.1.2 Buffers, solutions, and growth media

Buffers and growth media were prepared with Milli-Q H_2O, unless stated otherwise.

5.1.2.1 Buffers and solutions

APS solution	10% APS (*m/V*)
Capping buffer, 10x	500 mM Tris-HCl (pH 8), 50 mM KCl, 10 mM $MgCl_2$, 1 mM DTT
D_2O solutions	30 mM NH_4OAc in D_2O (pH 5, 6, 7, 8, 9, adjusted with glacial acetic acid or ammonium hydroxide)
DNase I buffer, 10x	100 mM Tris-HCl (pH 7.5), 25 mM $MgCl_2$, 1 mM $CaCl_2$
LC-MS/MS solvents	- 5, 10, 20 mM NH_4OAc (pH 4, 5, 6, 7, 7.5) - 30 mM NH4OAc (pH 7) - 5 mM NH_4OAc (pH 5.3) + 1% (*V/V*) ACN
Nuclease P1 buffer, 10x	200 mM NH_4OAc (pH 5.5), 2 mM $ZnCl_2$
Oligo d(T)25 binding buffer	20 mM Tris-HCl (pH 7.5), 1.0 M LiCl, 2 mM EDTA
Oligo d(T)25 washing buffer	10 mM Tris-HCl (pH 7.5), 0.15 M LiCl, 1 mM EDTA
Phosphatase buffer, 10x	50 mM Tris (pH 8), 10 mM $MgCl_2$
TBE buffer, 10x	1 M tris boric acid, 20 mM EDTA (pH 8.3)

5.1.2.2 Growth media

HEK cell culture medium	DMEM supplemented with 10% (*V/V*) FBS and 10 U/mL penicillin-streptomycin

5.1.3 Enzymes and kits

Benzonase® Nuclease (> 250 U/µL)	Sigma Aldrich (Steinheim, Germany)
Bovine intestine phosphatase	Sigma Aldrich (Steinheim, Germany)
Cfr42I (SacII) (10 U/µL)	Thermo Fisher Scientific (Waltham, USA)
DNase I, RNase-free (1 U/µL)	Thermo Fisher Scientific (Waltham, USA)
FastAP thermosensitive Alkaline Phosphatase (1 U/µL)	Thermo Fisher Scientific (Waltham, USA)
Faustovirus Capping Enzyme (FCE)	New England Biolabs (Ipswich, USA)
HiScribe T7 High Yield RNA Synthesis Kit	New England Biolabs (Ipswich, USA)
MEGAclear™ Kit	Thermo Fisher Scientific (Waltham, USA)
Nuclease P1 from *Penicillium citrinum* (lyophilized)	Sigma Aldrich (Steinheim, Germany)
RNA Clean&Concentrator-5 Kit	Zymo Research (Freiburg, Germany)
Snake Venom Phosphodiesterase from *Crotalus adamanteus venom* (lyophilized)	Worthington Biochemical Corporation (Lakewood, USA)
T7 RNA Polymerase Mix (HiScribe T7 kit)	New England Biolabs (Ipswich, USA)
Vaccinia Capping Enzyme (VCE)	New England Biolabs (Ipswich, USA)

5.1.4 Oligonucleotides and plasmids

5.1.4.1 RNA oligonucleotides

Name	Sequence	Supplier
MH1505	pppGGCAGACACCAUGmGUGCACCUGACUCCUGAGGAGAAGU	Bio-Synthesis (Lewisville, USA)

5.1.4.2 Plasmids

Vector	Name	Implemented sequence
pUC 57	eGFP-GG	CTCCAGCTTTTGTTTAATACGACTCACTATAGGGCGAATTGGGTACC GGGCCCCCCCTCGAAGACAAGCTTCCTGCAGGTCGACTCTAGAG GATCCCGGGTACCGAGCTCGAATTCGGCTTCCACCATGGTGAGCA

		AGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAG
		CTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGA
		GGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCAT
		CTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGA
		CCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACC
		ACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCT
		ACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACA
		AGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAAC
		CGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATC
		CTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATA
		TCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGA
		TCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCAC
		TACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCC
		CGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCC
		CAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGC
		CGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGG
		CCGCCACCGC---GGTGGAGCTCCAGCTTTTGTT
pUC57	eGFP	CTCCAGCTTTTGTTTAATACGACTCACTATAAGGCGAATTGGGTACC
	-AG	GGGCCCCCCCTCGAAGACAAGCTTCCTGCAGGTCGACTCTAGAG
		GATCCCGGGTACCGAGCTCGAATTCGGCTTCCACCATGGTGAGCA
		AGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAG
		CTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGA
		GGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCAT
		CTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGA
		CCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACC
		ACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCT
		ACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACA
		AGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAAC
		CGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATC
		CTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATA
		TCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGA
		TCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCAC
		TACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCC
		CGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCC
		CAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGC

CGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGG
CCGCCACCGC---GGTGGAGCTCCAGCTTTTGTT

Plasmids including the oligo corresponding to the respective gene of interest were synthesized and cloned into the pUC57 backbone by GenScript (Piscataway, USA).

5.1.5 Cell lines

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

5.1.6 Disposables

B Braun™ Cutfix Stainless Steel Scalpels	Thermo Fisher Scientific (Waltham, USA)
Eppendorf tubes (1.5 mL)	Carl Roth (Karlsruhe, Germany)
Falcon tubes (15, 50 mL)	Sarstedt (Nümbrecht, Germany)
Inserts, conical, clear glass (0.1 mL)	neoLab Migge GmbH (Heidelberg, Germany)
NanosepR 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)
Pipette tips (with filter, sterile, RNase/ DNase-free)	Greiner Bio-One (Frickenhausen, Germany)
Screw-cap ND9 (blue, septum red rubber/ PTFE beige)	neoLab Migge GmbH (Heidelberg, Germany)
Serological pipettes (2, 5, 10, 25, 50 mL)	Sarstedt (Nümbrecht, Germany)
Short thread vials ND9 (1.5 mL)	neoLab Migge GmbH (Heidelberg, Germany)
Surgical scalpel, guarded, SWANN-MORTON	VWR (Darmstadt, Germany)

5.1.7 Instruments

Analytical balances	
Mettler Toledo PM460	Mettler Toledo (Gießen, Germany)
Sartorius Cubis Analytical Balance	Sartorius (Goettingen, Germany)
Sartorius Quintix 65-1S	Sartorius (Goettingen, Germany)
Centrifuges	
Eppendorf Centrifuge 5810R	Eppendorf (Hamburg, Germany)
Eppendorf Centrifuge 5430R	Eppendorf (Hamburg, Germany)
Sprout mini centrifuge	Biozym (Hessisch Oldendorf, Germany)
Thermo Scientific™ Megafuge 8R	Thermo Fisher Scientific (Waltham, USA)
Eukaryotic cell culture	
Heraeus BB15 CO2 incubator	Thermo Fisher Scientific (Waltham, USA)
Herasafe™ HS 12	Thermo Fisher Scientific (Waltham, USA)
Invitrogen™ Countess™ 3FL automated cell counter	Thermo Fisher Scientific (Waltham, USA)
Microscope DM IRB	Leica Microsystems (Wetzlar, Germany)
Gel electrophoresis	
Electrophoresis Power Supply – EPS 3500XL	GE Healthcare (Buckinghamshire, UK)
Biometra Eco-Maxi	Analytik Jena (Jena, Germany)
Typhoon TRIO+ variable mode imager	GE Healthcare (Chicago, USA)
High performance liquid chromatography & columns	
Agilent 1260 Infinity (II) LC	Agilent Technologies (Waldbronn, Germany)
Synergi Fusion RP C18 column (2 x 250 mm, 4 µm particle size, 80 Å pore size)	Phenomenex (Aschaffenburg, Germany)

Poroshell 120EC-C18 column (3 x 150 mm, 2.7 µm particle size)	Agilent Technologies (Waldbronn, Germany)

General equipment

NanoDrop™ ND-2000	PeqLab (Erlangen, Germany)
pH-meter FiveEasy™ FE20	Mettler Toledo (Gießen, Germany)
pH electrode LE422	Mettler Toledo (Gießen, Germany)
Pipette boy Integra	VWR (Darmstadt, Germany)
TapeStation 4200	Agilent Technologies (Waldbronn, Germany)
Ultrapure water purification system Milli-QR	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)

Heatblocks & Thermoshaker

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

Mass Spectrometry

Agilent 6460A triple quadrupole	Agilent Technologies (Waldbronn, Germany)
Agilent 6470B triple quadrupole	Agilent Technologies (Waldbronn, Germany)
Genius XE-70, nitrogen generator	Peak Scientific (Düren, Germany)

5.1.8 Software/ KI

Adobe Illustrator 2024 V5	Adobe Inc. (San José, USA)
ChemBioDraw Ultra 14.0	Cambridge Soft/PerkinElmer (Waltham, USA)

Citavi 6	Swiss Academic Software (Wädenswil, Switzerland)
Deepl Write	DeepL SE (Cologne, Germany)
GraphPad Prism 9	GraphPad Software, Boston, Massachusetts USA
ImageJ V5	Created by Wayne Rasband (Madison, USA)
IrfanView	created by Irfan Skiljan
MassHunter software, version 10.0	Agilent Technologies (Waldbronn, Germany)
Microsoft Office 365	Microsoft (Redmont, USA)
Perplexity AI	Perplexity AI, Inc. (San Francisco, USA)
Typhoon scanner software	GE Healthcare (Chicago, USA)

5.2. Methods

5.2.1. Method optimization for absolute quantification of RNA cap modifications

In this project part, an Agilent 1260 Infinity (II) series HPLC connected to an Agilent 6470B triple quadrupole mass spectrometer interfaced with an Agilent Jet Stream electrospray ionization source was used.

5.2.1.1. Preparation of standard solutions

The cap analogs GpppA and m^7GpppA were obtained commercially from New England Biolabs. Several other cap analogs, including Gpppm^{6}A, GpppAm, Gpppm^{6}Am, m^7Gpppm^{6}A, m^7GpppAm, and m^7Gpppm^{6}Am, were synthesized by [redacted] as published previously.[255] The standards were weighted in and dissolved in nuclease-free water. The absorbance of each solution was measured at 260 nm using a Nanodrop 2000 spectrophotometer. Subsequent concentration calculations were based on previously published extinction coefficients.[201] Standard solutions were prepared by combining equal concentrations of the various cap dinucleotides.

5.2.1.2. Analyte specific mass parameters

Pure standard samples, each containing 10 pmol, were subjected to LC-MS/MS analysis. A constant mobile phase composition of 90% solvent A (20 mM ammonium acetate, pH 6, adjusted with glacial acetic acid) and 10% solvent B (LC-MS grade acetonitrile) was employed. The system operated at a flow rate of 0.35 mL/min, with the temperature maintained at 35 °C and a capillary loop was installed in place of an HPLC column. The MassHunter Optimizer software was used to automate the optimization process of the analyte specific mass parameters, including precursor and product ion *m/z*, fragmentor voltages, and collision energies. The MS was operated in full scan mode to determine the precursor ion. Subsequently, various fragmentor voltages ranging from 50 to 250 V were applied and investigated using the SIM mode and the voltages yielding maximum abundance were selected. Next, a product ion scan was conducted using the previously identified precursor ion *m/z* and the optimized fragmentor voltages. Various collision energies in a range of 10 to 90 V were applied in the collision cell, while the MS was operated in SRM mode. The most abundant resulting product ions, along with their respective collision energies, were automatically selected. The optimized parameters are summarized in Supplement, Table S1.

5.2.1.3. HPLC optimization

Various chromatographic conditions were systematically evaluated to optimize the analysis. These included adjusting the mobile phase composition (ammonium acetate concentration and pH), column temperature, and gradient profiles. Table 1 provides a comprehensive overview of the HPLC and MS parameters tested during each optimization step.

Table 1. HPLC and ion source parameters for each optimization step

Optimized parameter	Ammonium acetate concentration [mM] of solvent A	pH of solvent A	Column temperature [°C]	HPLC gradients (Table 2)	Ion source settings (Table 3)
Starting point	20	6	20	1	A
Ion strength	5, 10, 20	6	20	1	A
pH	5	4, 5, 6, 7	20	1	A
Column temperature	5	7	15, 20, 25, 35	2	A
Ion source optimization	5	7	15	2	A, B
Optimized method	5	7	15	3	B

Standard solutions containing a mixture of cap dinucleotides, ranging from 10 fmol to 1 pmol, were injected. The mobile phase consisted of solvent A (5, 10, or 20 mM ammonium acetate, with pH adjusted to 4, 5, 6, 7, or 7.5 using acetic acid or ammonium hydroxide) and solvent B (acetonitrile). The flow rate was set at 0.35 mL/min, with different gradient profiles applied as detailed in Tables 1 and 2. Chromatographic separation was performed using a Poroshell 120EC-C18 column (3 x 150 mm, 2.7 µm particle size) at various temperatures (15, 20, 25, or 35 °C). Mass spectrometric detection was carried out in MRM mode. The previously optimized analyte specific MS parameters were employed along with the source parameters described in Table 3. Data analysis was conducted using MassHunter software.

Table 2. HPLC gradients

Gradient 1			Gradient 2			Gradient 3		
Time [min]	A [%]	B [%]	Time [min]	A [%]	B [%]	Time [min]	A [%]	B [%]
0	98	2	0	100	0	0	100	0
1	98	2	0.1	100	0	0.1	100	0
6	85	15	0.11	99	1	0.11	98	2
7	50	50	1.5	99	1	2	98	2
7.5	50	50	2	98	2	4	96	4
8.5	98	2	4.5	98	2	6	85	15
15	98	2	5	96	4	8	70	30
			6	96	4	10	10	90
			8	70	30	10.5	10	90
			9.5	10	90	11.5	100	0
			10.5	10	90	19	100	0
			11.5	100	0			
			20	100	0			

5.2.1.4. Ion source parameters

A mixture containing 1 pmol of each cap dinucleotide standard was subjected to analysis and the optimized HPLC parameters were implemented. The MassHunter Source Optimizer Software was utilized to systematically optimize various ion source parameters, including capillary voltage, drying gas flow, drying gas temperature, nozzle voltage, sheath gas flow, sheath gas temperature, and nebulizer pressure. These parameters were tested at varying levels and the final settings were chosen based on signal intensity, thereby ensuring compatibility with all analyzed cap dinucleotide standards. Table 3 provides an overview of the ion source parameters.

Table 3. Ion source settings

	Starting parameters A	Optimized parameters B
Capillary voltage [V]	3000	2000
Drying gas flow [L/min]	8	4
Drying gas temperature [°C]	350	350
Nozzle voltage [V]	0	0
Sheath gas flow [L/min]	12	11
Sheath gas temperature [°C]	350	350
Nebulizer pressure [psi]	50	60

5.2.1.5. Simultaneous quantification of cap and other (modified) nucleosides

Two mixtures were subjected to LC-MS/MS analysis. The first contained 10 fmol each of the investigated cap dinucleotides and of various modified nucleosides, including m^7G, Cm, m^2A, m^6A, and $m^{6,6}A$. The second mixture contained 10 pmol of each main nucleoside (C, U, G, and A). The measurement was performed as described in chapter 5.2.1.3., using the parameters of the optimized method, detailed in Table 1. The main nucleosides were detected UV spectroscopically at 254 nm. The cap dinucleotides and modified nucleosides were analyzed by mass spectrometry, the MS settings for the modified nucleosides are listed in Table 4.

Table 4. Mass spectrometer settings for modified nucleosides comprising precursor and product ion mass-to-charge ratios (*m/z*), fragmentor voltages, and collision energies.

Compound	Precursor Ion [*m/z*]	Product Ion [*m/z*]	Fragmentor voltage [V]	Collision Energy [V]	^{13}C-labeled precursor ion [*m/z*]	^{13}C-labeled product ion [*m/z*]
Cm	258	112	80	13		
m^2A	282	150	110	21		
m^5C	258	126	75	13	268	131
m^6A	282	150	110	21	293	156
$m^{6,6}A$	296	164	115	21		
m^7G	298	166	80	13	309	172
3'*O*-Me-m^7G	312	166	80	12		
Ψ	245	209	85	9	254	218

5.2.1.6. Instrument detection limits

Instrument detection limits were determined according to technical guidelines from Agilent Technologies.[268,269] Briefly, multiple consecutive injections of an appropriate quantity of analyte were performed. This quantity should produce an *S/N* between 5 and 10. The IDL was then calculated using the relative standard deviation of the signal intensity.

For the starting point method, 12 successive injections of 5 fmol of the cap dinucleotide standard mixture were carried out. For the optimized method, a lower concentration of 2 fmol was utilized for the same procedure. The used LC-MS/MS parameters are listed in Table 1.

5.2.2. Biological sample extraction and RNA isolation

5.2.2.1. HEK cells

HEK cells were cultured by Lukas Gleue as described previously.[255]

Briefly, HEK293 cells were grown in DMEM supplemented with 10% FBS and 10 U/mL Penicillin-Streptomycin at 37 °C in a 5% CO_2 atmosphere. After washing with DPBS, the cells were detached using Trypsin-EDTA. For RNA extraction, the detached cells were resuspended in DMEM, centrifuged at 400 x g for 5 min, and the supernatant was discarded. The obtained

pellets were washed with DPBS, then resuspended in 1 mL TRI Reagent® (per $5x10^6$ cells) and stored at −20 °C.

5.2.2.2. Mouse tissue

Mice were sacrified and brain regions were dissected by Prof. Kristina Friedland as described previously.[255]

Two male B6.129S mice (B6.129S-Trpc6tm1Lbi/Mmjax), aged between 39 and 41 weeks, were used in the study. For details of housing conditions, see Brandscheid et al.[313] All procedures were in accordance with the European Council Directive on the Ethical Treatment of Animals and were approved by the local authorities (LUA-Rhineland-Palatinate). The mice underwent decapitation while under isoflurane anesthesia for humane euthanasia. Brains were then removed and regions of interest, including the hippocampus, cortex, and cerebellum, were dissected, snap frozen in liquid nitrogen, and stored at −80 °C.

5.2.2.3. CMTR1 k.o. cells

A CMTR1-deficient HEK293T cell line was generated using CRISPR/Cas9 technology by [redacted]. For control purposes, a CMTR1 knockout clone was stably reconstituted with a retroviral vector expressing CMTR1. Extracted total RNA from these cell lines was kindly provided for further analysis. Experimental details are provided in the references.[272,273]

5.2.2.4. Total RNA isolation

Total RNA was extracted from either cell pellets or tissue using TRI Reagent® according to the manufacturer's protocol. For cell pellets, 1 mL of TRI Reagent® was added for every $5x10^6$ cells. Mouse brain tissues were mixed with 5 mL of TRI Reagent® and thoroughly homogenized with a pestle, except for hippocampus tissue, which was resuspended in only 1 mL of TRI Reagent®. In a next step, 200 µL of chloroform per mL of TRI Reagent® was added, and the mixture was incubated at room temperature for 2 min before centrifugation (13000 x g for 15 min at 4 °C). The upper aqueous phase was transferred to a new tube, and the chloroform extraction process was repeated. The aqueous phase was again transferred to a new tube, and 0.5 mL of isopropyl alcohol per mL of TRI Reagent®, along with 1 µL of

glycogen, was added for precipitation. After incubating for 10 min at room temperature, another centrifugation step (13000 x g for 30 min at 4 °C) was conducted. The supernatant was discarded, and the RNA pellet was washed with 1 mL of ice cold 75% (*V/V*) ethanol. After a final centrifugation step (13000 x g for at least 45 min at 4 °C), the supernatant was removed, and the RNA pellet was air-dried and dissolved in nuclease-free water. Determination and assessment of RNA concentration and integrity were conducted using a Nanodrop 2000 UV-VIS spectrometer and an Agilent TapeStation 4200.

5.2.2.5. Isolation of poly(A)-tailed RNA

The purification of poly(A)-tailed RNA from total RNA was carried out following the manufacturer's guidelines for Dynabeads (Thermo Scientific), with some modifications made to the protocol. 100 µL of prewashed oligo d(T)25 magnetic beads (NEB) were used per 60 µg of total RNA. The RNA was incubated with the beads in a binding buffer (20 mM Tris-HCl, pH 7.5, 1.0 M LiCl, 2 mM EDTA) for 5 min while rotating at room temperature. After binding, the tube was placed on a magnet to separate the beads, and the supernatant was carefully discarded. Two washing steps were performed by mixing the beads with washing buffer (10 mM Tris-HCl, pH 7.5, 0.15 M LiCl, 1 mM EDTA) before removing the supernatant each time. The RNA was then eluted from the beads by adding 100 µL of nuclease-free water and heating at 70 °C for 2 min. Following this, the tube was placed on a magnet again, and the supernatant containing the poly(A) RNA was transferred to a new tube. The procedure was repeated for a second round of purification. Finally, the aqueous eluate containing poly(A) RNA underwent further purification and concentration using the RNA Clean & Concentrator-5 Kit, following the manufacturer's instructions.

5.2.3. LC-MS/MS analysis of cap dinucleotides in biological samples

For method validation, 2 µg of total RNA from HEK cells were spiked with 400 fmol of each cap dinucleotide. In separate experiments, distinct amounts of poly(A) RNA were utilized, including up to 31.25 µg from HEK cells, 1-3 µg from mouse brain tissue, and 500 ng from both the CMTR1 k.o. and control sample. All samples were prepared for absolute quantification via LC-MS/MS. Each sample was supplemented with a SILIS cap mixture containing 400 fmol of D_3C-m^7Gpppm^{6}A, D_3C-m^7GpppAm, and D_3C-m^7Gpppm^{6}Am. The samples were then hydrolyzed to nucleotide level, while leaving the cap triphosphate bridge intact, by incubation with 0.6 U nuclease P1 in 1/10 volume of 10x nuclease P1 buffer (2 mM $ZnCl_2$, 200 mM ammonium

acetate, pH 5.5) for 1 h at 37 °C. Following this, the nucleotide mixtures were further hydrolyzed to nucleosides by incubation with 1 U FastAP in 1/10 volume of 10x phosphatase buffer (10 mM $MgCl_2$, 50 mM Tris, pH 8) for an additional 1 h at 37 °C. Subsequently, each digested sample was spiked with 5 ng of previously digested ^{13}C-labeled nucleosides from *S. cerevisiae*, brought to a specific volume using a mixture of both digestion buffers, and subjected to LC-MS/MS analysis.

The LC-MS/MS parameters of the optimized method (see Table 1, 2 and 3) were applied. The UV signal was recorded at 254 nm by a diode array detector to detect main nucleosides photometrically. The MS was operated in MRM mode, detecting cap dinucleotides and modified nucleosides, depending on the specific experiment, with corresponding analyte-specific parameters provided in Table S1 and Table 4. Absolute quantification was performed as described by Thüring *et al*.[248] Briefly, external calibration dilutions for both cap dinucleotides and nucleosides were prepared. These dilutions were spiked with 500 fmol of a SILIS cap mixture and 5 ng ^{13}C labeled nucleosides from *S. cerevisiae*, as was done with the samples. Following this, the dilutions were adjusted to the appropriate volume using digestion buffer before being injected into the LC-MS. The resulting standard curves facilitated the calculation of cap dinucleotides and modified nucleoside quantities in the sample, accounting for SILIS corrections. Of note, in cases where a structurally identical SILIS was not available for specific cap dinucleotides, D_3C-m7Gpppm^{6}A was utilized. For mouse brain samples and CMTR1 k.o. samples, including their controls, the quantities of cap dinucleotides were normalized to the amount of injected RNA, expressed as cap dinucleotide per 1000 nucleotides. Consequently, an additional standard curve of main nucleosides was measured to determine the sum of the amount of the four main nucleosides within a sample.

5.2.4. Deuterium exchange

Cap dinucleotide standards were weighed in and dissolved in D_2O. The solutions were incubated in heating blocks at various temperatures and aliquots were withdrawn at specified time points. The solutions were stored at −20 °C until further use.

For pH experiments, aqueous solutions containing 30 mM ammonium acetate were prepared. pH values were adjusted to 5, 6, 7, 8, or 9 using acetic acid or ammonium hydroxide and the cap dinucleotide standards were solved and diluted in these prepared solutions.

5.2.5. LC-MS/MS analysis of deuterated cap dinucleotides

Of note, this subproject was initiated prior to the optimization of the LC-MS/MS method for cap dinucleotides. Thus, this described method differs from the previously presented optimized LC-MS/MS method. In this project part, an Agilent 1260 Infinity (II) series HPLC connected to an Agilent 6460A triple quadrupole mass spectrometer interfaced with an Agilent Jet Stream electrospray ionization source was used.

5 pmol of cap dinucleotide standard in D_2O was subjected to LC-MS/MS analysis. The mobile phases consisted of solvent A (30 mM ammonium acetate at pH 7, adjusted with ammonium hydroxide) and solvent B (acetonitrile). A gradient was applied starting with 100% solvent A for 1 min, followed by a linear increase to 30% solvent B at 7 min and to 50% after 9 min. These conditions were maintained for 1 min before returning to 100% solvent A conditions within 2.5 min and re-equilibrating the column for 7 min. The flow rate was set to 0.5 mL/min and a Poroshell 120EC-C18 column (3 x 150 mm, 2.7 µm particle size) was used at a temperature of 25 °C. When the column was exchanged through a capillary loop, a constant mobile phase composition of 80% solvent A and 20% solvent B was employed at a flow rate of 0.5 mL/min for 3 min. The mass spectrometer was operated in the positive ion mode and the ESI parameters were set as follows: gas temperature 350 °C, gas flow 8 L/min, nebulizer 50 psi, sheath gas temperature 350 °C, sheath gas flow 12 L/min, capillary voltage 3000 V, and nozzle voltage 0 V. The investigated mass transitions are listed in Table 5 and were measured operating in neutral loss scan. Data analysis was conducted using MassHunter software.

Table 5. Mass spectrometer settings and investigated mass transitions (precursor ion *m/z* → product ion *m/z*)

	Mass transitions				Fragmentor voltage [V]	Collision energy [V]
	Non-deuterated	Only A/Am/m^6Am-deuterated	Only G/m^7G-deuterated	Doubly-deuterated		
GpppA	773 → 136	774 → 137	774 → 136	775 → 137	135	70
GpppAm	787 → 136	788 → 137	788 → 136	789 → 137	135	70
Gpppm^{6}Am	801 → 150	802 → 151	802 → 150	803 → 151	135	70
m^{7}GpppA	787 → 136	788 → 137	788 → 136	789 → 137	135	70

5.2.6. Synthesis of *in vitro* transcribed RNA

5.2.6.1. Plasmid DNA linearization

A sequence coding for eGFP mRNA (AG as first nucleotides after the promotor sequence when using CleanCap® reagent, otherwise GG) was designed, containing the T7 promotor sequence and a cleavage site of the restriction enzyme SacII (chapter 5.1.4.2). The respective sequences were synthesized and cloned into the circular plasmid pUC57 vector by GenScript, USA.

The plasmid DNA (pDNA) was linearized using the restriction enzyme SacII according to the manufacturer's protocol, with an overnight incubation at 37 °C. Following this, the linearized pDNA was purified using phenol-chloroform extraction. The solution was mixed with chloroform and DNA phenol in a 2:1:1 ratio. After mixing, the upper aqueous phase was transferred to a new tube, and the extraction process was repeated. The aqueous phase was then extracted twice with an equal volume of chloroform, followed by a final extraction with the same volume of diethyl ether to remove any residual phenol. Each extraction step involved vortexing, followed by centrifugation to achieve phase separation, allowing for the transfer of the upper aqueous phase to a new tube.

5.2.6.2. Ammonium acetate ethanol precipitation

Next, ammonium acetate ethanol precipitation was carried out. The solution was mixed with 1/10 volume of 5 M ammonium acetate, 1 µL glycogen, and 3 volumes of ice-cold absolute ethanol. The mixture was then incubated at −80 °C for two hours. After centrifugation (12000 x g, 4 °C, 45 min), the supernatant was removed, and the DNA pellet was subsequently washed with ethanol 75%. This was followed by a final centrifugation step (12000 x g, 4 °C, 30 min). After removing the supernatant again, the DNA pellet was air-dried and subsequently dissolved in nuclease-free water.

5.2.6.3. *In vitro* transcription

In vitro transcription was performed using the HiScribe T7 High Yield RNA Synthesis Kit (NEB, USA) following the manufacturer's instructions.

Without capping, 0.5 µg of the linearized cDNA "eGFP-GG" was mixed with 1/10 volume of 10x T7 reaction buffer and 2 µL of T7 RNA polymerase mix. CTP, UTP, GTP, and ATP were

added to a final concentration of 10 mM, respectively, resulting in a total reaction volume of 20 µL. When performing co-transcriptional capping using ARCA cap analog, 8 mM of the cap analog was added to the IVT reaction, while the concentration of GTP was reduced to 2 mM, achieving an ARCA:GTP ratio of 4:1. For co-transcriptional capping with CleanCap® cap analog, 0.5 µg of the linearized cDNA "eGFP-AG" was mixed with 1/20 volume of 10x T7 reaction buffer and 4 µL of T7 RNA polymerase mix. The cap analog was added to a final concentration of 4 mM and CTP, UTP, GTP, and ATP were added to a final concentration of 5 mM, respectively, resulting in a total reaction volume of 40 µL. The mixtures were incubated at 37 °C for 2 h. The DNA template was digested by addition of 0.1 U DNase I and 1/10 volume of 10x DNase I buffer and subsequent incubation at 37 °C for 30 min. The IVT products were purified using the MEGAClear™ Kit according to manufacturer's instructions.

Negative controls were prepared by adding the cap analog only after DNA template digestion but before purification with the MEGAClear™ Kit.

5.2.7. Enzymatic capping

400 pmol of MH1505, the 5'-triphosphate containing oligonucleotide of 38 nt length, was mixed with nuclease-free water, incubated at 65 °C for 5 min and immediately placed on ice. Subsequently, 1/10 volume of 10x capping buffer was added, as well as 25 U FCE. Additionally, SAM was added to a final concentration of 0.1 mM and GTP was added to a final concentration of 0.5 mM, resulting in a total reaction volume of 20 µL. A negative control was prepared by omitting the enzyme. The mixture was incubated at 37 °C for 135 min. Purification was performed using the RNA Clean & Concentrator-5 Kit according to manufacturer's instructions.

1 µg of gel-purified, uncapped eGFP mRNA (see chapter 5.2.8) was mixed with nuclease-free water, incubated at 65 °C for 5 min and immediately placed on ice. Subsequently, 1/10 volume of 10x capping buffer was added, as well as 10 U VCE. Additionally, SAM was added to a final concentration of 0.2 mM and GTP was added to a final concentration of 0.5 mM, resulting in a total reaction volume of 20 µL. A negative control was prepared by omitting the enzyme. The mixture was incubated at 37 °C for 60 min. Purification was performed using the RNA Clean & Concentrator-5 Kit according to manufacturer's instructions.

5.2.8. Purification via denaturing PAGE

For eGFP mRNA, a 6% denaturing PAGE gel was prepared by mixing 100 µL of a gel solution containing commercially available Rotiphorese Sequencing solutions gel concentrate, gel diluent, and gel buffer concentrate in a 2.4:6.6:1 ratio, with 200 µL APS 10% solution and 20 µL TEMED solution. After polymerization, the gel was loaded with RNA samples that had been previously mixed with 1/6 volume of Gel Loading Dye, Purple (6X). The gel was run in 1x TBE buffer at 18 W for 5.5 h. For oligonucleotides, a 20% denaturing PAGE gel was used (the gel solution was prepared in a ratio of 8:1:1), and run in 1x TBE buffer at 15 W for 5.5 h.

The gels were stained with GelRed and scanned using a Typhoon imager with laser excitation of 532 nm and emission filter of 610 nm. The bands of interest were excised with a scalpel using a printed image of the gel as template. After mashing the gel pieces, 300 µL of 0.5 M ammonium acetate solution was added, and the mixture was shaken overnight at 15 °C and 600 rpm. Subsequently, the mixture was filtered through Nanosep® centrifugal filters to remove gel pieces, and the RNA was precipitated using ammonium acetate/ethanol, as described in chapter 5.2.6.2.

Densitometric analysis for comparison of band intensities was performed using ImageJ.

5.2.9. LC-MS/MS analysis of cap-related nucleoside modifications

Up to 400 ng of purified IVT-RNA were digested to nucleoside level including the hydrolyzation of the cap triphosphate bridge by incubation with a digestion mixture containing 0.6 U nuclease P1, 0.2 U snake venom phosphodiesterase, 0.2 U bovine intestine phosphatase, 10 U benzonase, 200 ng pentostatin, and 500 ng tetrahydrouridine in phosphatase buffer (10 mM $MgCl_2$, 50 mM Tris, pH 8) for 2 h at 37 °C. Subsequently, up to 50 ng of the digested samples were spiked with 0.5 ng of previously digested ^{13}C-labeled nucleosides from *S. cerevisiae* and subjected to LC-MS/MS analysis. In this project part, an Agilent 1260 Infinity (II) series HPLC connected to an Agilent 6470B triple quadrupole mass spectrometer interfaced with an Agilent Jet Stream electrospray ionization source was used. A C18 reverse HPLC column (Synergi™ 4 µM particle size, 80 Å pore size, 250 × 2.0 mm; Phenomenex) was used and the column temperature was maintained at 35 °C. The solvents consisted of buffer A (5 mM NH_4OAc, pH 5.3 adjusted with glacial acetic acid) and buffer B (acetonitrile, LC-MS grade) and a gradient was applied (Table 6) at a constant flow rate of 0.35 mL/min. The UV signal was recorded at 254 nm by a diode array detector to detect main nucleosides photometrically. The MS was operated in MRM mode, detecting the nucleoside modifications m^7G and 3'*O*-Me-m^7G, with

analyte-specific parameters provided in Table 4. The ESI parameters were set as follows: gas temperature 300 °C, gas flow 7 L/min, nebulizer 60 psi, sheath gas temperature 400 °C, sheath gas flow 12 L/min, capillary voltage 3000 V, and nozzle voltage 0 V. Absolute quantification was performed as described by Thüring *et al*.[248] Briefly, external calibration dilutions for m^7G and 3'*O*-Me-m^7G, spiked with 0.5 ng ^{13}C labeled nucleosides from *S. cerevisiae*, were prepared. Notably, the 3'*O*-Me-m^7G standard was obtained by digestion of the cap analog ARCA. The resulting standard curves were used to calculate modified nucleoside quantities in the samples, with SILIS correction applied only for m^7G. The determined amounts of modified nucleoside were then related to the amount of injected RNA molecules. To determine the amount of injected RNA molecules in the samples, an additional standard curve for the main nucleosides was measured. The quantity of each main nucleoside was divided by its number in the RNA sequence, and this calculation was performed for each main nucleoside to obtain an average value.

Table 6. HPLC gradient for nucleoside analysis

Time [min]	A [%]	B [%]
0	100	0
10	92	8
20	60	40
23	100	0
30	100	0

List of Publications

(1) Ripp, A.*, Krämer, M.*, Barth, V., Moser, P., Haas, T. M., Singh, J., Huck, T., Gleue, L., Friedland, K., Helm, M., Jessen, H. J. The P(III)-amidite based synthesis of stable isotope labeled mRNA-cap-structures enables their sensitive quantitation from brain tissue. *Angewandte Chemie*; 10.1002/ange.202414537 (2024).

* These authors contributed equally to this work.

(2) Becker, M. A., Meiser, N., Schmidt-Dengler, M., Richter, C., Wacker, A., Schwalbe, H., Hengesbach, M. m6A Methylation of Transcription Leader Sequence of SARS-CoV-2 Impacts Discontinuous Transcription of Subgenomic mRNAs. *Chemistry – A European Journal* **30,** e202401897; 10.1002/chem.202401897 (2024).

(3) Boulet, M., Gilbert, G., Renaud, Y., Schmidt-Dengler, M., Plantié, E., Bertrand, R., Nan, X., Jurkowski, T., Helm, M., Vandel, L., Waltzer, L. Adenine methylation is very scarce in the Drosophila genome and not erased by the ten-eleven translocation dioxygenase. *eLife* **12**; 10.7554/eLife.91655 (2023).

(4) Koch, J., Neuberger, M., Schmidt-Dengler, M., Xu, J., Coutinho Carneiro, V., Ellinger, J., Kriegmair, M. C., Nuhn, P., Erben, P., Michel, M. S., Helm, M., Rodríguez-Paredes, M., Nientiedt, M., Lyko, F. Reinvestigating the clinical relevance of the m6A writer METTL3 in urothelial carcinoma of the bladder. *iScience* **26,** 107300; 10.1016/j.isci.2023.107300 (2023).

(5) Cavallin, I., Bartosovic, M., Skalicky, T., Rengaraj, P., Demko, M., Schmidt-Dengler, M. C., Drino, A., Helm, M., Vanacova, S. HITS-CLIP analysis of human ALKBH8 reveals interactions with fully processed substrate tRNAs and with specific noncoding RNAs. *RNA* **28,** rna.079421.122; 10.1261/rna.079421.122 (2022).

(6) Biedenbänder, T., de Jesus, V., Schmidt-Dengler, M., Helm, M., Corzilius B., Fürtig, B. RNA modifications stabilize the tertiary structure of tRNAfMet by locally increasing conformational dynamics. *Nucleic acids research* **50,** 2334–2349; 10.1093/nar/gkac040 (2022).

(7) Helm, M., Schmidt-Dengler, M. C., Weber, M. & Motorin, Y. General Principles for the Detection of Modified Nucleotides in RNA by Specific Reagents. *Advanced Biology* **5**; 10.1002/adbi.202100866 (2021).

(8) Siebenaller, C., Schlösser, L., Junglas, B., Schmidt-Dengler, M., Jacob, D., Hellmann, N., Sachse, C., Helm, M., Schneider, D. Binding and/or hydrolysis of purine-based nucleotides is not required for IM30 ring formation. *FEBS letters* **595,** 1876–1885; 10.1002/1873-3468.14140 (2021).

Bibliography

1. Suzanne Clancy. Chemical structure of RNA. *Nature education* **7** (2008).
2. Cobb, M. Who discovered messenger RNA? *Current biology: CB* **25,** R526-32; 10.1016/j.cub.2015.05.032 (2015).
3. Ramanathan, A., Robb, G. B. & Chan, S.-H. mRNA capping: biological functions and applications. *Nucleic acids research* **44,** 7511–7526; 10.1093/nar/gkw551 (2016).
4. Shatkin, A. J. Capping of eucaryotic mRNAs. *Cell* **9,** 645–653; 10.1016/0092-8674(76)90128-8 (1976).
5. Furuichi, Y. & Shatkin, A. J. Viral and cellular mRNA capping: past and prospects. *Advances in virus research* **55,** 135–184; 10.1016/S0065-3527(00)55003-9 (2000).
6. Furuichi, Y. Discovery of m(7)G-cap in eukaryotic mRNAs. *Proceedings of the Japan Academy. Series B, Physical and biological sciences* **91,** 394–409; 10.2183/pjab.91.394 (2015).
7. Wei, C., Gershowitz, A. & Moss, B. N6, O2'-dimethyladenosine a novel methylated ribonucleoside next to the 5' terminal of animal cell and virus mRNAs. *Nature* **257,** 251–253; 10.1038/257251a0 (1975).
8. Kastern, W. H. & Berry, S. J. Non-methylated guanosine as the 5' terminus of capped mRNA from insect oocytes. *Biochemical and biophysical research communications* **71,** 37–44; 10.1016/0006-291x(76)90246-1 (1976).
9. Galloway, A. & Cowling, V. H. mRNA cap regulation in mammalian cell function and fate. *Biochimica et biophysica acta. Gene regulatory mechanisms* **1862,** 270–279; 10.1016/j.bbagrm.2018.09.011 (2019).
10. Statello, L., Guo, C.-J., Chen, L.-L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* **22,** 96–118; 10.1038/s41580-020-00315-9 (2021).
11. Wu, H., Yang, L. & Chen, L.-L. The Diversity of Long Noncoding RNAs and Their Generation. *Trends in genetics : TIG* **33,** 540–552; 10.1016/j.tig.2017.05.004 (2017).
12. Mattick, J. S. *et al.* Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat Rev Mol Cell Biol* **24,** 430–447; 10.1038/s41580-022-00566-8 (2023).
13. Cai, X., Hagedorn, C. H. & Cullen, B. R. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA (New York, N.Y.)* **10,** 1957–1966; 10.1261/rna.7135204 (2004).
14. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal* **23,** 4051–4060; 10.1038/sj.emboj.7600385 (2004).
15. Xie, M. *et al.* Mammalian 5'-capped microRNA precursors that generate a single microRNA. *Cell* **155,** 1568–1580; 10.1016/j.cell.2013.11.027 (2013).

16. Shang, R., Lee, S., Senavirathne, G. & Lai, E. C. microRNAs in action: biogenesis, function and regulation. *Nature reviews. Genetics* **24,** 816–833; 10.1038/s41576-023-00611-y (2023).

17. Ozata, D. M., Gainetdinov, I., Zoch, A., O'Carroll, D. & Zamore, P. D. PIWI-interacting RNAs: small RNAs with big functions. *Nature reviews. Genetics* **20,** 89–108; 10.1038/s41576-018-0073-3 (2019).

18. Podvalnaya, N. *et al.* piRNA processing by a trimeric Schlafen-domain nuclease. *Nature* **622,** 402–409; 10.1038/s41586-023-06588-2 (2023).

19. Parry, H. D., Scherly, D. & Mattaj, I. W. 'Snurpogenesis': The transcription and assembly of U snRNP components. *Trends in Biochemical Sciences* **14,** 15–19; 10.1016/0968-0004(89)90083-2 (1989).

20. Zhang, R. *et al.* Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly. *Cell* **146,** 384–395; 10.1016/j.cell.2011.06.043 (2011).

21. Mattaj, I. W. Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell* **46,** 905–911; 10.1016/0092-8674(86)90072-3 (1986).

22. Matera, A. G. & Wang, Z. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol* **15,** 108–121; 10.1038/nrm3742 (2014).

23. Fischer, U. & Lührmann, R. An essential signaling role for the m3G cap in the transport of U1 snRNP to the nucleus. *Science (New York, N.Y.)* **249,** 786–790; 10.1126/science.2143847 (1990).

24. Potužník, J. F. & Cahova, H. If the 5' cap fits (wear it) - Non-canonical RNA capping. *RNA biology* **21,** 1–13; 10.1080/15476286.2024.2372138 (2024).

25. Watkins, N. J. *et al.* Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA (New York, N.Y.)* **4,** 1549–1568; 10.1017/s1355838298980761 (1998).

26. Weinstein, L. B. & Steitz, J. A. Guided tours: from precursor snoRNA to functional snoRNP. *Current opinion in cell biology* **11,** 378–384; 10.1016/S0955-0674(99)80053-2 (1999).

27. Tycowski, K. T., Smith, C. M., Shu, M. D. & Steitz, J. A. A small nucleolar RNA requirement for site-specific ribose methylation of rRNA in Xenopus. *Proceedings of the National Academy of Sciences of the United States of America* **93,** 14480–14485; 10.1073/pnas.93.25.14480 (1996).

28. Terns, M. P., Grimm, C., Lund, E. & Dahlberg, J. E. A common maturation pathway for small nucleolar RNAs. *The EMBO journal* **14,** 4860–4871; 10.1002/j.1460-2075.1995.tb00167.x (1995).

29. Maxwell, E. S. & Fournier, M. J. The small nucleolar RNAs. *Annual review of biochemistry* **64,** 897–934; 10.1146/annurev.bi.64.070195.004341 (1995).

30. Matera, A. G., Terns, R. M. & Terns, M. P. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* **8,** 209–220; 10.1038/nrm2124 (2007).

31. Ohira, T. & Suzuki, T. Precursors of tRNAs are stabilized by methylguanosine cap structures. *Nature chemical biology* **12,** 648–655; 10.1038/nchembio.2117 (2016).

32. Singh, R. & Reddy, R. Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. *Proceedings of the National Academy of Sciences of the United States of America* **86,** 8280–8283; 10.1073/pnas.86.21.8280 (1989).

33. Shimba, S., Buckley, B., Reddy, R., Kiss, T. & Filipowicz, W. Cap structure of U3 small nucleolar RNA in animal and plant cells is different. gamma-Monomethyl phosphate cap structure in plant RNA. *Journal of Biological Chemistry* **267,** 13772–13777; 10.1016/S0021-9258(18)42281-8 (1992).

34. Shumyatsky, G. P., Tillib, S. V. & Kramerov, D. A. B2 RNA and 7SK RNA, RNA polymerase III transcripts, have a cap-like structure at their 5' end. *Nucleic acids research* **18,** 6347–6351; 10.1093/nar/18.21.6347 (1990).

35. Shumyatsky, G., Wright, D. & Reddy, R. Methylphosphate cap structure increases the stability of 7SK, B2 and U6 small RNAs in Xenopus oocytes. *Nucleic acids research* **21,** 4756–4761; 10.1093/nar/21.20.4756 (1993).

36. Bhattacharya, R., Perumal, K., Sinha, K., Maraia, R. & Reddy, R. Methylphosphate cap structure in small RNAs reduces the affinity of RNAs to La protein. *Gene expression* **10,** 243–253; 10.3727/000000002783992398 (2002).

37. Chen, Y. G., Kowtoniuk, W. E., Agarwal, I., Shen, Y. & Liu, D. R. LC/MS analysis of cellular RNA reveals NAD-linked RNA. *Nature chemical biology* **5,** 879–881; 10.1038/nchembio.235 (2009).

38. Cahová, H., Winz, M.-L., Höfer, K., Nübel, G. & Jäschke, A. NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. *Nature* **519,** 374–377; 10.1038/nature14020 (2015).

39. Winz, M.-L. *et al.* Capture and sequencing of NAD-capped RNA sequences with NAD captureSeq. *Nature protocols* **12,** 122–149; 10.1038/nprot.2016.163 (2017).

40. Zhang, H. *et al.* NAD tagSeq reveals that NAD+-capped RNAs are mostly produced from a large number of protein-coding genes in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **116,** 12072–12077; 10.1073/pnas.1903683116 (2019).

41. Zhang, H. *et al.* Use of NAD tagSeq II to identify growth phase-dependent alterations in E. coli RNA NAD+ capping. *Proceedings of the National Academy of Sciences of the United States of America* **118**; 10.1073/pnas.2026183118 (2021).

42. Hu, H. *et al.* SPAAC-NAD-seq, a sensitive and accurate method to profile NAD+-capped transcripts. *Proceedings of the National Academy of Sciences of the United States of America* **118**; 10.1073/pnas.2025595118 (2021).

43. Sharma, S., Yang, J., Favate, J., Shah, P. & Kiledjian, M. NADcapPro and circNC: methods for accurate profiling of NAD and non-canonical RNA caps in eukaryotes. *Communications biology* **6,** 406; 10.1038/s42003-023-04774-6 (2023).

44. Vvedenskaya, I. O. *et al.* CapZyme-Seq Comprehensively Defines Promoter-Sequence Determinants for RNA 5' Capping with NAD<sup. *Molecular cell* **70,** 553-564.e9; 10.1016/j.molcel.2018.03.014 (2018).

45. Wang, J. *et al.* Quantifying the RNA cap epitranscriptome reveals novel caps in cellular and viral RNA. *Nucleic acids research* **47,** e130; 10.1093/nar/gkz751 (2019).

46. Nübel, G., Sorgenfrei, F. A. & Jäschke, A. Boronate affinity electrophoresis for the purification and analysis of cofactor-modified RNAs. *Methods (San Diego, Calif.)* **117,** 14–20; 10.1016/j.ymeth.2016.09.008 (2017).

47. Grudzien-Nogalska, E., Bird, J. G., Nickels, B. E. & Kiledjian, M. “NAD-capQ” detection and quantitation of NAD caps. *RNA* **24,** 1418–1425; 10.1261/rna.067686.118 (2018).

48. Frindert, J. *et al.* Identification, Biosynthesis, and Decapping of NAD-Capped RNAs in B. subtilis. *Cell reports* **24,** 1890-1901.e8; 10.1016/j.celrep.2018.07.047 (2018).

49. Morales-Filloy, H. G. *et al.* The 5’ NAD Cap of RNAIII Modulates Toxin Production in Staphylococcus aureus Isolates. *Journal of bacteriology* **202**; 10.1128/jb.00591-19 (2020).

50. Gomes-Filho, J. V. *et al.* Identification of NAD-RNA species and ADPR-RNA decapping in Archaea. *Nature communications* **14,** 7597; 10.1038/s41467-023-43377-x (2023).

51. Walters, R. W. *et al.* Identification of NAD+ capped mRNAs in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* **114,** 480–485; 10.1073/pnas.1619369114 (2017).

52. Jiao, X. *et al.* 5’ End Nicotinamide Adenine Dinucleotide Cap in Human Cells Promotes RNA Decay through DXO-Mediated deNADding. *Cell* **168,** 1015-1027.e10; 10.1016/j.cell.2017.02.019 (2017).

53. Wang, Y. *et al.* NAD+-capped RNAs are widespread in the Arabidopsis transcriptome and can probably be translated. *Proceedings of the National Academy of Sciences of the United States of America* **116,** 12094–12102; 10.1073/pnas.1903682116 (2019).

54. Bird, J. G. *et al.* Highly efficient 5’ capping of mitochondrial RNA with NAD+ and NADH by yeast and human mitochondrial RNA polymerase. *eLife* **7**; 10.7554/eLife.42179 (2018).

55. Bird, J. G. *et al.* The mechanism of RNA 5′ capping with NAD+, NADH and desphospho-CoA. *Nature* **535,** 444–447; 10.1038/nature18622 (2016).

56. Höfer, K. *et al.* Structure and function of the bacterial decapping enzyme NudC. *Nature chemical biology* **12,** 730–734; 10.1038/nchembio.2132 (2016).

57. Sharma, S. *et al.* Xrn1 is a deNADding enzyme modulating mitochondrial NAD-capped RNA. *Nature communications* **13,** 889; 10.1038/s41467-022-28555-7 (2022).

58. Abele, F. *et al.* A Novel NAD-RNA Decapping Pathway Discovered by Synthetic Light-Up NAD-RNAs. *Biomolecules* **10**; 10.3390/biom10040513 (2020).

59. Wolfram-Schauerte, M. *et al.* A viral ADP-ribosyltransferase attaches RNA chains to host proteins. *Nature* **620,** 1054–1062; 10.1038/s41586-023-06429-2 (2023).

60. Zhang, Y. *et al.* Extensive 5’-surveillance guards against non-canonical NAD-caps of nuclear mRNAs in yeast. *Nature communications* **11,** 5508; 10.1038/s41467-020-19326-3 (2020).

61. Wolfram-Schauerte, M. & Höfer, K. NAD-capped RNAs - a redox cofactor meets RNA. *Trends in Biochemical Sciences* **48,** 142–155; 10.1016/j.tibs.2022.08.004 (2023).

62. Kowtoniuk, W. E., Shen, Y., Heemstra, J. M., Agarwal, I. & Liu, D. R. A chemical screen for biological small molecule-RNA conjugates reveals CoA-linked RNA. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 7768–7773; 10.1073/pnas.0900528106 (2009).

63. Löcherer, C., Bühler, N., Lafrenz, P. & Jäschke, A. Staphylococcus aureus Small RNAs Possess Dephospho-CoA 5'-Caps, but No CoAlation Marks. *Non-coding RNA* **8**; 10.3390/ncrna8040046 (2022).

64. Miller, J. R. *et al.* Phosphopantetheine adenylyltransferase from Escherichia coli: investigation of the kinetic mechanism and role in regulation of coenzyme A biosynthesis. *Journal of bacteriology* **189,** 8196–8205; 10.1128/JB.00732-07 (2007).

65. Sapkota, K. *et al.* Post-transcriptional capping generates coenzyme A-linked RNA. *RNA biology* **21,** 1–12; 10.1080/15476286.2023.2288740 (2024).

66. František Potužník, J. *et al.* Diadenosine Tetraphosphate (Ap4 A) Serves as a 5' RNA Cap in Mammalian Cells. *Angew Chem Int Ed* **63,** e202314951; 10.1002/anie.202314951 (2024).

67. Julius, C. & Yuzenkova, Y. Bacterial RNA polymerase caps RNA with various cofactors and cell wall precursors. *Nucleic acids research* **45,** 8282–8290; 10.1093/nar/gkx452 (2017).

68. Doamekpor, S. K. *et al.* DXO/Rai1 enzymes remove 5'-end FAD and dephospho-CoA caps on RNAs. *Nucleic acids research* **48,** 6136–6148; 10.1093/nar/gkaa297 (2020).

69. Sherwood, A. V. *et al.* Hepatitis C virus RNA is 5'-capped with flavin adenine dinucleotide. *Nature* **619,** 811–818; 10.1038/s41586-023-06301-3 (2023).

70. McLennan, A. G. Dinucleoside polyphosphates-friend or foe? *Pharmacology & therapeutics* **87,** 73–89; 10.1016/S0163-7258(00)00041-3 (2000).

71. Hudeček, O. *et al.* Dinucleoside polyphosphates act as 5'-RNA caps in bacteria. *Nature communications* **11,** 1052; 10.1038/s41467-020-14896-8 (2020).

72. Luciano, D. J. & Belasco, J. G. Np4A alarmones function in bacteria as precursors to RNA caps. *Proceedings of the National Academy of Sciences of the United States of America* **117,** 3560–3567; 10.1073/pnas.1914229117 (2020).

73. Izaurralde, E. *et al.* A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78,** 657–668; 10.1016/0092-8674(94)90530-4 (1994).

74. Edery, I. & Sonenberg, N. Cap-dependent RNA splicing in a HeLa nuclear extract. *Proceedings of the National Academy of Sciences of the United States of America* **82,** 7590–7594; 10.1073/pnas.82.22.7590 (1985).

75. Konarska, M. M., Padgett, R. A. & Sharp, P. A. Recognition of cap structure in splicing in vitro of mRNA precursors. *Cell* **38,** 731–736; 10.1016/0092-8674(84)90268-x (1984).

76. Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C. & Mattaj, I. W. A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes & development* **10,** 1683–1698; 10.1101/gad.10.13.1683 (1996).

77. Görnemann, J., Kotovic, K. M., Hujer, K. & Neugebauer, K. M. Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Molecular cell* **19,** 53–63; 10.1016/j.molcel.2005.05.007 (2005).

78. Pabis, M. *et al.* The nuclear cap-binding complex interacts with the U4/U6·U5 tri-snRNP and promotes spliceosome assembly in mammalian cells. *RNA* **19,** 1054–1063; 10.1261/rna.037069.112 (2013).

79. Topisirovic, I., Svitkin, Y. V., Sonenberg, N. & Shatkin, A. J. Cap and cap-binding proteins in the control of gene expression. *Wiley interdisciplinary reviews. RNA* **2,** 277–298; 10.1002/wrna.52 (2011).

80. Flaherty, S. M., Fortes, P., Izaurralde, E., Mattaj, I. W. & Gilmartin, G. M. Participation of the nuclear cap binding complex in pre-mRNA 3' processing. *Proceedings of the National Academy of Sciences of the United States of America* **94,** 11893–11898; 10.1073/pnas.94.22.11893 (1997).

81. Narita, T. *et al.* NELF interacts with CBC and participates in 3' end processing of replication-dependent histone mRNAs. *Molecular cell* **26,** 349–365; 10.1016/j.molcel.2007.04.011 (2007).

82. Cheng, H. *et al.* Human mRNA export machinery recruited to the 5' end of mRNA. *Cell* **127,** 1389–1400; 10.1016/j.cell.2006.10.044 (2006).

83. Ohno, M., Segref, A., Bachi, A., Wilm, M. & Mattaj, I. W. PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell* **101,** 187–198; 10.1016/S0092-8674(00)80829-6 (2000).

84. Hanson, F. M., Hodgson, R. E., Oliveira, M. I. R. de, Allen, K. E. & Campbell, S. G. Regulation and function of eIF2B in neurological and metabolic disorders. *Bioscience reports* **42**; 10.1042/BSR20211699 (2022).

85. Preiss, T. & Hentze, M. W. From factors to mechanisms: translation and translational control in eukaryotes. *Current opinion in genetics & development* **9,** 515–521; 10.1016/S0959-437X(99)00005-2 (1999).

86. Wells, S. E., Hillner, P. E., Vale, R. D. & Sachs, A. B. Circularization of mRNA by eukaryotic translation initiation factors. *Molecular cell* **2,** 135–140; 10.1016/S1097-2765(00)80122-7 (1998).

87. Smith, R. W. P., Blee, T. K. P. & Gray, N. K. Poly(A)-binding proteins are required for diverse biological processes in metazoans. *Biochemical Society transactions* **42,** 1229–1237; 10.1042/BST20140111 (2014).

88. Maquat, L. E., Tarn, W.-Y. & Isken, O. The pioneer round of translation: features and functions. *Cell* **142,** 368–374; 10.1016/j.cell.2010.07.022 (2010).

89. Ishigaki, Y., Li, X., Serin, G. & Maquat, L. E. Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* **106,** 607–617; 10.1016/s0092-8674(01)00475-5 (2001).

90. Lejeune, F., Ishigaki, Y., Li, X. & Maquat, L. E. The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. *The EMBO journal* **21,** 3536–3545; 10.1093/emboj/cdf345 (2002).

91. Hosoda, N., Lejeune, F. & Maquat, L. E. Evidence that poly(A) binding protein C1 binds nuclear pre-mRNA poly(A) tails. *Molecular and cellular biology* **26,** 3085–3097; 10.1128/MCB.26.8.3085-3097.2006 (2006).

92. Chiu, S.-Y., Lejeune, F., Ranganathan, A. C. & Maquat, L. E. The pioneer translation initiation complex is functionally distinct from but structurally overlaps with the steady-

state translation initiation complex. *Genes & development* **18,** 745–754; 10.1101/gad.1170204 (2004).

93. Sato, H. & Maquat, L. E. Remodeling of the pioneer translation initiation complex involves translation and the karyopherin importin beta. *Genes & development* **23,** 2537–2550; 10.1101/gad.1817109 (2009).
94. Furuichi, Y., LaFiandra, A. & Shatkin, A. J. 5'-Terminal structure and mRNA stability. *Nature* **266,** 235–239; 10.1038/266235a0 (1977).
95. Hsu, C. L. & Stevens, A. Yeast cells lacking 5'--3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. *Molecular and cellular biology* **13,** 4826–4835; 10.1128/mcb.13.8.4826-4835.1993 (1993).
96. Grudzien-Nogalska, E. & Kiledjian, M. New insights into decapping enzymes and selective mRNA decay. *Wiley interdisciplinary reviews. RNA* **8**; 10.1002/wrna.1379 (2017).
97. Song, M.-G. & Kiledjian, M. 3' Terminal oligo U-tract-mediated stimulation of decapping. *RNA* **13,** 2356–2365; 10.1261/rna.765807 (2007).
98. Li, Y., Song, M. & Kiledjian, M. Differential utilization of decapping enzymes in mammalian mRNA decay pathways. *RNA* **17,** 419–428; 10.1261/rna.2439811 (2011).
99. Grudzien-Nogalska, E., Jiao, X., Song, M.-G., Hart, R. P. & Kiledjian, M. Nudt3 is an mRNA decapping enzyme that modulates cell migration. *RNA* **22,** 773–781; 10.1261/rna.055699.115 (2016).
100. Song, M.-G., Bail, S. & Kiledjian, M. Multiple Nudix family proteins possess mRNA decapping activity. *RNA* **19,** 390–399; 10.1261/rna.037309.112 (2013).
101. Jurado, A. R., Tan, D., Jiao, X., Kiledjian, M. & Tong, L. Structure and function of pre-mRNA 5'-end capping quality control and 3'-end processing. *Biochemistry* **53,** 1882–1898; 10.1021/bi401715v (2014).
102. Xiang, S. *et al.* Structure and function of the 5'--3' exoribonuclease Rat1 and its activating partner Rai1. *Nature* **458,** 784–788; 10.1038/nature07731 (2009).
103. Jiao, X. *et al.* Identification of a quality-control mechanism for mRNA 5'-end capping. *Nature* **467,** 608–611; 10.1038/nature09338 (2010).
104. Chang, J. H. *et al.* Dxo1 is a new type of eukaryotic enzyme with both decapping and 5'-3' exoribonuclease activity. *Nature structural & molecular biology* **19,** 1011–1017; 10.1038/nsmb.2381 (2012).
105. Jiao, X., Chang, J. H., Kilic, T., Tong, L. & Kiledjian, M. A mammalian pre-mRNA 5' end capping quality control mechanism and an unexpected link of capping to pre-mRNA processing. *Molecular cell* **50,** 104–115; 10.1016/j.molcel.2013.02.017 (2013).
106. Leung, D. W. & Amarasinghe, G. K. When your cap matters: structural insights into self vs non-self recognition of 5' RNA by immunomodulatory host proteins. *Current opinion in structural biology* **36,** 133–141; 10.1016/j.sbi.2016.02.001 (2016).
107. Li, D. & Wu, M. Pattern recognition receptors in health and diseases. *Signal transduction and targeted therapy* **6,** 291; 10.1038/s41392-021-00687-0 (2021).
108. Devarkar, S. C. *et al.* Structural basis for m7G recognition and 2'-O-methyl discrimination in capped RNAs by the innate immune receptor RIG-I. *Proceedings of the National*

Academy of Sciences of the United States of America **113,** 596–601; 10.1073/pnas.1515152113 (2016).

109. Schuberth-Wagner, C. *et al.* A Conserved Histidine in the RNA Sensor RIG-I Controls Immune Tolerance to N1-2'O-Methylated Self RNA. *Immunity* **43,** 41–51; 10.1016/j.immuni.2015.06.015 (2015).
110. Hyde, J. L. & Diamond, M. S. Innate immune restriction and antagonism of viral RNA lacking 2′-O methylation. *Virology* **479-480,** 66–74; 10.1016/j.virol.2015.01.019 (2015).
111. Diamond, M. S. IFIT1: A dual sensor and effector molecule that detects non-2'-O methylated viral RNA and inhibits its translation. *Cytokine & growth factor reviews* **25,** 543–550; 10.1016/j.cytogfr.2014.05.002 (2014).
112. Kumar, P. *et al.* Inhibition of translation by IFIT family members is determined by their ability to interact selectively with the 5'-terminal regions of cap0-, cap1- and 5'ppp- mRNAs. *Nucleic acids research* **42,** 3228–3245; 10.1093/nar/gkt1321 (2014).
113. Habjan, M. *et al.* Sequestration by IFIT1 impairs translation of 2'O-unmethylated capped RNA. *PLoS pathogens* **9,** e1003663; 10.1371/journal.ppat.1003663 (2013).
114. Daffis, S. *et al.* 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature* **468,** 452–456; 10.1038/nature09489 (2010).
115. Niedzwiecka, A. *et al.* Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5' cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins. *Journal of molecular biology* **319,** 615–635; 10.1016/S0022-2836(02)00328-5 (2002).
116. Worch, R. *et al.* Specificity of recognition of mRNA 5' cap by human nuclear cap-binding complex. *RNA (New York, N.Y.)* **11,** 1355–1363; 10.1261/rna.2850705 (2005).
117. Muthukrishnan, S., Moss, B., Cooper, J. A. & Maxwell, E. S. Influence of 5'-terminal cap structure on the initiation of translation of vaccinia virus mRNA. *Journal of Biological Chemistry* **253,** 1710–1715; 10.1016/S0021-9258(17)34923-2 (1978).
118. Inesta-Vaquera, F. & Cowling, V. H. Regulation and function of CMTR1-dependent mRNA cap methylation. *Wiley interdisciplinary reviews. RNA* **8**; 10.1002/wrna.1450 (2017).
119. Muthukrishnan, S., Morgan, M., Banerjee, A. K. & Shatkin, A. J. Influence of 5'-terminal m7G and 2'—O-methylated residues on messenger ribonucleic acid binding to ribosomes. *Biochemistry* **15,** 5761–5768; 10.1021/bi00671a012 (1976).
120. Kuge, H., Brownlee, G. G., Gershon, P. D. & Richter, J. D. Cap ribose methylation of c-mos mRNA stimulates translation and oocyte maturation in Xenopus laevis. *Nucleic acids research* **26,** 3208–3214; 10.1093/nar/26.13.3208 (1998).
121. Jin, H., Shi, Z., Zhou, T. & Xie, S. Regulation of m6Am RNA modification and its implications in human diseases. *Journal of molecular cell biology* **16**; 10.1093/jmcb/mjae012 (2024).
122. Mauer, J. *et al.* Reversible methylation of m6Am in the 5' cap controls mRNA stability. *Nature* **541,** 371–375; 10.1038/nature21022 (2017).
123. Boulias, K. *et al.* Identification of the m6Am Methyltransferase PCIF1 Reveals the Location and Functions of m6Am in the Transcriptome. *Molecular cell* **75,** 631-643.e8; 10.1016/j.molcel.2019.06.006 (2019).

124. Akichika, S. *et al.* Cap-specific terminal N6-methylation of RNA by an RNA polymerase II-associated methyltransferase. *Science (New York, N.Y.)* **363**; 10.1126/science.aav0080. (2019).

125. Sendinc, E. *et al.* PCIF1 Catalyzes m6Am mRNA Methylation to Regulate Gene Expression. *Molecular cell* **75,** 620-630.e9; 10.1016/j.molcel.2019.05.030 (2019).

126. Abbas, Y. M. *et al.* Structure of human IFIT1 with capped RNA reveals adaptable mRNA binding and mechanisms for sensing N1 and N2 ribose 2'-O methylations. *Proceedings of the National Academy of Sciences of the United States of America* **114,** E2106-E2115; 10.1073/pnas.1612444114 (2017).

127. Despic, V. & Jaffrey, S. R. mRNA ageing shapes the Cap2 methylome in mammalian mRNA. *Nature* **614,** 358–366; 10.1038/s41586-022-05668-z (2023).

128. Drazkowska, K. *et al.* 2'-O-Methylation of the second transcribed nucleotide within the mRNA 5' cap impacts the protein production level in a cell-specific manner and contributes to RNA immune evasion. *Nucleic acids research* **50,** 9051–9071; 10.1093/nar/gkac722 (2022).

129. Dohnalkova, M. *et al.* Essential roles of RNA cap-proximal ribose methylation in mammalian embryonic development and fertility. *Cell reports* **42,** 112786; 10.1016/j.celrep.2023.112786 (2023).

130. Shatkin, A. J. & Manley, J. L. The ends of the affair: capping and polyadenylation. *Nature structural biology* **7,** 838–842; 10.1038/79583 (2000).

131. McCracken, S. *et al.* 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes & development* **11,** 3306–3318; 10.1101/gad.11.24.3306 (1997).

132. Cho, E. J., Takagi, T., Moore, C. R. & Buratowski, S. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes & development* **11,** 3319–3326; 10.1101/gad.11.24.3319 (1997).

133. Mao, X., Schwer, B. & Shuman, S. Yeast mRNA cap methyltransferase is a 50-kilodalton protein encoded by an essential gene. *Molecular and cellular biology* **15,** 4167–4174; 10.1128/MCB.15.8.4167 (1995).

134. Gu, M., Rajashankar, K. R. & Lima, C. D. Structure of the Saccharomyces cerevisiae Cet1-Ceg1 mRNA capping apparatus. *Structure (London, England : 1993)* **18,** 216–227; 10.1016/j.str.2009.12.009 (2010).

135. Sun, H., Zhang, M., Li, K., Bai, D. & Yi, C. Cap-specific, terminal N6-methylation by a mammalian m6Am methyltransferase. *Cell research* **29,** 80–82; 10.1038/s41422-018-0117-4 (2019).

136. Bélanger, F., Stepinski, J., Darzynkiewicz, E. & Pelletier, J. Characterization of hMTr1, a human Cap1 2'-O-ribose methyltransferase. *The Journal of biological chemistry* **285,** 33037–33044; 10.1074/jbc.M110.155283 (2010).

137. Werner, M. *et al.* 2'-O-ribose methylation of cap2 in human: function and evolution in a horizontally mobile family. *Nucleic acids research* **39,** 4756–4768; 10.1093/nar/gkr038 (2011).

138. Borden, K., Culjkovic-Kraljacic, B. & Cowling, V. H. To cap it all off, again: dynamic capping and recapping of coding and non-coding RNAs to control transcript fate and biological

activity. *Cell cycle (Georgetown, Tex.)* **20,** 1347–1360; 10.1080/15384101.2021.1930929 (2021).

139. Decroly, E., Ferron, F., Lescar, J. & Canard, B. Conventional and unconventional mechanisms for capping viral mRNA. *Nature reviews. Microbiology* **10,** 51–65; 10.1038/nrmicro2675 (2011).

140. Cong, P. & Shuman, S. Mutational analysis of mRNA capping enzyme identifies amino acids involved in GTP binding, enzyme-guanylate formation, and GMP transfer to RNA. *Molecular and cellular biology* **15,** 6222–6231; 10.1128/MCB.15.11.6222 (1995).

141. Hodel, A. E., Gershon, P. D., Shi, X. & Quiocho, F. A. The 1.85 A structure of vaccinia protein VP39: a bifunctional enzyme that participates in the modification of both mRNA ends. *Cell* **85,** 247–256; 10.1016/S0092-8674(00)81101-0 (1996).

142. Ahola, T. & Kääriäinen, L. Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP. *Proceedings of the National Academy of Sciences of the United States of America* **92,** 507–511; 10.1073/pnas.92.2.507 (1995).

143. Plotch, S. J., Bouloy, M., Ulmanen, I. & Krug, R. M. A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* **23,** 847–858; 10.1016/0092-8674(81)90449-9 (1981).

144. Walker, A. P. & Fodor, E. Interplay between Influenza Virus and the Host RNA Polymerase II Transcriptional Machinery. *Trends in microbiology* **27,** 398–407; 10.1016/j.tim.2018.12.013 (2019).

145. Fujimura, T. & Esteban, R. The cap-snatching reaction of yeast L-A double-stranded RNA virus is reversible and the catalytic sites on both Gag and the Gag domain of Gag-Pol are active. *Molecular microbiology* **111,** 395–404; 10.1111/mmi.14161 (2019).

146. Wolff, J. A. *et al.* Direct gene transfer into mouse muscle in vivo. *Science (New York, N.Y.)* **247,** 1465–1468; 10.1126/science.1690918 (1990).

147. Sahin, U. *et al.* COVID-19 vaccine BNT162b1 elicits human antibody and TH1 T cell responses. *Nature* **586,** 594–599; 10.1038/s41586-020-2814-7 (2020).

148. Corbett, K. S. *et al.* SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. *Nature* **586,** 567–571; 10.1038/s41586-020-2622-0 (2020).

149. Warminski, M., Mamot, A., Depaix, A., Kowalska, J. & Jemielity, J. Chemical Modifications of mRNA Ends for Therapeutic Applications. *Accounts of chemical research* **56,** 2814–2826; 10.1021/acs.accounts.3c00442 (2023).

150. Sahin, U., Karikó, K. & Türeci, Ö. mRNA-based therapeutics—developing a new class of drugs. *Nature reviews. Drug discovery* **13,** 759–780; 10.1038/nrd4278 (2014).

151. Flemmich, L., Bereiter, R. & Micura, R. Chemical Synthesis of Modified RNA. *Angew Chem Int Ed* **63**; 10.1002/anie.202403063 (2024).

152. Esquiaqui, J. M., Sherman, E. M., Ye, J.-D. & Fanucci, G. E. Conformational Flexibility and Dynamics of the Internal Loop and Helical Regions of the Kink-Turn Motif in the Glycine Riboswitch by Site-Directed Spin-Labeling. *Biochemistry* **55,** 4295–4305; 10.1021/acs.biochem.6b00287 (2016).

153. Ensinger, M. J., Martin, S. A., Paoletti, E. & Moss, B. Modification of the 5'-terminus of mRNA by soluble guanylyl and methyl transferases from vaccinia virus. *Proceedings of the National Academy of Sciences of the United States of America* **72,** 2525–2529; 10.1073/pnas.72.7.2525 (1975).

154. Martin, S. A., Paoletti, E. & Moss, B. Purification of mRNA guanylyltransferase and mRNA (guanine-7-) methyltransferase from vaccinia virions. *Journal of Biological Chemistry* **250,** 9322–9329 (1975).

155. Paterson, B. M. & Rosenberg, M. Efficient translation of prokaryotic mRNAs in a eukaryotic cell-free system requires addition of a cap structure. *Nature* **279,** 692–696; 10.1038/279692a0 (1979).

156. Muttach, F., Muthmann, N. & Rentmeister, A. Review: Synthetic mRNA capping. *Beilstein journal of organic chemistry* **13,** 2819–2832; 10.3762/bjoc.13.274 (2017).

157. Pasquinelli, A. E., Dahlberg, J. E. & Lund, E. Reverse 5' caps in RNAs made in vitro by phage RNA polymerases. *RNA (New York, N.Y.)* **1,** 957–967 (1995).

158. Stepinski, J., Waddell, C., Stolarski, R., Darzynkiewicz, E. & Rhoads, R. E. Synthesis and properties of mRNAs containing the novel "anti-reverse" cap analogs 7-methyl(3'-O-methyl)GpppG and 7-methyl (3'-deoxy)GpppG. *RNA (New York, N.Y.)* **7,** 1486–1495 (2001).

159. Jemielity, J. *et al.* Novel "anti-reverse" cap analogs with superior translational properties. *RNA (New York, N.Y.)* **9,** 1108–1122; 10.1261/rna.5430403 (2003).

160. Ishikawa, M., Murai, R., Hagiwara, H., Hoshino, T. & Suyama, K. Preparation of eukaryotic mRNA having differently methylated adenosine at the 5'-terminus and the effect of the methyl group in translation. *Nucleic Acids Symposium Series* **53,** 129–130; 10.1093/nass/nrp065 (2009).

161. Sikorski, P. J. *et al.* The identity and methylation status of the first transcribed nucleotide in eukaryotic mRNA 5' cap modulates protein expression in living cells. *Nucleic acids research* **48,** 1607–1626; 10.1093/nar/gkaa032 (2020).

162. Henderson, J. M. *et al.* Cap 1 Messenger RNA Synthesis with Co-transcriptional CleanCap® Analog by In Vitro Transcription. *Current protocols* **1,** e39; 10.1002/cpz1.39 (2021).

163. Depaix, A. *et al.* Preparation of RNAs with non-canonical 5' ends using novel di- and trinucleotide reagents for co-transcriptional capping. *Frontiers in molecular biosciences* **9,** 854170; 10.3389/fmolb.2022.854170 (2022).

164. Grudzien-Nogalska, E., Jemielity, J., Kowalska, J., Darzynkiewicz, E. & Rhoads, R. E. Phosphorothioate cap analogs stabilize mRNA and increase translational efficiency in mammalian cells. *RNA (New York, N.Y.)* **13,** 1745–1755; 10.1261/rna.701307 (2007).

165. Kuhn, A. N. *et al.* Phosphorothioate cap analogs increase stability and translational efficiency of RNA vaccines in immature dendritic cells and induce superior immune responses in vivo. *Gene therapy* **17,** 961–971; 10.1038/gt.2010.52 (2010).

166. Sahin, U. *et al.* Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* **547,** 222–226; 10.1038/nature23003 (2017).

167. Sahin, U. *et al.* An RNA vaccine drives immunity in checkpoint-inhibitor-treated melanoma. *Nature* **585,** 107–112; 10.1038/s41586-020-2537-9 (2020).

168. Kozarski, M. *et al.* Towards superior mRNA caps accessible by click chemistry: synthesis and translational properties of triazole-bearing oligonucleotide cap analogs. *RSC advances* **13,** 12809–12824; 10.1039/D3RA00026E (2023).

169. Wojcik, R. *et al.* Novel N7-Arylmethyl Substituted Dinucleotide mRNA 5' cap Analogs: Synthesis and Evaluation as Modulators of Translation. *Pharmaceutics* **13**; 10.3390/pharmaceutics13111941 (2021).

170. Grzela, R. *et al.* N2 modified dinucleotide cap analogs as a potent tool for mRNA engineering. *RNA* **29,** 200–216; 10.1261/rna.079460.122 (2023).

171. Kore, A. R., Shanmugasundaram, M., Charles, I., Vlassov, A. V. & Barta, T. J. Locked nucleic acid (LNA)-modified dinucleotide mRNA cap analogue: synthesis, enzymatic incorporation, and utilization. *J. Am. Chem. Soc.* **131,** 6364–6365; 10.1021/ja901655p (2009).

172. Issur, M., Bougie, I., Despins, S. & Bisaillon, M. Enzymatic synthesis of RNAs capped with nucleotide analogues reveals the molecular basis for substrate selectivity of RNA capping enzyme: impacts on RNA metabolism. *PloS one* **8,** e75310; 10.1371/journal.pone.0075310 (2013).

173. Ohno, H. *et al.* Versatile strategy using vaccinia virus-capping enzyme to synthesize functional 5' cap-modified mRNAs. *Nucleic acids research* **51,** e34; 10.1093/nar/gkad019 (2023).

174. van Dülmen, M., Muthmann, N. & Rentmeister, A. Chemo-Enzymatic Modification of the 5' Cap Maintains Translation and Increases Immunogenic Properties of mRNA. *Angew Chem Int Ed* **60,** 13280–13286; 10.1002/anie.202100352 (2021).

175. Warminski, M. *et al.* Photoactivatable mRNA 5' Cap Analogs for RNA-Protein Crosslinking. *Advanced science (Weinheim, Baden-Wurttemberg, Germany)* **11,** e2400994; 10.1002/advs.202400994 (2024).

176. Bollu, A. *et al.* Light-Activated Translation of Different mRNAs in Cells via Wavelength-Dependent Photouncaging. *Angew Chem Int Ed* **62,** e202209975; 10.1002/anie.202209975 (2023).

177. Klöcker, N. *et al.* Photocaged 5' cap analogues for optical control of mRNA translation in cells. *Nature chemistry* **14,** 905–913; 10.1038/s41557-022-00972-7 (2022).

178. Inagaki, M. *et al.* Cap analogs with a hydrophobic photocleavable tag enable facile purification of fully capped mRNA with various cap structures. *Nature communications* **14,** 2657; 10.1038/s41467-023-38244-8 (2023).

179. Gunawardana, D., Domashevskiy, A. V., Gayler, K. R. & Goss, D. J. Efficient preparation and properties of mRNAs containing a fluorescent cap analog: Anthraniloyl-m(7)GpppG. *Translation (Austin, Tex.)* **3,** e988538; 10.4161/21690731.2014.988538 (2015).

180. Ziemniak, M. *et al.* Synthesis and evaluation of fluorescent cap analogues for mRNA labelling. *RSC advances* **3**; 10.1039/C3RA42769B (2013).

181. Jemielity, J. *et al.* Synthesis of biotin labelled cap analogue—incorporable into mRNA transcripts and promoting cap-dependent translation. *Organic & biomolecular chemistry* **10,** 8570–8574; 10.1039/C2OB26060C (2012).

182. Warminski, M. *et al.* Amino-Functionalized 5' Cap Analogs as Tools for Site-Specific Sequence-Independent Labeling of mRNA. *Bioconjugate chemistry* **28,** 1978–1992; 10.1021/acs.bioconjchem.7b00291 (2017).

183. Mamot, A., Sikorski, P. J., Warminski, M., Kowalska, J. & Jemielity, J. Azido-Functionalized 5' Cap Analogues for the Preparation of Translationally Active mRNAs Suitable for Fluorescent Labeling in Living Cells. *Angew Chem Int Ed* **56,** 15628–15632; 10.1002/anie.201709052 (2017).

184. Cozens, C., Pinheiro, V. B., Vaisman, A., Woodgate, R. & Holliger, P. A short adaptive path from DNA to RNA polymerases. *Proceedings of the National Academy of Sciences of the United States of America* **109,** 8067–8072; 10.1073/pnas.1120964109 (2012).

185. Wojtczak, B. A. *et al.* Clickable trimethylguanosine cap analogs modified within the triphosphate bridge: synthesis, conjugation to RNA and susceptibility to degradation. *RSC Adv.* **6,** 8317–8328; 10.1039/C5RA25684D (2016).

186. Walczak, S. *et al.* A novel route for preparing 5' cap mimics and capped RNAs: phosphate-modified cap analogues obtained via click chemistry. *Chemical science* **8,** 260–267; 10.1039/C6SC02437H (2017).

187. Chen, H. *et al.* Chemical and topological design of multicapped mRNA and capped circular RNA to augment translation. *Nature biotechnology*; 10.1038/s41587-024-02393-y (2024).

188. Heckle, W. L., Fenton, R. G., Wood, T. G., Merkel, C. G. & Lingrel, J. B. Methylated nucleosides in globin mRNA from mouse nucleated erythroid cells. *Journal of Biological Chemistry* **252,** 1764–1770 (1977).

189. Bellows, E., Fray, R. G., Knight, H. M. & Archer, N. The expanding role of cap-adjacent modifications in animals. *Front. RNA Res.* **2**; 10.3389/frnar.2024.1485307 (2024).

190. Kruse, S. *et al.* A novel synthesis and detection method for cap-associated adenosine modifications in mouse mRNA. *Scientific reports* **1,** 126; 10.1038/srep00126 (2011).

191. Haussmann, I. U. *et al.* CMTr cap-adjacent 2'-O-ribose mRNA methyltransferases are required for reward learning and mRNA localization to synapses. *Nature communications* **13,** 1209; 10.1038/s41467-022-28549-5 (2022).

192. Dix, T. C. *et al.* CMTr mediated 2'-O-ribose methylation status of cap-adjacent nucleotides across animals. *RNA* **28,** 1377–1390; 10.1261/rna.079317.122 (2022).

193. Benak, D., Kolar, F., Zhang, L., Devaux, Y. & Hlavackova, M. RNA modification m6Am: the role in cardiac biology. *Epigenetics* **18,** 2218771; 10.1080/15592294.2023.2218771 (2023).

194. Linder, B. *et al.* Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nature methods* **12,** 767–772; 10.1038/nmeth.3453 (2015).

195. Koh, C. W. Q., Goh, Y. T. & Goh, W. S. S. Atlas of quantitative single-base-resolution N6-methyl-adenine methylomes. *Nature communications* **10,** 5636; 10.1038/s41467-019-13561-z (2019).

196. Sun, H. *et al.* m6Am-seq reveals the dynamic m6Am methylation in the human transcriptome. *Nature communications* **12,** 4778; 10.1038/s41467-021-25105-5 (2021).

197. Zhang, M., Sun, H., Li, K., Xiao, Y. & Yi, C. m6Am RNA modification detection by m6Am-seq. *Methods (San Diego, Calif.)* **203,** 242–248; 10.1016/j.ymeth.2021.10.001 (2022).

198. Cai, W. M. *et al.* A Platform for Discovery and Quantification of Modified Ribonucleosides in RNA: Application to Stress-Induced Reprogramming of tRNA Modifications. *Methods in enzymology* **560,** 29–71; 10.1016/bs.mie.2015.03.004 (2015).

199. Muthmann, N., Špaček, P., Reichert, D., van Dülmen, M. & Rentmeister, A. Quantification of mRNA cap-modifications by means of LC-QqQ-MS. *Methods (San Diego, Calif.)* **203,** 196–206; 10.1016/j.ymeth.2021.05.018 (2022).

200. Galloway, A. *et al.* CAP-MAP: cap analysis protocol with minimal analyte processing, a rapid and sensitive approach to analysing mRNA cap structures. *Open biology* **10,** 190306; 10.1098/rsob.190306 (2020).

201. Wang, J. *et al.* A systems-level mass spectrometry-based technique for accurate and sensitive quantification of the RNA cap epitranscriptome. *Nature protocols* **18,** 2671–2698; 10.1038/s41596-023-00857-0 (2023).

202. Suzuki, T., Ikeuchi, Y., Noma, A., Suzuki, T. & Sakaguchi, Y. Mass spectrometric identification and characterization of RNA-modifying enzymes. *Methods in enzymology* **425,** 211–229; 10.1016/S0076-6879(07)25009-8 (2007).

203. Beverly, M., Dell, A., Parmar, P. & Houghton, L. Label-free analysis of mRNA capping efficiency using RNase H probes and LC-MS. *Analytical and bioanalytical chemistry* **408,** 5021–5030; 10.1007/s00216-016-9605-x (2016).

204. Lobue, P. A., Jora, M., Addepalli, B. & Limbach, P. A. Oligonucleotide analysis by hydrophilic interaction liquid chromatography-mass spectrometry in the absence of ion-pair reagents. *Journal of chromatography. A* **1595,** 39–48; 10.1016/j.chroma.2019.02.016 (2019).

205. Yoluç, Y. *et al.* Instrumental analysis of RNA modifications. *Critical reviews in biochemistry and molecular biology* **56,** 178–204; 10.1080/10409238.2021.1887807 (2021).

206. Bélanger, J. M., Jocelyn Paré, J. R. & Sigouin, M. Chapter 2 High performance liquid chromatography (HPLC): Principles and applications. In *Instrumental Methods in Food Analysis* (Elsevier1997), Vol. 18, pp. 37–59.

207. Žuvela, P. *et al.* Column Characterization and Selection Systems in Reversed-Phase High-Performance Liquid Chromatography. *Chemical reviews* **119,** 3674–3729; 10.1021/acs.chemrev.8b00246 (2019).

208. Sarker, S. D. & Nahar, L. Applications of High Performance Liquid Chromatography in the Analysis of Herbal Products. In *Evidence-Based Validation of Herbal Medicine* (Elsevier2015), pp. 405–425.

209. Dvořák, M. *et al.* Equivalent peak resolution: characterization of the extent of separation for two components based on their relative peak overlap. *Electrophoresis* **36,** 646–654; 10.1002/elps.201400475 (2015).

210. Practical HPLC Method Development. *Anal. Chem.* **66,** 853A-853A; 10.1021/ac00089a717 (1994).

211. Dolan, J. W. Why Do Peaks Tail? *LCGC North America* **21,** 612–616 (2003).

212. Birajdar, A. S. New Method Development by HPLC and Validation as per ICH Guideline. *Acta Scientific Pharmaceutical Sciences (ISSN: 2581-5423)* **4,** 55–60 (2020).

213. Kromidas, S. *Optimization in HPLC. Concepts and Strategies* (Wiley-VCH, Weinheim, Germany, 2021).

214. Kazakevic, Y. & LoBrutto, R. *HPLC for pharmaceutical scientists* (John Wiley & Sons, Hoboken, New Jersey, 2007).

215. Dolan, J. W. Back to Basics: The Role of pH in Retention and Selectivity. *LCGC Europe* **30,** 30–33 (2017).

216. Lupo, S., Kahler & Ty. New Advice on an Old Topic: Buffers in Reversed-Phase HPLC. *LCGC North America* **35,** 424–433 (2017).

217. Snyder, L. R., Kirkland, J. J. & Glajch, J. L. *Practical HPLC method development.* 2nd ed. (Wiley, New York, 1997).

218. Agrafiotou, P., Ràfols, C., Castells, C., Bosch, E. & Rosés, M. Simultaneous effect of pH, temperature and mobile phase composition in the chromatographic retention of ionizable compounds. *Journal of chromatography. A* **1218,** 4995–5009; 10.1016/j.chroma.2010.12.119 (2011).

219. Dolan, J. W. Temperature selectivity in reversed-phase high performance liquid chromatography. *Journal of chromatography. A* **965,** 195–205; 10.1016/s0021-9673(01)01321-8 (2002).

220. Kromidas, S. (ed.). *The HPLC expert. Possibilities and limitations of modern high performance liquid chromatography* (Wiley-VCH, Weinheim, Germany, 2016).

221. McKern, N. M., Edskes, H. K. & Shukla, D. D. Purification of hydrophilic and hydrophobic peptide fragments on a single reversed phase high performance liquid chromatographic column. *Biomedical chromatography : BMC* **7,** 15–19; 10.1002/bmc.1130070105 (1993).

222. Antia, F. D. & Horváth, C. High-performance liquid chromatography at elevated temperatures: examination of conditions for the rapid separation of large molecules. *Journal of chromatography. A* **435,** 1–15; 10.1016/S0021-9673(01)82158-0 (1988).

223. Konermann, L. Addressing a Common Misconception: Ammonium Acetate as Neutral pH “Buffer” for Native Electrospray Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **28,** 1827–1835; 10.1007/s13361-017-1739-3 (2017).

224. Pitt, J. J. Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *The Clinical biochemist. Reviews* **30,** 19–34 (2009).

225. Amjad, M. & Sarfraz, S. A Comprehensive Review on HPLC Method Development, Validation, Optimization in Pharmaceuticals. *OAJWX* **6,** 1–9; 10.23880/oajwx-16000188 (2023).

226. Atkins, P. Looking with Light: Breaking Down Liquid Chromatography Method Development. *Cannabis Science and Technology* **4,** 17–28 (2021).

227. Tome, T., Žigart, N., Časar, Z. & Obreza, A. Development and Optimization of Liquid Chromatography Analytical Methods by Using AQbD Principles: Overview and Recent Advances. *Org. Process Res. Dev.* **23,** 1784–1802; 10.1021/acs.oprd.9b00238 (2019).

228. Ganorkar, S. B. & Shirkhedkar, A. A. Design of experiments in liquid chromatography (HPLC) analysis of pharmaceuticals: analytics, applications, implications and future prospects. *Reviews in Analytical Chemistry* **36**; 10.1515/revac-2016-0025 (2017).

229. Latrous, L. Optimization and Validation in Liquid Chromatography Using Design of Experiments. *Chemistry Africa* **5,** 437–458; 10.1007/s42250-022-00344-1 (2022).

230. Thorsteinsdóttir, U. A. & Thorsteinsdóttir, M. Design of experiments for development and optimization of a liquid chromatography coupled to tandem mass spectrometry bioanalytical assay. *Journal of mass spectrometry : JMS* **56,** e4727; 10.1002/jms.4727 (2021).

231. Zhang, J. *et al.* Analytical Procedure Development and Proposed Established Conditions: A Case Study of a Mass Spectrometry Based NDSRI Analytical Procedure. *Journal of pharmaceutical sciences* **113,** 3028–3033; 10.1016/j.xphs.2024.07.022 (2024).

232. Park, G., Kim, M. K., Go, S. H., Choi, M. & Jang, Y. P. Analytical Quality by Design (AQbD) Approach to the Development of Analytical Procedures for Medicinal Plants. *Plants (Basel, Switzerland)* **11**; 10.3390/plants11212960 (2022).

233. Agilent Technologies. *Agilent 6400 Series Triple Quadrupole LC/MS Ssystem. Concepts Guide. The Big Picture* (2012).

234. Kebarle, P. & Verkerk, U. H. Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass spectrometry reviews* **28,** 898–917; 10.1002/mas.20247 (2009).

235. Zhou, S. & Hamburger, M. Effects of solvent composition on molecular ion response in electrospray mass spectrometry: Investigation of the ionization processes. *Rapid Comm Mass Spectrometry* **9,** 1516–1521; 10.1002/rcm.1290091511 (1995).

236. Cech, N. B. & Enke, C. G. Practical implications of some recent studies in electrospray ionization fundamentals. *Mass spectrometry reviews* **20,** 362–387; 10.1002/mas.10008 (2001).

237. Liigand, J., Laaniste, A. & Kruve, A. pH Effects on Electrospray Ionization Efficiency. *Journal of the American Society for Mass Spectrometry* **28,** 461–469; 10.1007/s13361-016-1563-1 (2017).

238. Constantopoulos, T. L., Jackson, G. S. & Enke, C. G. Effects of salt concentration on analyte response using electrospray ionization mass spectrometry. *Journal of the American Society for Mass Spectrometry* **10,** 625–634; 10.1016/S1044-0305(99)00031-8 (1999).

239. Kostiainen, R. & Kauppila, T. J. Effect of eluent on the ionization process in liquid chromatography-mass spectrometry. *Journal of chromatography. A* **1216,** 685–699; 10.1016/j.chroma.2008.08.095 (2009).

240. Juraschek, R., Dülcks, T. & Karas, M. Nanoelectrospray—more than just a minimized-flow electrospray ionization source. *Journal of the American Society for Mass Spectrometry* **10,** 300–308; 10.1016/S1044-0305(98)00157-3 (1999).

241. Tang, X., Bruce, J. E. & Hill, H. H. Characterizing electrospray ionization using atmospheric pressure ion mobility spectrometry. *Anal. Chem.* **78,** 7751–7760; 10.1021/ac0613380 (2006).

242. Prabhu, G. R. D., Ponnusamy, V. K., Witek, H. A. & Urban, P. L. Sample Flow Rate Scan in Electrospray Ionization Mass Spectrometry Reveals Alterations in Protein Charge State Distribution. *Anal. Chem.* **92,** 13042–13049; 10.1021/acs.analchem.0c01945 (2020).

243. Mordehai, A. & Fjeldsted, J. Agilent Jet Stream Thermal Gradient Focusing Technology. *Agilent Technologies* (2009).

244. Bruins, A. P. Mechanistic aspects of electrospray ionization. *Journal of chromatography. A* **794,** 345–357; 10.1016/S0021-9673(97)01110-2 (1998).

245. Pedro, L., van Voorhis, W. C. & Quinn, R. J. Optimization of Electrospray Ionization by Statistical Design of Experiments and Response Surface Methodology: Protein-Ligand Equilibrium Dissociation Constant Determinations. *Journal of the American Society for Mass Spectrometry* **27,** 1520–1530; 10.1007/s13361-016-1417-x (2016).

246. Szerkus, O., Yumba Mpanga, A., Siluk, D. & Kaliszan, Roman, Markuszewski, Michal J. Optimization of the Electrospray Ionization Source with the Use of the Design of Experiments Approach for the LC–MS-MS Determination of Selected Metabolites in Human Urine. *Spectroscopy Supplements* **14,** 8–16 (2016).

247. Tandem Mass Spectrometry. In *Fundamentals of Contemporary Mass Spectrometry,* edited by C. Dass (Wiley2007), pp. 119–150.

248. Thüring, K., Schmid, K., Keller, P. & Helm, M. Analysis of RNA modifications by liquid chromatography-tandem mass spectrometry. *Methods (San Diego, Calif.)* **107,** 48–56; 10.1016/j.ymeth.2016.03.019 (2016).

249. Pomerantz, S. C. & McCloskey, J. A. Analysis of RNA hydrolyzates by liquid chromatography-mass spectrometry. *Methods in enzymology* **193,** 796–824; 10.1016/0076-6879(90)93452-Q (1990).

250. Kellner, S. *et al.* Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. *Nucleic acids research* **42,** e142; 10.1093/nar/gku733 (2014).

251. Asadi-Atoi, P., Barraud, P., Tisne, C. & Kellner, S. Benefits of stable isotope labeling in RNA analysis. *Biological chemistry* **400,** 847–865; 10.1515/hsz-2018-0447 (2019).

252. Cole, M. D. & Cowling, V. H. Specific regulation of mRNA cap methylation by the c-Myc and E2F1 transcription factors. *Oncogene* **28,** 1169–1175; 10.1038/onc.2008.463 (2009).

253. Habjan, M. & Pichlmair, A. Cytoplasmic sensing of viral nucleic acids. *Current opinion in virology* **11,** 31–37; 10.1016/j.coviro.2015.01.012 (2015).

254. Stępiński, J., Jemielity, J., Lewdorowicz, M., Jankowska-Anyszka, M. & Darżynkiewicz, E. Catalytic efficiency of divalent metal salts in dinucleoside 5',5'-triphosphate bond formation. In *Collection Symposium Series* (Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Prague, 2002), pp. 154–158.

255. Ripp, A. *et al.* The P(III)-amidite based synthesis of stable isotope labeled mRNA-cap-structures enables their sensitive quantitation from brain tissue. *Angewandte Chemie*; 10.1002/ange.202414537 (2024).

256. Pichlmair, A. *et al.* IFIT1 is an antiviral protein that recognizes 5'-triphosphate RNA. *Nature immunology* **12,** 624–630; 10.1038/ni.2048 (2011).

257. Kageyama Kaneshima, A., Motoyama, A. & Takayama, M. Influence of Solvent Composition and Surface Tension on the Signal Intensity of Amino Acids in Electrospray Ionization Mass Spectrometry. *Mass spectrometry (Tokyo, Japan)* **8,** A0077; 10.5702/massspectrometry.A0077 (2019).

258. Hertler, J. *et al.* Synthesis of point-modified mRNA. *Nucleic acids research* **50,** e115; 10.1093/nar/gkac719 (2022).

259. Gao, S., Zhang, Z.-P. & Karnes, H. T. Sensitivity enhancement in liquid chromatography/atmospheric pressure ionization mass spectrometry using derivatization and mobile phase additives. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **825,** 98–110; 10.1016/j.jchromb.2005.04.021 (2005).

260. Mallet, C. R., Lu, Z. & Mazzeo, J. R. A study of ion suppression effects in electrospray ionization from mobile phase additives and solid-phase extracts. *Rapid Comm Mass Spectrometry* **18,** 49–58; 10.1002/rcm.1276 (2004).

261. *European pharmacopoeia 11.6, 2.2.46 Chromatographic Separation Techniques* (European Directorate for the Quality of Medicines: Council of Europe, Strasbourg, France, 2024).

262. Corfù, N. A., Tribolet, R. & Sigel, H. Comparison of the self-association properties of the 5'-triphosphates of inosine (ITP), guanosine (GTP), and adenosine (ATP). Further evidence for ionic interactions in the highly stable dimeric H2(ATP)2(4-) stack. *European journal of biochemistry* **191,** 721–735; 10.1111/j.1432-1033.1990.tb19181.x (1990).

263. Jones, E. L., Mlotkowski, A. J., Hebert, S. P., Schlegel, H. B. & Chow, C. S. Calculations of pKa Values for a Series of Naturally Occurring Modified Nucleobases. *The journal of physical chemistry. A* **126,** 1518–1529; 10.1021/acs.jpca.1c10905 (2022).

264. Law, A. S., Hafen, P. S. & Brault, J. J. Liquid chromatography method for simultaneous quantification of ATP and its degradation products compatible with both UV-Vis and mass spectrometry. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **1206,** 123351; 10.1016/j.jchromb.2022.123351 (2022).

265. Waters Knowledge Base. How do I determine column void volume? - WKB28079. Article number: 28079. Available at https://support.waters.com/KB_Chem/Columns/WKB28079_How_do_I_determine_column_void_volume (2022), retrieved on 23.12.2024.

266. Ribar, D., Lukšič, M. & Kralj Cigić, I. Towards an accurate method for column void volume determination using liquid chromatography-mass spectrometry. *Journal of chromatography. A* **1706,** 464245; 10.1016/j.chroma.2023.464245 (2023).

267. Luo, H. & Cheng, Y.-K. A comparative study of void volume markers in immobilized-artificial-membrane and reversed-phase liquid chromatography. *Journal of chromatography. A* **1103,** 356–361; 10.1016/j.chroma.2005.12.036 (2006).

268. Sheehan, T. L. & Yost, R. A. What's the Most Meaningful Standard for Mass Spectrometry: Instrument Detection Limit or Signal-to-Noise Ratio? *Spectroscopy* **13,** 16–22 (2015).

269. Parra, N. P. & Taylor, L. Why Instrument Detection Limit (IDL) Is a Better Metric for Determining The Sensitivity of Triple Quadrupole LC/MS Systems. *Agilent Technologies* (2014).

270. Slomovic, S., Laufer, D., Geiger, D. & Schuster, G. Polyadenylation of ribosomal RNA in human cells. *Nucleic acids research* **34,** 2966–2975; 10.1093/nar/gkl357 (2006).

271. Kuai, L., Fang, F., Butler, J. S. & Sherman, F. Polyadenylation of rRNA in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 8581–8586; 10.1073/pnas.0402888101 (2004).

272. Wolter, S. Die biologische Funktion der N1-2´O-Methylierung eukaryotischer mRNA durch die Methyltransferase CMTr1. Universitäts- und Landesbibliothek Bonn, 2021.

273. Wolter, S. *et al.* mRNA N1-2'O-methylation by CMTR1 affects NVL2 mRNA splicing. *bioRxiv,* 2023.09.28.560047; 10.1101/2023.09.28.560047 (2023).

274. Jansen, R. S., Rosing, H., Schellens, J. H. M. & Beijnen, J. H. Mass spectrometry in the quantitative analysis of therapeutic intracellular nucleotide analogs. *Mass spectrometry reviews* **30,** 321–343; 10.1002/mas.20280 (2011).

275. Vela, J. E., Olson, L. Y., Huang, A., Fridland, A. & Ray, A. S. Simultaneous quantitation of the nucleotide analog adefovir, its phosphorylated anabolites and 2'-deoxyadenosine triphosphate by ion-pairing LC/MS/MS. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **848,** 335–343; 10.1016/j.jchromb.2006.10.063 (2007).

276. Bapiro, T. E. *et al.* A novel method for quantification of gemcitabine and its metabolites 2',2'-difluorodeoxyuridine and gemcitabine triphosphate in tumour tissue by LC-MS/MS: comparison with (19)F NMR spectroscopy. *Cancer chemotherapy and pharmacology* **68,** 1243–1253; 10.1007/s00280-011-1613-0 (2011).

277. Holmgren, E., Carlsson, H., Goede, P. & Crescenzi, C. Determination and characterization of organic explosives using porous graphitic carbon and liquid chromatography-atmospheric pressure chemical ionization mass spectrometry. *Journal of chromatography. A* **1099,** 127–135; 10.1016/j.chroma.2005.08.088 (2005).

278. Bapiro, T. E., Richards, F. M. & Jodrell, D. I. Understanding the Complexity of Porous Graphitic Carbon (PGC) Chromatography: Modulation of Mobile-Stationary Phase Interactions Overcomes Loss of Retention and Reduces Variability. *Anal. Chem.* **88,** 6190–6194; 10.1021/acs.analchem.6b01167 (2016).

279. Keith, J. M., Ensinger, M. J. & Moss, B. HeLa cell RNA (2'-O-methyladenosine-N6-)-methyltransferase specific for the capped 5'-end of messenger RNA. *Journal of Biological Chemistry* **253,** 5033–5039; 10.1016/S0021-9258(17)34652-5 (1978).

280. Schweizer, M. P., Chan, S. I., Helmkamp, G. K. & Ts'o, P. O. P. An Experimental Assignment of the Proton Magnetic Resonance Spectrum of Purine. *J. Am. Chem. Soc.* **86,** 696–700; 10.1021/ja01058a032 (1964).

281. Bullock, F. J. & Jardetzky, O. Proton Magnetic Resonance Studies of Purines and Pyrimidines. XII. An Experimental Assignment of Peaks in Purine Derivatives 1a. *J. Org. Chem.* **29,** 1988–1990; 10.1021/jo01030a082 (1964).

282. Gregson, J. M. *et al.* Structure of the archaeal transfer RNA nucleoside G*-15 (2-amino-4,7-dihydro- 4-oxo-7-beta-D-ribofuranosyl-1H-pyrrolo[2,3-d]pyrimidine-5-carboximi dam ide (archaeosine)). *Journal of Biological Chemistry* **268,** 10076–10086; 10.1016/S0021-9258(18)82174-3 (1993).

283. Dal Magro, C. *et al.* A Vastly Increased Chemical Variety of RNA Modifications Containing a Thioacetal Structure. *Angewandte Chemie (International ed. in English)* **57,** 7893–7897; 10.1002/anie.201713188 (2018).

284. Bessler, L. *et al.* A New Bacterial Adenosine-Derived Nucleoside as an Example of RNA Modification Damage. *Angewandte Chemie (International ed. in English)* **62,** e202217128; 10.1002/anie.202217128 (2023).

285. Tomasz, M., Olson, J. & Mercado, C. M. Mechanism of the isotopic exchange of the C-8 hydrogen of purines in nucleosides and in deoxyribonucleic acid. *Biochemistry* **11,** 1235–1241; 10.1021/bi00757a019 (1972).

286. Shelton, K. R. & Clark, J. M. A proton exchange between purines and water and its application to biochemistry. *Biochemistry* **6,** 2735–2739; 10.1021/bi00861a013 (1967).

287. Abdel-Hamid, M., Novotny, L. & Hamza, H. Stability study of selected adenosine nucleosides using LC and LC/MS analyses. *Journal of pharmaceutical and biomedical analysis* **22,** 745–755; 10.1016/S0731-7085(99)00276-9 (2000).

288. Hulett, H. R. Non-enzymatic hydrolysis of adenosine phosphates. *Nature* **225,** 1248–1249; 10.1038/2251248a0 (1970).

289. Moeller, C., Schmidt, C., Guyot, F. & Wilke, M. Hydrolysis rate constants of ATP determined in situ at elevated temperatures. *Biophysical Chemistry* **290,** 106878; 10.1016/j.bpc.2022.106878 (2022).

290. Leibrock, E., Bayer, P. & Lüdemann, H. D. Nonenzymatic hydrolysis of adenosinetriphosphate (ATP) at high temperatures and high pressures. *Biophysical Chemistry* **54,** 175–180; 10.1016/0301-4622(94)00134-6 (1995).

291. Tomasz, M. Extreme lability of the C-8 proton: a consequence of 7-methylation of guanine residues in model compounds and in DNA and its analytical application. *Biochimica et biophysica acta* **199,** 18–28; 10.1016/0005-2787(70)90690-8 (1970).

292. Pelletier, J., Schmeing, T. M. & Sonenberg, N. The multifaceted eukaryotic cap structure. *Wiley interdisciplinary reviews. RNA* **12,** e1636; 10.1002/wrna.1636. (2021).

293. Biedenbänder, T. *et al.* RNA modifications stabilize the tertiary structure of tRNAfMet by locally increasing conformational dynamics. *Nucleic acids research* **50,** 2334–2349; 10.1093/nar/gkac040 (2022).

294. Walker, S. E. & Lorsch, J. RNA purification-precipitation methods. *Methods in enzymology* **530,** 337–343; 10.1016/B978-0-12-420037-1.00019-1 (2013).

295. Zhang, J. *et al.* Recent Advances and Innovations in the Preparation and Purification of In Vitro-Transcribed-mRNA-Based Molecules. *Pharmaceutics* **15**; 10.3390/pharmaceutics15092182 (2023).

296. Lenk, R. *et al.* Understanding the impact of in vitro transcription byproducts and contaminants. *Frontiers in molecular biosciences* **11,** 1426129; 10.3389/fmolb.2024.1426129 (2024).

297. Beckert, B. & Masquida, B. Synthesis of RNA by In Vitro Transcription. In *RNA,* edited by H. Nielsen (Humana Press, Totowa, NJ, 2011), Vol. 703, pp. 29–41.

298. Martin, C. T., Muller, D. K. & Coleman, J. E. Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry* **27,** 3966–3974; 10.1021/bi00411a012 (1988).

299. Hadas, Y. *et al.* Optimizing Modified mRNA In Vitro Synthesis Protocol for Heart Gene Therapy. *Molecular therapy. Methods & clinical development* **14,** 300–305; 10.1016/j.omtm.2019.07.006 (2019).

300. Wu, M. Z., Asahara, H., Tzertzinis, G. & Roy, B. Synthesis of low immunogenicity RNA with high-temperature in vitro transcription. *RNA (New York, N.Y.)* **26,** 345–360; 10.1261/rna.073858.119 (2020).

301. Ziegenhals, T. *et al.* Formation of dsRNA by-products during in vitro transcription can be reduced by using low steady-state levels of UTP. *Frontiers in molecular biosciences* **10,** 1291045; 10.3389/fmolb.2023.1291045 (2023).

302. Mu, X., Greenwald, E., Ahmad, S. & Hur, S. An origin of the immunogenicity of in vitro transcribed RNA. *Nucleic acids research* **46,** 5239–5249; 10.1093/nar/gky177 (2018).

303. Zhang, Q. *et al.* Recent Methods for Purification and Structure Determination of Oligonucleotides. *International journal of molecular sciences* **17**; 10.3390/ijms17122134 (2016).

304. Davies, S. R. *et al.* Purity assessment of organic calibration standards using a combination of quantitative NMR and mass balance. *Analytical and bioanalytical chemistry* **407,** 3103–3113; 10.1007/s00216-014-7893-6 (2015).

305. Shoolery, J. N. Some quantitative applications of 13C NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy* **11,** 79–93; 10.1016/0079-6565(77)80003-4 (1977).

306. Giancaspro, G. *et al.* The qNMR Summit 5.0: Proceedings and Status of qNMR Technology. *Anal. Chem.* **93,** 12162–12169; 10.1021/acs.analchem.1c02056 (2021).

307. Ammann, G., Berg, M., Dalwigk, J. F. & Kaiser, S. M. Pitfalls in RNA Modification Quantification Using Nucleoside Mass Spectrometry. *Accounts of chemical research* **56,** 3121–3131; 10.1021/acs.accounts.3c00402 (2023).

308. Anderson, B. R. *et al.* Incorporation of pseudouridine into mRNA enhances translation by diminishing PKR activation. *Nucleic acids research* **38,** 5884–5892; 10.1093/nar/gkq347 (2010).

309. Grudzien, E. *et al.* Novel cap analogs for in vitro synthesis of mRNAs with high translational efficiency. *RNA (New York, N.Y.)* **10,** 1479–1487; 10.1261/rna.7380904 (2004).

310. Fuchs, A.-L., Neu, A. & Sprangers, R. A general method for rapid and cost-efficient large-scale production of 5′ capped RNA. *RNA (New York, N.Y.)* **22,** 1454–1466; 10.1261/rna.056614.116 (2016).

311. Vlatkovic, I. *et al.* Ribozyme Assays to Quantify the Capping Efficiency of In Vitro-Transcribed mRNA. *Pharmaceutics* **14**; 10.3390/pharmaceutics14020328 (2022).

312. Karikó, K., Muramatsu, H., Ludwig, J. & Weissman, D. Generating the optimal mRNA for therapy: HPLC purification eliminates immune activation and improves translation of nucleoside-modified, protein-encoding mRNA. *Nucleic acids research* **39,** e142; 10.1093/nar/gkr695 (2011).

313. Brandscheid, C. *et al.* Altered Gut Microbiome Composition and Tryptic Activity of the 5xFAD Alzheimer's Mouse Model. *Journal of Alzheimer's disease : JAD* **56,** 775–788; 10.3233/JAD-160926 (2017).

Supplement

Table S1. Mass spectrometer settings for analyzed cap dinucleotides comprising precursor and product ion mass-to-charge ratios (*m/z*), as well as optimized values for fragmentor voltages, and collision energies.

Compound	Precursor ion [*m/z*]	Product ion [*m/z*]	Fragmentor voltage [V]	Collision energy [V]	$(D_3C)^7$G-labeled precursor ion [*m/z*]	$(D_3C)^7$G-labeled product ion [*m/z*]
GpppA	773	136	145	66		
Gpppm^{6}A	787	150	155	82		
GpppAm	787	136	180	58		
Gpppm^{6}Am	801	150	165	82		
m^{7}GpppA	787	136	130	82		
m^{7}Gpppm^{6}A	801	150	175	90	804	150
m^{7}GpppAm	801	136	165	62	804	136
m^{7}Gpppm^{6}Am	815	150	165	74	818	150

Table S2. Instrument detection limits for cap dinucleotides

Compound	IDL of initial method [fmol]	IDL of optimized method [fmol]	Ratio of initial method/ optimized method
GpppA	1.9	0.9	2.1
Gpppm^{6}A	2.9	0.9	3.2
GpppAm	2.7	0.9	3.0
Gpppm^{6}Am	3.6	1.2	3.0
m^{7}GpppA	3.2	1.0	3.2
m^{7}Gpppm^{6}A	3.7	1.1	3.4
m^{7}GpppAm	2.4	1.0	2.4
m^{7}Gpppm^{6}Am	2.5	1.0	2.5

Table S3. Use of artificial intelligence

Artificial intelligence tool	Used for	Why	When
DeepL Write	Reformulation of own text drafts	To improve English language, grammar, and style	Throughout the entire work
Perplexity	Assessment of comprehensibility and use of scientific language of own text drafts	To improve English language, grammar, and style	For selected text passages throughout the entire work
	Reformulation and rephrasing of own text drafts		
	Proposals to shorten and condense own text drafts	To achieve a more compact writing style	Abstract and conclusion
	Summary of provided literature	To efficiently identify key points and find entry points for summarizing complex concepts	Introduction chapter 1.3 and 1.5

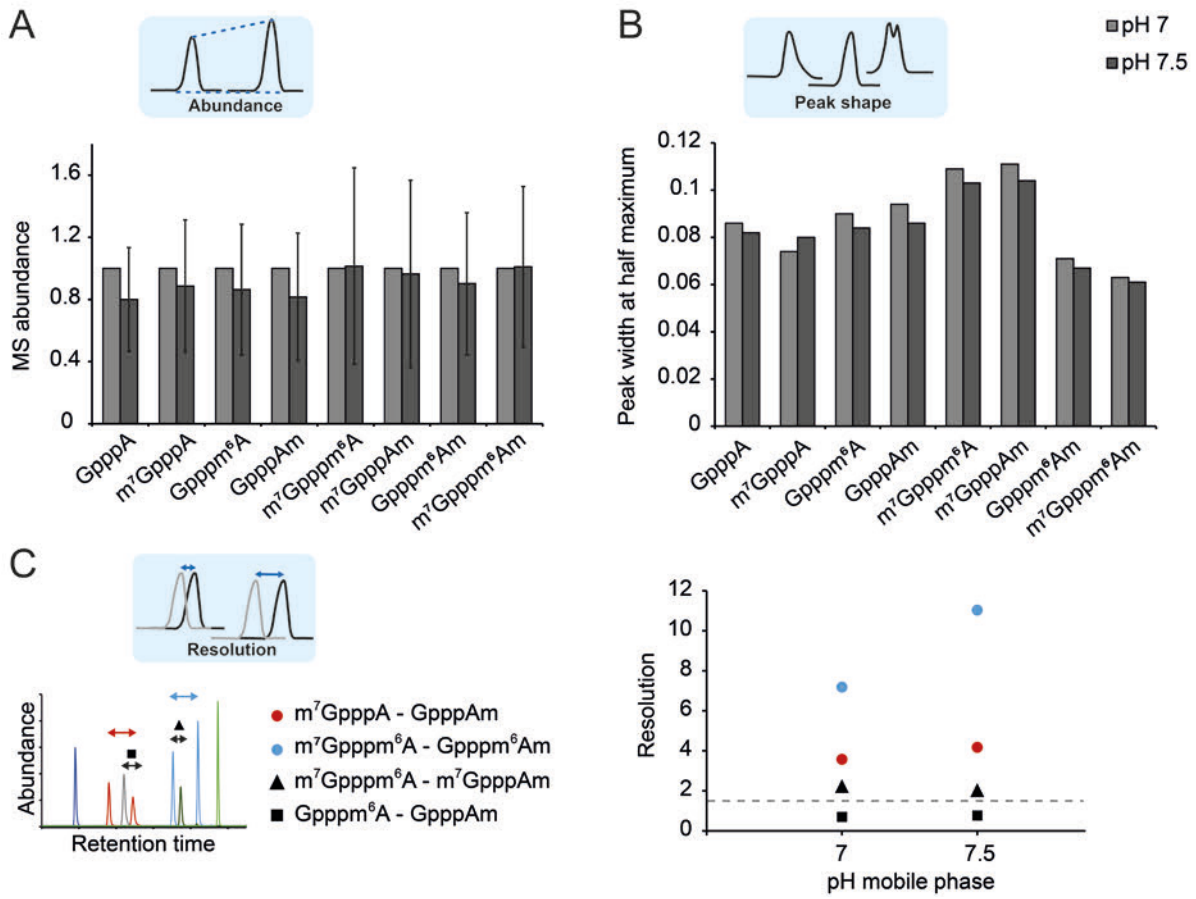

Figure S1. A.) Peak areas (expressed as MS abundance) of cap dinucleotides analyzed using aqueous mobile phases at pH 7 and 7.5. Measurements were conducted in three replicates on three different days. For each replicate, peak areas were normalized to the corresponding peak area of each cap dinucleotide measured at pH 7. This normalization accounts for potential daily variations in MS performance. The average of three replicates is shown. B.) Investigation of peak shapes as determined by peak width at half maximum at aqueous mobile phase pH values of 7 and 7.5. The peak widths at half maximum were determined from one representative technical replicate out of the three used to investigate MS abundances. C.) Resolution of the peak pairs: m^7GpppA + GpppAm (red), m^7Gpppm^{6}A + Gpppm^{6}Am (light blue), m^7Gpppm^{6}A + m^7GpppAm (black triangle), and Gpppm^{6}A + GpppAm (black square), analyzed using aqueous mobile phases at pH 7 and 7.5. The threshold of 1.5 is indicated by a gray dashed line.

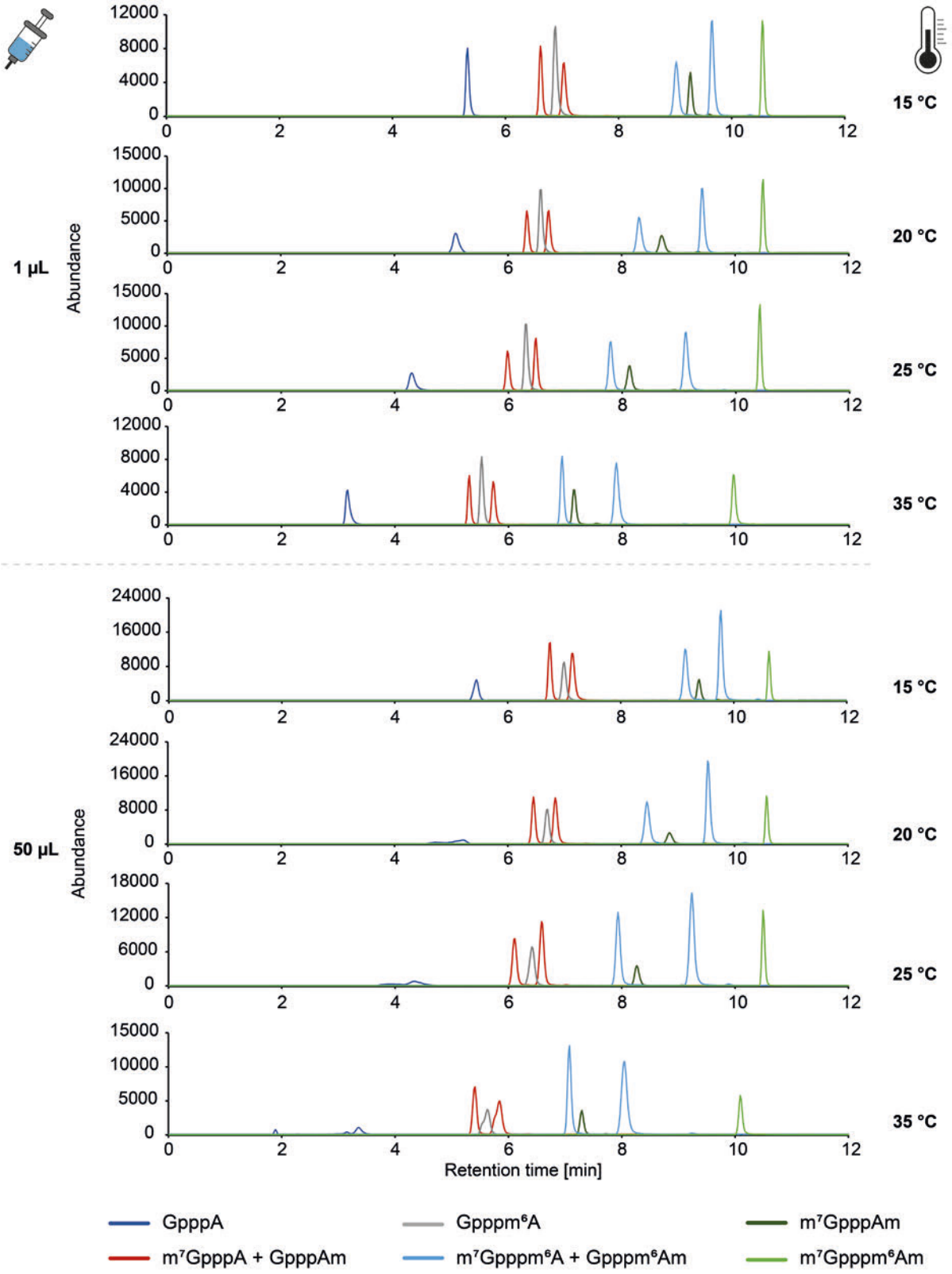

Figure S2. Overlay of extracted ion chromatograms of cap dinucleotides resulting from the injection of same amounts of a mixture of cap dinucleotide standards in different injection volumes and at different column temperatures. The upper four chromatograms were measured injecting 1 µL of the mixture of cap dinucleotide standards, the lower four injecting 50 µL.

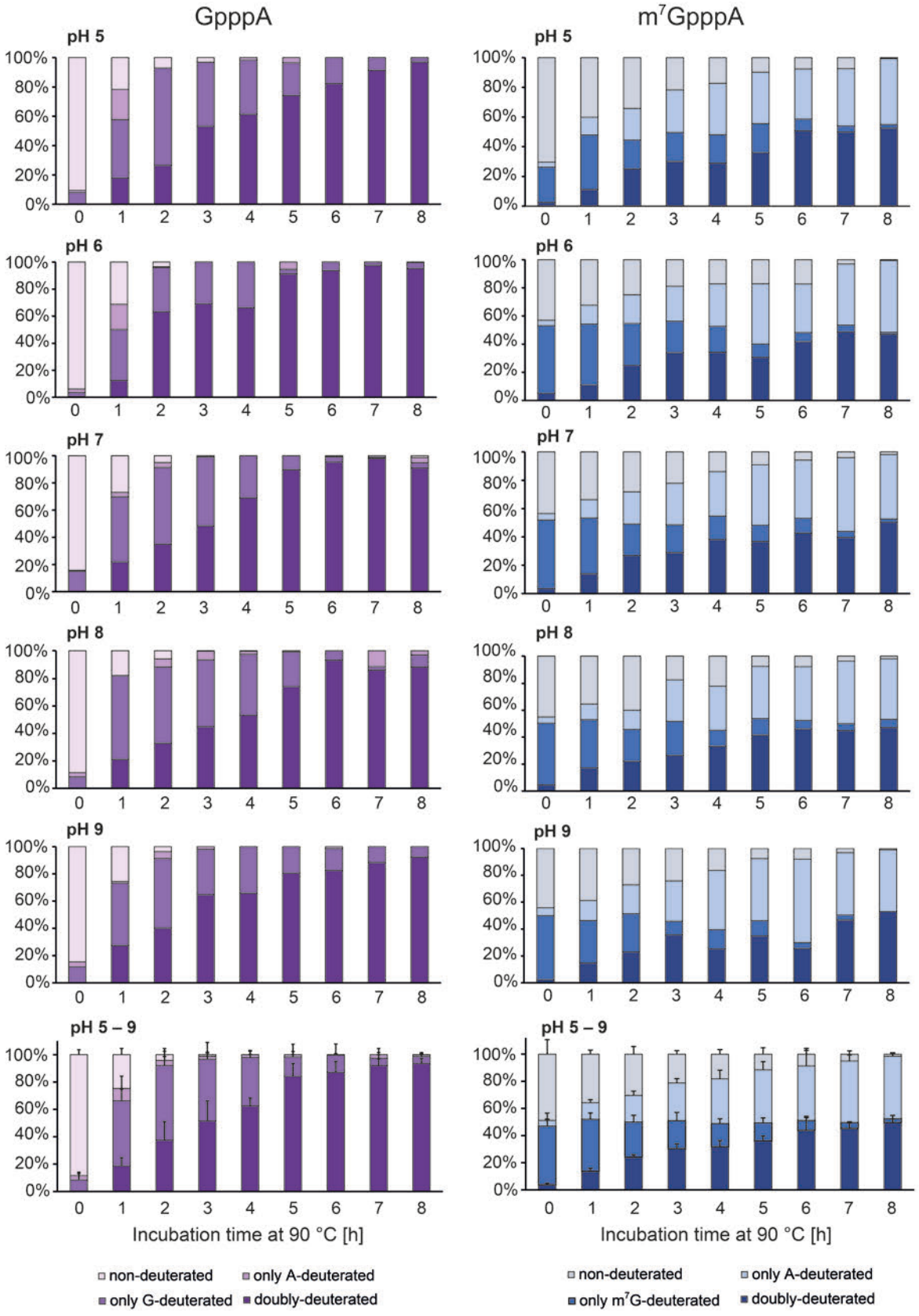

Figure S3. Deuteration pattern of GpppA and m^7GpppA after incubation in D_2O solution with pH values between 5-9 at 90 °C for varying time periods. The bottom graph depicts mean and standard deviation of the upper graphs, respectively.

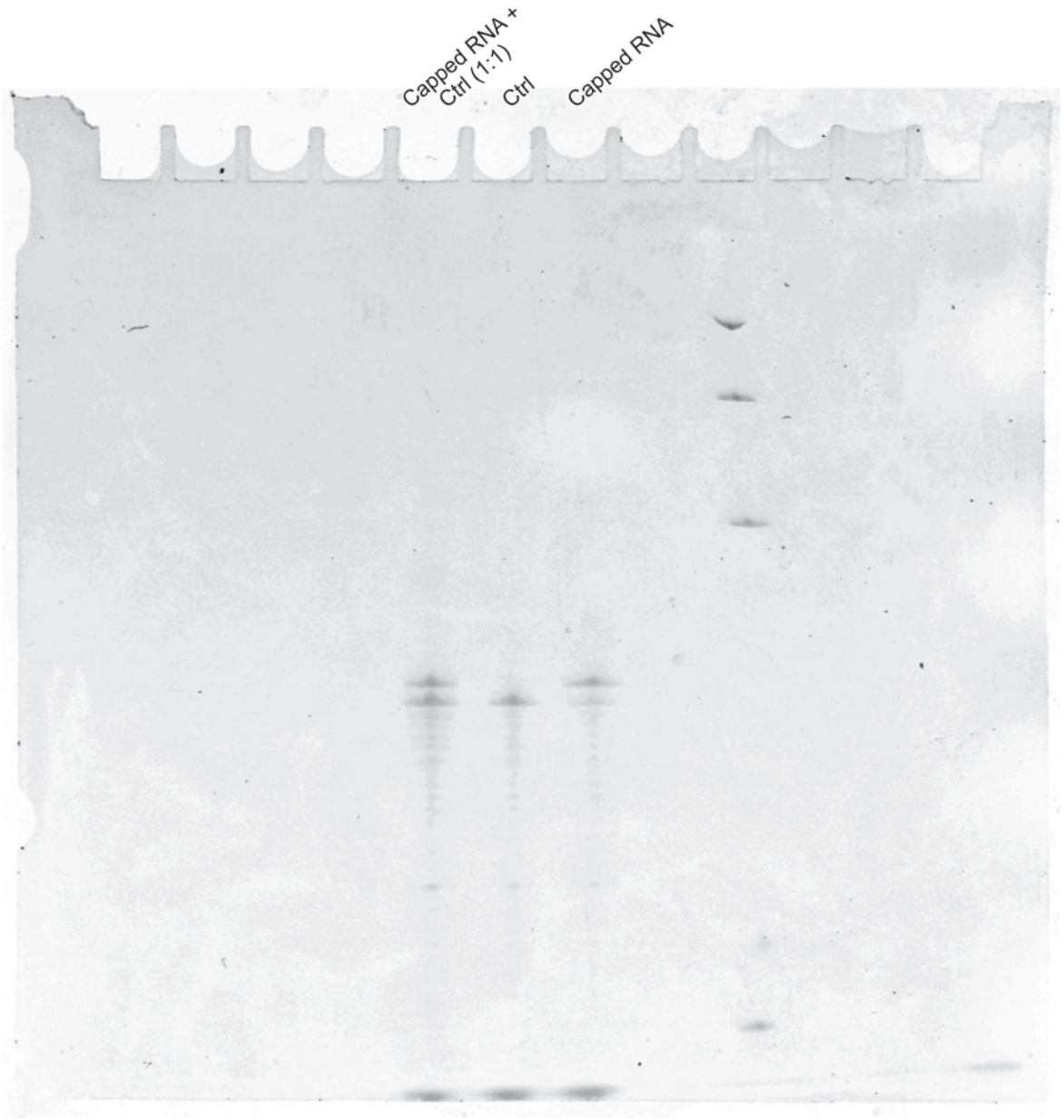

Figure S4. Complete gel image of the section depicted in Figure 3.21. A. Denaturing PAGE (20%, GelRed staining) of the capped RNA, the untreated control (Ctrl) and a 1:1 mixture of both.

Curriculum Vitae

[redacted]