Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Absolute quantification of exogenous stimuli-induced nucleic acid modification dynamics with LC-MS

Yasemin Yoluç

aus

Hagen, Deutschland

2022

Danksagung

An erster Stelle möchte ich mich bei Prof. Dr. Stefanie Kaiser bedanken, für all das was ich lernen durfte und die Verantwortungen die mir übertragen wurden. Schon als ich 2017 als wissenschaftliche Hilfskraft zum (damals noch) AK Kellner kam begeisterte mich die Welt der RNA Analytik. Der herzliche Umgang in der Gruppe und das Interesse an Massenspektrometrie führten mich 2018 wieder zum AK Kaiser und dadurch zum Beginn meiner Promotion. Stefanie, ich möchte dir besonders für deine Unterstützung und für die Motivation meine eigenen Ideen zu verfolgen bedanken. In den vergangenen 3.5 Jahren gab es viel Wandel in unserer Gruppe, vor allem mit dem Umzug des Labors, war ich mit vielen Änderungen und Herausforderungen konfrontiert, an denen ich gewachsen bin. Ich danke dir für diese schöne Zeit des Lernens und der Entwicklung.

Außerdem möchte ich Dr. Pavel Kielkowski für die Erstellung des Zweitgutachtens danken. Sowie den anderen Mitgliedern meines Prüfungskomitees, Prof. Thomas Carell, Prof. Konstantin Karaghiosoff, Prof. Axel Imhoff und Prof. Lena Daumann.

Ein großer Dank geht an meine lieben Kollegen, die ich alle ins Herz geschlossen habe und mit denen ich wundervolle Erinnerungen an meine Promotion erschaffen habe. Ich erinnere mich zu gern an meine Rückkehr von einem 8-tägigen New York Trip im Herbst 2019 zurück, bei dem ich zu meiner Freundin sagte: „Oh, übermorgen ist endlich Montag und ich kann wieder in's Labor. Hab's echt vermisst!" Dass ich mich so auf die Arbeit freute, war zum Teil meiner Freude an der Wissenschaft geschuldet aber sicherlich auch dem herzlichen und freundschaftlichen Miteinander unter den Kollegen.

Ich möchte mich bei Dr. Felix Hagelskamp bedanken, nicht nur für den allmorgendlichen Gossip und die kleinen Rage-Mode-Times, sondern vor allem für seine hilfsbereite und freundliche Art, mit der er mir stets zu Rat und Tat stand. Außerdem, bedanke ich mich bei Dr. Matthias Heiß, der mir das Leben nicht nur durch das ein oder andere Flötenspiel und Cocktails versüßte, sondern vor allem durch seine Bereitschafft sich mit mir in meine Experimente reinzufuchsen. Ja, und dann war da noch der liebe Dr. Valentin Reichle, muss ich da noch mehr sagen?! Durch sein sonniges Gemüt und einer großen Portion Humor brachte er uns alle zum Lachen. Oft teilten wir unsere 5 Minute, zum Freud oder Leid der anderen. Wessen Anwesenheit sicherlich zur Freude aller führte, war Paria Asadi Atoi. Durch deine Besonnene Art, hast du mir in mancher

verzweifelten Minute wirklich gutgetan. Ich danke dir für die innigen Gespräche. Und letztlich, Dr. Kayla Borland, neben unseren regelmäßigen Isotope-Girls-Nights, an denen wir definitiv den amerikanischen Traum in Sachen Essen lebten, genoss ich die Zeit mit dir sehr und bin dankbar dafür, dass du dein breitgefächertes Wissen in RNA Biologie und Massenspektrometrie mit mir teiltest.

Wen ich hier definitiv nicht außer Acht lassen darf sind die lieben Carellos. Stefan Wiedemann und Jonas Feldmann, es war so schön mit euch! Jede Woche wartete ich nur darauf, dass es wieder Donnerstag wurde und wir wieder schön beisammensaßen. Auch die legendären Partys in Haus F waren jedes Mal eine große Freude und haben mit Sicherheit ein großes Zutragen zu meinen schönen Erinnerungen an die Promotion.

Umso schmerzlicher war der Umzug nach Frankfurt. Zum Glück kam da der Gregor, der den Übergang einfacher machte. Vor allem, der Umzug nach Frankfurt hat uns zusammengeschweißt. Ob Umzug, Laboraufbau, Publikationen oder die ein oder andere Party, durch dick und dünn sind wir gegangen und ich danke dir dafür, dass ich mich stets auf dich verlassen konnte und du mir mit deinem Rat stets zur Seite standst. Doch auch in Frankfurt erwarteten mich herzallerliebste Kollegen. Nurilein, du hast dem Felix tatsächlich den Rang abgelaufen in Sachen Gossip und warst somit eine prächtige Nachfolgerin. Auch für den ein oder anderen Lachflash hast du gesorgt und das Ragen wurde durch dich zum Rangen. Dann ist da noch der Max, das Schmusekätzchen, ich danke dir besonders für die zahlreichen Memes die du erstellt hast und mich dadurch immer zum Lachen gebracht hast. Und Hagen, der immer wusste wo die nächste Party steil ging und stets für den Aperol Nachschub sorgte. Ich wünsche euch, der neuen Ära des AK Kaisers alles liebe!

Außerdem möchte ich mich bei meinen Studenten bedanken, vor allem möchte ich hier Erik van de Logt und Hurina Hu betonen, deren Masterarbeiten ich betreut habe. Eure Arbeit hat maßgeblich zu meinen Ergebnissen beigetragen.

Der größte Dank geht an meine Eltern. Ich bin so dankbar für eure Unterstützung, dass ihr mich habt ausfliegen lassen, damit ich meine Träume verwirklichen und meine Ziele erreichen kann. Ich danke euch für die finanzielle Unterstützung, die es mir erlaubt hat in München zu studieren, mein Auslandssemster in Oxford zu absolvieren und meine Masterarbeit in Halle anzufertigen. Ich danke euch für die mentale Unterstützung, die

mich an schweren Tagen aufgebaut hat. Ihr seid stets meine Stütze gewesen. Vielen lieben Dank dafür!

Sonsuz desteğiniz için minnettarım!

Table of content

List of publications

- **Yoluç Y.**, *et al.*, Systematic assessment of methylation damage in nucleic acids, ChemBioChem, **2022** *(in revision)*
- **Yoluç Y.**, *et al.*, The Stress-Dependent Dynamics of Saccharomyces cerevisiae tRNA and rRNA Modification Profiles, Genes, **2021**
- Heiss M., **Yoluç Y**, *et al.*, Quantification of Modified Nucleosides in the Context of NAIL-MS, Methods Mol Biol., **2021**
- **Yoluç Y.**, *et al.*, Instrumental analysis of RNA modifications, Crit Rev Biochem Mol Biol., **2021**
- Kaiser S., **Yoluç Y.**, *et al.*, Strategies to Avoid Artifacts in Mass Spectrometry-Based Epitranscriptome Analyses, Angew Chem Int Ed Engl., **2021**

Conference attendances

- **Poster:** New tools for RNA modification dynamics analysis, 3rd Symposium on Nucleic Acid Modifications in Mosbach, Germany, **2022**
- **Oral presentation & poster:** The Stress-Dependent Dynamics of *S. cerevisiae* tRNA and rRNA Modification Profiles, SPP 1784 Meeting in Mainz, Germany, **2021**
- **Oral presentation & poster:** The mechanisms of stress-induced tRNA modification reprogramming in *S. cerevisiae* by NAIL-MS, SPP1784 Meeting in Bremen, Germany, **2019**
- **Poster:** The mechanisms of stress-induced tRNA modification reprogramming in *S. cerevisiae* by NAIL-MS, 2nd Symposium on Nucleic Acid Modifications in Rehovot, Germany, **2019**

Abstract

Modifications of nucleic acids comply different functions and are involved in genome organization, cell differentiation, silencing, structure stability and enzyme recognition. Modification abundances can be regulated intrinsically, like the incorporation of cap modifications on viral RNA to evade the host immune response, but also extrinsically as a cause of damage, which can result in mutations or translational defects. Either way, modifications are highly dynamic. It is of great importance to trace and quantify these changes in order to understand the underlying mechanisms, which may offer a more divers applicability of RNA therapeutics and even facilitate the establishment of personalized medicine.

Mass Spectrometry is a common technique to examine nucleic acids. However, mass spectrometry *per se* offers solely a static insight into the versatile dynamics of nucleic acid modifications. In order to circumvent this obstacle, **N**ucleic **A**cid **I**sotope **L**abeling **c**oupled **M**ass **S**pectrometry (NAIL-MS) was developed. This powerful technique allows for absolute quantification on the one hand and on the other hand for examination of modification dynamics originating from endogenous or exogenous actuators.

In 2010, the stress-induced reprogramming of tRNA modification in *S. cerevisiae* was reported. However, the underlying mechanisms remained to be elucidated. Few years later, the dynamics of RNA modifications and mechanisms like dilution, degradation and (de-)modification could be identified by the application of NAIL-MS. The first part of my dissertation deals with the examination of the extent of damage-induced alterations on nucleic acids. Therefore, a novel biosynthetically produced stable isotope labeled internal standard (SILIS) was established, to avoid the interference of signals with isotopologues generated in the stable isotope labeled pulse-chase experiments. Furthermore, the L-methionine-[^{2}H$_3$]-methyl labeling in *S. cerevisiae* was optimized to achieve full efficient labeling and thus again avoiding signal interferences with isotopologues due to inefficient labeling. Additionally, the tandem size exclusion chromatography was developed, allowing the time efficient purification of 28S/25S, 18S rRNA and tRNA in a single step. The application of improved stable isotope labeling and the facilitated purification of RNA populations allowed for the examination of the stress-induced alterations in the RNA modification profile of *S. cerevisiae*. Thereby, the knowledge on stress-induced reprogramming of RNA modifications in yeast could be expanded. Original and new transcripts of 25S, 18S and tRNA could be discerned and

in addition endogenous methylation could be differentiated from damage induced methylation. It was shown, that stress-induced alterations occur on original tRNA transcripts, whereas new transcripts were not affected. Moreover, the fast decrease of damage-induced methylations on 25S, 18S rRNA and tRNA in *S. cerevisiae* was demonstrated. Additionally, the formation of base damage on 2'-O-methylated nucleosides in rRNA upon methyl methanesulfonate (MMS) exposure was detected and thereby novel damage products of MMS could be identified. Furthermore, the application of NAIL-MS was expanded to study the endogenous and damage-induced methylome on the genomic levels in *S. cerevisiae* and *E. coli*. In parallel to the aforementioned findings, the fast disappearance of damage-induced methylations in the genome and transcriptome of *S. cerevisiae* and *E. coli* was shown. Apart from that, m^7dG and m^7G could be identified as the main damage products in the genome and transcriptome of both organisms.

In parallel to prokaryotes and eukaryotes, the modifications in viral RNA are highly dynamic. RNA viruses have high mutation rates and their modification abundances can vary during infection. So, several mutants and variants of the RNA virus SARS-CoV-2 emerged since 2019. It is necessary to understand the characteristics of the viral genome and the differences in mutants and variants in order to identify novel drug targets and optimize the application of available therapeutics and vaccines. Our previous work on absolute quantification of nucleic acid modifications in various organisms showed the strength of our LC-MS based approach. In the course of this study it was aimed at investigating the viral RNA modification profile in the different mutants and variants of SARS-CoV-2. The absolute quantification of RNA modifications and the comparison to published reports lead to the assumption that observed modification densities are highly dependent on the cultivation and infection conditions as well as the purification method and verification of sample integrity is crucial for valid analysis.

As outlined above, less is known about the genome of SARS-CoV-2 in terms of internal modifications. While the cap modification of the 5' end of the SARS-CoV-2 genome is confirmed from many sides and is ascribed to regulate the host innate immune response and the viral replication. Hence, a better understanding of the viral capping mechanism is required in order to limit its contagiousness. Besides the interest in biological capping processes, the investigations on cap modifications become more relevant nowadays because of mRNA therapeutics. The cap modification on engineered

mRNA is necessary to prevent immunogenicity, improve intercellular stability and translation efficiency. Thus, therapeutic mRNA is engineered to resemble mature and processed eukaryotic mRNA, including the 5' cap and the 3' poly A tail. Currently, there are only a few published LC-MS methods for detection of cap modifications. Nevertheless, these methods include labor intensive sample preparation, long analyses times and have moderate sensitivity. In the course of my dissertation, the development and optimization of a time efficient and highly sensitive LC-MS method for absolute quantification of cap modifications is presented. It includes an extensive method development, optimizing chromatographic and mass spectrometric parameters under consideration of short analysis time, low detection and quantification limits. For absolute quantification of cap modifications, an *in vitro* transcribed cap-SILIS was generated. Furthermore, limits of detection and quantification as well as the dynamic range for size and amount of macromolecules to be analyzed were determined. The high sensitivity allows for the analysis of RNA from synthetic but also from biological sources. The time efficiency is aspirational for ecologic and economic reasons, thus making this method suitable for high throughput analyses and industry.

The identification and quantification of RNA modifications is getting more important with the significance of RNA therapeutics. In this work, efficient LC-MS based tools to study the extent of nucleic acid modifications are described. Insight into the stress-dependent regulation of the genome and transcriptome of common model organisms is given and a powerful method to quantify cap modifications is presented. These techniques can be used to study nucleic acid dynamics in clinical studies but also for quality control of RNA therapeutics.

1   Introduction

1.1   Biosynthesis, structure and function of DNA and RNA

All life is based on the concatenation of deoxyribonucleic acids (DNA) being tran-scribed into ribonucleic acids (RNA), which are translated into proteins. Already in the late 1920, Griffith conducted the first experiments, showing the transformation of ge-netic information.[1] About 20 years later, Avery, MacLeod and McCarty identified deoxy-ribonucleic acids to be the genetic information carrier. In 1953, the groundwork for our science today was set by Watson and Crick with the support of Rosalind Franklin. The double helical structure of DNA was elucidated.[2] Crick expanded his work and postulated the "*Central dogma of molecular biology*" and thereby paved the way for the understanding of transcription and translation.[3]

The canonical nucleobases thymine (T), uracil (U), adenine (A), cytosine (C) and gua-nine (G) form the building blocks for DNA and RNA. While A, C and G are present in both biomolecules, the fourth base differs. Thymine is incorporated into DNA and uracil into RNA (Figure 1A, B). In contrast to uracil, thymine has an additional methyl group at position 5. Nucleobases are connected to ribose by a glyosidic bond with the C1' position of the sugar. Also here DNA and RNA differ, while nucleobases of RNA are connected to ribose, nucleobases of DNA are connected to 2'-deoxyribose (Figure 1A, B).

The pathways for biosynthesis of nucleic acids are similar in microorganisms and mam-mals and can be separated in the *de novo* and salvage pathway. In the high energy consuming *de novo* pathway, ribonucleotides are biosynthetically generated from CO_2, glutamine, aspartate, glycine, format and glucose. The energetic favorable salvage pathway, recycles bases or nucleosides from catabolism to build new nucleotides. 2'-deoxyribonucleotides are generated by deoxygenation of the respective ribonucleo-tides.[4]

In DNA and RNA, nucleotides of one strand are connected by formation of phos-phodiester bonds between C5' and C3' of adjacent nucleotides. The strand is elon-gated at the C3' hydroxyl group of the ribose by the addition of nucleoside monophos-phate, transferred from a nucleoside triphosphate under cleavage of its two terminal phosphate groups. This phosphoanhydride cleavage provides energy for the extension of the strand with new nucleotides. The direction of extension along the phosphodiester

backbone defines the 5' → 3' orientation (Figure 1A, B). DNA and RNA can be present in double strands, which is attributed to the formation of hydrogen bonds between bases. Usually, cytosine pairs with guanine and adenine pairs with thymine or uracil, which is known as Watson-Crick base pairing (Figure 1C).
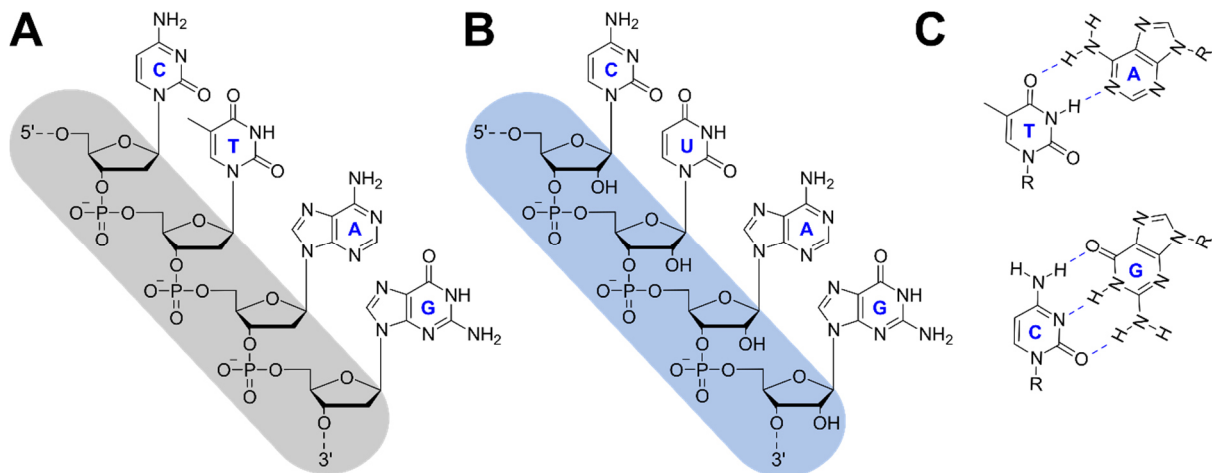

Figure 1: Structure of (A) DNA and (B) RNA, respective nucleobases in 5' → 3' orientation, cytosine (C), thymine (T)/ uracil (U), adenine (A) and guanine (G), the sugar-phosphodiester backbone is highlighted in colour (C) Watson-Crick base pairing of thymine (T) and adenine (A), cytosin (C) and guanine (G), the 2'-deoxyribose mojety is abbreviated by R

The biomolecules are further structured for condensation and/or to comply different functions. Prokaryotic DNA is double stranded and circular, it is compacted by the factor of 1000[5, 6] in the nucleoid via supercoiling and chromosomal proteins. The genome of bacteria ranges from 100 to 15 000 kb.[7] In contrast to eukaryotes, prokaryotes possess extrachromosomal DNA, called plasmids, which is also constituted in circular shape.

In eukaryotes, DNA forms a double helix by base pairing of two antiparallel strands. This double helix can assume different geometries. The physiological predominant conformation is the one described by Watson-Crick and is called B-form. Its crystal structure is illustrated in figure 2A. In a less hydrated or highly ionic environment, it is suggested that the rotation angles of the helices can differ based on the conformation of bases, and aberrant geometries can be assumed like the A- or Z-form. Further, the DNA double helix is wrapped around histone octamers, building nucleosomes (Figure 2B, C). These nucleosomes are stacked onto each other forming chromatin. Chromatin is coiled and further condensed to form chromosomes (Figure 2C). The genome of eukaryotes is up to four orders of magnitude larger than the genome of prokaryotes. The yeast genome was elucidated in 1996 and consists of almost 12 Mb, while the

human genome was analyzed in the early 2000s and reported to be composed of 6.2 Gb.[8, 9]

The genome of higher organisms is larger, as some genes are repeated but also due to expanded organism complexity. The complexity starts from cellular structure over manifold different types of cells to larger systems like organs. Despite the complexity and diversity of DNA, it has a sole function in every organism, the transmission of genetic information. This function can be subdivided into replication, which is the basis of cell division and transcription, which is the fundament for protein biosynthesis.

Another major class of biomolecules required for protein biosynthesis are RNA. There are different types of RNA, that comply various functions, and hence have different sequences, sizes and structures. In eukaryotes, mRNA is generated by the transcription of DNA into pre-mRNA in the nucleus. The pre-mRNA is synthesized by the RNA polymerase II and is further processed by splicing and the addition of the 5' cap as well as the poly A tail at the 3' end. Both of these accessory structures promote translation and protect the mRNA from degradation. Subsequently, the mature mRNA is exported into the cytoplasm to serve as a template for protein synthesis. mRNA accounts for 5% of the transcriptome in mammals[10], its length can be various and correlates with the size of the polypeptide it codes for. mRNA is known to build intramolecular base pairing interactions leading to the formation of hairpin loops. These structures play a major role in codon recognition and translational initiation.[11, 12]
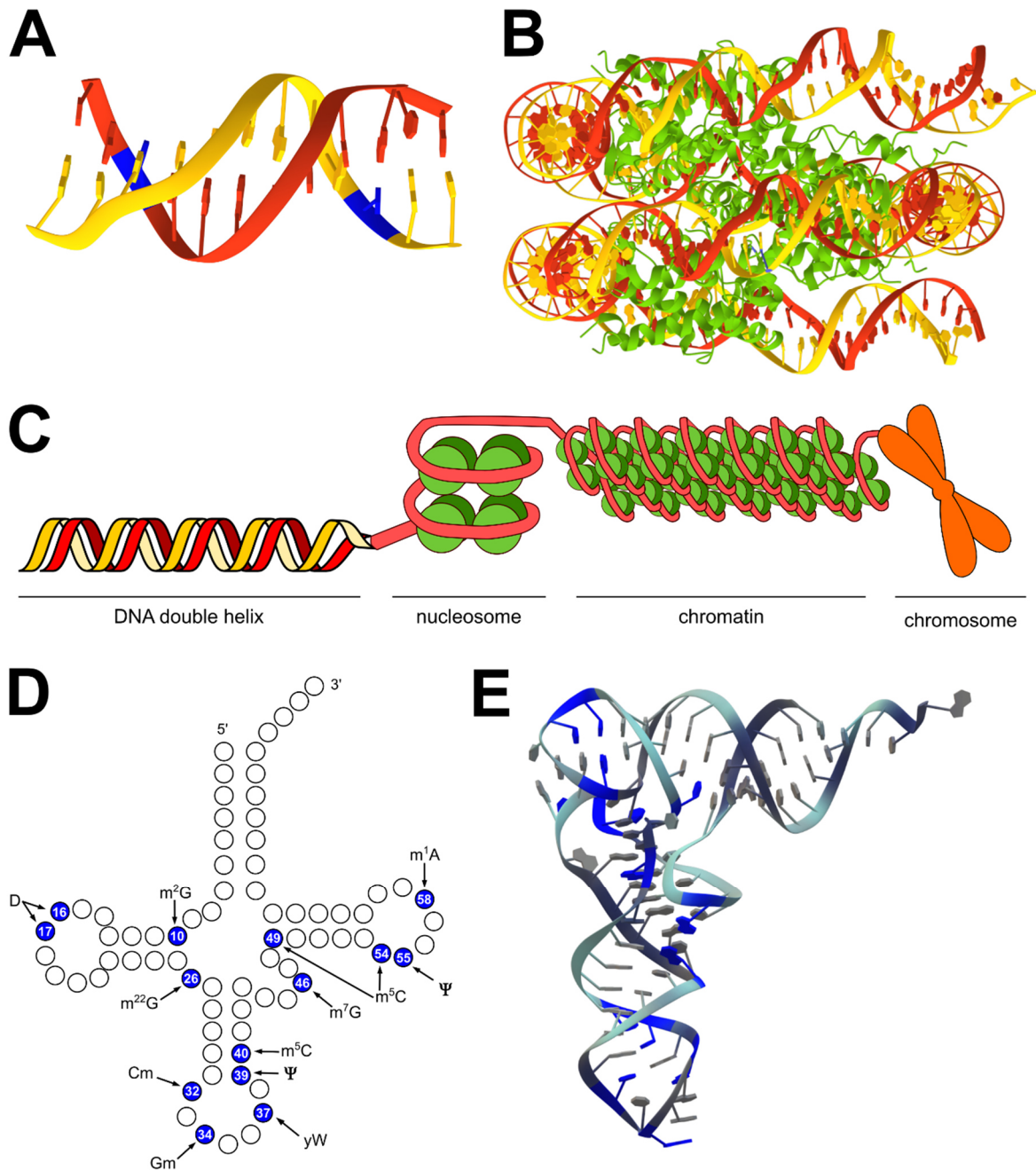
Figure 2: Illustration of DNA and RNA (A) Crystal structure of human B-DNA dodecamer, with modification m^{5}dC highlighted in blue. Illustration adapted from 4HLI in PDB[13] (B) Cryo-EM structure of nucleosome from *D. melanogaster*. Illustration adapted from 6PWE in PDB[14] (C) Scheme of DNA packing, double helical DNA is wrapped around histone octamer, forming a nucleosome. Nucleosomes are stacked to build chromatin, which is further condensed to chromosomes (D) 2D clover leave structure of tRNA^{Phe} from *S. cerevisiae*, modifications and their positions are highlighted in blue (E) Crystal structure of tRNA^{Phe} from *S. cerevisiae*, modifications are highlighted in blue. Illustration adapted from 1EHZ in PDB[15]

mRNA is decoded in the ribosome for protein synthesis. The ribosome can be divided in the large and small ribosomal subunit. The subunits consist of ribosomal proteins and ribosomal RNA (rRNA), which can be further classified. In prokaryotes the large subunit is composed of 23S rRNA with a length of 2900 nucleotides (nts) and 5S rRNA consisting of 120 nts, whereas the small subunit is built of 16S rRNA with a length of

1500 nts. The 23S, 16S and 5S rRNA are co-transcribed and subsequently released by cleavage, to assemble the ribosome together with ribosomal proteins. In eukaryotes the large subunit comprises the 28S rRNA (mammals) with a length of ~5000 nts or 25S rRNA (yeast) with a length of 3400 nts, further it includes the 5.8S and 5S rRNA. The small subunit is composed of the ribosomal proteins and the 18S rRNA with a length of ~2000 nts. In eukaryotes pre-rRNA is synthesized by RNA polymerase I and subsequently cleaved to 28S/25S, 18S and 5.8S rRNA. The 5S rRNA, which is part of the large ribosomal subunit is synthesized by the Polymerase III and brought to the ribosome. In mammals, rRNA accounts for 85% of the transcriptome.[16] Also rRNA is suggested to build intramolecular interactions, like formation of pseudoknots, hairpin and helical structures in order to arrange within the ribosome and to fulfill its function.[17-21] The mRNA and tRNA binding for peptide bond formation takes place in the cavity of the ribosome. While the mRNA is placed in the small subunit, the tRNA interacts with the large subunit. The large subunit contains three loops for tRNA interaction. One loop, the A (aminoacyl) site is occupied by an aminoacyl tRNA. Next, this tRNA is transferred to the second loop, the P (peptidyl) site, where the amino acid is transferred from the tRNA to the growing polypeptide chain. Lastly, the discharged tRNA with a free 3' end leaves the ribosome via the E (exit) site. While rRNA was long thought to have a sole function, the alignment of mRNA and tRNA, its catalytic role as peptidyl transferase was proven in 2000.[22]

Amino acids for protein biosynthesis are brought by transfer RNA (tRNA). Pre-tRNA is synthesized in the nucleus with extended 5' end 3' ends. The 3' end of mature tRNA ends with the base triplet CCA, on which the aminoacyl group is covalently bound. The size of tRNA ranges from 74 – 93 nts in *E. coli* and 76 - 90 nts in eukaryotes.[23, 24] The bases of tRNA are pairing to form four stems and three loops, which can be recognized in the clover leaf structure (Figure 2D). A prominent loop is the anticodon, which comprises the base triplet in position 34, 35 and 36. This base triplet pairs with the codon of the mRNA for decoding. In contrast to other types of RNA, tRNA is highly modified (Figure 2D, E; modifications highlighted in blue), with up to 13 modifications per tRNA molecule.[25] Modifications in tRNA attribute to the specific L-shape folding (Figure 2E) and to recognition by aminoacyl-tRNA synthetases, for correct amino acid charging to the cognate tRNA. In mammals, tRNA accounts for 10% of the transcriptome.[10]

Moreover, there are many more RNA types, like small interfering (siRNA), micro (miRNA) or small noncoding (snRNA) RNA, most of which are involved in the transcriptional and translational regulation.

However, RNA can also serve as genomic biomolecule. Unlike in the aforementioned organisms, the genome of viruses can consist of either DNA or RNA. Thereby, RNA viruses comprise 70% of all viruses.[26] To date two RNA viruses convulsed the human population. The Human Immunodeficiency Virus (HIV), which was perceived in 1981 for the first time, and just recently in 2019 the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), which caused a global pandemic leading to high mortality.[27] However, SARS-CoV-2 is not the first highly pathogen coronavirus. In 2002 the severe acute respiratory syndrome coronavirus (SARS-CoV) and in 2011 the Middle East respiratory syndrome-related coronavirus (MERS-CoV) were discovered. But there are many more RNA viruses, that we encounter, like Hepatitis-C-Virus (HCV), Ebolavirus (EV) or influenza virus. In contrast to other organisms, viruses do not possess an autonomous replication machinery, but exploit the host proteins. Further, the viral genome may be single-stranded (ss) or double-stranded (ss). Apart from that, the genome can occur in linear or circular shape and can be segmented, where each strand codes for a certain number of genes, or non-segmented, where one strand codes for all genes. Single-stranded genome viruses, which are prevalent, can be further classified into positive and negative-sense RNA viruses. The genome of positive-sense viruses resembles mRNA and thus can be directly translated by the host cell proteins, whereas negative-sense RNA needs to be transcribed into the complementary positive-sense RNA for translation. In contrast to DNA viruses, RNA viruses have high mutation rates.[28] While the DNA polymerases have a proofreading capability, RNA polymerases are lacking such a function, leading to deviating sequences and provoking mutations. Apart from that, the length of the viral genome can range from 1.7 kb, as for the hepatitis delta virus,[29] to up to 30 kb, as for SARS-CoV-2[30]. The viral genomic structures are quite divers, due to the great variances in size, sequence and base pairing. However, its folding and structure is of high relevance in any case to fulfill its function.[31]

1.2 Chemical modification of nucleic acids

Conrad Waddington introduced the term epigenetics. Epigenetics comprise a multitude of processes with heritable effects on gene expression without changing the sequence of bases.[32] The epigenome is controlled by chemical (de-)modifications of histones, non-histone proteins and nucleic acids. Up to today, 17 DNA modifications were reported in bacteria und eukaryotes.[33] In analogy to epigenetics, the field of RNA modifications is called epitranscriptomics. The canonical nucleosides can be modified with divers chemical groups and 170 RNA modifications were discovered to date.[23] In order to understand the location and type of modification the nomenclature is going to be introduced in the following. The atoms of nucleic acids are numbered in a uniform nomenclature. In the beginning it needs to be differentiated between nucleosides of DNA and RNA. The canonical bases of DNA are abbreviated with an additional d in front of the abbreviation for the base, except of thymine (dC, dG, dA, T). The numbering of purines (adenine, guanine) starts with the heteroatom the furthermost from the glyo-sidic bond and ends with the nitrogen at the glyosidic bond. Pyrimidines (thymine, uracil, cytosine) are numbered according to the IUPAC nomenclature, starting with the nitrogen atom at the glyosidic bond and further continuing to yield low numbers for the heteroatoms. Also the ribose is numbered according to the IUPAC nomenclature, in accordance to the fisher projection. For differentiation of atoms from base and ribose, the latter is specified by an apostrophe (Figure 3). The phosphate is connected to the ribose via the C5' or C3' atom and is termed ac-
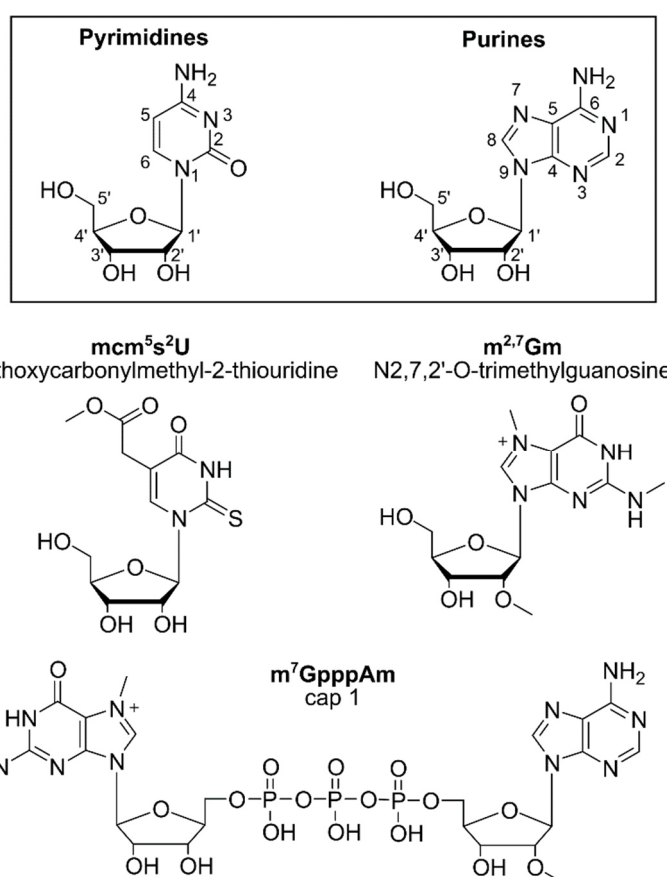

Figure 3: Nomenclature and examples for modification of nucleic acids. The numbering for pyrimidines and purines is displayed at the top. Exemplary modifications are illustrated in the center. The cap modification cap 1 is depicted on the bottom.

cordingly. For the designation of modified nucleosides, the chemical groups are abbreviated *(e.g.* m for methyl, ho for hydroxyl, s for sulfur, ac for acetyl, n for amino). A superscript number between the abbreviation of the group and the base indicates the position of the modification on the base. In case of multiple groups, the numbers are listed one after the other. If the modification is on the ribose, the abbreviation is set after the base. In contrast to modifications of base or ribose, the cap modifications are simply termed cap 0 or cap 1, according to Furuichi *et al.*[34] A distinction is made between the 2'-O-methylation of the first nucleoside of the sequence, defining m^7GpppN as cap 0 and $m^7GpppNm$ as cap 1.

The first discovered DNA modification m^5dC was found to regulate gene expression.[35] This modification is the most predominant in mammals and occurs at CpG dinucleotides. DNA methyltransferase 1 (DNMT1) is responsible for the methyl group transfer from S-Adenosyl methionine (SAM) to the nucleobase.[36] DNA methylation regulates genome organization, genomic imprinting, reprogramming, stability, cell differentiation, silencing, DNA repair and RNA splicing.[37-39] Further well studied DNA modifications are the oxidation products of m^5dC. So, hm^5dC was found to have a great impact on early development, while f^5dC and ca^5dC affect transcription.[40, 41] The most predominant DNA modification in prokaryotes is m^6dA. It is involved in the regulation of gene expression related to repair[42], cellular defense[43] and virulence[44]. In lower eukaryotes, m^6dA is involved in the regulation of gene expression[45], cell cycle and nucleosome positioning[46]. It is suggested that the abundance of m^6dA decreases with increasing complexity of the eukaryotic genome. While 0.3% m^6dA/A were detected in *C. elegans*, the levels in murine liver tissue were three orders of magnitude lower.[47-49] Further it was shown that levels of m^6dA decrease during embryogenesis of zebra fish and pigs and the abundance is five to six times lower in adult tissue than in oocytes. Thus its importance in development and differentiation is suggested.[50] Furthermore, descending levels of m^6dA have been correlated to cancer in human cell culture recently.[51] The distribution of modifications varies appreciably among species and is highly dynamic.[52, 53] The contribution of modifications in disease has been shown numerous times, like the involvement of m^5dC in cancer[54], cardiovascular[55], neurological[56] and aging related diseases[57], as well as metabolic[58] and reproductive disorders[59].

Also RNA modifications comply various functions, like structure stabilization, enzyme recognition and translation fidelity.[60, 61] Modifications alter the interaction of base pairs

in terms of base stacking and ionic effects,[62, 63] which can result in cancer or neurological diseases. However, the lack of modifications can lead to translational defects and can cause mutations. In 1951, the first and most abundant modification pseudouridine (Ψ) was discovered in cellular RNA.[64] Most divers and numerous modifications occur in tRNA with 13 modifications per molecule in average (10% -15%), contributing to the formation of its unique structure, efficiency and fidelity of coding as well as determination of isoacceptor identity.[25, 65] In rRNA, the most abundant modifications are 2'-O-methylations, besides Ψ and a few base methylations. In human rRNA up to 210 modifications sites were reported, resulting in a prevalence of 3%.[25] Besides the internal mononucleotidic modifications, the maturation of mRNA is entailed with the decoration of the 5' end with a cap and of the 3' end with a poly A tail. Both accessory structures protect the mRNA from degradation and promote translation. The cap modification is placed co-transcriptional in the nucleus and is the first step of mRNA processing.[66] The mRNA cap is an evolutionary and functional conserved structure. It was first described by Furuichi *et al.* in 1976[67] and its existence in virus[68], yeast and mammals[69] was confirmed later. In prokaryotes, mRNA begins with a 5' triphosphate purine, while the cap of eukaryotes is more complex. In eukaryotes the cap possesses an unusual structure, m^7G is connected to the first nucleotide of the sequence with a 5' to 5' triphosphate linkage and thus differs from the usual 5' to 3' directionality. Usually, the first encoded nucleotide has a 2'-O-methylation. Interestingly, viruses tailor their cap modification to match the host's cap, irrespective of their own genomic structure or replication state. The resemblance in cap modification proceeds the replication of the viral genome and aids to evade the host's immune response. Apart from cap modifications, the viral genome can be decorated with internal modifications too. The most prevalent modifications in this context are m^6A, m^5C, Ψ, I and 2'-O-methylations.[70] But these modifications are subject to dynamics and changes during infection, thus it was difficult to bring out their function for a long time. By now, few features could be determined, like the effect on viral replication, gene expression and the modulation of the innate immune system.[70, 71] However, it is expected that there may be many more functions, which remain to be elucidated.[72]

1.3 Dynamics of nucleic acid modifications

1.3.1 Endogenous modification

The most common modification on 2'-deoxynucleotides is the base methylation. This modification is introduced by the family of DNA methyltransferases (DNMT). DNMTs catalyze the transfer of a methyl group from the methyl donor S-Adenosyl methionine (SAM) on to the nucleobase. In mammals this family comprises 5 proteins. DNMT1 is responsible for the formation of m^5dC. DNMT2 was found to be barely active on DNA but to rather introduce modifications to the anticodon of tRNAs.[73, 74] DNTM3A and DNMT3B are called *de novo* methyltransferases and are involved in methylation during early development.[75] DNMT3L has no catalytic activity, but stimulates the *de novo* methylation by DNMT3A.[76] Also in *E. coli* DNA methylation is SAM-dependent and catalyzed by the two enzymes Dam and Dcm, resulting in the formation of m^6dA and m^5dC respectively.[77-79] The existence of m^5dC in the yeast genome is controversially discussed[80, 81] and no evidence is found for the existence of other genomic modifications.

RNA modifications are more divers and so are their enzymes. In the context of RNA, modification placing enzymes are called writers. The first discovered modification Ψ is incorporated by pseudouridine synthase (PUS). Ψ is a constitutional isomer of uridine, generated by cleavage of the C-N glycosidic bond of the base, rotation around the N3-C6 axis and the rebuild of a C-C glycosidic bond. ADAR is executing the deamination of adenosine to inosine, better known as A-to-I editing.[82, 83] There are enzymes introducing certain modifications, like methyltransferase-like (METTL) protein family in mammals, which catalyze the SAM-dependent methylation of RNA. The family of METTL proteins comprises numerous members, that methylate different types of RNA at various positions. Thus m^3C is formed by METTL8 in mRNA[84] and by METTL6 in tRNA[85]. A well-studied protein complex is METTL3/14, which forms a heterodimer complex to generate m^6A in mRNA and non-coding RNA.[86] The very prominent modification m^7G is formed by METTL1/WDR4.[87] There are also enzymes specifically modifying one type of RNA, like the tRNA methyltransferases, they are abbreviated with 'Trm' or 'Trmt' followed by a number (*e.g.* Trm8) or a letter (*e.g.* TrmD), depending on the enzymes kingdom of origin. In human, Trmt61A and Trmt61B introduce m^1A into tRNA. While in *S. cerevisiae*, Trm6/Trm61 catalyze the formation of m^1A[88] and Trm140 was found to form m^3C[89, 90]. Trm8 and its essential co-factor Trm82 catalyze the SAM-

dependent formation of m^{7}G in tRNA of *S. cerevisiae* and TrmB in tRNA of *E. coli*.[87, 91] In contrast to mammals and yeast, the formation of m^{1}A and m^{3}C in *E. coli* is limited to damage-induced methylation.[92]

A mRNA specific modification is the cap, introduced at the 5' end of nascent mRNA. The γ phosphate at the 5' end of the transcript is removed by RNA triphosphatase and is capped by the addition of GMP, transferred from GTP by guanosyltransferase. Further, the SAM-dependent methylation at position N7 is driven by the mRNA guanin-N-7-methyltransferase (RNMT) building cap 0.[66] In higher eukaryotes cap 1 can be formed, which is characterized by 2'-O-methylation of the first nucleotide of the sequence. In human, this methylation is conducted by the SAM-dependent methyl transferase hMTr1.[93] While Viruses usually exploit the host enzymes, they were found to be capable of encoding their individual capping enzymes. In SARS-CoV-2 16 non-structural proteins (nsp) were identified, some of which implement the capping of the viral genome. The heterodimer nsp10/nsp14 builds a complex with SAM to implement the N7-guanine methylation at the 5' end of the viral genome.[94] The presence of m^{7}G on the cap is a prerequisite for the following 2'-O-methylation of the first nucleotide of the sequence by the heterodimer nsp10/16.[95, 96]

1.3.2 Damage-induced modification

Nucleosides can be modified endogenously by enzymes, but also as a consequence of direct chemical modification of nucleic acids induced by exogenous factors, like UV radiation[97] or chemicals[98, 99], resulting DNA and RNA damage. DNA damage can lead to carcinogenesis[100-102] while RNA damage can cause defects in protein synthesis, which can induce neurodegenerative disorders[103-105]. Nevertheless, damage-induced modifications do not only bring about disadvantages. Goodman and Gilman pioneered in cancer therapy, describing the effect of chloroethylamines, better known as nitrogen mustard, on Hodkin's diease, lymphosarcoma and leukemia.[106] Exposition to chloroethylamines leads to cross linking of guanine to guanine or to adenine and results in blocked replication.[107] Today we still benefit from these findings, as alkylating agents are still used as cytostatics.

However, the introduction of modifications can impair the binding of functional proteins. Further, the methylation of position N3 of pyrimidines and N1 of purines prevents the regular Watson-Crick base pairing. Taken together the undesired modifications can cause cellular malfunction.[108-110] Common damage-induced modifications are the result of oxidation and methylation of nucleobases. Oxidative stress can arise from UV radiation, ionizing radiation or heat exposure and provokes the enhanced production of reactive oxygen species (ROS). As a consequence of increase in ROS level, nucleobases can be oxidized at the indicated sites in figure 4, and modifications like 8-oxo-guanine (8-oxo-G) und 5-hydroxy-cytosine (ho5C) are formed.[108, 111] Most exogenous methylation occurs through direct attack of the methylating agents on the nucleophilic positions of the nucleobases (Figure 4). Thereby, the methylation site is dependent on the reaction type. S_N1 reactions usually result in methylation of oxygens and nitrogens, whereas S_N2 reactions yield methylated nitrogens.[112] Prevalent methylation damage products are m7G, m6A, m3C and m1A.[112] However, the majority of these modifications occur also endogenously in the genome and transcriptome of prokaryotes and eukaryotes. As a consequence, it is difficult to distinguish between enzymatic and damage methylation with most detection techniques.

Figure 4: Damage sites of nucleobases, blue arrow for oxidation sites and red arrows for methylation sites, the ribose is abbreviated with Rib.

1.3.3 Active demodification

All organisms have developed various pathways to maintain the integrity of their genome and transcriptome. Numerous repair pathways have been postulated over the last decades, such as direct reverse base excision repair, mismatch repair, nucleotide excision repair, homologues and non-homologues end joining pathway.

The first active demethylation of the epigenome in mammalian cells was described for TET (ten-eleven translocation) family of mononuclear nonheme Fe^{II}-dependent dioxygenases which stepwise oxidize m5dC to hm5dC, f5dC and ca5dC.[113-116] These oxidized

products can be converted to cytosine through the base excision pathway by various proteins.[117]

Direct reversal repair in *E. coli* is conducted by the methyl transferase Ada. Ada demethylates the phosphodiester backbone and m^6dG at DNA lesion sites by transfer of the methyl group onto its own cysteine residue.[118] Besides its direct demethylation activity, it serves as initiator for adaptive response (Ada response) which provokes the expression of the repair enzymes AlkA and AlkB.[118, 119] Repair enzymes have the function of removing modifications in order to recover canonical nucleosides. In the context of RNA damage and repair these repair enzymes are referred to as eraser.

AlkA removes methylations on position N3 of adenine and positions N3 and N7 of guanine in bacterial DNA through glycosidic bond hydrolysis.[120] Subsequently, the apurinic/apyrimidic site is repaired by the base excision pathway.[119] 20 years ago, the alpha-ketoglutarate-dependent dioxygenase AlkB was found to repair methylation damage on DNA.[121, 122] AlkB oxidizes the methyl groups of m^1A and m^3C to hydroxymethyl derivatives, which dissociate to formaldehyde and recover adenine and cytosine in DNA and RNA.[121, 122] Additionally, AlkB was reported to demethylate m^1G[123], m^3T[123, 124] and m^6A[125] *in vitro.* Further, its ability to reverses ms^2C to s^2C by direct sulfur demethylation *in vitro* and *in vivo* in tRNA of *E. coli* has been shown.[126]

Up to now, 9 homologues of AlkB were described in human, namely AlkBH1-8 and FTO. Amongst these, demethylation activities on nucleobases were observed for AlkBH1, AlkBH2, AlkBH3, AlKBH5 and FTO. It was discovered that AlkBH1 oxidizes m^6A to hm^6A *in vitro* and *in vivo*.[127] Further, the substrates m^1A[128], m^3C[129, 130] and m^5C[131] were identified for AlkBH1 *in vitro* and *in vivo*. The functions of AlkBH2 and AlkBH3 were determined to be similar to those of AlkB in *E. coli*, oxidizing m^1A and m^3C.[132] It is indicated that AlkBH2 is prone to repair double stranded nucleic acids, whereas AlkBH3 favors single stranded nucleic acids and therefore also acts on RNA.[133] In 2013 the He lab identified the demethylating activity of AlkBH5 on m^6A in mRNA *in vitro* and *in vivo*.[134] In contrast to the other deoxygenases the formation of the intermediate oxidation products hm^6A and f^6A was not confirmed, potentially due to divergences in protein active site.[134, 135] In RNA FTO was shown to oxidize m^6A *in vitro*[136, 137] and *in vivo*[137], yielding hm^6A and f^6A, which can decompose to adenine under physiological conditions.[135] Further, FTO was found to selectively and preferentially demethylate the base of m^6Am rather than m^6A, at the first transcribed nucleotide

adjacent to the m^{7}G cap. Besides cap demethylation, the whole cap can be removed by *e.g.* DCP2, which hydrolyzes capped mRNA, releasing m^{7}GDP and initiating mRNA decay.[138] However, transcripts can be spared by the incorporation of further modifications, so m^{6}Am was reported to determine mRNA stability as a consequence of its resistance to the mRNA decapping enzyme DCP2.[139] Also nudix hydrolase 16 (NUDT16) is an RNA binding and decapping enzyme. It has diphosphatase activity and catalyzes the cleavage of cap modifications, leaving solely a 5' monophosphate on the transcript.[140]

Even though, yeast is amongst the most studied eukaryotes, less is known about its active demethylation activities on the genome and especially on the transcriptome. In *S. pombe* two AlkB homologues (Ofd1, Abh1) have been found, but neither oxidation nor demethylation of nucleic acids could be ascribed to their functions.[141, 142] Shivange *et al.* proposed the repair activity of Tpa1[143], an orthologue to Ofd1[144]. It was described to act comparable to AlkB[143], however this was disputed later[145]. A study in 2002 identified 40 genes in yeast that showed increased sensitivity towards MMS.[146] One of them is Mag1, the homologue of bacterial AlkA. Mag1 removes methylation damage on m^{7}dG and m^{3}dA in DNA *in vivo*.[147] Additionally, it has been presented that Mag1 recognizes the known substrates m^{1}dA and m^{3}dC of AlkB *in vitro*.[145] Besides Mag1, the apurinic endonuclease/3'-diesterase Apn1 is reported to repair DNA methylation through the base excision repair pathway.[148, 149] Most of the proteins involved in DNA repair have Fe-S clusters as their co-factors. Met18 composes a subunit of the cytosolic Fe-S protein assembly machinery.[150] Interestingly, the sensitivity of Met18 towards MMS was revealed.[146] Met18 is involved in oxidative stress response[151, 152], chromosome segregation[153], telomere length maintenance[154] and RNA polymerase II transcription,[155] it is also involved in nucleotides excision repair. In plants Met18 was identified to be involved in active DNA demethylation.[152]

1.3.4 tRNA modification dynamics during stress and growth in *S. cerevisiae*

As already outlined, the spectrum of RNA modifications is highly divers. The different types of modifications and the populations of RNA they are incorporated to were already introduced. Besides these variations, the abundance of modifications can alter within the same RNA molecule. The levels of RNA modifications are dynamic and adapt to environmental changes. The adaption of tRNA modification levels in the model organism *S. cerevisiae* during stress and growth was shown.[156-158] In 2010, Chan *et*

al. found signature changes in small RNA (<200 nts) of *S. cerevisiae* in response to different toxicants, like the increase of m^5C, Cm and $m^{22}G$ upon H_2O_2 exposure or the increase of m^7G upon MMS exposure.[156] Up to 90% of the purified small RNA consisted of tRNA, thus these findings were interpreted as stress-induced changes in tRNA modifications. The results were evaluated based on changes in modifications densities in comparison to untreated control samples, which derogates the interpretation of dynamics for modifications which occur as damage product but as well as endogenously (*e.g.* m^7G). The authors suggest that changes in modification densities may arise from modification of existing tRNA transcripts or changes in tRNA copy number.[156]

Furthermore, it was shown that tRNA modification densities do not only adapt to stress but also vary at different growth phases.[157, 158] My colleague Dr. Matthias Heiß succeeded in tracing tRNA modifications at different growth phases by the application of NAIL-MS (**N**ucleic **A**cid **I**sotope **L**abeling coupled **M**ass **S**pectrometry), which is based on changing metabolic labeling during different growth phases.[158] The classification of original and new transcripts as well as absolute quantities of modifications could be determined. Thereby, it was revealed that the supposed static level of m^7G can be traced back to decreasing levels of m^7G in original tRNA transcripts and increasing levels of m^7G in new tRNA transcripts as well as post-methylation of original tRNA.[158] It can be summarized, that tRNA modifications are highly dynamic and NAIL-MS offers a tool to investigate these dynamics and the underlying mechanisms.

1.4 Instrumental analysis of DNA and RNA modifications

1.4.1 Overview of detection methods

As described in the previous chapter DNA and RNA modifications can be manifold and dynamic, accounting for the phenotype, characterizing diseases and building the key to new emerging personalized medicine. Thus, powerful tools are needed to accurately determine the identity, sequence and quantity of nucleic acids and their modifications.

Over the last decades, different tools were developed to detect, localize and quantify nucleic acids and to understand their function. The analytical methods exploit the differences in physico-chemical properties of the various modifications. Below, sequencing, nuclear magnetic resonance (NMR) spectroscopy as well as mass spectrometry for analysis of nucleic acids are briefly expounded.

In 1977, Frederick Sanger published his work on a "*A new method for determining nucleotide sequences in DNA*" and thus established a method for DNA sequencing and put the ground stone for current sequencing techniques.[159] Over the last decades, more time and cost efficient sequencing techniques were developed. After the Sanger or first generation sequencing, the first automated second generation sequencing was established[160], followed by next-generation sequencing of second and third generation, which are the new frontiers in DNA and RNA modification analysis. In the early 2000s, the human genome was elucidated by sequencing in the framework of the Human Genome Project.[9] While this technique offers a tool to determine the base sequence the presence of modifications remains elusive. But this obstacle can be circumvented by chemical or antibody treatment, functionalizing or labeling specific modifications and hence allowing a direct readout of modifications.

While sequencing sheds light onto the sequence of nucleic acids, it leaves the location of modifications and the structure of the RNA molecule aside. Here NMR becomes important, already in the early 1970s tRNA was studied by NMR.[161-163] With this, structural information can be gained besides the detection of modifications. The three-dimensional folding and conformational stability can be examined[163, 164] as well as the impact of specific modifications on stability[165, 166] and the anticodon stem-loop conformation can be investigated[167-169].

The third technique for nucleic acid analysis is mass spectrometry. A broad spectrum of information can be gained with mass spectrometry. Thereby it can be distinguished

between three approaches. First, top-down-MS, here non-hydrolyzed RNA is analyzed and the location of modifications can be determined within the sequence context. The application of top-down-MS succeeded in the analysis of a full-length tRNA.[170, 171] Further, bottom-up-MS, is applied to partly hydrolyzed RNA and mass mapping can be conducted, if the sequence is known. This approach even enabled the complete mapping of mRNA, rRNA and tRNA.[172-177] Finally, nucleoside-MS of fully hydrolyzed RNA is conducted, where besides identity and quantity dynamics of nucleic acid modifications can be examined.[178]

These tools have diversely pronounced strengths and weaknesses that compensate for each other. On the basis of the diverse advantages of every technique, it is evident that the application of orthogonal methods is recommended in any case. The functionality, strengths and weaknesses of the listed methods are compared intensively in a review written by me.[179]

Instrumental analysis of RNA modifications, **Yoluç Y.**, *et al.*, Crit Rev Biochem Mol Biol., 2021

Declaration of contribution: The chapters "Advances in chemical labeling of modified RNA in the context of instrumental analysis", "Nucleoside MS" were researched and written by me. The manuscript was completed by cooperation of all authors.

1.4.2 LC-MS/MS for quantitative analysis of nucleic acids

Previously, three techniques were introduced for analysis of nucleic acids. In the following the quantitative mass spectrometric analysis of nucleic acids is going to be further elaborated. The soft ionization techniques of matrix assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) are especially suitable for the analysis of nucleic acids. MALDI is generally utilized for the ionization of oligonucleotides, however the resulting ions tend to be instable, as just a single charge is transferred.[180, 181] Whereas, ESI can transfer multiple charges and excels in its compatibly with capillary electrophoresis (CE) or high performance liquid chromatography (HPLC). Electrophoresis and chromatography separate the nucleotides/nucleosides, resulting from RNA hydrolysis, prior to ionization. The separation is advantageous for mixtures of analytes, as it facilitates the data evaluation and ameliorates ionization efficiency as well as sensitivity of the method. After separation and ionization, analytes enter the mass spectrometer. The detection is based on the mass-to-charge ratio (*m/z*) of each

analyte. Different types of mass spectrometers allow for different mass analysis and detection. While a time of flight (TOF) detector or Orbitrap can record spectra of high resolution, a triple quadrupole mass spectrometer (QQQ) excels in high sensitivity, which is of great importance for the detection of low abundant analytes. In the following the function and optimization of each section for quantitative LC-MS/MS analysis of nucleic acids is going to be described.

1.4.2.1 Chromatography

Chromatography is a technique to separate mixtures to their single components. All chromatographic systems consist of a stationary phase, with which the analytes interact to a different extent and are retained as well as a mobile phase, which carries the molecule through the system. The chromatographic mechanisms can be classified in adsorption and partition chromatography. Adsorption chromatography is based on the direct interaction of the analyte with the surface of the stationary phase. The analytes are separated according to their extent of interaction with the stationary phase. In partition chromatography, the system consists of two immiscible liquids, of which one is immobilized on an inert carrier and acts as stationary phase. The retention is based on the difference in solubility of analytes in the two liquids. Already in the 1950s, the presence of m^5dC in nucleic acids could be confirmed with paper chromatography.[182] While nucleosides can be also separated with thin-layer chromatography[183-188] or gas chromatography[189-192], the separation with liquid chromatography[158, 193-195] is the most common method today. There were also few approaches to separate nucleotides with CE, which is quite feasible, due to their negative charge.[196, 197] However, CE did not find its advent into nucleic acid analysis as the separation is based on size distribution, and single nucleotide/nucleoside sizes vary only slightly.[197]

For evaluation of the chromatographic separation, different chromatographic parameters have to be examined, which are going to be introduced in the following. For this purpose, an exemplary chromatogram is shown in figure 5.

The void time t_0 is the elution time of an analyte, which is not retained on the stationary phase. The void time of the column can be calculated based on the empty column volume, the pore and interstitial volume of the packing. However, it is suggested to determine the exact void volume with an analyte that is inert to the stationary phase, generally uracil is utilized for reversed phase (RP) and hexane for normal phase (NP)

columns. This approach includes the void volume of the whole chromatographic system and not only the column and thus is more precise. The retention time is described as the time an analyte takes to pass through the column. Here one needs to distinguish between the total retention time t_R and the netto retention time t'_R, which consists of the difference between the total retention time and the void time. Apart from that, one can determine the peak width at the baseline (w_h) as well as at 1/10 ($w_{0.1}$) or half maximum of the peak ($w_{0.5}$). Moreover, the leading (A) and trailing (B) widths at 1/10 maximum to the center line at peak maximum can be determined to evaluate the peak shape.
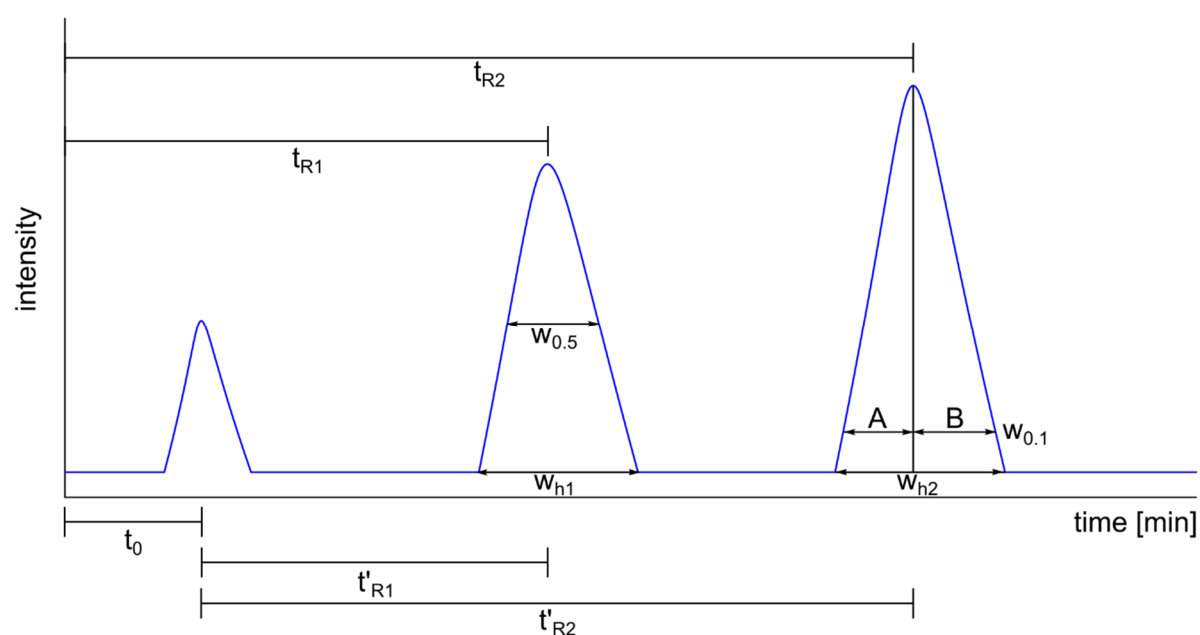

Figure 5: Exemplary chromatogram for explanation of chromatographic parameters t_0 = void time, t_R = total retention time, t'_R = netto retention time, w_h = peak width at baseline, $w_{0.1}$ = peak width at 1/10 maximum, $w_{0.5}$ = peak width at half maximum, A = leading width at 1/10 maximum to center line at peak maximum, B = trailing width at 1/10 maximum to center line at peak maximum

The retention factor (k) determines the retention of an analyte on the stationary phase and is calculated by the quotient of t'_R and t_0 (Equation 1).

$$(1) \qquad k = \frac{t_R - t_0}{t_0}$$

The retention factor is independent of the flowrate but can be adjusted by solvent strength and thus the mobile phase composition (aqueous/organic phase). Evident from equation 1, long retention times entail large retention factors. An optimal retention factor should be between 2 and 8.[198]

The selectivity (α) is a measure for the chemical separation of two peaks and is determined by the ratio of the respective netto retention times, where the later eluting compound (2) builds the dividend and the sooner eluting compound (1) the divisor (Equation 2).

$$\alpha = \frac{t'_{R2}}{t'_{R1}}$$

(2)

The selectivity is always larger than one and high selectivity indicates an appropriate separation of peaks. If α is one, the separation is poor and peaks are co-eluting. As the selectivity is based on chemical separation, the following parameters can be adjusted to optimize the separation. Firstly, the chemistry of the stationary phase itself, hence the physico-chemical interaction of analyte and stationary phase alters the selectivity, just as the utilization of different mobile phases. Also the pH of the mobile phase contributes to the selectivity, as analytes and stationary phase can be differently charged at different pH. Besides the pH the mobile phase can be modified by its composition but also by the addition of additives, like ion pairing reagents. Further, the selectivity is impacted by the temperature, but only if the retention mechanisms of the analytes differ. If the retention mechanisms of analytes are the same, *e.g.* due to structural similarity, the temperature cannot be used as a tool for selectivity optimization.

The column performance can be evaluated by the theoretical plate number (N), which is a measure for peak dispersion. This measure is adapted from the Nobel prize winners Martyn and Synge, who build the analogy between distillation and chromatography in 1941 and described the column efficiency with theoretical plates.[199] A plate characterizes the hypothetical stage for the analyte to reach the equilibrium between stationary and mobile phase. The quality of separation increases with the number of theoretical plates. The theoretical plate number is determined by the retention time and the peak width at half maximum ($w_{0.5}$) (Equation 3).

$$N = 5.54 * \left(\frac{t_R}{w_{0.5}}\right)^2$$

(3)

In first instance the theoretical plate number is dependent on the column specifications, namely its length, diameter and particle size. N increases proportional to the column length, but decreases with larger column diameter and particle size. Further the flow rate, temperature and viscosity of the mobile phase can impact N. While N decreases

with augmenting flow rate and viscosity, it increases at higher temperature, which results in reduced viscosity by implication.

A parameter derived from the theoretical plate number is the plate height (H). Martin and Synge described the plate height as thickness, the thicker or rather wider the plate, the lower the separation efficiency.[199] Briefly, H is a measure of column efficiency per unit length of the column. In contrast to the theoretical plate number the plate height is desired to be rather small. The plate height is the quotient of column length and theoretical plate number (Equation 4).

$$(4) \qquad\qquad H = \frac{L}{N}$$

The correlation of the plate height and the linear mobile phase velocity is defined by the van Deemter equation (Figure 6, Equation 5). The van Deemter equation describes a hyperbolic function, predicting the optimum velocity for maximum separation efficiency.

$$(5) \qquad\qquad H = A + \frac{B}{u} + C * u$$

The three summands characterize the motion or physical interaction of analyte molecules with the particles of the stationary phase and represent different factors to contribute to peak dispersion. The A-term (Eddy-diffusion/dispersion term) is independent of the flow rate (u). It describes the turbulent motion of analytes through the column, attributed to heterogeneity in particle size and geometry. The B-term (diffusion coefficient) defines the longitudinal motion of analyte molecules along the column in consideration of molecular diffusion and the resulting peak dispersion. The B-term is inversely correlated to the flow rate and is dependent on the mobile phase's temperature and viscosity. The C-term describes the resistance to mass transfer between mobile and stationary phase and is directly proportional to the flow rate. It describes the difference in motion of analytes that penetrate further or less into a particle pore and thus proceed slower or faster with the elution.
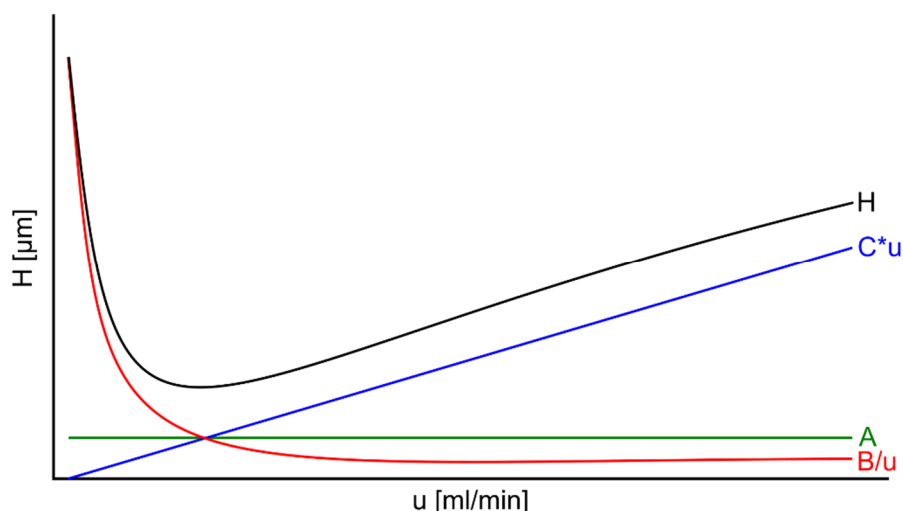
Figure 6: van Deemter Plot, A-term is visualized in green, B term is shown in red, C-term is depicted in blue while the respective plate height is extrapolated in black

Furthermore, the resolution (R_s) should be introduced which is a measure to validate the quality of separation of two peaks. The optimal resolution allows for baseline separation of adjacent peaks and is dependent on the difference in retention time and the sum of peak width at base line (w_h) (Equation 6).

$$(6) \qquad R_s = \frac{1.18 * (t_{R2} - t_{R1})}{w_{h1} + w_{h2}}$$

For ideal gaussian-shaped peaks with comparable height, a resolution of 1.5 is sufficient, whereas for non-symmetric peaks the resolution should be at least 1.8-2.[198] The resolution is strongly dependent on the previously introduced retention factor (k), selectivity (α) and theoretical plate number (N) and can be defined by an additional term (Equation 7).

$$(7) \qquad R_S = \frac{1}{4}\sqrt{N} * \frac{k_2}{1 + k_2} * \frac{\alpha - 1}{\alpha}$$

The strongest factor in this term is the selectivity, as it describes the variances in interaction of two analytes with the stationary phase under given conditions. The selectivity is optimized empirically by adjustment of mobile phase or temperature. If the resolution cannot be optimized by selectivity, it can be improved by ameliorated separation efficiency and thus an increase in the theoretical plate number, which can be achieved by adjustment of flow rate, mobile phase viscosity and with a longer column or a column of small homogenous particles. Lastly, the resolution can be optimized by adjustment of retention factor, this can be easily achieved by changes in mobile phase.

The resolution can be further increased by narrowing the peak dispersion. Ideally, peaks are narrow and symmetric. The peak shape can be evaluated by the asymmetry factor (As), which is described by the ratio of the trailing peak half at 1/10 maximum to center line at peak maximum to the leading peak half at 1/10 maximum to center line at peak maximum (Figure 5, Equation 8).

$$As = \frac{B}{A}$$

(8)

For symmetric peaks the factor equals 1. Peaks with a factors of >1 can be characterized as fronting peaks, and if the factor is <1, the peaks are tailing. Fronting can be a consequence of column overloading. Tailing can be caused by multiple factors, besides column overloading it can result from column aging or strong interactions of the analyte with the stationary phase. Column overloading can be avoided by reduction of sample amount or injection volume. The performance of an aged column can be improved by flushing with organic solvent to restore it or exchanging it by a new one. Further, the pH, ion strength or the proportion of the organic solvent of the mobile phase can be adjusted as well as the temperature to yield narrow and symmetric peaks.

Taken together, a good chromatographic separation is characterized by retention factors between 2-8, while the selectivity should be >1 to avoid co-elution. Furthermore, the resolution should be >1.5 for ideal, symmetric peaks and >2 for asymmetric peaks. Additionally, the peak shape should be narrow and close to symmetric, thus the asymmetry factor should be 1. All of these parameters should be adjusted under consideration of short analysis time.

1.4.2.2 Electrospray Ionization

In 1988, Fenn and his co-workers demonstrated the applicability of mass spectrometry with electrospray ionization (ESI) for large biomolecules[200], but the fundaments of ESI date back to the 1960s and are briefly described in the following.

The coupling of liquid chromatography and ESI is quite common for multiply chargeable small molecules and it prevailed in analysis of nucleic acids.[201, 202] The analyte containing eluent is flowing from the HPLC into the spray needle, to which a high voltage of 2-6 kV is applied (Figure 7). Here, the analytes are ionized. In positive ion mode, analytes are protonated by addition of hydrogen cations [M+H$^+$] or formation of adducts

with other cations such as Na^+, K^+ or NH_4^+. However, in negative ion mode, analytes are ionized by deprotonation or formation of adducts with anions like Cl^-.[203, 204] As illustrated in figure 7, the entrance of the mass analyzer is building the counter electrode to the spray needle, generating an electric field and accelerating ions towards the mass spectrometer. The droplet at the needle tip contains an excess of homogenously charged ions that are repulsive. The ion repulsion and the electric field, generated between the spray needle and the counter electrode, result in the formation of the so called Taylor cone at the needle tip.[205] As a result of electrospray based dispersion, the liquid cone is disintegrating and single droplets with high ion density are formed (Figure 7). In meantime, the solvent is being evaporated, which decreases the droplet size and increases the ion density within the droplet. If the electrostatic repulsion of ions is larger than the cohesive force of the solvent, the parent droplets form offspring droplets by fission. The ion density in the offspring droplets increases by the continuous solvent evaporation until the Rayleigh limit is reached and the Coulomb explosion leads to disintegration of droplets. The ion evaporation is explained by two models: the ion evaporation model according to Thomson[206, 207] and the charge residue model according to Dole[208]. Under the terms of the ion evaporation model, it is assumed that the droplet size shrinks due to evaporation until the field strength at the surface is high enough, so ions can be ejected from the droplet. Whereas the charge residue model describes that droplets containing a single ion are formed and finally the remaining solvent is evaporated, yielding the sole gas phase ion.
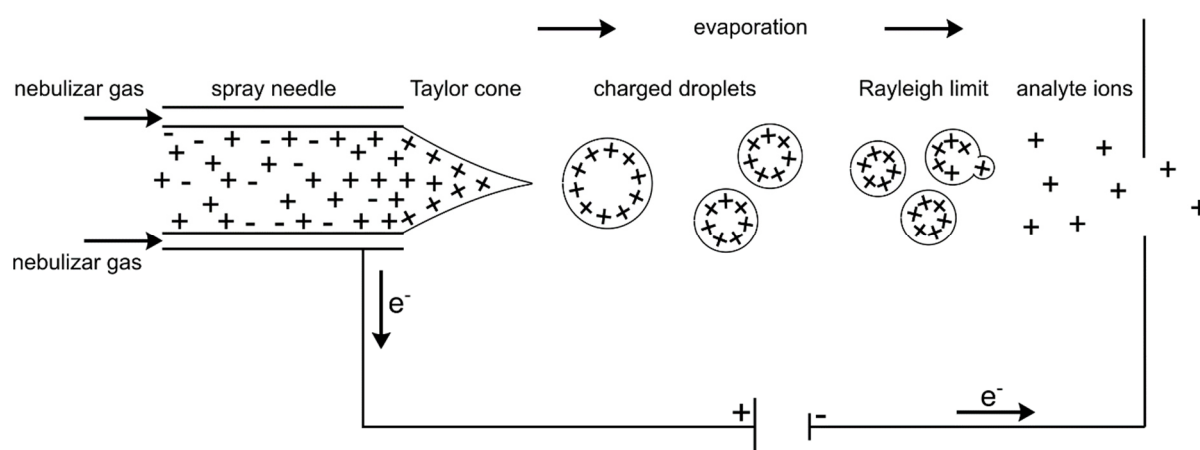
Figure 7: Schematic illustration of electrospray ionization (ESI) source, operated in positive ion mode, analyte containing solvent is eluting from the spray needle and a Taylor cone is built at the needle tip. Liquid is dispersed to droplets and ions are accelerated towards the mass spectrometer, droplet size is shrinking as result of solvent evaporation until Rayleigh limit is reached, ions are released from droplets and enter the mass spectrometer.

The ionization efficiency is highly dependent on the mobile phase composition and flow rate. In first instance, all solvent components have to be volatile and the addition of organic solvent facilitates the desolvation due to reduced surface tension. In any case, the aqueous solvent should contain a charge carrier, *e.g.* a salt, which is indispensable for charge separation at the needle tip. The selection of a suitable pH and ion strength can favor the formation of ions with a certain charge.[203, 209-211] Apart from the solvent composition, the flow rate has high impact on droplet size, spray stability and thus ionization efficiency.[203] The droplet size is decreasing with the flow rate.[212, 213] The ionization benefits from reduction in droplet size, as less droplet fission and solvent evaporation is necessary to release the gas phase ion. As a result, more analyte of the initial droplet can be proceeded for further analysis.

Besides solvent composition and flow rate, the selection of suitable parameters within the ion source are crucial for an effective ionization. The parameters to be optimized, are the capillary voltage, which is responsible for droplet charging, the flow rate and temperature of the drying gas, which are crucial for droplet formation as well as control of droplet size and desolvation. The nebulizer pressure impacts the nebulization process efficiency of the spray and in combination with the drying gas, it is responsible for ion formation. A further plane to ameliorate ionization, is the application of the sheath gas for pneumatic assistance during ESI. The sheath gas is applied concentrically around the spray needle. Its purpose is to narrow the spray, which results in more effective desolvation.[214] The additional thermal focusing by the sheath gas concentrates the ions spatial.[215] Further the nozzle voltage can be applied, which improves the ionization of apolar analytes.

ESI has the limit of saturation, even if the process is not fully understood, some suggestions should be introduced in the following. On the one hand it is argued, that there must be a maximal concentration of analyte in the droplet that can be charged[216], on the other hand it is elaborated that the saturation must be caused by strong molecule-molecule interactions on the limited droplet surface that impairs ion ejection and retains further ions in the droplet interior[217-219]. Either way, the impact of saturation should be determined by the upper limit of linearity.

1.4.2.3 Triple quadrupole mass spectrometry - QQQ

After the analytes were separated by chromatography and were ionized in the ion source, they are accelerated towards the mass spectrometer. Different monitoring modes can be applied with a triple quadruple mass spectrometer. In alignment with other types of mass spectrometers, a full scan can be conducted, within a mass range of 50 – 1000 Da, however the resolving power is lower in comparison to TOF or Orbi Trap devices. According to IUPAC, the resolving power is defined by the ratio of the *m/z* of an ion to the full width at half maximum (FWHM) of the respective peak (Equation 9).[220] Higher resolving power yields to higher mass accuracy, which increases the confidence in peak assignment.[221]

$$(9) \qquad Resolving\ Power = \frac{m/z}{FWHM}$$

In addition to the full scan mode, following scan modes can be conducted with a triple quadrupole mass spectrometer: (a) the product ion scan, selects precursor ions of a certain *m/z* in Q1, fragments those in the Q2 and scans the resulting product ions in Q3, (b) the precursor ion scan, scans precursor ions in Q1, that yield certain product ions, fragmented in Q2 and selected in Q3 (c) neutral loss scan, scans precursor ions in Q1 and product ions in Q3 that are connected by the loss of a certain neutral fragment, generated in Q2 (d) selected reaction monitoring (SRM), selects precursor ions of certain *m/z*, transmits them to Q2, where they are fragmented and transmitted to Q3, where product ions of certain *m/z* are selected (Figure 8). If Q1 and Q3 are set to select multiple ions with different *m/z*, the monitoring mode is called multiple reaction monitoring (MRM). Especially, SRM/MRM provide high sensitivity in contrast to the scan modes a)-c). The monitoring of specific mass transitions in SRM and MRM bring about noise reduction and thus an increase in signal-to-noise ratio. The high sensitivity that can be achieved with a triple quadrupole mass spectrometer by SRM/MRM make it very suitable for quantitative approaches.

Triple quadrupole mass spectrometers are used for analysis of nucleosides. The lone electron pair of nitrogen in the nucleobase can be easily protonated in the ion source, hence the operation in positive ion mode is common, however there are also methods in negative ion mode.[222, 223] In our approach, nucleosides are protonated in the ion source for generation of precursor ions (Figure 8, highlighted in blue). Next, a certain precursor ion is selected based on its *m/z* in Q1 (Figure 8, highlighted in yellow) and

subsequently fragmented in the Q2 (Figure 8, highlighted in orange). For the majority of nucleosides, the fragmentation results in the cleavage of the glyosidic bond, yielding the neutral ribose and a charged nucleobase. Whereas, Ψ fragments differently. Here, the carbon-carbon bond is more stable than the usual carbon-nitrogen glycosidic bond and hence the fragmentation is not resulting in cleavage between nucleobase and ribose but in ring-opening and fragmentation of the ribose.[224] After fragmentation, the product ion, is selected in Q3 (Figure 8, highlighted in yellow) and finally the ions are detected by an electron multiplier (Figure 9, highlighted in green). Analogously, cap modifications can be analyzed with the Cap MS method, which was developed and optimized by me. In contrast to the Nucleoside MS method, the device is operated in negative ion mode, in favor of the negative charge resulting from the phosphate ester. The principle is shown in figure 8, analogously to Nucleoside MS.
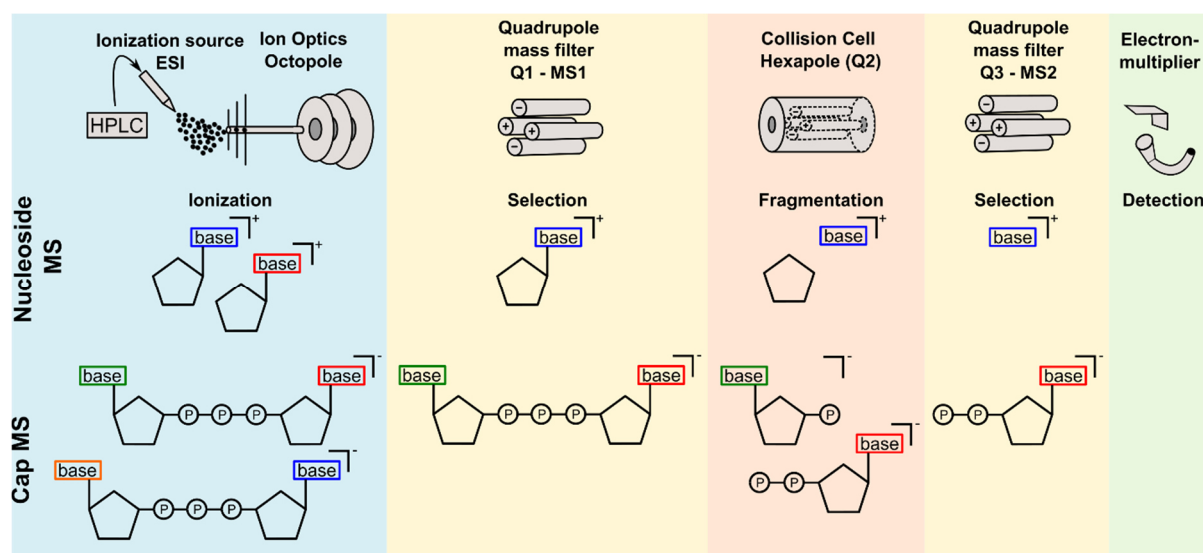

Figure 8: Principle of nucleoside and cap mass spectrometry (MS). Instrumentation for sensitive detection of modified nucleosides by a triple quadrupole MS. After separation of the RNA hydrolysate via HPLC, the eluent is ionized in an electrospray (ESI) ion source and the nucleoside is ionized by protonation and the cap analogue by deprotonation. The first quadrupole selects for the precursor ion. The collision cell fragments the nucleoside or cap modification and the charged nucleobase or nucleoside diphosphate is selected in Q3 and subsequently enters the detector and a signal is recorded. Figure adapted from Yoluç *et al.*[179]

The analysis of nucleosides with a triple quadrupole allows for quantification of nucleosides in attomole amounts, nonetheless there are various screws to adjust in order to fine tune the parameters and to exploit the limits of sensitivity. The detection sensitivity can be improved by the introduction of dynamic multiple reaction monitoring (dMRM). Here analytes are not only specified by their mass transitions but also by retention times. In this monitoring only a certain mass transition is measured in a certain time window, reducing the number of MRM scans and resulting in improved sensitivity, as the cycle time, in which a specific mass transition is monitored, can be reduced while

the dwell time, in which data is acquired for certain mass transition within one cycle, can be prolonged. Further, the voltages within the mass spectrometer can be optimized for high sensitivity. The ion source is under atmospheric pressure, whereas the mass spectrometer is under high vacuum with values of up to 5×10^{-5} Torr. The fragmentor voltage is responsible for this acceleration, and has to be high enough to escort the ions towards the ion optics without inducing an in-source fragmentation of the ions. In course of this, it should be mentioned that the sampling efficiency is not solely depended on the ionization efficiency but also on the transfer of ions through the various compartments of the mass spectrometer.[225, 226] The cell accelerator voltage (CAV) is responsible for the motion of ions through the mass spectrometer. It determines the speed of the product ion leaving the collision cell, and is especially important in a dMRM scan, to avoid the persistence of fragmentation products in Q2 from previous scans, while ions of the new scan are already entering the collision cell. Further, it should be aimed at determining the optimal collision energy for each analyte, in order to generate the desired product ion in high abundance.

1.4.2.4 Quantitative mass spectrometry

While mass spectrometry is *per se* not quantitative, quantification is still possible with the aid of internal standards.[227] Therefore, a calibration curve for each analyte is generated and the same amount of internal standard is added to each dilution (Figure 9A). It is of great importance that the internal standard is as similar as possible to the analyte, so the physico-chemical attributes are comparable. It is advised to utilize isotopologues as internal standards, so the properties are the same and the molecules co-elute and have the same ionization and fragmentation behavior. Thereby, effects of ion suppression or deviations in inter-sampling efficiencies can be eradicated and the analyte amount in the sample can be determined with the calibration curve. In general, the areas of integrated MS signals from analyte in the calibration is set in relation to the area of the respective internal standard, which yields the nucleoside isotope factor (NIF_{cal}).[228] This factor is plotted against the expected concentration of the calibration. The resulting slope corresponds to the relative response factor for the nucleoside (rRNF) (Figure 9B). Furthermore, the same amount of internal standard as in the calibration is added to the samples (Figure 9C). The amount of analyte in the sample can be assessed by the determination of the NIF_{sample}. Therefore, the ratio of area of integrated MS signal from analyte in the sample to the area of the respective internal

standard in the sample is built, resulting in NIF_{sample}. This value is divided by the rRNF to determine the molar amount of analyte in the sample (nuc [pmol]) (Figure 9D).
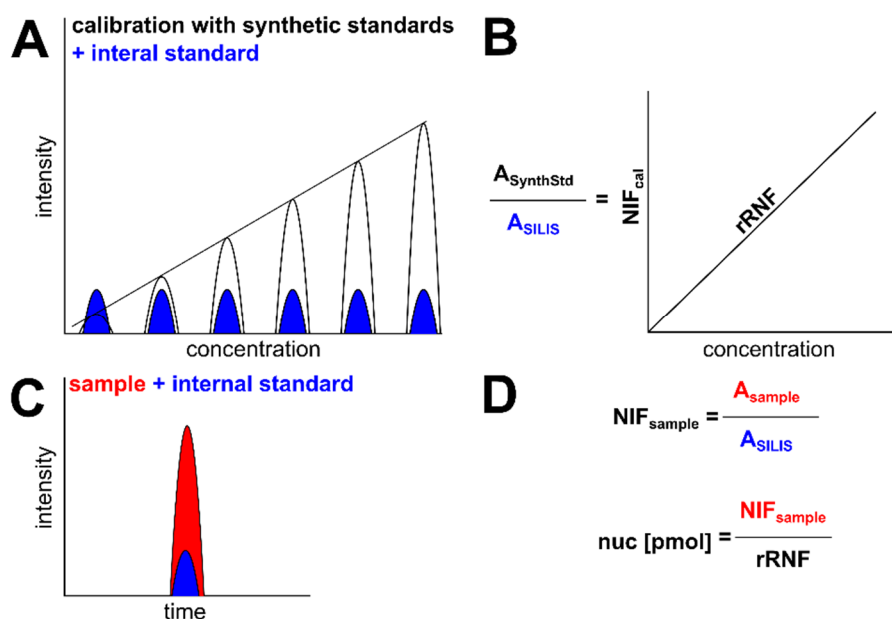

Figure 9: Quantification of nucleosides with LC-MS/MS (A) calibration curve with equal amounts of internal standard (B) peak for analyte and internal standards from sample (C) determination of the nucleoside isotope factor (NIF) and the relative response factors for the nucleosides (rRNF) from the calibration (D) determination of the nucleoside isotope factor (NIF) and the molar analyte amount in the sample, according to Kellner *et al.*[228]

1.4.2.5 Isotope labeling

Already in the 1950s isotope labeling was utilized to shed light onto bimolecular mechanisms like the Calvin cycle[229] or the confirmation of the semi conservative replication of DNA by the Meselson-Stahl-experiments[230]. Especially the application of isotope labeling with amino acids in cell culture (SILAC), opened a new frontier in protein analytics with mass spectrometry.[231, 232] Isotope labeling can also be applied to nucleic acids, thereby the dynamics and function of oxidation products of m^5C were traced in mouse embryonic stem cells.[114, 233] The techniques to examine isotope labeled biomolecules were expanded to NMR for structural elucidation and investigation of tRNA dynamics.[234] Furthermore, isotope labeled RNA *in vivo* paved the way for discovery of novel modifications, like hm^5Cm[235], $msms^2i^6A$[236], ms^2C[126] and the dynamics of hm^5C and f^5C[237].

The isotope labeling of DNA and RNA in various organisms is especially feasible and facilitates the examination of intracellular mechanisms. So did my doctoral supervisor, Prof. Dr. Stefanie Kaiser herself lay the first founding stones for absolute quantification

of nucleic acids and together with her team she pioneered in establishing and applying metabolic labeling strategies for multiple organism and developed the NAIL-MS technique (**N**ucleic **A**cid **I**sotope **L**abeling coupled **M**ass **S**pectrometry).[158, 178, 238] The strength of NAIL-MS lies in its multitude of application areas (Figure 10). This technique is used for modification discovery[126], multiplexing[239] and quantification[98, 158]. However, its largest application is the investigation of dynamics by
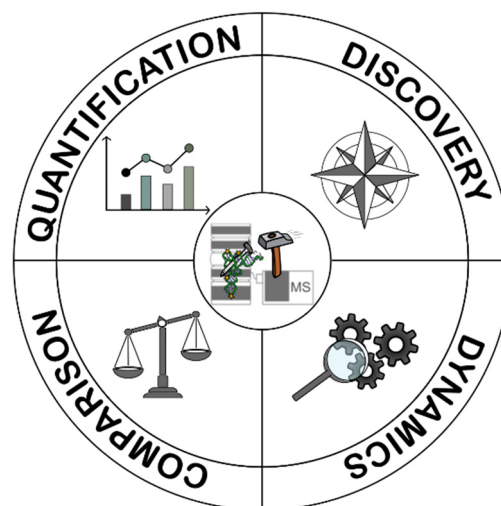

Figure 10: Application Areas of NAIL-MS, figure taken from Yoluç *et al.*[179]

combination of metabolic labeling with pulse-chase experiment where the isotope labeled medium is changed at a certain time point, allowing the differentiation between original and new transcripts. Further, it makes possible to differentiate between the mechanisms of degradation, dilution and demodification and it also helps to understand cellular mechanisms by identification of endogenous or exogenous incorporation of modifications and their removal.[98, 126, 158, 240]

Moreover, the use of isotope labeled internal standard (SILIS) is advantageous for quantification, as outlined in the previous chapter. In contrast to artificial SILIS, the biosynthesis of isotopologues by metabolic labeling with isotope labeled nutrients enables the access to a great variety of modifications at once.[241, 242] The SILIS enables the absolute quantification of all modifications occurring in an organism, without the utilization of SILIS only a relative quantification is possible.[228] The production of SILIS relies on the efficient and mono-isotopic labeling of DNA and RNA. Initially, SILIS was produced by cultivation of *S. cerevisiae* in commercially available ^{13}C rich growth medium for stable isotope labeling of nucleobases. Additionally, $^{13}C_6$-glucose was added for ribose labeling and L-methionine-[$^2H_3$]-methyl to obtain deuterated methyl groups. However, only partial CD_3 labeling was achieved and 20% of the methylated nucleosides possessed an $^{13}CH_3$ methylation (Figure 11A). This partly labeled SILIS caused interferences with isotopologues from NAIL-MS experiments and remained to be improved for efficient labeling and divers applicability of NAIL-MS.

In 2017, Heiss *et al.* reported a strategy for metabolic labeling in *S. cerevisiae* to distinguish the pre-existing tRNA pool from new emerging tRNA transcripts. Thus enabling the examination of dilution by new transcripts and the examination of tRNA modifications during growth. Furthermore, the extent of post-methylation on existing canonicals in tRNA could be studied by selection of a sophisticated approach for metabolic stable isotope labeling.[158] In first instance, various isotope labeled media were tested to achieve efficient stable isotope labeling of nucleosides. While the utilization of ^{15}N-glutamine amide and ^{15}N-aspartic acid resulted in low labeling efficiency, ^{15}N$_2$-uracil yielded efficient labeling of N1 and N3 of pyrimidines. The utilization of ^{13}C$_6$-glucose resulted in full ribose labeling of purines and pyrimidines as well as additional ^{13}C labeling of C8 in purine bases. Moreover, the addition of L-methionine-[^{2}H$_3$]-methyl, lead to CD$_3$ labeling of methylated nucleosides. In order to study the fate of tRNA modifications, yeast was initially cultivated in cultivated in ^{13}C$_6$-glucose and ^{15}N$_2$-uracil labeled medium (Figure 11B, pre-existing). For experimental initiation, the medium was exchanged to L-methionine-[^{2}H$_3$]-methyl containing medium, resulting in deuterated methyl groups but unlabeled nucleobase and ribose of nucleosides from new tRNA transcripts (Figure 11B, new). Furthermore, this selection of isotope labeled medium allowed for identification of post-methylation of pre-existing tRNA transcripts by CD$_3$ methylation of the ^{13}C, ^{15}N labeled nucleosides (Figure 11B, post-metylation). In summary, the application of NAIL-MS helped to identify fluctuations in modification densities depending on growth phase. However, this approach is limited by the congruent mass transition of inefficient labeled SILIS and post-methylation of nucleosides (mass spectra in figure 11). For a comprehensive analysis of the methylome the optimization in methyl labeling strategy should be considered.
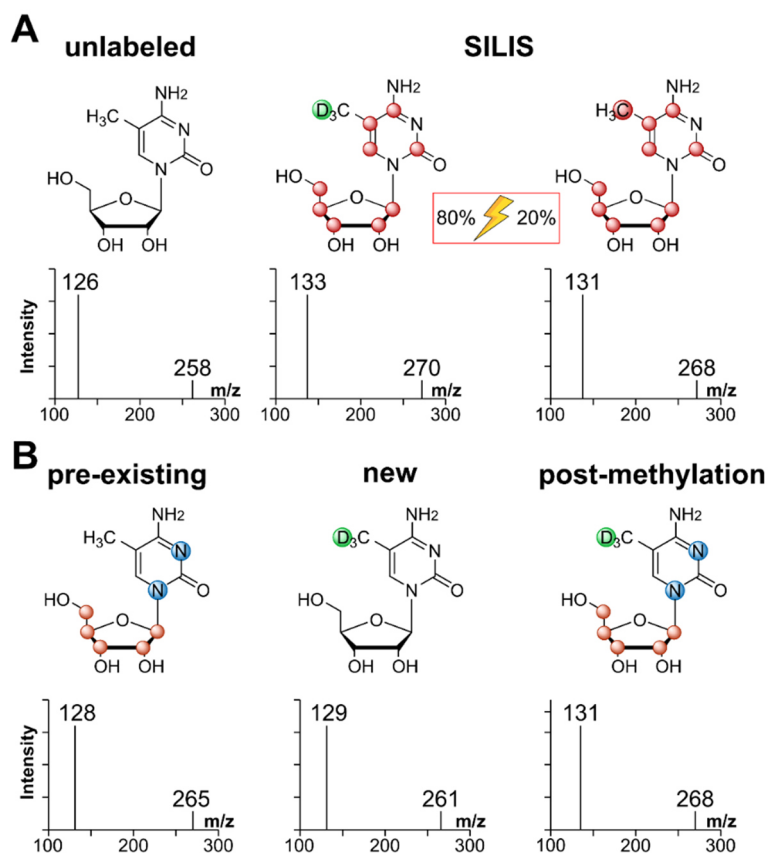
Figure 11: Isotopologues of m⁵C in *S. cerevisiae* (A) structural formula and mass spectra of unlabeled m⁵C and its isotopologues from SILIS (B) structural formula and mass spectra of unlabeled m⁵C and its isotopologues from NAIL-MS described in Heiss et al.[243], red balls indicate ^{13}C labeling, blue balls indicate ^{15}N labeling and green balls indicate CD_3 labeling

2 Aim of the project

Nucleic acid modifications are highly dynamic. Organisms adapt their modification density to environmental changes like infections, psychological stress and chemicals. An earlier study proposed the reprogramming of tRNA modifications in yeast upon exposure to methylating, oxidizing and reductive agents. The resulting variances in modification densities resulted in translational changes. These findings underlined the impact of RNA modifications on cellular mechanisms and thus it is of great importance to study the fate and effect of these nucleic acid modifications.

Intrigued, by the findings in stress-induced tRNA reprogramming in *S. cerevisiae*, the impact of dilution by new transcripts, the degradation of damaged RNA as well as the effects of active (de-)modification should be investigated. Therefore, the discernibility of original and new transcripts as well as of enzymatic and damage-induced modifications is required.

Nucleic acid modifications can be analyzed with mass spectrometry, but this technique is limited by its static insight into modification dynamics. In order to assess these dynamics with mass spectrometry NAIL-MS was developed. The metabolic stable isotope labeling enables the investigation of modification dynamics by the utilization of variable isotope labeled media. The combination of pulse-chase experiments and NAIL-MS allows to trace the fate of the aforementioned different RNA transcripts and their modifications

However, the application of NAIL-MS requires full efficient labeling and the uniqueness of isotopologues emerging from the different stages of stable isotope labeled pulse-chase experiments and SILIS. Therefore, a novel SILIS, clearly differentiable from isotopologues emerging from pulse-chase experiments needs to be developed, in order to avoid signal interferences of isotopologues. For the same reason, the L-methionine-[^{2}H$_3$]-methyl labeling has to be optimized, to achieve full efficient labeling in order to clearly distinguish the endogenous methylome from damage-induced methylation. After succeeding in establishment of the desired labeling, the extent of stress-induced tRNA and rRNA modification changes in *S. cerevisiae* should be examined. Besides the investigations on the epitranscriptom, it was aimed at determining quantities of endogenous and damage-induced methylations in the genome of *E. coli*, *S. cerevisiae*

and HEK cells by expanding the combination of pulse-chase experiments and NAIL-MS.

In 2019, the RNA virus SARS-CoV-2 was discovered. Since then this virus developed several mutants and variants. In parallel to prokaryotes and eukaryotes, the modification of the viral genome is highly dynamic. But in contrast to the aforementioned organisms, less is known about the underlying mechanisms as their identification is complicated by the multitude of modification origin, since the viral genome is prone to mutation, the modification densities can vary during infection and the virus can exploit the host enzymes to regulate its modification profile. With the spreading of the RNA virus SARS-CoV-2, the modification abundances in the genome of the mutants D614 and G164 as well as if the variants alpha, beta and delta of SARS-CoV-2 should be quantified and compared in order to gain a better understanding of this novel virus and the changes in its modification profile. A key feature of viral contagiousness, is the evasion of the host innate immune response, by mimicking the host's mRNA cap modification. In parallel to the absolute quantification of internal modification in the genome of SARS-CoV-2, a time efficient and highly sensitive LC-MS/MS method for absolute quantification of cap modifications per RNA molecule should be developed. This method might find its applicability in the investigation of biological capping mechanisms but also in determination of capping inhibitor effectiveness and in quality control of mRNA therapeutics. It is of great importance to identify the characteristics of the viral genome and the differences in mutants and variants in order to identify novel drug targets and optimize the application of available therapeutics and vaccines.

3   Results and Discussion

3.1   Damage to nucleic acids

Synopsis

The canonical nucleotides cytidine, uridine/ thymidine, guanosine and adenosine can be modified with various functional groups. Besides endogenous introduction and/or removal of modifications by enzymes, they can be (de-)modified as a consequence of stress exposure. Modifications of nucleic acids comply different functions like gene regulation, RNA structure stabilization, translation fidelity and enzyme recognition. It is concluded that modifications of nucleic acids contribute considerably to cellular regulation and therefore it is of great interest to understand the mechanisms of (de-)modification, to identify the targets of modification and to trace their fate.

Conventional tools, like sequencing and mass spectrometric analysis are confined to a static view, neglecting the dynamics of cellular mechanism. Additionally, these methods disregard the cause of modification and thus the differentiation between endogenous and damage-induced modification. The strengths and weaknesses of current instrumental analytical methods are discussed in my review.[179] With nucleic acid isotope labeling coupled mass spectrometry (NAIL-MS) we developed a powerful tool to identify the cause of modification and to trace the dynamics of nucleic acid modifications by combining pulse-chase experiments with NAIL-MS. This method relies on metabolic labeling of nucleic acids, generating isotopologues. The metabolic labeling allows to distinguish between original and new replicates or transcripts as well as enzymatic or damage-induced modifications.

In 2010, Chan *et al*. reported the stress-induced reprogramming of tRNA modifications in *S. cerevisiae*. Further, the impact of modifications on translational processes could be shown. However, the underlying mechanisms could not be elucidated. With the powerful tool NAIL-MS in hand I aimed at studying the stress-induced modification dynamics and at recording the quantities of damage-induced modifications aside from endogenous modifications over time.

In order to determine absolute quantities, a biosynthetically produced stable isotope labeled internal standard (SILIS) was generated. First, the SILIS production was optimized to yield an internal standard that is not inferring with the isotopologues gener-

ated in the NAIL-MS pulse-chase experiments, which is described in detail in a methodical book chapter, I contributed to.[243] As it was focused on identifying macromolecular targets of damage, a time- and sample-efficient tandem-SEC-method was established to separate 25S, 18S rRNA and tRNA from the same sample, which is also outlined in the methodical book chapter.[243]

The application of the above mentioned techniques revealed that methyl methanesulfonate (MMS) generates methylated nucleosides (m^7G, m^1A, m^3C and m^6A) by direct methylation of canonical nucleobases in 25S, 18S rRNA and tRNA of *S. cerevisiae*. In addition, novel damage products were identified. It was found that, nucleobases of 2'-O-methylated rRNA were also damaged by direct methylation. Further, the fate of enzymatic as well as damage-induced methylations was monitored. It was specified that damage is occurring only on original transcripts, while new transcripts are spared. In addition, the decrease in abundance of damage-induced methylation within a short recovery time was determined, while the level of endogenous methylation was maintained.[99]

Intrigued by this preceding study, I was interested in the dynamics of the methylome on the genomic level. Therefore, the established NAIL-MS techniques for the organisms *E. coli*, *S. cerevisiae* and HEK cells were extended to the DNA level. This study revealed that the major damage product on genomic and transcriptomic level in all studied organisms is N-7-methyl(deoxy)guanosine (m^7dG/m^7G). However, the extent of damage varied amongst the model organisms, and less damage was recorded in organisms of higher complexity. In line with my previous study, the rapid removal of damage-induced methylations on the genome and transcriptome could be shown as well as the involvement of certain enzymes in damage removal in *E. coli* and *S. cerevisiae*.[244]

Instrumental analysis of RNA modifications, **Yoluç Y.**, *et al.*, Crit Rev Biochem Mol Biol., 2021

Declaration of contribution: The chapters "Advances in chemical labeling of modified RNA in the context of instrumental analysis", "Nucleoside MS" were researched and written by me. The manuscript was completed by cooperation of all authors.

Open access under Creative Commons Attribution License

Quantification of Modified Nucleosides in the Context of NAIL-MS, Heiss M, **Yoluç Y.**, *et al.*, Methods Mol Biol., 2021

Declaration of contribution: The described size exclusion tandem chromatography and the preparation of the SILIS^{Gen2} was developed and established by me. All other described methods have been established and optimized over the years by all authors and other laboratory members. The preparation of the manuscript was completed by cooperation of all authors.

Copy of publication with permission of the journal: https://www.springer.com/series/7651

The Stress-Dependent Dynamics of *Saccharomyces cerevisiae* tRNA and rRNA Modification Profiles, **Yoluç Y.**, *et al.*, Genes, 2021

Declaration of contribution: All experiments, were planned and performed by me or under my supervision by the Master student Erik van de Logt. The manuscript was prepared by Prof. Stefanie Kaiser and myself.

Open access under Creative Commons Attribution License

Systematic assessment of methylation damage in nucleic acids, **Yoluç Y.**, *et al.*, ChemBioChem, 2022

Declaration of contribution: All experiments in *S. cerevisiae* and *E. coli* were planned and performed by me or under my supervision by the Master student Hurina Hu. Experiments in HEK cells were planned and performed by my former colleague Dr. Matthias Heiß. The manuscript was prepared by Prof. Stefanie Kaiser and myself.

-In revision -

Taylor & Francis
Taylor & Francis Group

REVIEW ARTICLE

🔓 OPEN ACCESS   Check for updates

Instrumental analysis of RNA modifications

Yasemin Yoluç[a] (ID), Gregor Ammann[a] (ID), Pierre Barraud[b] (ID), Manasses Jora[c] (ID), Patrick A. Limbach[c] (ID), Yuri Motorin[d] (ID), Virginie Marchand[e] (ID), Carine Tisné[b] (ID), Kayla Borland[a] and Stefanie Kellner[a,f] (ID)

[a]Department of Chemistry, Ludwig Maximilians University, Munich, Germany; [b]Expression génétique microbienne, UMR 8261, CNRS, Institut de biologie physico-chimique, IBPC, Université de Paris, Paris, France; [c]Department of Chemistry, University of Cincinnati, Cincinnati, OH, USA; [d]Université de Lorraine, CNRS, UMR7365 IMoPA, Nancy, France; [e]Université de Lorraine, CNRS, INSERM, Epitranscriptomics and RNA Sequencing Core facility, UM S2008, IBSLor, Nancy, France; [f]Institute of Pharmaceutical Chemistry, Goethe-University, Frankfurt, Germany

ABSTRACT
Organisms from all domains of life invest a substantial amount of energy for the introduction of RNA modifications into nearly all transcripts studied to date. Instrumental analysis of RNA can focus on the modified residues and reveal the function of these epitranscriptomic marks. Here, we will review recent advances and breakthroughs achieved by NMR spectroscopy, sequencing, and mass spectrometry of the epitranscriptome.

GRAPHICAL ABSTRACT
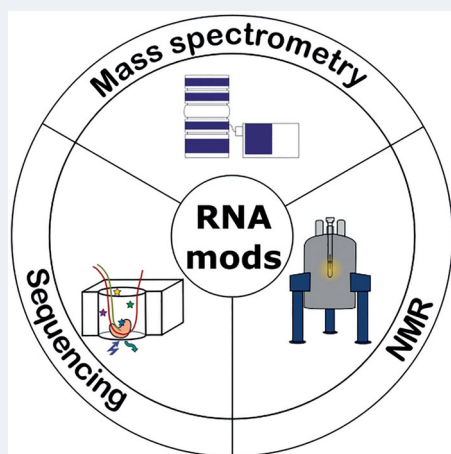
Overview

RNA is a dominant macromolecule in all organisms. Its monomeric units are the canonical nucleosides cytidine, uridine, guanosine, and adenosine which are interconnected by a 3′-5′-phosphate backbone. After transcription, enzymes target the RNA and introduce a variety of modifications onto nucleobases or the ribose 2′-OH. Overall, more than 150 unique RNA modifications have been identified in many RNA types and throughout all domains of life.

The detection, localization, and quantification of these RNA modifications are crucial for understanding the

elusive function of RNA modifications. Here, we review the most common techniques for instrumental analysis of modified nucleosides and discuss recent advances in the field as well as the strengths and weaknesses of every approach. As shown in Figure 1, NMR spectroscopy is mostly suited to study the impact of RNA modifications on RNA structure and stability in addition to recent progress made in RNA modification dynamics. One of the fastest developing fields is RNA sequencing, which allows localization of modifications at nucleotide resolution. Many RNA modifications, such as thiolation, amino acid addition, and even methylations lead to a change in the chemical
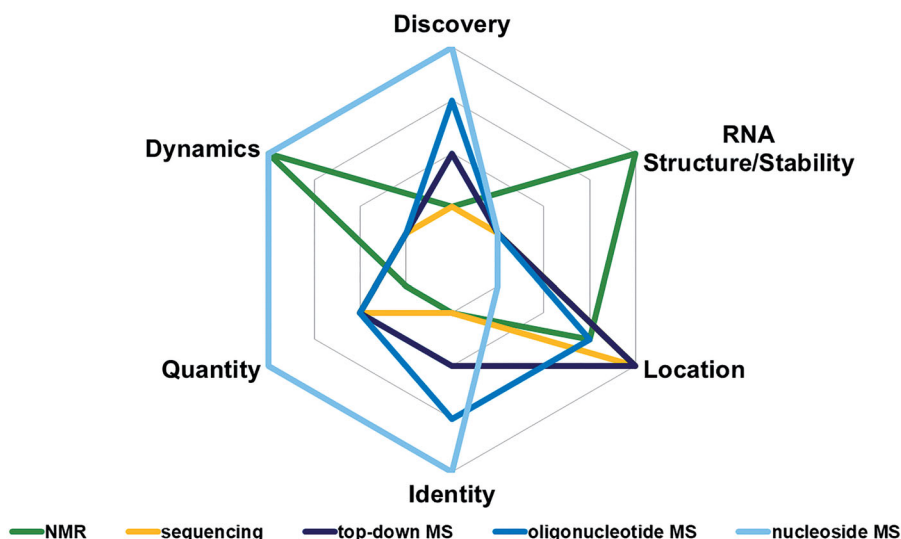
Figure 1. Strengths (outside) and weaknesses (inside) of current methods for RNA modification analysis. Legend: Discovery (and chemical characterization of novel RNA modifications), (impact of RNA modifications on) RNA Structure/Stability, Location (of modification within the sequence), Identity (chemical structure of modification, especially isomer discrimination), Quantity (of modification), Dynamics (of RNA modifications). See colour version of this figure at www.tandfonline.com/ibmg.

reactivity of the RNA. The resulting differential reactivity of modified nucleosides is exploited for their detection by sequencing but also by mass spectrometry, the last field discussed in this review. Mass spectrometry (MS) of RNA modifications can be performed on full-length RNA (top-down MS), partial RNA hydrolysates (oligonucleotide MS), or complete hydrolysates (nucleoside MS).

Interestingly, NMR, MS, and sequencing analyses complement each other as they annihilate the major pitfalls of the other. For example, mass spectrometry is ideally suited to unambiguously identify the chemical nature of the modified nucleoside even with minute amounts of sample material. Unfortunately, nucleoside MS depends on the complete enzymatic digestion of the RNA and thus all sequence information and the location of the modified nucleoside remains unknown. Sequencing on the other hand is, if carefully conducted, perfectly able to pinpoint the location of a modified nucleoside within the RNA but it does not deliver any information about the chemical nature of it. Thus, MS and sequencing are orthogonal techniques that benefit from each other while NMR analysis adds information on the structural impact of an RNA modification. In this work, we discuss the basic principles of NMR, sequencing, and MS analyses, with the main focus on the most recent advances in these fields.

NMR spectroscopy in the analysis of RNA modifications

The use of NMR in the study of RNA modifications goes back to the origins of the study of biological macromolecules by NMR (Crawford et al. 1971; Koehler and Schmidt 1973; Kan et al. 1974). The large chemical diversity contained in modified nucleotides, as compared with the homogeneous chemistry of the canonical RNA nucleotides, is associated in NMR with a large variety of signals, which often are located in regions of the spectra that are empty of canonical RNA signals. These signals, appearing at particular positions, and most of the time isolated from each other, are thereby relatively easy to recognize. Pioneering studies have used the isolated signals arising from nucleotide modifications in tRNAs as molecular probes to explore the three-dimensional folding and the conformational stability of these molecules (Kan et al. 1974; Kastrup and Schmidt 1978). In addition, signals from tRNA modifications have played a leading role in the chemical shift assignment process for this family of RNAs. Some isolated signals in the upfield part of the spectra, mostly methyl groups, were indeed used as independent and unambiguous signals for the assignment of the imino groups in the downfield part of the spectra (Hilbers et al. 1983; Roy and Redfield 1983; Hare et al. 1985; Heerschap et al. 1985; Choi and Redfield 1986). These signals include the methyl group of 5-methyluridine (m^{5}U54) (~1.00 ppm), 5-methylcytidine (m^{5}C49) (~1.55 ppm), 7-methylguanosine (m^{7}G46) (~3.75 ppm), 1-methyladenosine (m^{1}A58) (~3.00 ppm), and 2-methylguanosine (m^{2}G10) (~2.75 ppm) that are found in many tRNAs including the yeast tRNA^{Phe} (Agris et al. 1986).

After these original studies, the investigation of RNA modifications with NMR has expanded beyond the study of tRNA structure and stability. Occasionally, NMR

has been used for the identification of modifications contained in a given RNA (Sakamoto et al. 1993; Tisné et al. 2000; Gaudin et al. 2003; Wurm et al. 2010; Ranaei-Siadat et al. 2013). But NMR has been and is still mostly used in detailed inspections dealing with the characterization of the specific effects caused by RNA modifications. For instance, NMR combined with UV-melting curves has been used to decipher the contribution of specific modifications in RNA stability (Kumar and Davis 1997; Strebitzer et al. 2018). Similarly, the ability of NMR to assess the dynamic behavior of macromolecules at atomic levels was used to demonstrate that the presence of m^{6}A perturbs the hybridization kinetics of RNAs, and slows down the annealing of m^{6}A-containing RNAs (Shi et al. 2019). In addition, several structural studies have analyzed the role and the effect of modifications in the conformation of tRNA anticodon stem-loops (Stuart et al. 2003; Durant et al. 2005; Vendeix et al. 2012), overall showing that modifications stabilize and restrict the conformations of the anticodon domain in tRNAs (Agris 2008). In another example, deuterium exchange NMR experiments were used to show that the dynamics and flexibility of *Escherichia coli* tRNA^{Val} are significantly affected by post-transcriptional modifications, with some base pairs being 20 times more stable in the native tRNA than in the unmodified one, although the local and global structure of this tRNA seems rather unchanged in the presence or absence of modifications (Vermeulen et al. 2005). Finally, several NMR studies have also investigated in detail the structural effects produced by specific modifications, such as Ψ (Davis and Poulter 1991), m^{6}A (Zhou, Parisien, et al. 2016), D (Dyubankova et al. 2015), m^{1}A and m^{1}G (Zhou, Kimsey, et al. 2016), further expanding the repertoire of NMR applications associated with the study of RNA modifications.

Recently, a novel approach using NMR has been introduced to analyze RNA modifications. Following this methodology, NMR offers a way to monitor tRNA maturation in a non-disruptive and continuous manner thereby providing information on the temporality of tRNA modification events (Barraud et al. 2019). The strategy relies on the introduction of ^{15}N isotope-labeled tRNAs into unlabeled cell extracts containing the cellular enzymatic activities and on the use of isotope-editing in NMR experiments to only detect the tRNA of interest within the complex cell extract environment. With this experimental setup, RNA modification events are directly monitored in a time-resolved fashion, by measuring successive NMR experiments on a single sample directly incubated in the NMR spectrometer (Figure 2). This methodology consists of three main steps. In the first step, a ^{15}N-labeled tRNA sample and an unlabeled cell extract are produced. The ^{15}N-labeled tRNA sample is produced by *in vitro* transcription with ^{15}N-labeled nucleotides and purified by ion-exchange chromatography (Catala et al. 2020b). This tRNA contains no post-transcriptional modifications. The unlabeled cellular extract is produced from the lysis of cell culture. To preserve as far as possible the cellular enzymatic activities in the extract, cell lysis is performed under gentle conditions and in the presence of anti-proteases (Gato et al. 2021). In the second step, the NMR signature of each individual modification is obtained from the NMR chemical shift assignments of different tRNA samples with different modification content. The comparison of the NMR spectra of the unmodified tRNA (Figure 2(B)) with the NMR spectra of the modified tRNA provides the means to identify the NMR signature of each post-transcriptional modification (Catala et al. 2020a). In the third step, the NMR monitoring of tRNA maturation is performed. The ^{15}N-labeled tRNA sample is mixed with the unlabeled cell extract to yield the *in extract* NMR sample (Figure 2(A)). This sample is incubated directly in the NMR spectrometer and series of NMR experiments are measured in order to monitor tRNA maturation events in a time-resolved fashion (Figure 2(C)). In the NMR spectra, the progressive appearance of new signals and the correlated disappearance of signals from the unmodified tRNA sample are the signature of chemical modifications being introduced in the initial transcript. These changes in the NMR spectra can be analyzed in the light of the NMR signature of the individual modifications previously identified. Since NMR spectra are measured as a time course series, this methodology enables the identification of early and latter modifications events. For instance, in the yeast tRNA^{Phe}, this methodology demonstrated that in the T-arm, modifications Ψ55, T54 and m^{1}A58 are introduced in a defined sequential order (Figure 2(C)), which is controlled by a strong circuit of modifications (Ψ55 → T54 → m^{1}A58) (Barraud et al. 2019). Modifications circuits can be complicated to identify (Han and Phizicky 2018; Barraud and Tisné 2019), and this NMR approach adds to the available tools for the analysis of potential cross-talks between modification events. Overall, NMR spectroscopy provides the means to observe sequential orders in the introduction of modifications along the tRNA maturation pathway (Figure 2(C)) and to identify modification circuits.
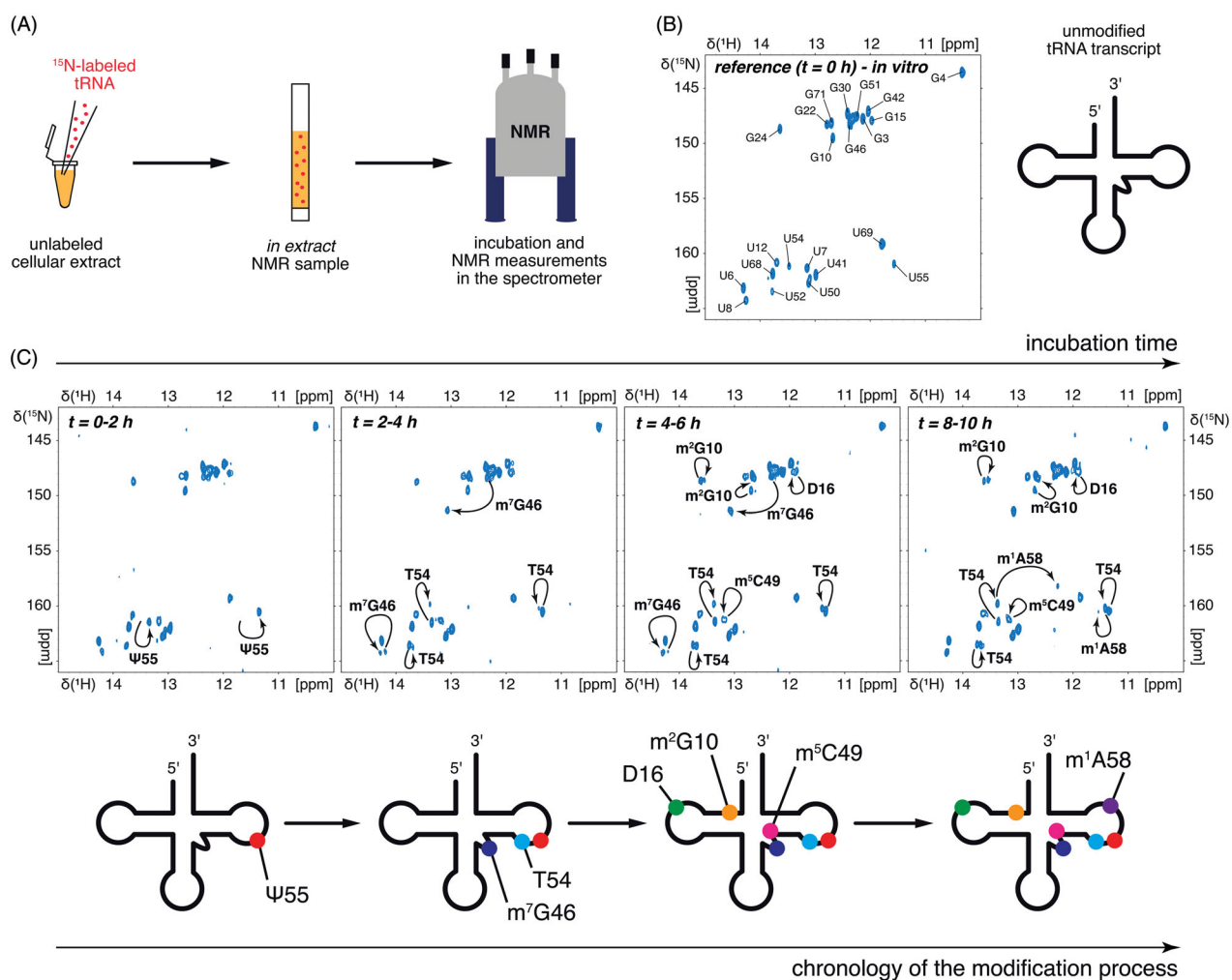
Figure 2. Time-resolved NMR monitoring of tRNA modifications. (A) A ^{15}N-labeled tRNA transcript (represented as red dots) is introduced in an unlabeled cellular extract (in orange). This mix is transferred to an NMR tube to yield an in extract NMR sample. The tRNA maturation sample is then directly incubated in the NMR spectrometer. (B) Imino (^{1}H, ^{15}N) correlation spectrum of a ^{15}N-labeled yeast tRNA^{Phe} measured *in vitro* to provide a reference spectrum (left part) and schematic representation of this unmodified tRNA transcript (right part). (C) Successive NMR measurements provide time-resolved information on the tRNA maturation process. Top part: Imino (^{1}H, ^{15}N) correlation spectra of ^{15}N-labeled yeast tRNA^{Phe} measured during a continuous incubation at 30 °C in yeast extract. Each NMR spectrum measurement spreads over 2 h (incubation time indicated on each spectrum). NMR chemical shift changes (indicated with arrows on the NMR spectra) are identified and linked to specific modification events. Bottom part: Schematic view of the sequential order of the introduction of modifications in yeast tRNA^{Phe} corresponding to the NMR spectra depicted above. The acquisition of NMR spectra in a continuous and time-resolved fashion enables the identification of a sequential order in the introduction of post-transcriptional modifications and thereby gives access to the chronology of the modification process. See colour version of this figure at www.tandfonline.com/ibmg.

Advances in chemical labeling of modified RNA in the context of instrumental analysis

Many RNA modifications, such as thiolation, amino acid addition, and even methylations lead to a change in the chemical reactivity of the RNA. Thus, early analyses focused on the differential reactivity of modified nucleosides for their detection. With the advent of instrumental analysis, the benefit of these chemical reactions was more and more forgotten and they have

been rarely used (Kellner et al. 2010; Heiss and Kellner 2017). Only those compatible with modern detection techniques such as mass spectrometry (Durairaj and Limbach 2008) or sequencing (Motorin and Helm 2019) are now commonly used as they aid in the detection of certain RNA modifications. For example, pseudouridine (Ψ) and inosine (I) are similar to their canonical precursor (U and A, respectively) in their mass and base-pairing abilities. Due to isomerization, Ψ gains additional functionalities which react with acrylonitrile (Figure
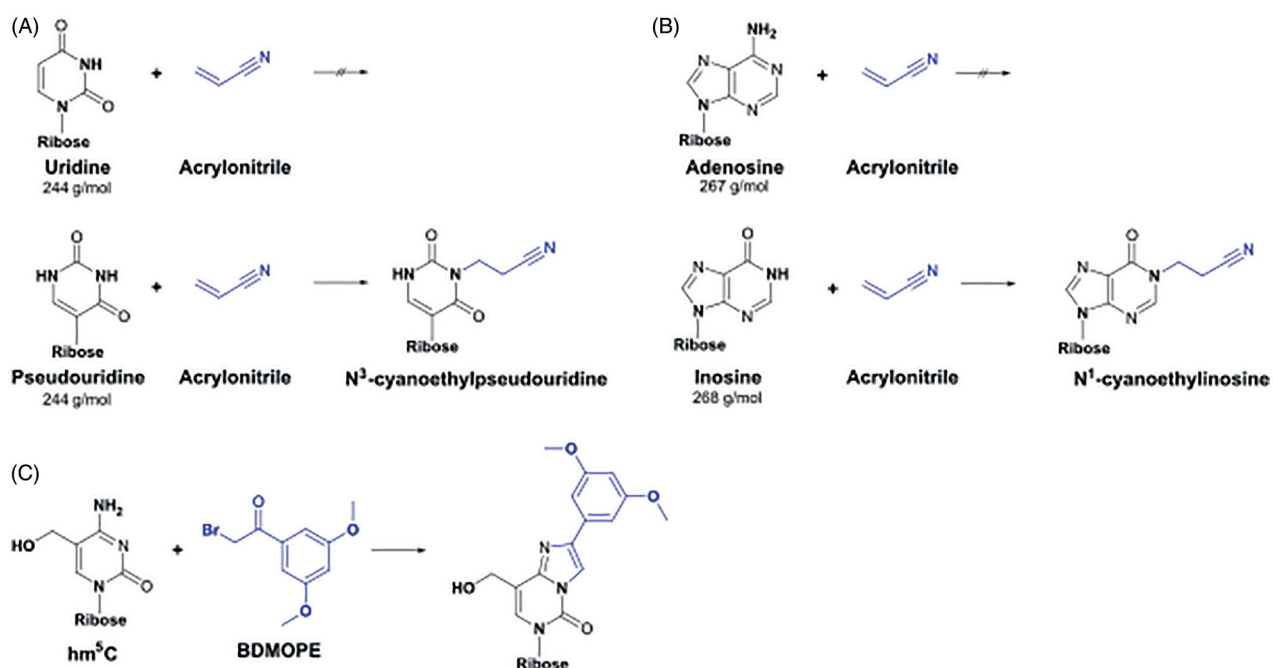
Figure 3. Selected chemical labeling strategies for detection of modified nucleosides. (A, B) Reaction of uridine and adenosine with acrylonitrile is not observed (upper panels), whereas pseudouridine and inosine are successfully cyanoethylated (lower panels). Modified nucleotides are afterwards detected through mass spectrometry and sequencing. (C) 5-hydroxymethylcytidine (hm5C) reacts with 2-bromo-1-(3,4-dimeth oxyphenyl)-ethanone (BDMOPE) which allows highly sensitive mass spectrometric detection of hm5C (Feng et al. 2020). See colour version of this figure at www.tandfonline.com/ibmg.

3(A)) (Mengel-Jørgensen and Kirpekar 2002) and thus facilitate detection by, for example, mass spectrometry (Emmerechts et al. 2005; Hossain and Limbach 2007; Durairaj and Limbach 2008) and even sequencing (Bakin and Ofengand 1993; 1998) becomes possible. Similarly, A-to-I editing can be visualized by cyanoethylation of inosine (Figure 3(B)) (Sakurai et al. 2010). Further developments of acrylonitrile-based reagents have led to the detection of inosine through fluorescence and bioaffinity enrichment (Knutson et al. 2018) and through inosine chemical erasing sequencing, ICE-Seq (Okada et al. 2019). The most trusted methods for RNA modification analysis through sequencing are based on the use of specific chemical reagents. This allows one to compare an untreated (mock) sample with a treated one and exclude (or at least reduce) random noise in peak calling. Such derivatization methods were developed for numerous naturally modified nucleotides in RNA, like inosine (Suzuki et al. 2015), 5-methylcytosine (Tuorto et al. 2012; Schaefer et al. 2017; Khoddami et al. 2019), pseudouridine (Carlile et al. 2014; Lovejoy et al. 2014; Schwartz et al. 2014; Li et al. 2015), 1-methyladenosine (Dominissini et al. 2016; Safra et al. 2017), 4-acetylcytidine (Thomas et al. 2018; Sas-Chen et al. 2020), and 7-methylguanosine (Enroth et al. 2019; Pandolfini et al. 2019).

Another progress was recently achieved for the sensitive detection of modified cytidines. So far, the quantification of these modifications appeared to be challenging due to their low ionizability by mass spectrometry. Despite several attempts by chemical derivatization (Tang et al. 2015; Huang et al. 2016; Guo et al. 2017) and selective enrichment, it was not possible to quantify these nucleosides easily. In 2020, the Feng lab reported the use of bromoacetonylated benzene reagents for efficient labeling of cytidines and many of their naturally occurring modifications (Feng et al. 2020). After the addition of the reagent to the N3 and N4 of cytidines (Figure 3(C)), the new chemical properties of the product are exploited in the context of LC-MS detection. The hydrophobicity leads to better retention on a reverse-phase (RP) column and thus the analytes reach the mass spectrometer at a higher organic solvent ratio which increases detection sensitivity. Furthermore, the tertiary amines are easily protonated in, for example, electrospray ionization and thus Feng et al. could show a 40- to 460-fold increased detection sensitivity for modified cytidines after the reaction with their chemical label (lower limit of detection (LLOD) of 0.06 fmol for m5Cm, 0.17 fmol for hm5Cm, 0.22 fmol for f5Cm and 0.06 fmol for ca5Cm). In this work, the authors quantified the abundance of modified cytidines in several human cell lines and even carcinoma tissue without utilizing a stable

isotope-labeled standard (SILIS). Due to the chemical simplicity of the labeling reagent, its synthesis with stable isotopes should be an easily possible and thus simple access to synthetically produced cytidine modification SILISs. From our perspective, the method appears to have a great potential for sensitive and accurate quantification of modified cytidines.

Sequencing

Deep sequencing protocols for high-throughput analysis of DNA are now widespread and indispensable for comprehensive projects in modern biology. The output of current DNA deep sequencing machines is now sufficient for a complete analysis of one human genome (or even several genomes) in only few hours. Such extraordinary performance is also employed for RNA analysis, like whole-transcriptome sequencing, mRNA alternative splicing, or single-cell transcriptome studies.

Post-transcriptional RNA modifications can be detected in RNA using various methods and approaches exploiting the chemical and physico-chemical properties of these non-canonical RNA nucleotides. In addition to classical RNA techniques like 5′/3′ and specific internal labeling, methods based on next-generation sequencing of the second (NGS) and third (NNGS) generation become more and more popular. These approaches generally provide single-nucleotide resolution for identification of the modified RNA position but maybe less accurate in the identification of the exact nature of the modified residue, due to a rather generic treatment used during the library preparation step. The most popular and reliable methods using NGS analysis rely on various specific chemical treatments applied to alter particular modified RNA residues and thus to make them detectable either as RT-stop or mis-incorporation of nucleotides into cDNA. For the moment, NNGS approaches are mostly using ion-current profiles through a nanopore or kinetics of dNTP incorporation in a PacBio chip to deduce the presence of unusual modified nucleotide (Figure 4).

Except for very recent attempts of direct RNA sequencing by nanopores (see below) all other protocols for RNA modification analysis use the conversion of RNA into cDNA sequence by primer extension with reverse transcriptase (RT). This RT step is almost inevitably followed by PCR amplification, to obtain a sufficient amount of amplicons. Current technologies for DNA (direct RNA) deep sequencing belong to two distinct generations: Next Generation Sequencing (second generation, NGS), also called massive parallel sequencing which uses amplified real or virtual DNA clusters (Illumina/Ion Torrent), and the third generation technologies (NNGS), based on single-molecule sequencing analysis (PacBio/Oxford Nanopore) (Figure 4).

Current sequencing technology for RNA modification analysis

The first protocols developed for the analysis of RNA modifications used a traditional RT-based RNA sequencing for detection of RT-stops due to impeded primer extension (Maden 2001; Motorin et al. 2007; Behm-Ansmant et al. 2011). With the development of deep-sequencing technologies, such low-throughput approaches were progressively replaced by next-generation sequencing, the most popular technology remains Solexa (now Illumina) sequencing-by-synthesis (SBS) protocol using fluorophores. The very first commercial Solexa/Illumina devices (like Genome Analyzer IIx released in 2009) were able to provide very short reads (with a maximum of 36 nt in length) and thus had limited applications for RNA modification analysis, which generally requires longer sequences to analyze. In addition, the sequencing output was extremely modest. When the HiSeq series of sequencers (HiSeq2000 and 1000) were released in 2011, their performance became fully compatible with the analysis of RNA modifications, first by analysis of RT signatures (Ryvkin et al. 2013). Other competitor technologies that emerged in the DNA deep sequencing field had very limited use in the analysis of RNA modifications, mostly due to high levels of intrinsic sequencing errors. Only one application of Ion Torrent sequencing was described for the original RiboMethSeq protocol developed by H. Nielsen's lab (Birkedal et al. 2015), but the use of Illumina sequencing is recommended. The current evolution of the NGS field left the space only for various Illumina devices employing SBS technologies differing by the number of fluorophores, mostly four (for older HiSeq/MiSeq series) and two (for newer NextSeq/NovaSeq devices). The currently available low-end Illumina sequencer (iSeq100) is not truly compatible with RNA modification analysis due to its very low throughput (4 mln reads/run), and more productive machines, at least MiniSeq or MiSeq (25 mln reads/run) are required. Optimal performance is achieved with NextSeq series (NextSeq550 and NextSeq2000) which are able to generate 400–1000 millions reads/run since RNA modification analysis routinely requires 15-25 mln of raw reads per sample, or more (up to 100–200 mln) in case of whole-transcriptome RNA modification analysis. Even if the sequencing
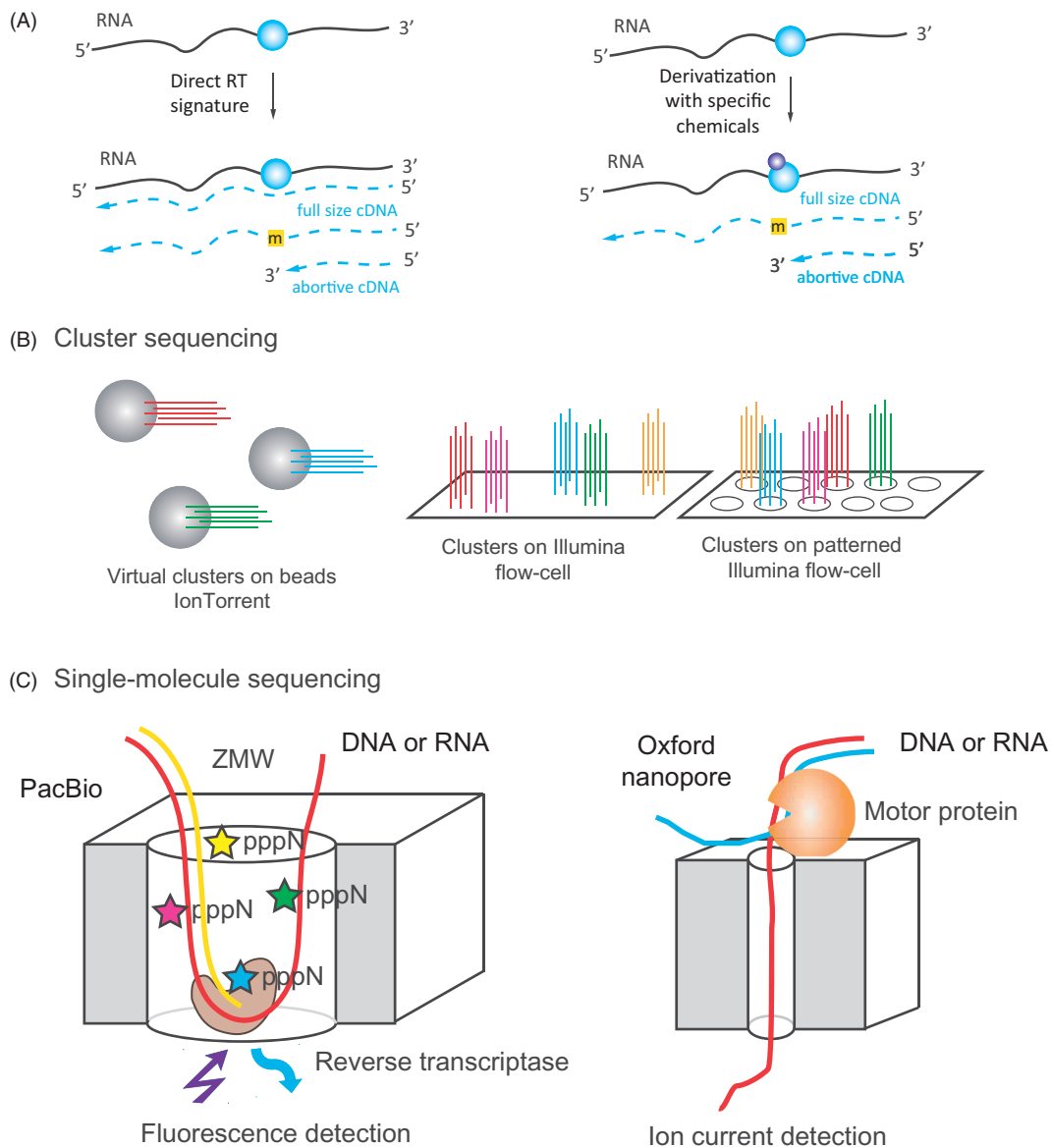
Figure 4. (A) RNA modifications leave RT-signature traces in cDNA during primer extension (RT) step. Those may be "mutations" or more complex profiles including nucleotide mis-incorporations, abortive cDNA synthesis or both. (B) Cluster sequencing in second generation (NGS) technologies. Clusters of identical DNAs are either formed by emulsion PCR on the surface of sequencing beads (currently used by Ion Torrent), or as physical clusters on the surface of a standard or patterned flow cell (Illumina). The signal is highly amplified since all DNA strands in the cluster have the same sequence. (C) Single molecule sequencing by PacBio ZMW chip or nanopore. ZMW well contains a single RT molecule attached to its bottom and generates fluorescent signal only during retention of the fluorescent nucleotide in the enzyme active site. For nanopore sequencing, DNA (or RNA for direct RNA sequencing) are pulled through a nanopore, generating alterations of ion current. RNA modification basecalling may be indirect (using errors/misincorporations observed in cDNA), or direct, by analysis of ion traces. See colour version of this figure at www.tandfonline.com/ibmg.

quality of the newest two colors NextSeq/NovaSeq machines is somehow lower, compared to the four colors HiSeq, the use of patterned flow cell limits overloading issues and improves the quality of the raw reads. The only current competitor technology is not second-generation sequencing, but single-molecule nanopore sequencing, as discussed below.

Analysis of RNA modifications by NGS

Different principles are currently used for the detection of RNA modifications in the epitranscriptome using NGS and NNGS. We classify them in:

1. analysis of RNA-signatures (natural/enhanced or chemically induced) visible in sequencing profiles

2. treatment-induced cleavage of the RNA phosphodiester chain followed by selective ligation of sequencing adapters
3. affinity-based enrichment protocols exploiting specificity of polyclonal or monoclonal antibodies (ab) and specific enzymes installing modifications in RNA.

In many instances, the developed protocols use a combination of these different principles (like ab-driven enrichment followed by specific chemical treatment). Several recent reviews provide more detailed information on these subjects (Helm and Motorin 2017; Hartstock and Rentmeister 2019; Krogh and Nielsen 2019; Linder and Jaffrey 2019; Motorin and Helm 2019; Zhao et al. 2020).

Naturally existing RT signatures of modified nucleotides

A natural RT signature consists of an altered reading of the modified nucleotide during primer extension by RNA-dependent DNA polymerase (reverse transcriptase, RT). Depending on the nature of the nucleotide such a signature may represent a "mutation" compared to the expected reference sequence, or constitute a more complex profile composed of mis-incorporations at different proportions in addition to the abortive RT products ending at the modified nucleotide (Ryvkin et al. 2013; Hauenschild et al. 2015; Tserovski et al. 2016). Such RT signatures can be altered or manipulated by pretreatment of the RNA template (e.g. by demodification/removal of modified residues) (Cozen et al. 2015; Zheng et al. 2015), by the choice of particular conditions of the primer extension reaction (Incarnato et al. 2017; Kristen et al. 2020), or by the use of non-natural dNTP substrate(s) (Hong et al. 2018) and RT active site mutants (Aschenbrenner and Marx 2016; Aschenbrenner et al. 2018). In vivo metabolic labeling with SAM analogs also allowed the incorporation of reactive chemical groups at positions of certain modifications (instead of methyl groups) and thus also helps to alter RT signature of the RNA template (Hartstock et al. 2018; Shu et al. 2020).

Chemically induced RT signatures or RT stops

The most trusted methods for RNA modification analysis are based on the use of specific chemical reagents able to distinguish modified RNA residues from unmodified counterparts (Behm-Ansmant et al. 2011; Heiss and Kellner 2017) as discussed in the Section "Advances in chemical labeling of modified RNA in the context of instrumental analysis".

Chemically induced cleavage of the ribose-phosphate backbone and selective ligation

Another application consists of a specific chemical cleavage of the RNA ribose-phosphate backbone at an RNA modification by one specific reagent or combination of reagents. In some instances, enhanced cleavage (positive signal) is in fact replaced by increased protection against cleavage (negative signal), thus allowing to distinguish modified and unmodified residue. The most popular applications of this approach are:

1. detection of 2'O-methylations (Nm) by RiboMethSeq (Birkedal et al. 2015; Marchand et al. 2016) and Nm-Seq/RibOxiSeq (Dai et al. 2017; Zhu et al. 2017)
2. detection of 7-methylguanosine (m^7G) (Lin, Liu, et al. 2018; 2019), 3-methylcytosine (m^3C)/dihydrouridine (D)/5-hydroxycytosine (ho^5C) by AlkAnilineSeq (Marchand et al. 2018)
3. mapping and quantification of pseudouridines (Ψ) by HydraPsiSeq (Marchand et al. 2020). A selective RNA protection against enzymatic cleavage was implemented for m^6A detection by MazF endonuclease (Garcia-Campos et al. 2019).

Antibody-based enrichment methods (MeRIP-Seq, i/miCLIP)

The use of the specific antibodies for the detection of RNA modifications was already proposed and successfully implemented in the late 70s (reviewed in (Feederle and Schepers 2017)). This development is still ongoing and, in few instances, highly specific antibodies can be obtained (Matsuzawa et al. 2019). However, the majority of antibodies against RNA modified nucleotides/nucleosides have poor affinity and specificity (Mishima et al. 2015) and enrichment factors for modified RNA are only very modest (Slama et al. 2019). Taking these considerations into account, it is not surprising that multiple artifacts in RNA modification mapping result from antibody cross-reactivity, uncertain specificity, and low enrichment (discussed in (Grozhik et al. 2019; McIntyre et al. 2020)). Despite these limitations, RNA modification-specific antibodies are widely used to RIP and CLIP protocols applied to RNA modification mapping.

Analysis of RNA modifications by NNGS (single molecule sequencing)

The use of single-molecule sequencing approaches (NNGS or third-generation deep sequencing) is an

attractive alternative to the classical cluster sequencing protocols. Indeed, cluster sequencing involves amplification steps, providing only an average picture of modifications in a population of RNA molecules. To get information about the exact combination of modifications in a given RNA molecule (individual modification pattern), a single molecule analysis should be performed (Xu and Seki 2020).

The proof of principle for the analysis of RNA modifications (namely m^{6}A) by single-molecule sequencing was established already in 2013, by using PacBio SMRT technology. HIV-1 and AMV RT were loaded to zero-mode waveguides (ZMWs) arrays and extension of DNA primer on m^{6}A-modified RNA template was monitored (Vilfan et al. 2013). Even if the precision of the RNA sequencing remains limited, the analysis of the reverse transcriptase kinetics can be used to discriminate RNA base modifications.

More recently, single-molecule direct RNA sequencing by nanopores (Oxford Nanopores) was used for the detection of RNA modifications. It was demonstrated that m^{6}A RNA modifications can be detected with high accuracy, in the form of systematic errors and decreased base-calling qualities (Liu et al. 2019). Analysis of raw ion current profiles for direct MinION nanopore sequencing of full-length 16S rRNA revealed 7-methylguanosine (m^{7}G) as well as pseudouridine modifications (Smith et al. 2019). The major challenge in the field of direct RNA sequencing and RNA modification mapping by nanopores consists of appropriate data analysis softwares and algorithms. Analysis can be either done by standard base-calling and identification of "sequencing signatures" (Lorenz et al. 2020), or by extremely laborious, but direct analysis of ion current traces (Cozzuto et al. 2020; Ding et al. 2020; Jenjaroenpun et al. 2021).

Mass spectrometry

Mass spectrometry is a valuable tool to analyze RNA modifications. Low-resolution instruments such as quadrupoles are usually very sensitive and can detect attomole amounts of low abundance modifications. High-resolution MS, such as time-of-flight (TOF) or orbi-trap instruments, can detect the exact mass of a modification and allow for elemental composition prediction and subsequent fragmentation of the modification, which can provide further structural information. The analysis of RNA can be conducted in a variety of ways and when combined, a broad spectrum of information becomes available. Depending on the enzymatic

treatment of the RNA prior to MS, we discuss 3 major types of MS analyses:

1. Top-down analysis of non-hydrolyzed RNA allows for total mass analysis and modifications can be identified and their location can be determined specifically within the sequence context
2. Bottom-up MS of partially hydrolyzed RNA allows for mass mapping and provides some sequence context for modifications but requires a sequence to map it back to
3. Nucleoside MS of fully hydrolyzed RNA can provide information on the chemical identity of modifications even at incredibly low abundance

Top-down MS

Top-down MS features the injection of undigested RNA into the mass spectrometer. This idea first came up in the early 1990s (McLuckey et al. 1992; Limbach et al. 1995) when oligonucleotides were "sequenced" using mass spectrometers and when MS was used to determine the molecular mass of tRNA isoacceptors and 5S rRNA and to detect potential modifications (Limbach et al. 1995). Later, in the mid to late 2000s approximately 20 nt long oligonucleotides were analyzed (Kellersberger et al. 2004; Huang et al. 2008) and in 2010 the length was extended to 61 nt (Taucher and Breuker 2010). At around that time, the first studies analyzing full-length tRNA by QqTOF (Huang et al. 2010) and FT-ICR (Taucher and Breuker 2012) mass spectrometers appeared. The study using the FT-ICR mass spectrometer achieved full sequence coverage by combining Electron Detachment Dissociation (EDD) and Collision-Activated Dissociation (CAD) experiments. Using the nomenclature for nucleotide fragments (McLuckey et al. 1992), CAD results mainly in complementary (i.e. the sum of their masses equals the mass of the intact RNA) c- and y-type fragment ions along with undesired base-loss fragments and secondary fragments where neither the 3′- nor the 5′-terminus is present (Figure 5 (A–C)).

The EDD pathway produces non-complementary d- and w-type fragment ions, whose masses add up to the mass of the intact tRNA + 97.98 Da (+H$_3$PO$_4$) (Taucher and Breuker 2010, 2012). This information can be combined to sequence RNAs and to locate the mass-altering post-transcriptional modifications present in the sample. Other dissociation techniques for oligonucleotides that were investigated in the past include UV-Photo Dissociation (Smith and Brodbelt 2010, 2011), Infrared Multiphoton Dissociation (Smith and Brodbelt 2011),
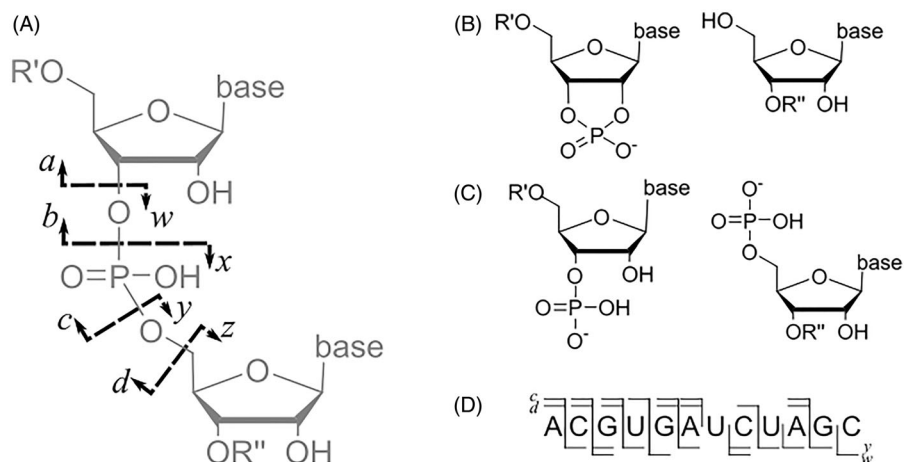
Figure 5. (A) Nomenclature of top-down fragment ions (McLuckey et al. 1992). (B) Complementary c (left) and y (right) type fragment ions. (C) Non-complementary d (left) and w (right) type fragment ions. (D) Possible cleavage map of a 12 nt long RNA.

and Electron Photodetachment Dissociation (Yang and Håkansson 2009; Smith and Brodbelt 2010), Electron Capture Dissociation (Mo and Håkansson 2006) and Negative Electron Transfer (Huang and McLuckey 2011; Gao and McLuckey 2013). An important measure for assessing top-down techniques is called sequence coverage. It describes the number of cleavage sites where at least one resulting product ion could be detected and is typically reported as a sequence cleavage map (Figure 5(D)).

In 2020, the Breuker lab published a report of the novel dissociation technique Radical Transfer Dissociation (RTD) (Calderisi et al. 2020). In RTD, the reagent cobalt(III)hexamine is concomitantly introduced into the mass spectrometer to induce the formation of RNA radical ions, which dissociate into d- and w-type fragments upon collisional activation. At the same time, c- and y-type ions are produced through the well-known CAD phosphodiester bond cleavage. In their publication, they report full sequence coverage of up to 39 nt long RNAs. The proposed mechanism for radical dissociation neither involves the nucleobase nor the 2'-OH group, consequently, RTD might be especially useful for the analysis of modified RNA. In addition, base-loss and internal fragmentation are reduced under RTD. Therefore, it is especially recommended for mapping of labile RNA modifications such as 5-formylcytidine and wybutosine and its derivatives. Moreover, RTD has comparably low requirements for instrumentation: Measurements can be conducted on any mass spectrometer that is equipped with an ESI source and a collision cell (Calderisi et al. 2020).

More recently in 2020, the Coon lab used fluoranthene cations as a Negative Electron Transfer reagent and an infrared laser to excite precursor ions (activated-ion negative electron transfer dissociation, AI-NETD)

(Peters-Clarke et al. 2020). Similar to what has been demonstrated for other electron-based MS/MS approaches, d- and w-type ions predominate in the MS/MS spectra generated by AI-NETD and only minute levels of base-loss and internal fragment ions are observed. The application of this method leads to full sequence coverage of the 21 nt long luciferase antisense siRNA when all fragment ion types are considered. A major benefit of this approach is the independency of FT-ICR instrumentation and thus it is more accessible. Another advantage highlighted by the authors is the lower complexity of spectra generated by their approach in comparison to traditional collision-based dissociation methods, as this method produces fewer secondary and base loss fragments. However, while full sequence coverage of 21 nt luciferase antisense siRNA was obtained by AI-NETD, its performance on longer RNA and more complex mixtures remains to be investigated. RNA-protein complexes can also be investigated using top-down analysis as recently reviewed (Schneeberger and Breuker 2017).

To summarize, top-down analysis of RNA is a technology that is still being developed. Its major benefits comprise the ability to perform de-novo sequencing, identify mass-altering modifications of nucleobases as well as localize these modifications. Furthermore, laborious digestion steps are not necessary. However, the fact that equal-mass modifications (e.g. m^1A and m^6A) cannot be distinguished from each other is a major disadvantage. In addition, the analysis of mass-silent modifications (such as pseudouridine) is challenging. Moreover, pure RNA is required for the top-down analysis, as the analysis of mixtures is laborious or even impossible. Further, the analysis and interpretation of top-down MS spectra are less straightforward due to high spectral complexity, especially as currently there is

no software available to support top-down MS analysis of RNA. With the influx of oligonucleotide-based vaccines and therapeutics exiting clinical trials and entering the pharmaceutical market, we expect an increased interest in the analysis of full-length RNA and thus finally software tools for data processing will be developed.

Modification mapping of oligonucleotides by LC-MS/MS

In the bottom-up approach, a single RNA or a mixture of RNA sequences are digested to oligonucleotides amenable to separation and subsequent analysis by tandem mass spectrometry (MS/MS). The general bottom-up approach for RNA modification mapping has not changed since introduced by McCloskey and colleagues over 25 years ago (Kowalak et al. 1993). Moreover, the key experimental factors and considerations that impact the success of this approach are well known. These include the approach for generating oligonucleotides from the RNA sample; chromatographic methods for the separation of oligonucleotide mixtures; MS and MS/MS approaches to generate sequencing ladders from oligonucleotides; and the data analysis steps that can simplify both interpretation of MS/MS data, RNA sequence reconstruction and subsequent mapping of modifications onto those RNA sequences.

In general, a review of advances in each of these areas finds that there have been incremental improvements within each step that – when combined – have led to major enhancements in the overall analytical strategy. These enhancements now permit the complete mapping of RNA modifications onto messenger RNA (mRNA) sequences, total transfer RNA (tRNA) mixtures from organisms, and ribosomal RNAs (rRNAs) (Jiang et al. 2019; Jora et al. 2019; Yu et al. 2019; Thakur et al. 2020; Wein et al. 2020). While these are impressive achievements, the field could benefit from a radical rethinking of what a bottom-up approach to RNA modification mapping could entail and achieve. In particular, given the advances in NGS and NNGS methods including the goal of single copy RNA sequencing and modification placement, which are built off a completely different technology platform, the inherent advantages of MS for directly detecting an intrinsic property of the molecule (i.e. the *m/z* or mass of the molecule) argues for continued investments and advancements in this platform for RNA modification mapping.

Digestion of RNA sample into oligonucleotides
Traditionally, RNA is enzymatically digested to oligonucleotides using site-specific ribonucleases (RNases). The most commonly used RNases are RNase T1 (guanosine-, N^2-methylguanosine-, and inosine-specific), RNase A (pyrimidine-, and pseudouridine-specific), and RNase U2 (purine-selective). In general, RNases are selected based on the desirable generation of relatively longer (5-15 nucleotides long) and sequence-unique oligonucleotides. While the use of single RNases has been proven effective throughout the years, the inherent challenge has been that a single RNase limits the analyst's ability to accurately reconstruct the entire RNA sequence. Not surprisingly then, much of the recent work in this area has focused on identifying and implementing new RNases that can be used in combination to increase sequence coverage through the generation of overlapping digestion products (Figure 6).

In a study mapping *E. coli* total tRNA, Addepalli and coauthors (Thakur et al. 2020) demonstrated that the combination of RNases T1, MC1 (uridine-specific), and cusativin (cytidine-specific) led to an overall ≥75% sequence coverage. They also showed that for *E. coli* 23S rRNA, sequence coverage was enhanced from ~25% (single RNases) to 85% when the same set of RNases were combined. In a similar fashion, Hua et al. have shown that similar sequence coverage is achieved when RNases T1, colicin E5 (cleaves between guanosine and uridine), and mazF (cleaves at ACA sequences) are used in parallel to map mRNAs (Jiang et al. 2019). An alternative approach was the development of a nonspecific RNase U2 mutant, which led to complete sequence coverage of modified RNAs through the generation of overlapping digestion products in a single-pot digestion reaction (Solivio et al. 2018).

While these technical achievements demonstrate new analytical approaches that can be taken, most of the RNases employed in these seminal studies (i.e. MC1, cusativin, colicin E5, and U2 mutant) were produced in-house. As a result, such approaches are commonly plagued by the lack of batch-to-batch or interlaboratorial reproducibility of in-house expressed and purified RNases. Thus, in addition to illustrating the efficacy of bottom-up MS-based approaches to locate modified nucleosides within different RNA species, these studies strongly advocate for the batch production and commercialization of reproducible complementary RNases.

If the analysis of multiple tRNAs is desired it needs to be taken into account that lots of tRNA isoacceptors do have similar sequences and therefore similar or even equal masses, which hampers sequencing by mass spectrometry, as isotopic pattern overlays may occur often. In order to overcome this obstacle, different isoacceptors can be separated by two-dimensional polyacrylamide gel (2D gel). This way $tRNA_{GCU}^{Ser}$ and $tRNA_{CAG}^{Leu}$
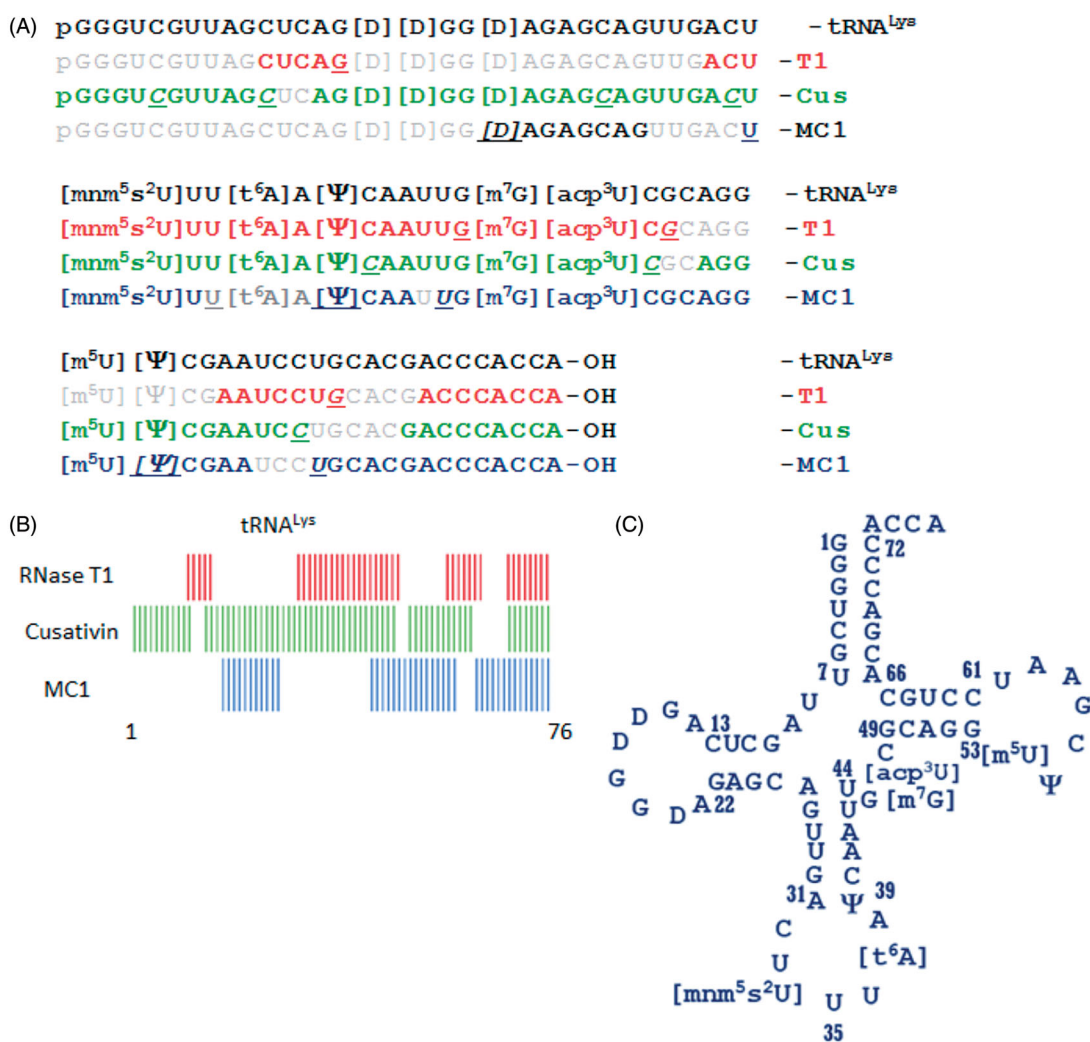
Figure 6. RNA modification and sequence mapping of *E. coli* tRNA^{Lys} from the digestion products of RNases T1, cusativin, and MC1. (A) The observed digestion products of each ribonuclease are matched against the known sequence of tRNA^{Lys}. The cleavage sites for each enzyme is underlined in an oligomer. Note the overlapping regions between the sequences of digestion products. (B) A schematic view of looking at the overlaps between the observed digestion products of the three ribonuacleases. (C) The clover-leaf model of 2D structure of *E. coli* tRNA^{Lys} is depicted. Figure reproduced with permission from (Thakur et al. 2020). See colour version of this figure at www.tandfonline.com/ibmg.

were sequenced. However, a major weakness is a necessity for ethidium bromide staining – a hazardous substance, which renders the technique obsolete (Fradin et al. 1975; Gaston and Limbach 2014). To obtain specific short oligonucleotide fragments, the iso-acceptors are digested with certain RNAses in an "in gel digestion". Subsequently, these oligonucleotides are separated on a nano-ion-pair reversed-phase high-performance liquid chromatography (nano-IP-RPHPLC) (Taoka et al. 2010) and are analyzed by MS/MS. This method enables the identification of tRNA isoacceptors by using genomic databases. Mass-silent modifications such as pseudouridine cannot be detected by this method, thus, CMCT derivatization is recommended prior to analysis. Moreover, methyl groups cannot be localized in modified nucleotides. However, if the tRNA

sequence is known, it is possible to predict the respective methyl group position.

While conventional mapping of modifications by MS is done through the use of RNase-based digestion, Zhang et al. have conceptually demonstrated the sequencing and modification mapping of RNA through sequence ladders generated by acidic digestion (50% v v^{-1} formic acid) (Bjorkbom et al. 2015; Zhang, Shi, et al. 2019). While the efficacy of the approach on more labile modifications and more complex RNA mixtures has not been demonstrated yet, their 3′- and 5′-end labeling strategies (through the addition of tags such as biotin, cyanine3, and cyanine5) (Zhang, Shi, et al. 2019) may pave the way for a multiplexed qualitative and quantitative analysis of modified oligonucleotides by bottom-up MS approaches. Looking forward, alternative

approaches generating longer oligonucleotides that remain amenable to mass spectrometry sequencing and analysis would greatly benefit the field. As the read length increases, the analyst's confidence in modification mapping accuracy improves. Further, strategies such as end-labeling could lead to significant enhancements in throughput as compared to what is currently achievable by more conventional approaches.

HPLC separation of oligonucleotides

In addition to developments in sample preparation strategies, separation approaches employed for modification mapping have also been improved. In general, oligonucleotides are subject to chromatography prior to MS analysis. To date, the most common chromatography approach is the use of reversed-phase high-performance liquid chromatography (RP-HPLC). In general, the typical RP-HPLC mobile phases – both aqueous and organic – can be selected for compatibility with MS conditions. Moreover, standard RP-HPLC stationary phases and columns have been optimized for various flow rates amenable to electrospray ionization.

Due to the relatively high hydrophilicity conferred by the negatively charged phosphate backbone, the use of ion-pairing reagents as mobile phase modifiers is needed to achieve chromatographic retention and selectivity of oligonucleotides during MS analysis (Biba et al. 2017; Li et al. 2018). Such a chromatographic approach is named ion-pair reversed-phase liquid chromatography (IP-RP-LC). Different alkylamine modifiers have been used during IP-RP-LC of oligonucleotides, and broadly investigated by the Bartlett lab (McGinnis et al. 2013; Li et al. 2018). Some of their recent studies have led to models able to predict optimal ion-pairing reagents based on sample hydrophobicity (McGinnis et al. 2013), as well as a better understanding of the retention mechanism under ion-paring conditions (Li et al. 2018). Nonetheless, the use of alkylamine triethylamine (TEA) and the counter ion hexafluoroisopropanol (HFIP), developed by Apffel et al. (Apffel et al. 1997), is still a method of choice for oligonucleotide analysis by LC-MS (Baldridge et al. 2018; Lin, Miyauchi, et al. 2018). While TEA contributes to excellent chromatographic performance, HFIP helps with the ionization of oligonucleotides, improving the sensitivity of oligonucleotide analysis (Biba et al. 2017). These chromatographic conditions under nanoflow-based LC-MS analysis have been used to map modifications in nanogram aliquots of complex RNA samples (Yamauchi et al. 2016; Taoka et al. 2018; Nakayama et al. 2019; Wein et al. 2020).

One drawback of IP-RP-LC is that the mobile phase modifiers (alkylamines/HFIP) are often linked to ion suppression, observed when subsequent LC-MS analyses are conducted in positive ion mode (e.g. during proteomics analyses). Such contamination concerns have led to the common practice of having dedicated LC and MS systems for oligonucleotide analysis, limiting the number of groups interested in this field. Due to the costly burdens of dedicated instruments, alternatives to IP-RP-LC of oligonucleotide have been investigated. The Kellner lab has shown that modifications on tRNA transcripts may be mapped using similar conditions used for nucleoside analysis (Hagelskamp et al. 2020). In their approach, chromatographic separation of modified oligonucleotides is achieved using an octadecylsilane (C18)-derived RP-LC column (Fusion-RP), and mobile phases consisting of water, acetonitrile, and ammonium acetate. The mass spectrometric analysis is done in positive ion mode (as opposed to the traditionally used negative ion mode). While the efficacy of this method to chromatographically resolve complex oligonucleotide mixtures (e.g. total tRNA digests) is yet to be demonstrated, it is a suitable alternative for the study of synthetic or purified RNA species. Such an approach led to identifying that A34 of $tRNA_{AAC}^{Val}$ is deaminated by the human adenosine deaminase tRNA specific enzyme 2/3 (ADAT2/3) (Hagelskamp et al. 2020). It also supported the discovery that *E. coli*'s alpha-ketoglutarate-dependent hydroxylase AlkB is an eraser with higher specificity toward m^1A1408 of the 16S rRNA (Hagelskamp et al. 2020).

Among the alternatives being explored to IP-RP methods, hydrophilic interaction liquid chromatography (HILIC) represents a promising alternative. The Limbach lab has shown that a performance comparable to IP-RP-LC in terms of chromatographic retention, selectivity, and resolution, as well as mass spectrometric sensitivity, is feasible through HILIC-MS (Lobue, Jora, et al. 2019). In this approach, water-, acetonitrile-, and ammonium acetate-based mobile phases are used. However, the column used in their study (HILICpak VN-50 2D) is only commercially available in relatively large particle size (5 μm) and limited column dimensions (2.0- or 4.6-mm internal diameter, by 150-mm length), thus limiting the chromatographic and mass spectrometric performance of the approach. Another recent study has demonstrated that similar chromatographic performance (under similar mobile phase conditions) may be achieved during oligonucleotide analysis using another commercially available HILIC column (BEH amide) (Demelenne et al. 2020). This column is available in a wider range of column and particle sizes, thus expanding possibilities in terms of method development and optimization. Continued efforts around HILIC-based

approaches, and other alternative LC methods, are needed to identify those LC-MS conditions that can achieve the performance found in current IP-RP methods while minimizing concerns about instrument carry-over, contamination, and sample suppression that are often shared anecdotally as reasons to avoid LC-MS analyses of oligonucleotides.

Independent of the LC method, modification mapping of complex RNA samples (such as total tRNA) by LC-MS is challenging. More often than not, oligonucleotides generated from RNase digestion of RNA mixtures are challenging to completely resolve chromatographically. Further, another challenge is the finite dynamic range (i.e. difference between the most abundant and least abundant species) of mass spectrometers (usually up to 10^4–10^5) (Zubarev 2013) that impact the ability to detect co-eluting oligonucleotides. To overcome such challenges, labor-intensive offline methods that minimize sample complexity are traditionally employed prior to RNase treatment and LC-MS analysis (Kumazawa et al. 1992; Suzuki and Suzuki 2007). In a more elegant fashion, online approaches that explore the orthogonality of the different LC modes employed during oligonucleotide analyses have been recently developed (Goyon and Zhang 2020; Li et al. 2020). More specifically, one-dimensional anion exchange chromatography (AEX) or IP-RP-LC methods have been coupled to HILIC, IP-RP-LC, or RP-LC as the second LC dimension to facilitate the characterization of synthetic oligonucleotide impurities (Goyon and Zhang 2020; Li et al. 2020). Considering the demonstrated potential of two-dimensional-LC (2D-LC) approaches to resolve synthetic oligonucleotides, we anticipate that 2D-LC approaches have the potential to facilitate modification mapping in complex biological samples. In this sense, a comprehensive investigation regarding the orthogonality of the different LC modes employed for modification mapping, as well as investigation of the retention mechanisms driving the separation of oligonucleotides in these different LC modes is advised.

MS/MS sequencing of oligonucleotides

The information generated through tandem mass spectrometry (MS/MS) plays a pivotal role in successful modification mapping. Among MS/MS tools, collision-induced dissociation (CID) is the most widespread fragmentation technique used. Under CID, precursor ions are accelerated by an electric field, and through collisions with a neutral gas (generally He, Ar, or N_2) bond cleavage is enabled, generating fragment ions. During oligonucleotide analysis, CID leads to fragmentation within the phosphodiester backbone of oligonucleotides in a ladder-fashion, generating c-, y-, a-, and w-type fragment ions, with c and y being the most abundant ions (Figure 5). Characteristic sequence ladders and mass shifts observed in these fragment ions help locate modification(s) within a given oligonucleotide precursor ion.

Mapping positional isomers such as 1-methyladenosine (m^{1}A), and N^6-methyladenosine (m^{6}A) within an RNA sequence by LC/MS-MS is not a straightforward task. Moreover, while pseudouridine (Ψ) may be easily characterized during nucleoside analysis, it adds another layer of complexity during modification mapping experiments. Because it is isomeric with uridine (U), pseudouridine cannot be directly identified based on its mass (or m/z) during oligonucleotide analysis by LC-MS/MS. Derivatization strategies mentioned earlier have been implemented for differentiating pseudouridine from uridine (Patteson et al. 2001; Mengel-Jørgensen and Kirpekar 2002; Durairaj and Limbach 2008). Even though derivatization facilitates mapping, this strategy requires extra experimental steps and is targeted to one modification (or class) at a time. Pomerantz and McCloskey have explored the characteristic C-glycosidic bond present in Ψ and the absence of "a-base" ions at the Ψ site to facilitate mapping of Ψ-containing oligonucelotides through MS/MS (Pomerantz and McCloskey 2005). In a similar fashion, the Taoka lab developed a pseudo-MS^3-based mapping strategy that allows direct identification and mapping of ribonucleoside isomers. In their approach, two rounds of LC-MS/MS analyses are performed. In the first round, traditional oligonucleotide analysis is carried out. As such, modifications may be mapped, but the nature of the isomer present in a given oligonucleotide may not be identified. In the second round, in-source CID (in which c-, y-, w-, a-type ions are generated) and higher-energy collisional dissociation (HCD, in which ribonucleoside fingerprints are generated) are performed. In this step, the nature of the isomers (mapped within a given oligonucleotide in the first step) is identified. Such an approach has been used to identify Ψ in the human spliceosomal snRNAs (Yamauchi et al. 2016) and in the rRNA of *Saccharomyces cerevisiae* (Taoka et al. 2016), as well as to map modifications present in the rRNAs of the human 80S ribosome (Taoka et al. 2018) and in the rRNAs of *Leishmania donovani* (Nakayama et al. 2019).

The nature of MS/MS spectra generated through collision-based dissociations (CID and HCD) are highly dependent on the charge state and length of the oligonucleotide. Moreover, they may be convoluted due to the occurrence of ions resulting from internal

fragmentation and neutral loss of labile nucleobases (Schurch 2016), thus impacting data processing and sequence coverage. Alternative MS/MS strategies employing electrons or photons have been investigated throughout the years (Smith and Brodbelt 2010; Taucher and Breuker 2010; Huang and McLuckey 2011). Recently the Coon lab proposed activated-ion negative electron transfer dissociation (AI-NETD) as an alternative tool for oligonucleotide mapping and the Breuker lab developed a fragmentation approach termed radical transfer dissociation (RTD) (Calderisi et al. 2020). These two approaches are discussed in detail in the chapter *Top-down MS*.

It is undeniable that these MS/MS strategies facilitate modification mapping. Nonetheless, challenges still exist when it comes to improving the acquisition of MS/MS spectra of low abundance modified oligonucleotides. MS/MS analysis of oligonucleotides is traditionally done using data-dependent acquisition (DDA) mode (Cao and Limbach 2015). Here, precursor ions detected in the first MS stage are sequentially selected (usually the top 3–10 most abundant ions) for MS/MS analysis. Consequently, each MS/MS acquired spectrum has a precursor ion *m/z* linked to it, thus facilitating data interpretation. Because ions are introduced to the mass spectrometer for a limited amount of time (i.e. chromatographic peak width), and the co-elution issues mentioned above, it is not always possible to acquire MS/MS spectra for all oligonucleotides present in a given sample. To enhance the acquisition of MS/MS spectra of modified oligonucleotides, the Limbach lab has implemented an exclusion list-based protocol (Cao and Limbach 2015). As such, *m/z*s corresponding to unmodified oligonucleotides are excluded from MS/MS analysis. This exclusion list-based approach improved RNA modification mapping of bacterial and archaeal total tRNAs by 10–25% when compared to standard DDA. What remains to be shown is whether this or other DDA-based strategies will enable modification mapping of more complex RNA mixtures, including those from mammalian samples.

In addition to DDA, MS/MS spectra may also be acquired using data-independent acquisition (DIA) mode (Fernández-Costa et al. 2020). As such, all precursor ions detected at a given time within a predefined *m/z* range are concomitantly fragmented, thus resulting in a multiplexed MS/MS spectrum. Despite the challenges associated with the deconvolution of DIA-based MS/MS spectra, this strategy has been largely employed for protein sequencing (including mapping of post-translational modification) by LC-MS/MS (Zhang et al. 2020). Nonetheless, it has yet to be explored for RNA

modification mapping. Considering the potential of the approach to facilitate modification mapping in complex samples (e.g. human total tRNA), developments in this area are advised.

Software to analyze LC-MS/MS data

A few significant software and bioinformatics resources have been developed to facilitate modification mapping (Figure 7) (Rozenski and McCloskey 2002; Matthiesen and Kirpekar 2009; Nakayama et al. 2009; Nyakas et al. 2013; Sample et al. 2015; Paulines et al. 2019). Improved performance has been achieved by two recently developed open-source tools, RNAModMapper (RAMM) (Yu et al. 2017; Lobue, Yu, et al. 2019) and NucleicAcidSearchEngine (NASE) (Wein et al. 2020). Both software platforms offer algorithms capable of interpreting and mapping modifications in large LC-MS/MS datasets obtained during analysis of complex oligonucleotide mixtures (Baldridge et al. 2018; Solivio et al. 2018; Yu et al. 2019; Thakur et al. 2020; Wein et al. 2020). RAMM is a user-friendly tool that maps modifications via two independent modes: fixed, for targeted modifications/positions; and variable, in which modifications are mapped in an untargeted fashion. Data interpretation is done based on RNA/DNA sequence inputs, through a two-component scoring function, with user-defined and instrument-based scoring thresholds. NASE is an OpenMS-based (Rost et al. 2016) search engine that presents extra functionalities such as high-quality data visualization and quantification capabilities. Because it supports decoy RNA sequences as input in addition to the target sequence, false discovery rates are minimized in such a tool. Moreover, the incorporation of *m/z* corrections due to salt adducts (traditionally observed during oligonucleotide analysis (Sutton and Bartlett 2020)), and mass defects resulting from instrumental selection of higher abundance isotopologue peaks, significantly enhances the ability of NASE to identify modified oligonucleotides. Although these (and earlier developed) tools have paved the way to facilitate LC-MS/MS spectral interpretation of oligonucleotide data, manual review is still a necessary step. Thus, the community has yet a lot to gain from further developments on software platforms dedicated to MS-based modified oligonucleotide mapping.

Nucleoside MS

Unlike the previously described MS methods, nucleoside MS relies on complete enzymatic hydrolysis of RNA into its nucleoside building block. The hydrolysis is
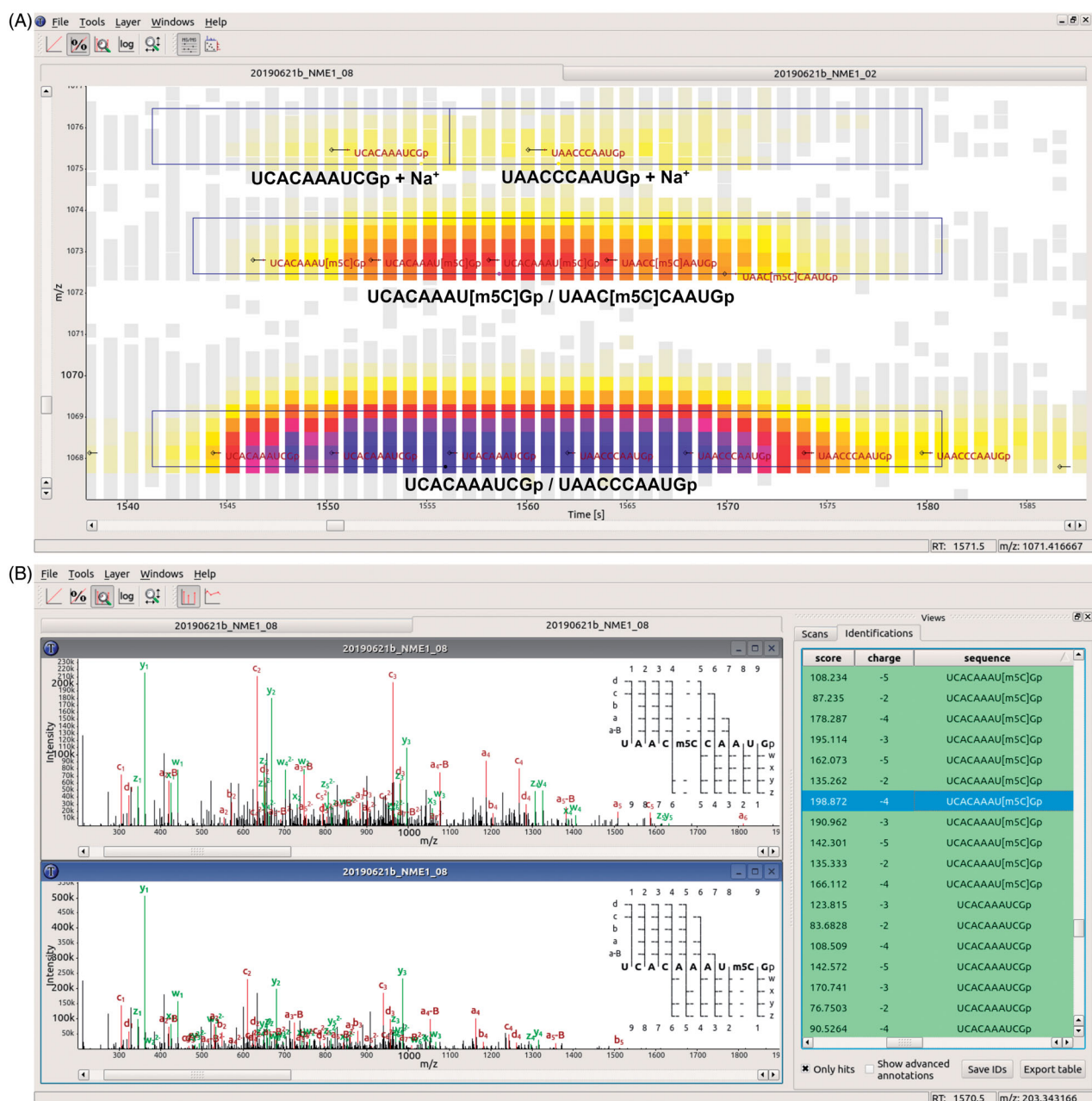
Figure 7. Interactive data visualization using TOPPView, showing data from the NCL1-treated NME1 sample. (a) MS1 view (RT-by-*m/z*) of a data section. LC-MS peaks are shown as small squares, colored according to their signal intensities. Small black diamonds and horizontal lines indicate MS/MS fragmentation events; oligonucleotide sequences identified by NASE from the MS/MS spectra are shown in dark red font. Black boxes outline features detected for label-free quantification, which have been annotated with the corresponding oligonucleotides. All oligonucleotides shown have a charge state of −3. (b) "Identification view" comparing two MS/MS spectra, identified by NASE as the sequences "UAAC[m5C]CAAUGp" and "UCACAAAU[m5C]Gp". Matching peaks between the acquired and theoretical spectra are annotated and highlighted in red and green. On the right in each spectrum plot, an ion coverage diagram shows which of the theoretical fragment ions of the sequence were matched in the MS2 spectrum (in any charge state). Figure reproduced with permission from (Wein et al. 2020). See colour version of this figure at www.tandfonline.com/ibmg.

commonly achieved using endonucleases such as nuclease P1, nuclease S1 or benzonase and supported by the use of phosphodiesterase 1 (Crain 1990; Cai et al. 2015). The resulting 5′-phosphorylated nucleotides are dephosphorylated using alkaline phosphatase and free nucleosides are released. The analysis of hydrolytic digests has several drawbacks. The first is the dependence of the result on the purity of the analyzed RNA. Contaminating RNA will obscure the quantitative result by either introducing unexpected RNA

modifications or by simply diluting the target RNA and thus the number of detectable RNA modifications. Thus, a quality control step, for example, (chip) gel electrophoresis or sequencing is crucial prior to hydrolysis to ensure comparable purity of samples. The second is the introduction of artifacts through the hydrolysis protocol. In the original protocol by Crain and colleagues (Crain 1990), the hydrolysis is performed in a two-step protocol first at pH 5 followed by a pH elevation to pH 8. These conditions may destroy labile RNA modifications such as cyclic N(6)-threonylcarbamoyladenosine (ct^{6}A) which undergoes epimerization under mild alkaline conditions (Matuszewski et al. 2017). Other artifacts such as isocytidines emerge through amination/imination of carbonothiolated nucleosides during RNA hydrolysis (Jora et al. 2020). Furthermore, not all enzymes used for RNA hydrolysis are capable of cleaving modified nucleotides. For example, the He lab recently demonstrated that nuclease S1 alone is not capable of cleaving m^{7}G from the mRNA cap. Only in the presence of phosphodiesterase 1, the complete m^{7}G cap is cleaved and m^{7}G is released (Zhang, Liu, et al. 2019).

In the past, nucleoside hydrolysate separation by thin-layer chromatography (TLC) of ^{32}P radioactively labeled nucleotides (Gupta et al. 1976; Stanley and Vassilenko 1978; Gupta and Randerath 1979; Keith 1995; Zhao and Yu 2004; Grosjean et al. 2007), or gas chromatography after nucleoside derivatization have been reported (Gehrke and Ruyle 1968; Lakings and Gehrke 1971; Gehrke and Patel 1976; Gehrke et al. 1980). Today, the resulting nucleoside mixture is most commonly separated using liquid chromatography (LC) before detection by UV absorption or MS (LC-UV or LC-MS). The field has recently seen substantial advances, especially in the context of separation techniques but also in terms of absolute quantification and analysis of RNA modification dynamics.

Advances in the separation of nucleosides
Most established separation systems rely on the use of reverse phase columns as described elsewhere (Pomerantz and McCloskey 1990). While these systems allow efficient separation and analysis of modified nucleosides from many native samples within less than 10 min (Heiss et al. 2017; Reichle, Kaiser, et al. 2018; Borland et al. 2019), recent advances to improve the separation efficiency have been made.

For example, new stationary phases, based on Hydrophilic interaction liquid chromatography (HILIC), have been reported. For native RNA hydrolysates, HILIC allows separation of nucleosides within 30 min (Lobue,

Jora, et al. 2019). Analysis of modified nucleosides from urine has a great potential as a diagnostic tool (Bryzgunova and Laktionov 2015) and here HILIC has also been used for efficient separation and sensitive detection of cytidine derivatives (Guo et al. 2018).

In 2018 Sarin et al. established a sensitive capillary nanoflow liquid chromatography mass spectrometry (nLC-MS) for absolute quantification of ribonucleosides. While the separation by liquid chromatography of some ribonucleosides remains challenging by the polar attributes of the analytes, the utilization of porous graphitic carbon (PGC) as column material can overcome this obstacle. This material is stable over broad temperature and pH ranges and possesses a polyaromatic scaffold which allows the separation of highly polar, charged, and structural similar analytes (Sarin et al. 2018).

Advances in absolute quantification
As described above, top-down and oligonucleotide MS are commonly done through negative ionization (e.g. deprotonation) of the phosphate backbone before analysis on high-resolution MS instruments. Due to the absence of phosphates after complete RNA hydrolysis to the nucleoside building block, ionization of nucleosides exploits the basic character of the nucleobases which are easily ionized through protonation. Thus, nucleoside MS is performed in positive ion mode. Quantification of small molecules is commonly done on low-resolution mass spectrometers consisting of two quadrupoles for ion selection interconnected through a collision cell that fragments the selected precursor ion (in the past another quadrupole). The principle is shown in Figure 8(A). Such triple quadrupole MS systems (QQQ-MS) are highly sensitive as the double selection of target analytes leads to substantial noise reduction and thus clear signals for quantification are received.

Even though a mass spectrometer might be a sensitive device for RNA modification analysis in the sub-femtomol range, its non-quantitative nature constitutes a drawback. To allow precise and accurate absolute quantification, stable isotope-labeled internal standards are pivotal for analysis. In the literature, the abbreviations SILIS (stable isotope-labeled internal standard) or ISTD/IS (internal standard) are commonly used, but the term SILIS is preferable as it is more clearly defined. In general, an IS is necessary to account for fluctuations in instrumental detection efficiency. In the case of mass spectrometry, fluctuations are caused by instrument runtime in between cleaning procedures, LC buffer composition, ion load of samples, and other mostly
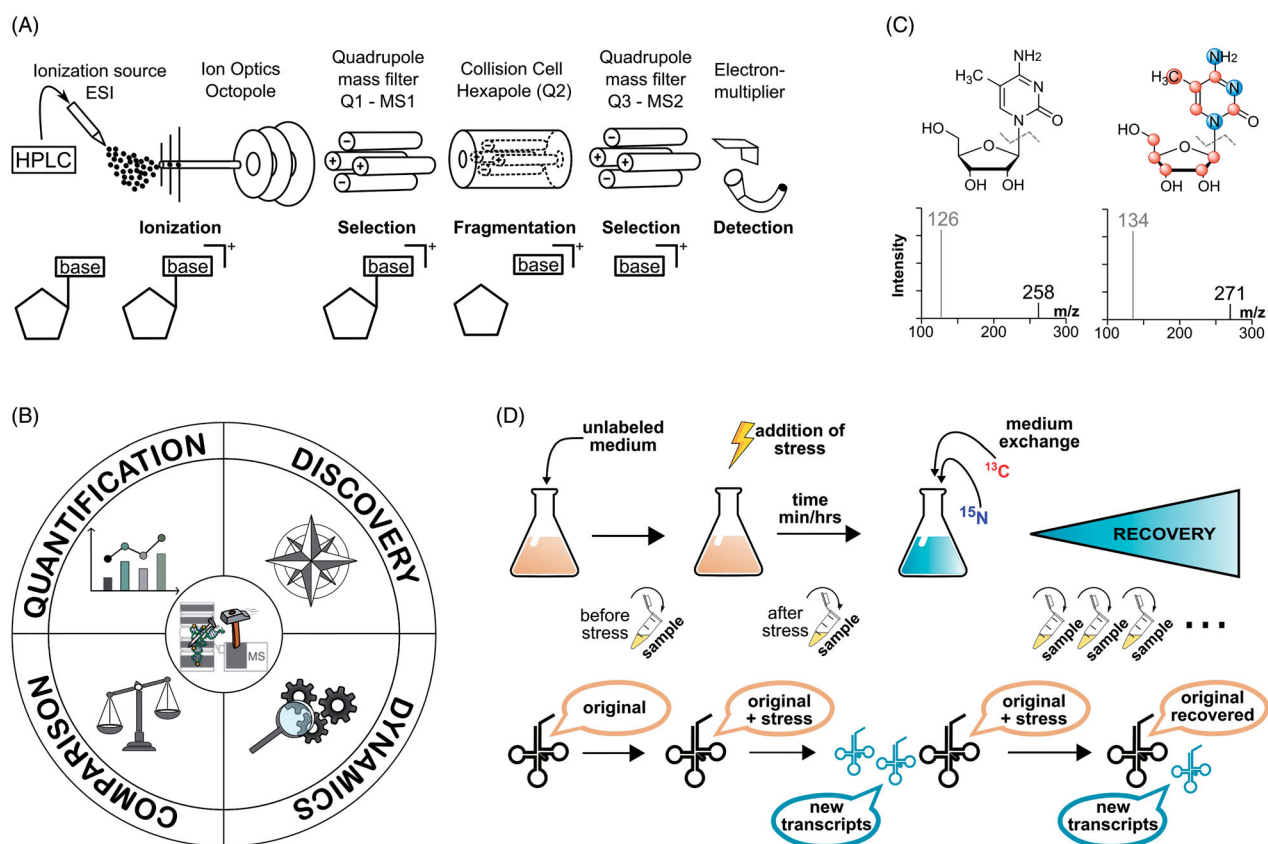
Figure 8. Principle and recent advances in nucleoside mass spectrometry (MS). (A) Instrumentation for sensitive detection of modified nucleosides by a triple quadrupole MS. After separation of the RNA hydrolysate via HPLC, the effluent is ionized in an electrospray ion source (ESI) and the nucleoside is ionized by protonation. The first quadrupole selects for the nucleoside ion. The collision cell fragments the nucleoside and the charged nucleobase remains. After another selection, the nucleobase enters the detector and a signal is recorded. (B) Chemical structure of 5-methylcytidine and its mass spectrum through targeted MS/MS analysis. The left structure is the natural isotopologue, the right structure an isotopologue where all carbon and nitrogen atoms have been exchanged to stable isotopes (here ^{13}C and ^{15}N). (C) Overview of different techniques utilizing stable isotope labeling for RNA modification analysis through mass spectrometry (NAIL-MS, nucleic acid isotope labeling coupled mass spectrometry). (D) Exemplary workflow of a dynamics NAIL-MS experiment. Cells are grown in, for example, unlabeled medium and stressed. After stress exposure the medium is exchanged to a labeled variant. Newly transcribed RNA and original RNA can be easily distinguished through their differential labeling. See colour version of this figure at www.tandfonline.com/ibmg.

uncontrollable factors. An IS is ideally identical to the analyte of interest in terms of detection efficiency but is yet sufficiently different to allow its discrimination from the actual analyte. Isotopologues are molecules that only differ in their isotope composition but retain identical physicochemical properties and thus behave identically on LC-MS systems. Many stable isotopes, especially ^{13}C and ^{15}N, fulfill this requirement and thus SILIS are now established aides in quantitative RNA modification analysis (Figure 8 (B)).

Here, synthetically produced SILIS (Brandmayr et al. 2012) and biosynthetically produced SILIS (Kellner, Ochel, et al. 2014; Borland et al. 2019) can be distinguished. Over the last decade, achievements have been made in terms of biosynthetically produced SILIS. While only 11 modifications could be used for quantification

in a SILIS in 2014 (Kellner, Ochel, et al. 2014) today up to 26 modifications produced in different organisms can be utilized as SILIS (Borland et al. 2019). In addition, there is now evidence that it is of no relevance in which organism the SILIS was produced as it can be used for quantification of RNA modifications in any other organism. With the help of SILIS whole modification profiles can be assessed (Borland et al. 2019).

Making RNA modification dynamics visible with NAIL-MS

With the rise of biosynthetic production of stable isotope-labeled RNA for mass spectrometry purposes, a new technique for tracing RNA modification dynamics emerged.

Nucleic acid isotope labeling coupled mass spectrometry (NAIL-MS) relies on the efficient and mono-isotopic labeling of RNA and is mainly used to follow changes of the epitranscriptome in a certain RNA population. However, NAIL-MS is also used in the context of RNA modification discovery (Dumelin et al. 2012; Kellner, Neumann, et al. 2014; Kellner, Ochel, et al. 2014; Thiaville et al. 2016; Dal Magro et al. 2018) quantification or multiplexing (Reichle, Kaiser, et al. 2018) (Figure 8(C)).

Dynamic NAIL-MS analyses are also referred to as pulse-chase experiments. Before experiment initiation, the cells of interest are grown in a medium containing controllable stable isotope sources. For example, in *E. coli* (Reichle, Kaiser, et al. 2018; Reichle, Weber, et al. 2018; 2019) and *S. cerevisiae* (Heiss et al. 2017) glucose can be used as either $^{13}C_6$- or $^{12}C_6$-glucose, while human cells require supplementation with stable isotope-labeled nucleosides or nucleobases for efficient labeling (Heiss et al. 2021). To assess the impact of a pulse on the epitranscriptome, the medium is exchanged to contain the other isotopologue of the nutrient. Through intelligent experiment design, it is possible to assess the impact of the pulse on the original transcripts alongside the modification kinetics of nascent RNA (concept shown in Figure 8(D)). Here, a maximal labeling efficiency of close to 100% is necessary. In some cases, the epitranscriptome of nascent RNA is not of interest and thus a lower starting labeling efficiency of around ~50% is acceptable.

NAIL-MS enables the differentiation between RNA degradation, RNA modification, or demodification. An extensively studied pulse for NAIL-MS studies is methyl methanesulfonate (MMS), which methylates nucleophilic positions in RNA. In the context of RNA methylation damage repair, NAIL-MS was used to discriminate the origin of methylation marks in damaged RNAs, and subsequently the repair of damage through enzymatic demethylation was visualized for m^1A, m^3C, and ms^2C in living *E. coli* cells (Reichle, Kaiser, et al. 2018; Reichle, Weber, et al. 2018; 2019). Interestingly, NAIL-MS revealed only low abundant methylation damage products in human tRNA and no repair was observable in living human cells (Heiss et al. 2021). This result highlights a major pitfall of NAIL-MS. The time resolution depends on the speed of labeled nutrient uptake upon medium exchange and afterward on the metabolic processing speed inside the cell. Thus, a delay from medium exchange to isotope-labeled RNA detection through LC-MS can be observed. If a biological process is faster than this delay, it is not possible to resolve the mechanisms of the epitranscriptome change through

NAIL-MS. Although there are limitations to NAIL-MS, it is currently the only available technique to allow a multi-layered analysis of the epitranscriptome and its dynamics.

Summary and outlook

To determine the structure, function, and impact of RNA modifications various techniques are available. Many require the use of expensive instrumentation dedicated to RNA modification analysis and often collaborative efforts of experts in these analyses are needed to study a given epitranscriptomic mark. Currently, N6-methyladenosine (m^6A), is one of the best-studied RNA modifications, as it regulates the fate and function of messenger RNA and thus impacts gene expression. Initial studies on m^6A writer and erasers (enzymes adding and removing m^6A from mRNA) strongly depended on nucleoside MS analyses and antibody-based sequencing techniques (Jia et al. 2011; Zheng et al. 2013; Liu et al. 2014). Many advances discussed in this review have been made in the context of m^6A detection. Nucleotide resolution mapping of m^6A is now possible without using antibodies of questionable specificity (Mishima et al. 2015; Grozhik et al. 2019; Slama et al. 2019; McIntyre et al. 2020) by using Oxford Nanopore sequencing (Li et al. 2020) or the endonuclease mazF, which shows an m^6A dependent cleavage ability (Garcia-Campos et al. 2019). In addition, top-down MS (Hoernes et al. 2016) and oligonucleotide MS (Jiang et al. 2019) are now capable to analyze full-length mRNA, and thus mapping m^6A modifications through MS is possible. However, software limitations for data analyses have to be overcome before broader applicability of MS becomes available. Our understanding of the m^6A modification has further deepened through NMR studies, where an impact of m^6A on RNA annealing was shown (Shi et al. 2019).

The power of combining sequencing and mass spectrometric analyses was recently presented by the Waldor lab. They predicted modified nucleosides in tRNAs of the uncharted organism *Vibrio cholerae* by comparing RT signature profiles with *E. coli* tRNA profiles. After sequencing, two RT signatures stood out. One of them was analyzed through nucleoside and oligonucleotide MS which led to the discovery of a novel modified nucleoside, namely acetylated 3-amino-3-carboxypropyl ($acacp^3U$) (Kimura et al. 2020). The discovery of novel modified nucleosides is currently supported through the continuously increasing sensitivity of modern mass spectrometers. Yet, this advance is connected to the risk of mis-interpretation of MS data

and reporting lowly abundant signals as novel epitranscriptomic marks. In fact, careful analysis of such native signals through high-resolution mass spectrometry and metabolic isotope labeling is necessary to exclude artifacts (Jora et al. 2020).

Recent advances in CryoEM analysis of large protein and RNA-protein complexes now allow "direct" observation of complex structures at almost atomic resolution. Therefore, it is tempting to interpret extra electronic density as post-transcriptional RNA modifications (Natchiar et al. 2017; Golubev et al. 2020; Sas-Chen et al. 2020; Stojković et al. 2020). While it is certain that CryoEM brings exceptional insights in structural analysis of RNA-protein complexes, its real value for "direct" observation of RNA modifications remains questionable. CryoEM data should be considered with extreme care since many of the newly identified nucleotides in human 80S ribosome (Natchiar et al. 2017) may be artifacts from tightly bound water molecules and/or Mg2+ atoms and were not confirmed by any other approaches (Taoka et al. 2018; Enroth et al. 2019).

The reversibility of RNA modifications through RNA erasers has introduced the dimension of time into the instrumental analysis of the epitranscriptome. The observation of RNA modification and demodification processes inside living cells is now possible by utilizing stable isotope labeling of RNA in combination with NMR and MS. Both techniques were introduced by demonstrating tRNA^{Phe} maturation processes. In yeast, sequential orders in the introduction of modifications along the tRNA maturation pathway were found alongside of modification circuits (Figure 2) (Barraud et al. 2019). By NMR, sequential incorporation of Ψ55 followed by m^{7}G and m^{5}U and finally m^{5}C and m^{1}A was observed. In human HEK 293 cell tRNA^{Phe}, a slightly different sequential order was observed through NAIL-MS analysis. Ψ is quickly incorporated, potentially accompanied by m^{5}U, followed by m^{5}C and m^{1}A and finally, m^{7}G occurs (Heiss et al. 2021). This study also showed that anticodon-loop modifications form rather slowly which implies that structure stabilization by modified nucleosides is a key necessity and must thus happen early on, while ac-loop modifications are not immediately needed and are potentially placed on-demand. Earlier studies on RNA maturation were done with radioactive labeling of RNA, as was the only study showing actual RNA demethylation *in vivo* (Ougland et al. 2004). NAIL-MS in *E. coli* confirmed the activity of the m^{1}A demethylase AlkB and furthermore its demethylation capacity and preference for 3-methylcytidine

(m^{3}C) and 2-methylthiocytidine (ms^{2}C) was revealed (Reichle, Weber, et al. 2018; Reichle et al. 2019). With NAIL-MS demethylation kinetics inside living cells observable a temporal view on the epitranscriptome becomes available.

Our way of life is currently dominated by the outbreak of the SARS-CoV-2 coronavirus. This virus is an RNA virus and according to one NNGS report, it is highly modified (Kim et al. 2020). So far Oxford Nanopore sequencing is the only resource to study the epitranscriptome of the SARS-CoV-2 virus as MS analysis depends on purified viral RNA free of contaminating host RNA. The development of more appropriate data analysis software will further increase the use of modern sequencing techniques for modification assignments. The importance of understanding the viral epitranscriptome was recently shown for HIV where RiboMethSeq revealed the presence 2′-O-ribose methylations and its role in evasion of innate immune sensing (Ringeard et al. 2019).

To end the pandemic, a vaccine is urgently needed. mRNA-based vaccines against the SARS-CoV-2 spike protein are promising and are being intensively tested (Jackson et al. 2020; Mulligan et al. 2020). Vaccine RNA might also be modified, by, for example, 1-methyl-pseudouridine, which increases mRNA translation and dampens innate immune sensing (Karikó et al. 2008). If mRNA vaccines successfully clear all clinical trials, their production will start as billions of vaccine doses are required. Here, fast and reliable quality control of the produced mRNA and its epitranscriptome is needed. Thus, instrumental analysis of the epitranscriptome, especially top-down and oligonucleotide MS, is now more important than ever.

Disclosure statement

Funding

ORCID

Yasemin Yoluç   http://orcid.org/0000-0001-9535-5020
Gregor Ammann   http://orcid.org/0000-0003-2135-058X
Pierre Barraud   http://orcid.org/0000-0003-4460-8360
Manasses Jora   http://orcid.org/0000-0002-1999-1508
Patrick A. Limbach   http://orcid.org/0000-0003-1526-4546
Yuri Motorin   http://orcid.org/0000-0002-8018-334X
Virginie Marchand   http://orcid.org/0000-0002-8537-1139
Carine Tisné   http://orcid.org/0000-0001-5534-4650
Stefanie Kellner   http://orcid.org/0000-0003-3224-7502

References

Agris PF. 2008. Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. EMBO Rep. 9(7):629–635.

Agris PF, Sierzputowska-Gracz H, Smith C. 1986. Transfer RNA contains sites of localized positive charge: carbon NMR studies of [13C]methyl-enriched *Escherichia coli* and yeast tRNAPhe. Biochemistry. 25(18):5126–5131.

Apffel A, Chakel JA, Fischer S, Lichtenwalter K, Hancock WS. 1997. Analysis of oligonucleotides by HPLC-electrospray ionization mass spectrometry. Anal Chem. 69(7):1320–1325.

Aschenbrenner J, Marx A. 2016. Direct and site-specific quantification of RNA 2′-O-methylation by PCR with an engineered DNA polymerase. Nucleic Acids Res. 44(8):3495–3502.

Aschenbrenner J, Werner S, Marchand V, Adam M, Motorin Y, Helm M, Marx A. 2018. Engineering of a DNA polymerase for direct m6 A sequencing. Angew Chem Int Ed Engl. 57(2):417–421.

Bakin A, Ofengand J. 1993. Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. Biochemistry. 32(37):9754–9762.

Bakin AV, Ofengand J. 1998. Mapping of pseudouridine residues in RNA to nucleotide resolution. Methods Mol Biol. 77:297–309.

Baldridge KC, Jora M, Maranhao AC, Quick MM, Addepalli B, Brodbelt JS, Ellington AD, Limbach PA, Contreras LM. 2018. Directed evolution of heterologous tRNAs leads to reduced dependence on post-transcriptional modifications. ACS Synth Biol. 7(5):1315–1327.

Barraud P, Gato A, Heiss M, Catala M, Kellner S, Tisne C. 2019. Time-resolved NMR monitoring of tRNA maturation. Nat Commun. 10(1):3373.

Barraud P, Tisné C. 2019. To be or not to be modified: miscellaneous aspects influencing nucleotide modifications in tRNAs. IUBMB Life. 71(8):1126–1140.

Behm-Ansmant I, Helm M, Motorin Y. 2011. Use of specific chemical reagents for detection of modified nucleotides in RNA. J Nucleic Acids. 2011:408053.

Biba M, Foley JP, Welch CJ. 2017. Liquid chromatographic separation of oligonucleotides. In: Fanali S, Haddad PR, Poole CF, editors. Liquid chromatography, 2nd ed. Amsterdam (The Netherlands): Elsevier; p. 159–182.

Birkedal U, Christensen-Dalsgaard M, Krogh N, Sabarinathan R, Gorodkin J, Nielsen H. 2015. Profiling of ribose

methylations in RNA by high-throughput sequencing. Angew Chem Int Ed Engl. 54(2):451–455.

Bjorkbom A, Lelyveld VS, Zhang S, Zhang W, Tam CP, Blain JC, Szostak JW. 2015. Bidirectional direct sequencing of noncanonical RNA by two-dimensional analysis of mass chromatograms. J Am Chem Soc. 137(45):14430–14438.

Borland K, Diesend J, Ito-Kureha T, Heissmeyer V, Hammann C, Buck AH, Michalakis S, Kellner S. 2019. Production and application of stable isotope-labeled internal standards for RNA modification analysis. Genes. 10(1):26.

Brandmayr C, Wagner M, Brückl T, Globisch D, Pearson D, Kneuttinger AC, Reiter V, Hienzsch A, Koch S, Thoma I, et al. 2012. Isotope-based analysis of modified tRNA nucleosides correlates modification density with translational efficiency. Angew Chem Int Ed Engl. 51(44):11162–11165.

Bryzgunova OE, Laktionov PP. 2015. Extracellular nucleic acids in urine: sources, structure, diagnostic potential. Acta Naturae. 7(3):48–54.

Cai WM, Chionh YH, Hia F, Gu C, Kellner S, McBee ME, Ng CS, Pang YLJ, Prestwich EG, Lim KS, et al. 2015. A platform for discovery and quantification of modified ribonucleosides in RNA: application to stress-induced reprogramming of tRNA modifications. Methods Enzymol. 560:29–71.

Calderisi G, Glasner H, Breuker K. 2020. Radical transfer dissociation for de novo characterization of modified ribonucleic acids by mass spectrometry. Angew Chem Int Ed Engl. 59(11):4309–4313.

Cao X, Limbach PA. 2015. Enhanced detection of post-transcriptional modifications using a mass-exclusion list strategy for RNA modification mapping by LC-MS/MS. Anal Chem. 87(16):8433–8440.

Carlile TM, Rojas-Duran MF, Zinshteyn B, Shin H, Bartoli KM, Gilbert WV. 2014. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature. 515(7525):143–146.

Catala M, Gato A, Tisné C, Barraud P. 2020a. 1H, 15N chemical shift assignments of the imino groups of yeast tRNAPhe: influence of the post-transcriptional modifications. Biomol NMR Assign. 14(2):169–174.

Catala M, Gato A, Tisné C, Barraud P. 2020b. Preparation of yeast tRNA sample for NMR spectroscopy. Bio-protocol. 10(12):e3646.

Choi BS, Redfield AG. 1986. NMR study of isoleucine transfer RNA from *Thermus thermophilus*. Biochemistry. 25(7):1529–1534.

Cozen AE, Quartley E, Holmes AD, Hrabeta-Robinson E, Phizicky EM, Lowe TM. 2015. ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. Nat Methods. 12(9):879–884.

Cozzuto L, Liu H, Pryszcz LP, Pulido TH, Delgado-Tejedor A, Ponomarenko J, Novoa EM. 2020. MasterOfPores: a workflow for the analysis of Oxford Nanopore Direct RNA Sequencing Datasets. Front Genet. 11:211.

Crain PF. 1990. Preparation and enzymatic hydrolysis of DNA and RNA for mass spectrometry. Methods Enzymol. 193:782–790.

Crawford JE, Chan SI, Schweizer MP. 1971. NMR studies of organic solvent denatured yeast phenylalanyl transfer RNA at 220 MHZ a. Biochem Biophys Res Commun. 44(1):1–7.

Dai Q, Moshitch-Moshkovitz S, Han D, Kol N, Amariglio N, Rechavi G, Dominissini D, He C. 2017. Nm-seq maps 2′-O-

methylation sites in human mRNA with base precision. Nat Methods. 14(7):695–698.

Dal Magro C, Keller P, Kotter A, Werner S, Duarte V, Marchand V, Ignarski M, Freiwald A, Müller RU, Dieterich C, et al. 2018. A vastly increased chemical variety of RNA modifications containing a thioacetal structure. Angew Chem Int Ed Engl. 57(26):7893–7897.

Davis DR, Poulter CD. 1991. 1H-15N NMR studies of *Escherichia coli* tRNA(Phe) from hisT mutants: a structural role for pseudouridine. Biochemistry. 30(17):4223–4231.

Demelenne A, Gou MJ, Nys G, Parulski C, Crommen J, Servais AC, Fillet M. 2020. Evaluation of hydrophilic interaction liquid chromatography, capillary zone electrophoresis and drift tube ion-mobility quadrupole time of flight mass spectrometry for the characterization of phosphodiester and phosphorothioate oligonucleotides. J Chromatogr A. 1614:460716.

Ding H, Bailey AD, Jain M, Olsen H, Paten B. 2020. Gaussian mixture model-based unsupervised nucleotide modification number detection using nanopore sequencing readouts. Bioinformatics. 36(19):4928–4934.

Dominissini D, Nachtergaele S, Moshitch-Moshkovitz S, Peer E, Kol N, Ben-Haim MS, Dai Q, Di Segni A, Salmon-Divon M, Clark WC, et al. 2016. The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. Nature. 530(7591):441–446.

Dumelin CE, Chen Y, Leconte AM, Chen YG, Liu DR. 2012. Discovery and biological characterization of geranylated RNA in bacteria. Nat Chem Biol. 8(11):913–919.

Durairaj A, Limbach PA. 2008. Improving CMC-derivatization of pseudouridine in RNA for mass spectrometric detection. Anal Chim Acta. 612(2):173–181.

Durant PC, Bajji AC, Sundaram M, Kumar RK, Davis DR. 2005. Structural effects of hypermodified nucleosides in the Escherichia coli and human tRNALys anticodon loop: the effect of nucleosides s2U, mcm5U, mcm5s2U, mnm5s2U, t6A, and ms2t6A. Biochemistry. 44(22):8078–8089.

Dyubankova N, Sochacka E, Kraszewska K, Nawrot B, Herdewijn P, Lescrinier E. 2015. Contribution of dihydrouridine in folding of the D-arm in tRNA. Org Biomol Chem. 13(17):4960–4966.

Emmerechts G, Herdewijn P, Rozenski J. 2005. Pseudouridine detection improvement by derivatization with methyl vinyl sulfone and capillary HPLC-mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci. 825(2):233–238.

Enroth C, Poulsen LD, Iversen S, Kirpekar F, Albrechtsen A, Vinther J. 2019. Detection of internal N7-methylguanosine (m7G) RNA modifications by mutational profiling sequencing. Nucleic Acids Res. 47(20):e126.

Feederle R, Schepers A. 2017. Antibodies specific for nucleic acid modifications. RNA Biol. 14(9):1089–1098.

Feng Y, Ma C-J, Ding J-H, Qi C-B, Xu X-J, Yuan B-F, Feng Y-Q. 2020. Chemical labeling – assisted mass spectrometry analysis for sensitive detection of cytidine dual modifications in RNA of mammals. Anal Chim Acta. 1098:56–65.

Fernández-Costa C, Martínez-Bartolomé S, McClatchy DB, Saviola AJ, Yu N-K, Yates JR. 2020. Impact of the identification strategy on the reproducibility of the DDA and DIA results. J Proteome Res. 19(8):3153–3161.

Fradin A, Gruhl H, Feldmann H. 1975. Mapping of yeast tRNAs by two-dimensional electrophoresis on polyacrylamide gels. FEBS Lett. 50(2):185–189.

Gao Y, McLuckey SA. 2013. Electron transfer followed by collision-induced dissociation (NET-CID) for generating sequence information from backbone-modified oligonucleotide anions. Rapid Commun Mass Spectrom. 27(1):249–257.

Garcia-Campos MA, Edelheit S, Toth U, Safra M, Shachar R, Viukov S, Winkler R, Nir R, Lasman L, Brandis A, et al. 2019. Deciphering the "m6A Code" via antibody-independent quantitative profiling. Cell. 178(3):731–747.

Gaston KW, Limbach PA. 2014. The identification and characterization of non-coding and coding RNAs and their modified nucleosides by mass spectrometry. RNA Biol. 11(12):1568–1585.

Gato A, Catala M, Tisné C, Barraud P. 2021. A method to monitor the introduction of post-transcriptional modifications in tRNAs with NMR spectroscopy. Methods Mol Biol.

Gaudin C, Nonin-Lecomte S, Tisné C, Corvaisier S, Bordeau V, Dardel F, Felden B. 2003. The tRNA-like domains of E coli and A.aeolicus transfer-messenger RNA: structural and functional studies. J Mol Biol. 331(2):457–471.

Gehrke CW, Kuo KC, Zumwalt RW. 1980. Chromatography of nucleosides. J Chromatogr. 188(1):129–147.

Gehrke CW, Patel AB. 1976. Gas-liquid chromatography of nucleosides. Derivatization and chromatography. J Chromatogr. 123(2):335–345.

Gehrke CW, Ruyle CD. 1968. Gas-liquid chromatographic analysis of nucleic acid components. J Chromatogr. 38(4):473–491.

Golubev A, Fatkhullin B, Khusainov I, Jenner L, Gabdulkhakov A, Validov S, Yusupova G, Yusupov M, Usachev K. 2020. Cryo-EM structure of the ribosome functional complex of the human pathogen *Staphylococcus aureus* at 3.2 Å resolution. FEBS Lett. 594(21):3551–3567.

Goyon A, Zhang K. 2020. Characterization of antisense oligonucleotide impurities by ion-pairing reversed-phase and anion exchange chromatography coupled to hydrophilic interaction liquid chromatography/mass spectrometry using a versatile two-dimensional liquid chromatography setup. Anal Chem. 92(8):5944–5951.

Grosjean H, Droogmans L, Roovers M, Keith G. 2007. Detection of enzymatic activity of transfer RNA modification enzymes using radiolabeled tRNA substrates. Methods Enzymol. 425:55–101.

Grozhik AV, Olarerin-George AO, Sindelar M, Li X, Gross SS, Jaffrey SR. 2019. Antibody cross-reactivity accounts for widespread appearance of m1A in 5'UTRs. Nat Commun. 10(1):5126.

Guo M, Li X, Zhang L, Liu D, Du W, Yin D, Lyu N, Zhao G, Guo C, Tang D. 2017. Accurate quantification of 5-Methylcytosine, 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine in genomic DNA from breast cancer by chemical derivatization coupled with ultra performance liquid chromatography- electrospray quadrupole time of flight mass spectrometry analysis. Oncotarget. 8(53):91248–91257.

Guo C, Xie C, Chen Q, Cao X, Guo M, Zheng S, Wang Y. 2018. A novel malic acid-enhanced method for the analysis of 5-methyl-2'-deoxycytidine, 5-hydroxymethyl-2'-deoxycytidine, 5-methylcytidine and 5-

hydroxymethylcytidine in human urine using hydrophilic interaction liquid chromatography-tandem mass spectrometry. Anal Chim Acta. 1034:110–118.

Gupta RC, Randerath K. 1979. Rapid print-readout technique for sequencing of RNA's containing modified nucleotides. Nucleic Acids Res. 6(11):3443–3458.

Gupta RC, Randerath E, Randerath K. 1976. A double-labeling procedure for sequence analysis of picomole amounts of nonradioactive RNA fragments. Nucleic Acids Res. 3(11): 2895–2914.

Hagelskamp F, Borland K, Ramos J, Hendrick AG, Fu D, Kellner S. 2020. Broadly applicable oligonucleotide mass spectrometry for the analysis of RNA writers and erasers in vitro. Nucleic Acids Res. 48(7):e41.

Han L, Phizicky EM. 2018. A rationale for tRNA modification circuits in the anticodon loop. RNA. 24(10):1277–1284.

Hare DR, Ribeiro NS, Wemmer DE, Reid BR. 1985. Complete assignment of the imino protons of Escherichia coli valine transfer RNA: two-dimensional NMR studies in water. Biochemistry. 24(16):4300–4306.

Hartstock K, Nilges BS, Ovcharenko A, Cornelissen NV, Püllen N, Lawrence-Dörner AM, Leidel SA, Rentmeister A. 2018. Enzymatic or in vivo installation of propargyl groups in combination with click chemistry for the enrichment and detection of methyltransferase target sites in RNA. Angew Chem Int Ed Engl. 57(21):6342–6346.

Hartstock K, Rentmeister A. 2019. Mapping N6-maethyladenosine (m6 A) in RNA: established methods, remaining challenges, and emerging approaches. Chemistry. 25(14): 3455–3464.

Hauenschild R, Tserovski L, Schmid K, Thüring K, Winz ML, Sharma S, Entian KD, Wacheul L, Lafontaine DL, Anderson J, et al. 2015. The reverse transcription signature of N-1-methyladenosine in RNA-Seq is sequence dependent. Nucleic Acids Res. 43(20):9950–9964.

Heerschap A, Mellema JR, Janssen HG, Walters JA, Haasnoot CA, Hilbers CW. 1985. Imino-proton resonances of yeast tRNAPhe studied by two-dimensional nuclear Overhauser enhancement spectroscopy. Eur J Biochem. 149(3): 649–655.

Heiss M, Hagelskamp F, Marchand V, Motorin Y, Kellner S. 2021. Cell culture NAIL-MS allows insight into human tRNA and rRNA modification dynamics in vivo. Nat Commun. 12(1):389.

Heiss M, Kellner S. 2017. Detection of nucleic acid modifications by chemical reagents. RNA Biol. 14(9):1166–1174.

Heiss M, Reichle VF, Kellner S. 2017. Observing the fate of tRNA and its modifications by nucleic acid isotope labeling mass spectrometry: NAIL-MS. RNA Biol. 14(9):1260–1268.

Helm M, Motorin Y. 2017. Detecting RNA modifications in the epitranscriptome: predict and validate. Nat Rev Genet. 18(5):275–291.

Hilbers CW, Heerschap A, Haasnoot CA, Walters JA. 1983. The solution structure of yeast tRNAPhe as studied by nuclear Overhauser effects in NMR. J Biomol Struct Dyn. 1(1): 183–207.

Hoernes TP, Clementi N, Faserl K, Glasner H, Breuker K, Lindner H, Hüttenhofer A, Erlacher MD. 2016. Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewire the genetic code. Nucleic Acids Res. 44(2):852–862.

Hong T, Yuan Y, Chen Z, Xi K, Wang T, Xie Y, He Z, Su H, Zhou Y, Tan ZJ, et al. 2018. Precise antibody-independent m6A identification via 4SedTTP-involved and FTO-assisted strategy at single-nucleotide resolution. J Am Chem Soc. 140(18):5886–5889.

Hossain M, Limbach PA. 2007. Mass spectrometry-based detection of transfer RNAs by their signature endonuclease digestion products. RNA. 13(2):295–303.

Huang W, Lan MD, Qi CB, Zheng SJ, Wei SZ, Yuan BF, Feng YQ. 2016. Formation and determination of the oxidation products of 5-methylcytosine in RNA. Chem Sci. 7(8): 5495–5502.

Huang TY, Liu J, Liang X, Hodges BD, McLuckey SA. 2008. Collision-induced dissociation of intact duplex and single-stranded siRNA anions. Anal Chem. 80(22):8501–8508.

Huang TY, Liu J, McLuckey SA. 2010. Top-down tandem mass spectrometry of tRNA via ion trap collision-induced dissociation. J Am Soc Mass Spectrom. 21(6):890–898.

Huang T-y, McLuckey SA. 2011. Gas-phase ion/ion reactions of rubrene cations and multiply charged DNA and RNA anions. Int J Mass Spectrom. 304(2–3):140–147.

Incarnato D, Anselmi F, Morandi E, Neri F, Maldotti M, Rapelli S, Parlato C, Basile G, Oliviero S. 2017. High-throughput single-base resolution mapping of RNA 2′-O-methylated residues. Nucleic Acids Res. 45(3):1433–1441.

Jackson LA, Anderson EJ, Rouphael NG, Roberts PC, Makhene M, Coler RN, McCullough MP, Chappell JD, Denison MR, Stevens LJ, et al. 2020. An mRNA vaccine against SARS-CoV-2 - preliminary report. N Engl J Med. 383(20): 1920–1931.

Jenjaroenpun P, Wongsurawat T, Wadley TD, Wassenaar TM, Liu J, Dai Q, Wanchai V, Akel NS, Jamshidi-Parsian A, Franco AT, et al. 2021. Decoding the epitranscriptional landscape from native RNA sequences. Nucleic Acids Res. 49(2):e7.

Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, Yi C, Lindahl T, Pan T, Yang Y-G, et al. 2011. N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol. 7(12):885–887.

Jiang T, Yu N, Kim J, Murgo JR, Kissai M, Ravichandran K, Miracco EJ, Presnyak V, Hua S. 2019. Oligonucleotide sequence mapping of large therapeutic mRNAs via parallel ribonuclease digestions and LC-MS/MS. Anal Chem. 91(13): 8500–8506.

Jora M, Borland K, Abernathy S, Zhao R, Kelley M, Kellner S, Addepalli B, Limbach PA. 2020. Chemical amination/imination of carbonothiolated nucleosides during RNA hydrolysis. Angew Chem Int Ed. 60(8):3961–3966.

Jora M, Lobue PA, Ross RL, Williams B, Addepalli B. 2019. Detection of ribonucleoside modifications by liquid chromatography coupled with mass spectrometry. Biochim Biophys Acta Gene Regul Mech. 1862(3):280–290.

Kan LS, Ts'o POP, Haar Fvd, Sprinzl M, Cramer F. 1974. NMR study on the methyl and methylene proton resonances of tRNA Phe yeast. Biochem Biophys Res Commun. 59(1): 22–29.

Karikó K, Muramatsu H, Welsh FA, Ludwig J, Kato H, Akira S, Weissman D. 2008. Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. Mol Ther. 16(11):1833–1840.

Kastrup RV, Schmidt PG. 1978. 1H NMR of valine tRNA modified bases. Evidence for multiple conformations. Nucleic Acids Res. 5(1):257–269.

Keith G. 1995. Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography. Biochimie. 77(1-2):142–144.

Kellersberger KA, Yu E, Kruppa GH, Young MM, Fabris D. 2004. Top-down characterization of nucleic acids modified by structural probes using high-resolution tandem mass spectrometry and automated data interpretation. Anal Chem. 76(9):2438–2445.

Kellner S, Burhenne J, Helm M. 2010. Detection of RNA modifications. RNA Biol. 7(2):237–247.

Kellner S, Neumann J, Rosenkranz D, Lebedeva S, Ketting RF, Zischler H, Schneider D, Helm M. 2014. Profiling of RNA modifications by multiplexed stable isotope labelling. Chem Commun. 50(26):3516–3518.

Kellner S, Ochel A, Thüring K, Spenkuch F, Neumann J, Sharma S, Entian K-D, Schneider D, Helm M. 2014. Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. Nucleic Acids Research. 42(18):e142.

Khoddami V, Yerra A, Mosbruger TL, Fleming AM, Burrows CJ, Cairns BR. 2019. Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. Proc Natl Acad Sci USA. 116(14):6784–6789.

Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H. 2020. The architecture of SARS-CoV-2 transcriptome. Cell. 181(4): 914–921.

Kimura S, Dedon PC, Waldor MK. 2020. Comparative tRNA sequencing and RNA mass spectrometry for surveying tRNA modifications. Nat Chem Biol. 16(9):964–972.

Knutson SD, Ayele TM, Heemstra JM. 2018. Chemical labeling and affinity capture of inosine-containing RNAs using acrylamidofluorescein. Bioconjug Chem. 29(9):2899–2903.

Koehler KM, Schmidt PG. 1973. NMR study of the modified base resonances of tRNA tyr-coli. Biochem Biophys Res Commun. 50(2):370–376.

Kowalak JA, Pomerantz SC, Crain PF, McCloskey JA. 1993. A novel method for the determination of post-transcriptional modification in RNA by mass spectrometry. Nucleic Acids Res. 21(19):4577–4585.

Kristen M, Plehn J, Marchand V, Friedland K, Motorin Y, Helm M, Werner S. 2020. Manganese ions individually alter the reverse transcription signature of modified ribonucleosides. Genes. 11(8):950.

Krogh N, Nielsen H. 2019. Sequencing-based methods for detection and quantitation of ribose methylations in RNA. Methods. 156:5–15.

Kumar RK, Davis DR. 1997. Synthesis and studies on the effect of 2-thiouridine and 4-thiouridine on sugar conformation and RNA duplex stability. Nucleic Acids Res. 25(6):1272–1280.

Kumazawa Y, Yokogawa T, Tsurui H, Miura K, Watanabe K. 1992. Effect of the higher-order structure of tRNAs on the stability of hybrids with oligodeoxyribonucleotides: separation of tRNA by an efficient solution hybridization. Nucleic Acids Res. 20(9):2223–2232.

Lakings DB, Gehrke CW. 1971. Analysis of base composition of RNA and DNA hydrolysates by gas-liquid chromatography. J Chromatogr. 62(3):347–367.

Li L, Dai H, Nguyen AP, Gu W. 2020. A convenient strategy to clone modified/unmodified small RNA and mRNA for high throughput sequencing. RNA Eng. 26(2):218–227.

Li N, El Zahar NM, Saad JG, van der Hage ERE, Bartlett MG. 2018. Alkylamine ion-pairing reagents and the chromatographic separation of oligonucleotides. J Chromatogr A. 1580:110–119.

Li F, Su X, Bäurer S, Lämmerhofer M. 2020. Multiple heart-cutting mixed-mode chromatography-reversed-phase 2D-liquid chromatography method for separation and mass spectrometric characterization of synthetic oligonucleotides. J Chromatogr A. 1625:461338.

Li X, Zhu P, Ma S, Song J, Bai J, Sun F, Yi C. 2015. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. Nat Chem Biol. 11(8):592–597. eng.

Limbach PA, Crain PF, McCloskey JA. 1995. Characterization of oligonucleotides and nucleic acids by mass spectrometry. Curr Opin Biotechnol. 6(1):96–102.

Lin S, Liu Q, Jiang YZ, Gregory RI. 2019. Nucleotide resolution profiling of m(7)G tRNA modification by TRAC-Seq. Nat Protoc. 14(11):3220–3242.

Lin S, Liu Q, Lelyveld VS, Choe J, Szostak JW, Gregory RI. 2018. Mettl1/Wdr4-mediated m7G tRNA methylome is required for normal mrna translation and embryonic stem cell self-renewal and differentiation. Mol Cell. 71(2): 244–255.

Lin H, Miyauchi K, Harada T, Okita R, Takeshita E, Komaki H, Fujioka K, Yagasaki H, Goto Y-I, Yanaka K, et al. 2018. CO_2-sensitive tRNA modification associated with human mitochondrial disease. Nat Commun. 9(1):1875.

Linder B, Jaffrey SR. 2019. Discovering and mapping the modified nucleotides that comprise the epitranscriptome of mRNA. Cold Spring Harb Perspect Biol. 11(6):a032201.

Liu H, Begik O, Lucas MC, Ramirez JM, Mason CE, Wiener D, Schwartz S, Mattick JS, Smith MA, Novoa EM. 2019. Accurate detection of m6A RNA modifications in native RNA sequences. Nat Commun. 10(1):4079.

Liu J, Yue Y, Han D, Wang X, Fu Y, Zhang L, Jia G, Yu M, Lu Z, Deng X, et al. 2014. A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat Chem Biol. 10(2):93–95.

Lobue PA, Jora M, Addepalli B, Limbach PA. 2019. Oligonucleotide analysis by hydrophilic interaction liquid chromatography-mass spectrometry in the absence of ion-pair reagents. J Chromatogr A. 1595:39–48.

Lobue PA, Yu N, Jora M, Abernathy S, Limbach PA. 2019. Improved application of RNAModMapper - An RNA modification mapping software tool - For analysis of liquid chromatography tandem mass spectrometry (LC-MS/MS) data. Methods. 156:128–138.

Lorenz DA, Sathe S, Einstein JM, Yeo GW. 2020. Direct RNA sequencing enables m6A detection in endogenous transcript isoforms at base-specific resolution. RNA. 26(1): 19–28.

Lovejoy AF, Riordan DP, Brown PO. 2014. Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. PLoS One. 9(10): e110799.

Maden BE. 2001. Mapping 2′-O-methyl groups in ribosomal RNA. Methods. 25(3):374–382.

Marchand V, Ayadi L, Ernst FGM, Hertler J, Bourguignon-Igel V, Galvanin A, Kotter A, Helm M, Lafontaine DLJ, Motorin

Y. 2018. AlkAniline-Seq: profiling of m7 G and m3 C RNA modifications at single nucleotide resolution. Angew Chem Int Ed Engl. 57(51):16785–16790.

Marchand V, Blanloeil-Oillo F, Helm M, Motorin Y. 2016. Illumina-based RiboMethSeq approach for mapping of 2′-O-Me residues in RNA. Nucleic Acids Res. 44(16):e135.

Marchand V, Pichot F, Neybecker P, Ayadi L, Bourguignon-Igel V, Wacheul L, Lafontaine DLJ, Pinzano A, Helm M, Motorin Y. 2020. HydraPsiSeq: a method for systematic and quantitative mapping of pseudouridines in RNA. Nucleic Acids Res. 48(19):e110.

Matsuzawa S, Wakata Y, Ebi F, Isobe M, Kurosawa N. 2019. Development and validation of monoclonal antibodies against N6-methyladenosine for the detection of RNA modifications. PLoS One. 14(10):e0223197.

Matthiesen R, Kirpekar F. 2009. Identification of RNA molecules by specific enzyme digestion and mass spectrometry: software for and implementation of RNA mass mapping. Nucleic Acids Res. 37(6):e48.

Matuszewski M, Wojciechowski J, Miyauchi K, Gdaniec Z, Wolf WM, Suzuki T, Sochacka E. 2017. A hydantoin isoform of cyclic N6-threonylcarbamoyladenosine (ct6A) is present in tRNAs. Nucleic Acids Res. 45(4):2137–2149.

McGinnis AC, Grubb EC, Bartlett MG. 2013. Systematic optimization of ion-pairing agents and hexafluoroisopropanol for enhanced electrospray ionization mass spectrometry of oligonucleotides. Rapid Commun Mass Spectrom. 27(23): 2655–2664.

McIntyre ABR, Gokhale NS, Cerchietti L, Jaffrey SR, Horner SM, Mason CE. 2020. Limits in the detection of m6A changes using MeRIP/m6A-seq. Sci Rep. 10(1):6590.

McLuckey SA, Van Berkel GJ, Glish GL. 1992. Tandem mass spectrometry of small, multiply charged oligonucleotides. J Am Soc Mass Spectrom. 3(1):60–70.

Mengel-Jørgensen J, Kirpekar F. 2002. Detection of pseudouridine and other modifications in tRNA by cyanoethylation and MALDI mass spectrometry. Nucleic Acids Res. 30(23):e135.

Mishima E, Jinno D, Akiyama Y, Itoh K, Nankumo S, Shima H, Kikuchi K, Takeuchi Y, Elkordy A, Suzuki T, et al. 2015. Immuno-northern blotting: detection of RNA modifications by using antibodies against modified nucleosides. PLoS One. 10(11):e0143756.

Mo J, Håkansson K. 2006. Characterization of nucleic acid higher order structure by high-resolution tandem mass spectrometry. Anal Bioanal Chem. 386(3):675–681.

Motorin Y, Helm M. 2019. Methods for RNA modification mapping using deep sequencing: established and new emerging technologies. Genes. 10(1):35.

Motorin Y, Muller S, Behm-Ansmant I, Branlant C. 2007. Identification of modified residues in RNAs by reverse transcription-based methods. Methods Enzymol. 425: 21–53.

Mulligan MJ, Lyke KE, Kitchin N, Absalon J, Gurtman A, Lockhart S, Neuzil K, Raabe V, Bailey R, Swanson KA, et al. 2020. Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults. Nature. 586(7830):589–593.

Nakayama H, Akiyama M, Taoka M, Yamauchi Y, Nobe Y, Ishikawa H, Takahashi N, Isobe T. 2009. Ariadne: a database search engine for identification and chemical analysis of RNA using tandem mass spectrometry data. Nucleic Acids Res. 37(6):e47.

Nakayama H, Yamauchi Y, Nobe Y, Sato K, Takahashi N, Shalev-Benami M, Isobe T, Taoka M. 2019. Method for direct mass-spectrometry-based identification of monomethylated RNA nucleoside positional isomers and its application to the analysis of Leishmania rRNA. Anal Chem. 91(24):15634–15643.

Natchiar SK, Myasnikov AG, Kratzat H, Hazemann I, Klaholz BP. 2017. Visualization of chemical modifications in the human 80S ribosome structure. Nature. 551(7681): 472–477.

Nyakas A, Blum LC, Stucki SR, Reymond J-L, Schürch S. 2013. OMA and OPA-software-supported mass spectra analysis of native and modified nucleic acids. J Am Soc Mass Spectrom. 24(2):249–256.

Okada S, Ueda H, Noda Y, Suzuki T. 2019. Transcriptome-wide identification of A-to-I RNA editing sites using ICE-seq. Methods. 156:66–78.

Ougland R, Zhang CM, Liiv A, Johansen RF, Seeberg E, Hou YM, Remme J, Falnes P. 2004. AlkB restores the biological function of mRNA and tRNA inactivated by chemical methylation. Mol Cell. 16(1):107–116.

Pandolfini L, Barbieri I, Bannister AJ, Hendrick A, Andrews B, Webster N, Murat P, Mach P, Brandi R, Robson SC, et al. 2019. METTL1 Promotes let-7 MicroRNA Processing via m7G Methylation. Mol Cell. 74(6):1278–1290.

Patteson KG, Rodicio LP, Limbach PA. 2001. Identification of the mass-silent post-transcriptionally modified nucleoside pseudouridine in RNA by matrix-assisted laser desorption/ionization mass spectrometry. Nucleic Acids Res. 29(10): E49.

Paulines MJ, Wetzel C, Limbach PA. 2019. Using spectral matching to interpret LC-MS/MS data during RNA modification mapping. J Mass Spectrom. 54(11):906–914.

Peters-Clarke TM, Quan Q, Brademan DR, Hebert AS, Westphall MS, Coon JJ. 2020. Ribonucleic acid sequence characterization by negative electron transfer dissociation mass spectrometry. Anal Chem. 92(6):4436–4444.

Pomerantz SC, McCloskey JA. 1990. [44] Analysis of RNA hydrolyzates by liquid chromatography-mass spectrometry. Methods Enzymol. 193:796–824.

Pomerantz SC, McCloskey JA. 2005. Detection of the common RNA nucleoside pseudouridine in mixtures of oligonucleotides by mass spectrometry. Anal Chem. 77(15): 4687–4697.

Ranaei-Siadat E, Fabret C, Seijo B, Dardel F, Grosjean H, Nonin-Lecomte S. 2013. RNA-methyltransferase TrmA is a dual-specific enzyme responsible for C5-methylation of uridine in both tmRNA and tRNA. RNA Biol. 10(4):572–578.

Reichle VF, Kaiser S, Heiss M, Hagelskamp F, Borland K, Kellner S. 2018. Surpassing limits of static RNA modification analysis with dynamic NAIL-MS. Methods. 156:91–101.

Reichle VF, Petrov DP, Weber V, Jung K, Kellner S. 2019. NAIL-MS reveals the repair of 2-methylthiocytidine by AlkB in E. coli. Nat Commun. 10(1):5600.

Reichle VF, Weber V, Kellner S. 2018. NAIL-MS in E. coli determines the source and fate of methylation in tRNA. ChemBioChem. 19(24):2575–2583.

Ringeard M, Marchand V, Decroly E, Motorin Y, Bennasser Y. 2019. FTSJ3 is an RNA 2′-O-methyltransferase recruited by HIV to avoid innate immune sensing. Nature. 565(7740): 500–504.

Rost HL, Sachsenberg T, Aiche S, Bielow C, Weisser H, Aicheler F, Andreotti S, Ehrlich HC, Gutenbrunner P, Kenar E, et al. 2016. OpenMS: a flexible open-source software platform for mass spectrometry data analysis. Nat Methods. 13(9):741–748.

Roy S, Redfield AG. 1983. Assignment of imino proton spectra of yeast phenylalanine transfer ribonucleic acid. Biochemistry. 22(6):1386–1390.

Rozenski J, McCloskey JA. 2002. SOS: a simple interactive program for ab initio oligonucleotide sequencing by mass spectrometry. J Am Soc Mass Spectrom. 13(3):200–203.

Ryvkin P, Leung YY, Silverman IM, Childress M, Valladares O, Dragomir I, Gregory BD, Wang LS. 2013. HAMR: high-throughput annotation of modified ribonucleotides. RNA. 19(12):1684–1692.

Safra M, Sas-Chen A, Nir R, Winkler R, Nachshon A, Bar-Yaacov D, Erlacher M, Rossmanith W, Stern-Ginossar N, Schwartz S. 2017. The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. Nature. 551(7679):251–255.

Sakamoto K, Kawai G, Niimi T, Satoh T, Sekine M, Yamaizumi Z, Nishimura S, Miyazawa T, Yokoyama S. 1993. A modified uridine in the first position of the anticodon of a minor species of arginine tRNA, the argU gene product, from *Escherichia coli*. Eur J Biochem. 216(2):369–375.

Sakurai M, Yano T, Kawabata H, Ueda H, Suzuki T. 2010. Inosine cyanoethylation identifies A-to-I RNA editing sites in the human transcriptome. Nat Chem Biol. 6(10):733–740.

Sample PJ, Gaston KW, Alfonzo JD, Limbach PA. 2015. RoboOligo: software for mass spectrometry data to support manual and de novo sequencing of post-transcriptionally modified ribonucleic acids. Nucleic Acids Res. 43(10):e64.

Sarin LP, Kienast SD, Leufken J, Ross RL, Dziergowska A, Debiec K, Sochacka E, Limbach PA, Fufezan C, Drexler HCA, et al. 2018. Nano LC-MS using capillary columns enables accurate quantification of modified ribonucleosides at low femtomol levels. RNA. 24(10):1403–1417.

Sas-Chen A, Thomas JM, Matzov D, Taoka M, Nance KD, Nir R, Bryson KM, Shachar R, Liman GLS, Burkhart BW, et al. 2020. Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. Nature. 583(7817):638–643.

Schaefer M, Kapoor U, Jantsch MF. 2017. Understanding RNA modifications: the promises and technological bottlenecks of the 'epitranscriptome'. Open Biol. 7(5):170077.

Schneeberger EM, Breuker K. 2017. Native top-down mass spectrometry of TAR RNA in complexes with a wild-type tat peptide for binding site mapping. Angew Chem Int Ed Engl. 56(5):1254–1258.

Schurch S. 2016. Characterization of nucleic acids by tandem mass spectrometry - The second decade (2004-2013): from DNA to RNA and modified sequences. Mass Spectrom Rev. 35(4):483–523.

Schwartz S, Bernstein DA, Mumbach MR, Jovanovic M, Herbst RH, León-Ricardo BX, Engreitz JM, Guttman M, Satija R, Lander ES, et al. 2014. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. Cell. 159(1):148–162.

Shi H, Liu B, Nussbaumer F, Rangadurai A, Kreutz C, Al-Hashimi HM. 2019. NMR chemical exchange measurements reveal that N6-methyladenosine slows RNA annealing. J Am Chem Soc. 141(51):19988–19993.

Shu X, Cao J, Cheng M, Xiang S, Gao M, Li T, Ying X, Wang F, Yue Y, Lu Z, et al. 2020. A metabolic labeling method detects m6A transcriptome-wide at single base resolution. Nat Chem Biol. 16(8):887–895.

Slama K, Galliot A, Weichmann F, Hertler J, Feederle R, Meister G, Helm M. 2019. Determination of enrichment factors for modified RNA in MeRIP experiments. Methods. 156:102–109.

Smith SI, Brodbelt JS. 2010. Characterization of oligodeoxynucleotides and modifications by 193 nm photodissociation and electron photodetachment dissociation. Anal Chem. 82(17):7218–7226.

Smith SI, Brodbelt JS. 2011. Hybrid activation methods for elucidating nucleic acid modifications. Anal Chem. 83(1):303–310.

Smith AM, Jain M, Mulroney L, Garalde DR, Akeson M. 2019. Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. PLoS One. 14(5):e0216709.

Solivio B, Yu N, Addepalli B, Limbach PA. 2018. Improving RNA modification mapping sequence coverage by LC-MS through a nonspecific RNase U2-E49A mutant. Anal Chim Acta. 1036:73–79.

Stanley J, Vassilenko S. 1978. A different approach to RNA sequencing. Nature. 274(5666):87–89.

Stojković V, Myasnikov AG, Young ID, Frost A, Fraser JS, Fujimori DG. 2020. Assessment of the nucleotide modifications in the high-resolution cryo-electron microscopy structure of the *Escherichia coli* 50S subunit. Nucleic Acids Res. 48(5):2723–2732.

Strebitzer E, Rangadurai A, Plangger R, Kremser J, Juen MA, Tollinger M, Al-Hashimi HM, Kreutz C. 2018. 5-oxyacetic acid modification destabilizes double helical stem structures and favors anionic watson-crick like cmo5 U-G base pairs. Chemistry. 24(71):18903–18906.

Stuart JW, Koshlap KM, Guenther R, Agris PF. 2003. Naturally-occurring modification restricts the anticodon domain conformational space of tRNA(Phe). J Mol Biol. 334(5):901–918.

Sutton JM, Bartlett MG. 2020. Modeling cationic adduction of oligonucleotides using electrospray desorption ionization. Rapid Commun Mass Spectrom. 34(8):e8696.

Suzuki T, Suzuki T. 2007. Chaplet column chromatography: isolation of a large set of individual RNAs in a single step. Methods Enzymol. 425:231–239.

Suzuki T, Ueda H, Okada S, Sakurai M. 2015. Transcriptome-wide identification of adenosine-to-inosine editing using the ICE-seq method. Nat Protoc. 10(5):715–732.

Tang Y, Zheng SJ, Qi CB, Feng YQ, Yuan BF. 2015. Sensitive and simultaneous determination of 5-methylcytosine and its oxidation products in genomic DNA by chemical derivatization coupled with liquid chromatography-tandem mass spectrometry analysis. Anal Chem. 87(6):3445–3452.

Taoka M, Ikumi M, Nakayama H, Masaki S, Matsuda R, Nobe Y, Yamauchi Y, Takeda J, Takahashi N, Isobe T. 2010. In-gel digestion for mass spectrometric characterization of RNA from fluorescently stained polyacrylamide gels. Anal Chem. 82(18):7795–7803.

Taoka M, Nobe Y, Yamaki Y, Sato K, Ishikawa H, Izumikawa K, Yamauchi Y, Hirota K, Nakayama H, Takahashi N, et al.

2018. Landscape of the complete RNA chemical modifications in the human 80S ribosome. Nucleic Acids Res. 46(18):9289–9298.

Taoka M, Nobe Y, Yamaki Y, Yamauchi Y, Ishikawa H, Takahashi N, Nakayama H, Isobe T. 2016. The complete chemical structure of Saccharomyces cerevisiae rRNA: partial pseudouridylation of U2345 in 25S rRNA by snoRNA snR9. Nucleic Acids Res. 44(18):8951–8961.

Taucher M, Breuker K. 2010. Top-down mass spectrometry for sequencing of larger (up to 61nt) RNA by CAD and EDD. J Am Soc Mass Spectrom. 21(6):918–929.

Taucher M, Breuker K. 2012. Characterization of modified RNA by top-down mass spectrometry. Angew Chem Int Ed Engl. 51(45):11289–11292.

Thakur P, Estevez M, Lobue PA, Limbach PA, Addepalli B. 2020. Improved RNA modification mapping of cellular non-coding RNAs using C- and U-specific RNases. Analyst. 145(3):816–827.

Thiaville JJ, Kellner SM, Yuan Y, Hutinet G, Thiaville PC, Jumpathong W, Mohapatra S, Brochier-Armanet C, Letarov AV, Hillebrand R, et al. 2016. Novel genomic island modifies DNA with 7-deazaguanine derivatives. Proc Natl Acad Sci USA. 113(11):E1452–E1459.

Thomas JM, Briney CA, Nance KD, Lopez JE, Thorpe AL, Fox SD, Bortolin-Cavaille ML, Sas-Chen A, Arango D, Oberdoerffer S, et al. 2018. A chemical signature for cytidine acetylation in RNA. J Am Chem Soc. 140(40):12667–12670.

Tisné C, Rigourd M, Marquet R, Ehresmann C, Dardel F. 2000. NMR and biochemical characterization of recombinant human tRNA(Lys)3 expressed in Escherichia coli: identification of posttranscriptional nucleotide modifications required for efficient initiation of HIV-1 reverse transcription. RNA. 6(10):1403–1412.

Tserovski L, Marchand V, Hauenschild R, Blanloeil-Oillo F, Helm M, Motorin Y. 2016. High-throughput sequencing for 1-methyladenosine (m(1)A) mapping in RNA. Methods. 107:110–121.

Tuorto F, Liebers R, Musch T, Schaefer M, Hofmann S, Kellner S, Frye M, Helm M, Stoecklin G, Lyko F. 2012. RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. Nat Struct Mol Biol. 19(9):900–905.

Vendeix FAP, Murphy t, Frank V, Cantara WA, Leszczyńska G, Gustilo EM, Sproat B, Malkiewicz A, Agris PF. 2012. Human tRNA(Lys3)(UUU) is pre-structured by natural modifications for cognate and wobble codon binding through keto-enol tautomerism. J Mol Biol. 416(4):467–485.

Vermeulen A, McCallum SA, Pardi A. 2005. Comparison of the global structure and dynamics of native and unmodified tRNAval. Biochemistry. 44(16):6024–6033.

Vilfan ID, Tsai YC, Clark TA, Wegener J, Dai Q, Yi C, Pan T, Turner SW, Korlach J. 2013. Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. J Nanobiotechnology. 11:8.

Wein S, Andrews B, Sachsenberg T, Santos-Rosa H, Kohlbacher O, Kouzarides T, Garcia BA, Weisser H. 2020. A computational platform for high-throughput analysis of RNA sequences and modifications by mass spectrometry. Nat Commun. 11(1):926.

Wurm JP, Meyer B, Bahr U, Held M, Frolow O, Kötter P, Engels JW, Heckel A, Karas M, Entian K-D, et al. 2010. The ribosome assembly factor Nep1 responsible for Bowen-

Conradi syndrome is a pseudouridine-N1-specific methyl-transferase. Nucleic Acids Res. 38(7):2387–2398.

Xu L, Seki M. 2020. Recent advances in the detection of base modifications using the Nanopore sequencer. J Hum Genet. 65(1):25–33.

Yamauchi Y, Nobe Y, Izumikawa K, Higo D, Yamagishi Y, Takahashi N, Nakayama H, Isobe T, Taoka M. 2016. A mass spectrometry-based method for direct determination of pseudouridine in RNA. Nucleic Acids Res. 44(6):e59.

Yang J, Håkansson K. 2009. Characterization of oligodeoxynu-cleotide fragmentation pathways in infrared multiphoton dissociation and electron detachment dissociation by Fourier transform ion cyclotron double resonance. Eur J Mass Spectrom. 15(2):293–304.

Yu N, Jora M, Solivio B, Thakur P, Acevedo-Rocha CG, Randau L, de Crecy-Lagard V, Addepalli B, Limbach PA. 2019. tRNA modification profiles and codon-decoding strategies in Methanocaldococcus jannaschii. J Bacteriol. 201(9):e00690.

Yu N, Lobue PA, Cao X, Limbach PA. 2017. RNAModMapper: RNA modification mapping software for analysis of liquid chromatography tandem mass spectrometry data. Anal Chem. 89(20):10744–10752.

Zhang F, Ge W, Ruan G, Cai X, Guo T. 2020. Data-independent acquisition mass spectrometry-based proteomics and software tools: a glimpse in 2020. Proteomics. 20(17–18):1900276.

Zhang LS, Liu C, Ma H, Dai Q, Sun HL, Luo G, Zhang Z, Zhang L, Hu L, Dong X, et al. 2019. Transcriptome-wide mapping of internal N7-methylguanosine methylome in mammalian mRNA. Mol Cell. 74(6):1304–1316.

Zhang N, Shi S, Jia TZ, Ziegler A, Yoo B, Yuan X, Li W, Zhang S. 2019. A general LC-MS-based RNA sequencing method for direct analysis of multiple-base modifications in RNA mixtures. Nucleic Acids Res. 47(20):e125.

Zhao LY, Song J, Liu Y, Song CX, Yi C. 2020. Mapping the epigenetic modifications of DNA and RNA. Protein Cell. 11(11):792–808.

Zhao X, Yu YT. 2004. Detection and quantitation of RNA base modifications. RNA. 10(6):996–1002.

Zheng G, Dahl JA, Niu Y, Fedorcsak P, Huang CM, Li CJ, Vågbø CB, Shi Y, Wang WL, Song SH, et al. 2013. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol Cell. 49(1):18–29.

Zheng G, Qin Y, Clark WC, Dai Q, Yi C, He C, Lambowitz AM, Pan T. 2015. Efficient and quantitative high-throughput tRNA sequencing. Nat Methods. 12(9):835–837.

Zhou H, Kimsey IJ, Nikolova EN, Sathyamoorthy B, Grazioli G, McSally J, Bai T, Wunderlich CH, Kreutz C, Andricioaei I, et al. 2016. m(1)A and m(1)G disrupt A-RNA structure through the intrinsic instability of Hoogsteen base pairs. Nat Struct Mol Biol. 23(9):803–810.

Zhou KI, Parisien M, Dai Q, Liu N, Diatchenko L, Sachleben JR, Pan T. 2016. N(6)-methyladenosine modification in a long noncoding RNA hairpin predisposes its conformation to protein binding. J Mol Biol. 428(5):822–833.

Zhu Y, Pirnie SP, Carmichael GG. 2017. High-throughput and site-specific identification of 2′-O-methylation sites using ribose oxidation sequencing (RibOxi-seq). RNA. 23(8):1303–1314.

Zubarev RA. 2013. The challenge of the proteome dynamic range and its implications for in-depth proteomics. Proteomics. 13(5):723–726.

Chapter 18

Quantification of Modified Nucleosides in the Context of NAIL-MS

Matthias Heiss, Kayla Borland, Yasemin Yoluç, and Stefanie Kellner

Abstract

Recent progress in epitranscriptome research shows an interplay of enzymes modifying RNAs and enzymes dedicated for RNA modification removal. One of the main techniques to study RNA modifications is liquid chromatography-coupled tandem mass spectrometry (LC-MS/MS) as it allows sensitive detection of modified nucleosides. Although RNA modifications have been found to be highly dynamic, state-of-the-art LC-MS/MS analysis only gives a static view on modifications and does not allow the investigation of temporal modification placement. Here, we present the principles of nucleic acid isotope labeling coupled with mass spectrometry, termed NAIL-MS, which overcomes these limitations by stable isotope labeling in human cell culture and gives detailed instructions on how to label cells and process samples in order to get reliable results. For absolute quantification in the context of NAIL-MS, we explain the production of internal standards in detail. Furthermore, we outline the requirements for stable isotope labeling in cell culture and all subsequent steps to receive nucleoside mixtures of native RNA for NAIL-MS analysis. In the final section of this chapter, we describe the distinctive features of NAIL-MS data analysis with a special focus toward absolute quantification of modified nucleosides.

Key words Epitranscriptome, RNA modification, Stable isotope labeling, Mass spectrometry, tRNA, LC-MS/MS

1 Introduction

RNA fulfills major functions for example in translation or gene regulation and requires an extended set of building blocks to allow this functionality. For this purpose, RNA is chemically modified by dedicated enzymes at predefined positions. To study the impact and function of RNA modifications two techniques are primarily used. The first one is sequencing that often exploits the chemical reactivity of modified nucleosides [1, 2] or requires antibodies specific to a modification of interest [3]. The second technique is mass spectrometry coupled for example with complete enzymatic digestion of RNA to the nucleoside level. The resulting nucleoside mixture is analyzed by liquid chromatography-coupled tandem mass spectrometry (LC-MS/MS). Firstly, the nucleosides

are separated on the HPLC by their physicochemical attributes and then analyzed in the mass spectrometer. To allow absolute quantification of modified nucleosides, stable isotope-labeled internal standards (SILIS) are used. These can be produced synthetically [4] or biosynthetically. The SILIS is added to the calibration solutions and samples in defined amounts [5–7].

Sequencing and mass spectrometry are often used as orthogonal techniques to clearly define the modification status of RNA at a given time point. Both techniques do not allow the tracing of modified nucleosides over time and the mechanisms of their placement, their dynamics, and potential removal often remain elusive. This problem can be overcome by metabolic labeling of the RNA in a pulse-chase setup. This technique, termed nucleic acid isotope labeling-coupled mass spectrometry (NAIL-MS), relies on the supplementation with stable isotope-labeled nutrients (e.g., *E. coli* [8], yeast [9], human cells [10]) which results in the formation of nucleosides with a defined number of stable isotopes (isotopologues). An example of isotopologues resulting from NAIL-MS experiments is given in Fig. 1a. Isotopologues possess the same physicochemical attributes and consequently co-elute in liquid chromatography. Even though isotopologues are not separated on the HPLC, these nucleosides can be differentiated due to their different masses, and thus a signal can be assigned to the pulse or
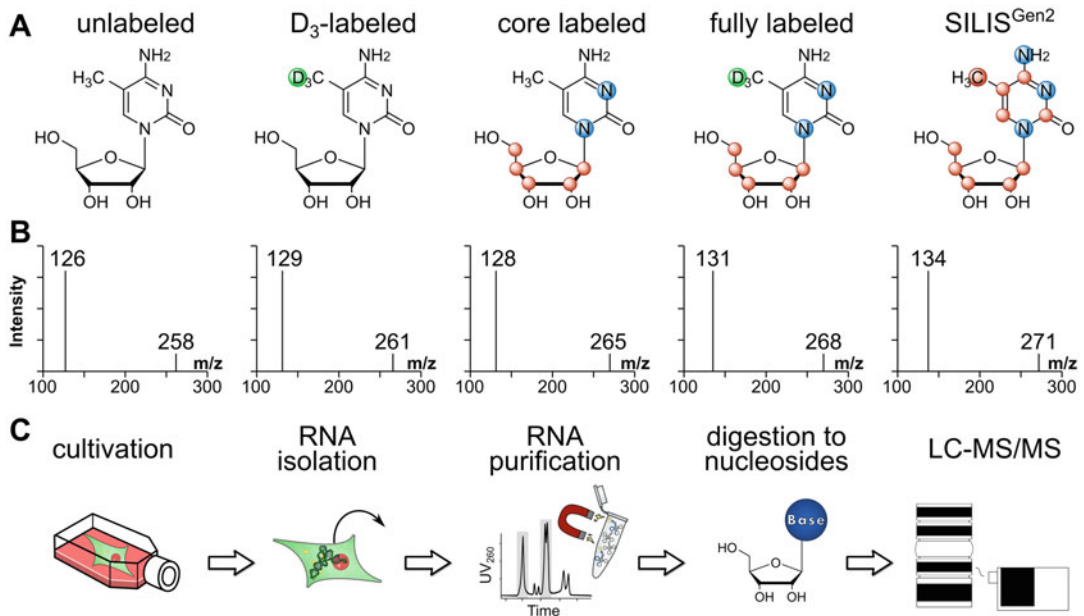
Fig. 1 Principles of NAIL-MS in cell culture. (**a**) Structure of 5-methylcytidine (m^{5}C) and various isotopologues which emerge in a human cell culture NAIL-MS experiment. (**b**) MS/MS spectra of the various m^{5}C isotopologues shown in A. (**c**) General workflow of a NAIL-MS experiment. First cells are cultured in stable isotope-labeled media. RNA is isolated and purified. After digestion to nucleosides, LC-MS/MS is performed

chase phase of the experiment. By the additional introduction of a labeled methyl group, the origin and fate of methylation marks, the most prominent type of RNA modifications, can be studied in a time-resolved fashion. To allow absolute quantification in the context of NAIL-MS a suitable SILIS is needed, which does not interfere with the signals from all emerging isotopologues of a nucleoside resulting from a NAIL-MS experiment. In a sophisticated setup, such as our recently reported study in human cell lines [10], more than five different isotopologues of the modified nucleosides can be reliably differentiated by mass spectrometry. Sensitive quantification is commonly achieved by tandem mass spectrometry (MS/MS). Here, the first mass analyzer selects the ionized nucleoside as the precursor ion and passes it on to the fragmentation chamber. The common fragmentation pattern of modified nucleosides is the cleavage of the glycosidic bond, which results in the formation of a neutral ribose moiety while the charge remains on the nucleobase (Fig. 1b). The charged nucleobase (product ion) is then selected in the second mass analyzer and enters the detector. A key problem of many mass spectrometers is the slow transition of the product ion into the second MS which might result in false signals if the product ion m/z of two isotopologues are identical. Thus, we recommend a labeling scheme which leads to nucleobase isotopologues. In this work, we give a complete workflow on human cell culture NAIL-MS (Fig. 1c) and an overview of modified nucleosides available for NAIL-MS analysis (Table 1) and point out the important steps for successful application of NAIL-MS experiments.

2 Materials

Prepare all solutions and buffers in ultrapure type I water. All steps starting from RNA isolation should be performed with RNase-free equipment and reagents. For LC-MS/MS analysis only use salts and reagents of LC-MS grade.

2.1 Stable Isotope Labeling in Cell Culture

1. Cell line of interest (here HEK293 cells are used).
2. L-Methionine-methyl-D_3 (98% atom, Sigma-Aldrich).
3. $^{13}C_5$, $^{15}N_2$-uridine (ribose-$^{13}C_5$, 98% atom; $^{15}N_2$, 96–98% atom, Cambridge Isotope Laboratories).
4. $^{15}N_5$-adenine ($^{15}N_5$, 98% atom, Cambridge Isotope Laboratories).
5. Growth medium: Dulbecco's modified Eagle media (DMEM) D0422, 10% dialyzed FBS, 0.584 g/L L-glutamine, 0.063 g/L cystine, 0.03 g/L methionine, 0.05 g/L uridine, and 0.014 g/L adenine.
6. Quenching medium: DMEM D0422, 10% dialyzed FBS.
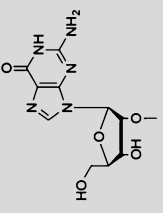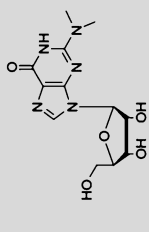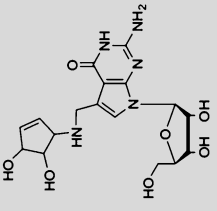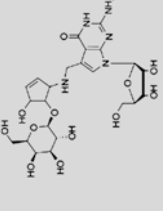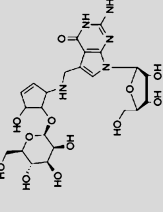
Table 1
Summary of nucleosides that can be quantified in cellular RNAs
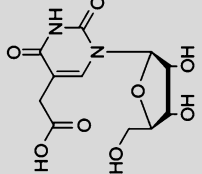
	C (1)	m³C (2)	m⁵C (2)	Cm (2)	ac⁴C (2)	U (1)
A						
B	244 → 112	258 → 126	258 → 126	258 → 112	286 → 154	245 → 113
C	n.a.	261 → 129	261 → 129	261 → 112	n.a.	n.a.
D	n.a.	265 → 128	265 → 128	265 → 114	n.a.	n.a.
E	251 → 114	268 → 131	268 → 131	268 → 114	293 → 156	252 → 115
F	253 → 116	270 → 133	270 → 133	270 → 116	297 → 160	254 → 117
G	256 → 119	271 → 134	271 → 134	271 → 119	300 → 163	256 → 119

	Ψ (2)	D (3)	m⁵U (2)	Um (2)	acp³U (6)	A (1)
A						
B	245 → 209	247 → 115	259 → 127	259 → 113	346 → 214	268 → 136
C	n.a.	n.a.	262 → 130	262 → 113	n.a.	n.a.
D	n.a.	n.a.	266 → 129	266 → 115	n.a.	n.a.
E	252 → 216	254 → 117	269 → 132	269 → 115	353 → 216	273 → 141
F	254 → 218	256 → 119	271 → 134	271 → 117	359 → 222	278 → 141
G	256 → 220	258 → 121	271 → 134	271 → 119	362 → 225	283 → 146

A	m¹A (2)	m⁶A (2)	Am (2)	m^{66}A (4)	i⁶A (7)	t⁶A (5)
B	282 → 150	282 → 150	282 → 136	296 → 164	336 → 204	413 → 281
C	285 → 153	285 → 153	285 → 136	302 → 156	n.a.	n.a.
D	287 → 155	287 → 155	287 → 141	301 → 155	n.a.	n.a.
E	290 → 158	290 → 158	290 → 141	307 → 161	341 → 209	418 → 286
F	295 → 158	295 → 158	295 → 141	312 → 175	351 → 214	428 → 291
G	298 → 161	298 → 161	298 → 146	313 → 176	356 → 219	434 → 297

A	I (2)	m¹I (5)	G (1)	m¹G (2)	m²G (1)	m⁷G (2)
B	269 → 137	283 → 137	284 → 152	298 → 166	298 → 166	298 → 166
C	n.a.	286 → 154	n.a.	301 → 169	301 → 169	301 → 169
D	n.a.	287 → 155	n.a.	302 → 170	302 → 170	302 → 170
E	273 → 141	290 → 158	294 → 157	305 → 173	305 → 173	305 → 173
F	279 → 142	296 → 145	294 → 157	311 → 174	311 → 174	311 → 174
G	283 → 146	298 → 161	299 → 162	314 → 177	314 → 177	314 → 177

(continued)

Table 1
(continued)

A	Gm (2)	m^{22}G (2)	Q (7)	GalQ (8)	ManQ (8)
B	298 → 152	312 → 180	410 → 295	572 → 295	572 → 295
C	301 → 152	318 → 186	n.a.	n.a.	n.a.
D	302 → 156	316 → 184	n.a.	n.a.	n.a.
E	305 → 156	322 → 190	n.a.	n.a.	n.a.
F	311 → 157	328 → 191	n.a.	n.a.	n.a.
G	314 → 162	329 → 192	n.a.	n.a.	n.a.

A	cm^{5}U (7)	ncm^{5}U (2)	mcm^{5}U (9)	ncm^{5}s^{2}U (9)	mcm^{5}s^{2}U (2)
B	303 → 171	302 → 170	317 → 185	318 → 186	333 → 201
C	n.a.	n.a.	320 → 188	n.a.	336 → 204
D	n.a.	n.a.	324 → 187	n.a.	340 → 203
E	310 → 173	309 → 172	327 → 190	325 → 188	343 → 206
F	314 → 177	303 → 176	331 → 194	329 → 192	347 → 210
G	316 → 179	316 → 179	331 → 194	332 → 195	347 → 210

Abbreviation of modified nucleoside (A), mass transitions for unlabeled (B), D$_3$-label (C), core label (D), full label (E), SILIS^{Gen1} (F), and SILIS^{Gen2} (G). n.a. = not applicable. Vendors: Sigma-Aldrich (1); Carbosynth, Newbury, UK (2); Apollo Scientific, Stockport, UK (3); Alfa Chemistry, Ronkonkoma, USA (4); Toronto Research Chemicals, Toronto, CA (5). Synthesis: Kellner lab (6), Dedon lab (7), Carell lab (8) [11], Helm lab (9)

2.2 SILIS Preparation	1. *Saccharomyces cerevisiae* (strain BY4741 MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0).
	2. $^{13}C_6$-glucose (≥99% atom, Eurisotop, Saarbruecken, Germany).
	3. ^{13}C, ^{15}N-rich growth medium for yeast (Silantes, Munich, Germany).
	4. TES buffer: 10 mM Tris–Cl, pH 7.5, 10 mM EDTA, 0.5% SDS.
	5. Acidic phenol: 25 g Phenol, 10 mL ultrapure water, pH 5. Store at -20 °C.
	6. Theophylline.
	7. Ultrapure water.
	8. 5 M NH_4OAc.
	9. Ethanol.
2.3 RNA Isolation and Purification	1. PBS.
	2. TRI Reagent.
	3. Chloroform.
	4. Isopropanol and 70% ethanol.
	5. Ultrapure water.
	6. HPLC system for SEC: Any isocratic system with UV detection, e.g., Agilent 1100.
	7. SEC columns: for tRNA; AdvanceBio SEC 300 Å, 2.7 μm, 7.8 × 300 mm and for rRNA; AdvanceBio SEC 1000 Å, 2.7 μm, 7.8 × 300 mm (Agilent).
	8. SEC buffer: 0.1 M NH_4OAc.
	9. 100 μM Biotinylated DNA oligonucleotide (ON). An example of a biotinylated ON used to purify tRNA$_{Phe}$: [Btn]AAATGG TGCCGAAACCCGGGATCGAACCAGGGT.
	10. B&W buffer: 5 mM Tris–HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl.
	11. SSC buffer (20×): 3 M NaCl, 300 mM trisodium citrate, pH 7.0.
	12. SSC buffer (5×, 1×, 0.1×): Dilute SSC buffer (20×) in water accordingly.
	13. Dynabeads® MyOne™ Streptavidin T1.
2.4 Digestion (See Table 2)	1. Alkaline phosphatase.
	2. Phosphodiesterase I.
	3. Benzonase.
	4. Tetrahydrouridine.

Table 2
Master mix for RNA digestion to nucleoside level

Compound	Stock		Goal	Volume
MgCl$_2$	20 mM	→	1 mM	1.75 µl
Tris, pH = 8	100 mM	→	5 mM	1.75 µl
Benzonase	1 U/µL	→	2 U	2 µl
CIP (Alk. Phos.)	1 U/µL	→	2 U	2 µl
SPD (PDE1)	0.1 U/µL	→	0.2 U	2 µl
Pentostatin	1 mg/mL	→	1 µg	1 µl
THU	5 mg/mL	→	5 µg	1 µl
BHT	10 mM	→	10 nmol	1 µl
H$_2$O				2.5 µl

Multiply volume by sample number to prepare the proper amount of master mix. Pentostatin, THU, and BHT are added to avoid deamination and oxidation of nucleosides

5. Butylated hydroxytoluene.

6. Pentostatin.

7. 20 mM MgCl$_2$.

8. 100 mM Tris, pH 8.

9. 96-Well filter plate (10 kDa MWCO).

2.5 LC-MS

1. High-resolution LC-MS: For example Dionex Ultimate 3000 HPLC system coupled with LTQ Orbitrap XL.

2. High-sensitivity LC-MS/MS: For example Agilent 1290 Infinity II with diode array detector (DAD) and G6470A triple-quadrupole, electrospray ionization (ESI-MS) (Agilent).

3. Chromatography: Phenomenex®, Synergi Fusion-RP 100 Å, 2.5 µm, 150 × 2.0 mm column (Phenomenex).

4. LC-MS buffer/aqueous solvent (A): 5 mM NH$_4$OAc, pH 5.3. Dissolve 0.3854 g NH$_4$OAc in ultrapure water, add 65 µL glacial acetic acid, and bring to a final volume of 1 L in ultrapure water.

5. Organic solvent (B): Acetonitrile.

6. LC-MS vials: With 200 µL insert.

7. Unlabeled synthetic nucleosides for calibration (Table 1).

3   Methods

3.1   General Design of NAIL-MS Experiments

NAIL-MS experiments can be subdivided into several categories depending on the purpose of the experiment. The two major categories are comparative and pulse-chase NAIL-MS. Comparative NAIL-MS experiments are comparable to SILAC proteomics experiments and are highly useful for method validation. Through pulse-chase NAIL-MS experiments, the dynamics of the epitranscriptome are investigated (Fig. 2). It is possible to use these NAIL-MS experiments without the addition of a SILIS and determine the relative changes of RNA modifications. However, we recommend the production and addition of a SILIS in order to receive absolute values for modified nucleosides in NAIL-MS experiments.
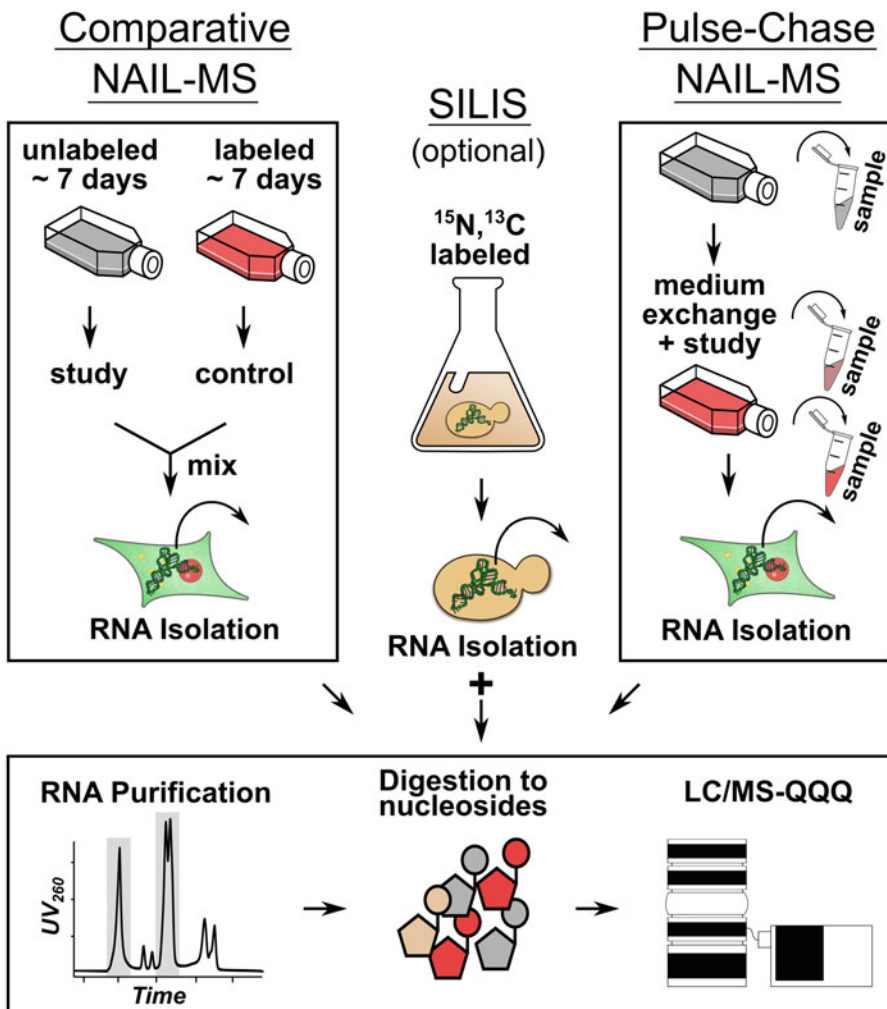

Fig. 2 Experimental procedure of comparative NAIL-MS (left) and pulse-chase NAIL-MS experiments (right). Although they differ in cell handling, the downstream processing is comparable. Gray and red represent media with different stable isotope labeling. The production of SILIS is shown in the middle and its addition is recommended for absolute quantification of NAIL-MS experiments

3.1.1 Considerations for SILIS Production

If a biosynthetic SILIS is used as an internal standard (ISTD), for many modifications complete labeling of all nucleosides is crucial. To circumvent an overlap with isotopologues of the NAIL-MS samples the mass increase should be as high as possible. We recommend *S. cerevisiae* for the production of a eukaryotic SILIS. Cultivation in ^{15}N- and ^{13}C-labeled Silantes complete growth medium leads to the m/z values shown in Table 1. This SILIS^{Gen2} labeling ensures a mass difference of at least 2 Dalton (Da) to any NAIL-derived nucleosides, which is important for reliable differentiation by the mass spectrometer. Our previously reported SILIS^{Gen1} [7] is less suitable for NAIL-MS experiments as it contains some residual nucleosides with incompatible labeling (*see* **Note 1**).

3.1.2 Considerations for Comparative NAIL-MS

Choosing the correct labeling strategy is crucial for comparative NAIL-MS experiments. The chosen medium must lead to distinct mass differences for nucleosides emerging from each culture (*see* **Note 2**). The goal is to enable mixing of the culture of interest (e.g., knockout (KO) strain, chemical treatment) and the control culture at the stage of cell lysis, followed by co-processing to reduce purification bias.

3.1.3 Considerations for Pulse-Chase NAIL-MS

For pulse-chase experiments the same principles as described for comparative NAIL-MS apply. It is important that the exchange of medium A by medium B results in isotopologues that differ by at least 2 Da. Only then it is possible to reliably distinguish between RNA molecules already existent before the experiment's initiation and RNA molecules transcribed after the experiment's initiation (*see* **Note 3**). This allows the detailed study of modification dynamics in vivo.

***3.2 Preparation of SILIS (10×) in* S. cerevisiae**

1. For preparation of SILIS^{Gen2} in *S. cerevisiae*, prepare 5 mL of ^{13}C, ^{15}N Silantes-rich growth medium supplemented with 1% (w/w) ^{13}C-glucose. (here: 250 μL sterile ^{13}C$_6$-glucose stock solution (200 g/L) to 4.75 mL ^{13}C, ^{15}N Silantes-rich growth medium), inoculate with a single-cell colony, and cultivate yeast overnight (30 °C, 250 rpm) in a shaking incubator.

2. Dilute the culture to OD 0.1 with fresh ^{13}C, ^{15}N medium supplemented with ^{13}C$_6$-glucose. We recommend a final culture volume of 100 mL. Continue cultivation for 2 days.

3. Split the culture into 50 mL aliquots and harvest the cells by centrifugation (3000 × *g*, 5 min, 4 °C). After discarding the supernatant, wash the pellet with 5 mL ultrapure water, transfer the suspension to clean tubes, and centrifuge again (3000 × *g*, 5 min, 4 °C).

4. After discarding the supernatant, resuspend each pellet in 4 mL TES buffer and add 4 mL acidic phenol to the suspension (*see* **Note 4**).

5. Incubate the sample at 65 °C for 1 h and vigorously vortex the mixture every 15 min.

6. After 1 h of incubation, place the sample on ice for 5 min, and then centrifuge the sample ($3000 \times g$, 5 min, 4 °C) to induce phase separation.

7. Transfer the upper aqueous phase into a clean centrifugation tube and add 4 mL acidic phenol. Vortex this mixture vigorously for 20 s and place on ice for 5 min. Centrifuge the sample at $3000 \times g$ for 5 min at 4 °C.

8. Again, transfer the upper aqueous phase into a clean centrifugation tube and add 4 mL chloroform. Vortex this mixture vigorously and keep it on ice for 5 min. Subsequently, centrifuge the sample ($3000 \times g$, 5 min, 4 °C).

9. Aliquot 1 mL of the resulting upper aqueous phase into 5 mL reaction tubes. Perform an ethanol precipitation with NH_4OAc followed by an ethanol wash.

10. Dissolve the resulting RNA in 250 μL ultrapure water. The RNA isolation procedure for yeast is summarized in Fig. 3.

11. The total RNA SILIS can be purified to tRNA and rRNA SILIS by SEC as will be described in Subheading 3.5.

12. After purification and precipitation of tRNA or rRNA, digest 3 μg of the RNA to nucleosides as will be described in Subheading 3.7.
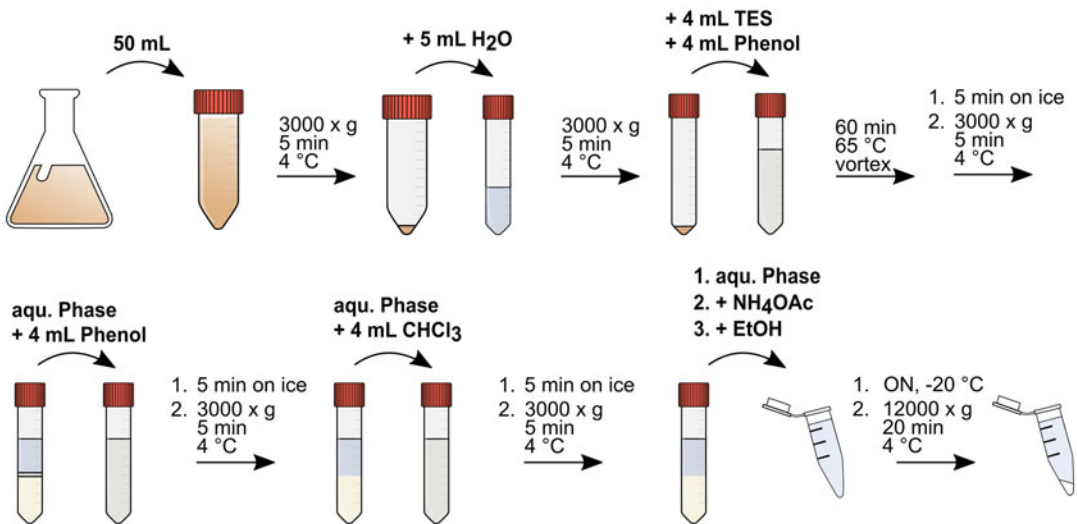

Fig. 3 Yeast RNA isolation procedure used for the preparation of SILIS. Yeast cells are harvested by centrifugation. The RNA is isolated by hot phenol/chloroform extraction, followed by ethanol precipitation with each step outlined

13. The SILIS (10×) should contain 10 mM theophylline as an external standard (*see* **Note 5**). Therefore, prepare a 100 mM theophylline stock in water and add 15 μL of it to 35 μL of RNA digest and add 100 μL of LC-MS buffer, resulting in 150 μL of SILIS (10×).

3.3 Stable Isotope Labeling of RNA in Cell Culture

For stable isotope labeling of HEK293 cells we chose DMEM D0422, which lacks methionine (and cystine) and thus allows complete labeling of methyl groups [10]. Cells grown in DMEM should be kept at 5–10% CO_2 for proper pH adjustment. Labeling of other cell lines and use of alternative media are also possible (*see* **Note 6**).

1. Prepare stock solutions of 29.2 g/L glutamine (50×) and 5.0 g/L uridine (100×) in water and freeze in aliquots. Prepare stock solutions of 15.0 g/L methionine (500×) and 0.7 g/L adenine (50×) in water and 78.8 g/L cystine in 1 M HCl and store at 4 °C.

2. Prepare the growth medium using DMEM D0422, dialyzed FBS, and stock solutions of glutamine (50×), methionine (500×), cystine (1250×), uridine (100×), and adenine (50×). For example for the preparation of 50 mL of fully labeled media mix 42.4 mL DMEM D0422 with 5 mL dialyzed FBS, 1 mL glutamine, 100 μL CD_3-labeled methionine, 40 μL cystine, 500 μL $^{15}N_2,^{13}C_5$-labeled uridine, and 1 mL $^{15}N_5$-labeled adenine (Fig. 4). Otherwise, uridine, adenine, and methionine are either added as unlabeled or labeled compounds depending on the desired labeling (*see* **Note 7**).

3. To prevent incomplete labeling, quenching medium should be used for trypsin deactivation during splitting procedures. It consists of DMEM D0422 and dialyzed FBS only and thereby prevents the carryover of (un)labeled compounds into the new cell culture flask (*see* **Note 8**).
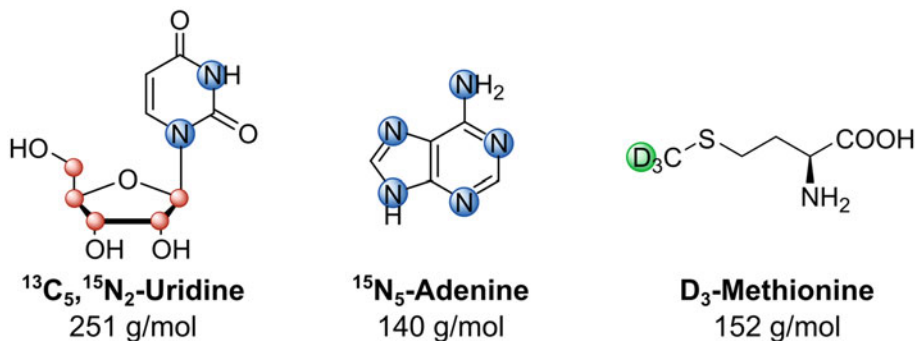

Fig. 4 Compounds used for stable isotope labeling in cell culture. Structures of $^{13}C_5,^{15}N_2$-uridine (left), $^{15}N_5$-adenine (middle), and CD_3-methionine (right) are shown

4. For complete labeling of HEK293 cultures, cells should be cultivated in the growth medium for at least 7 days (including at least two splitting steps with fresh medium). LC-MS/MS signals of stable isotope-labeled nucleosides are already detectable after 1 h of labeling. 50% labeling is achieved after ~2 days of labeling.

3.4 RNA Isolation from Human Cells

All steps starting from RNA isolation should be performed with RNase-free reagents and equipment.

1. After aspiration of the medium, wash cells carefully with PBS (*see* **Note 9**).

2. After aspirating PBS, cells are directly harvested and lysed in culture flasks using TRI Reagent. We suggest using 1 mL per 8×10^6 cells ($\triangleq$ confluent T25 flask of HEK293). Thoroughly pipette up and down and transfer the cell suspension into a 1.5 mL tube.

3. Vortex for 20 s and then incubate for 5 min at room temperature (*see* **Notes 10** and **11**).

4. Add $\frac{1}{5}$ of the volume of TRI Reagent used for cell lysis of chloroform (e.g., 200 μL chloroform to 1 mL cell suspension in TRI Reagent) and mix thoroughly until the whole suspension becomes uniformly opaque.

5. Leave the mixture at room temperature for 5 min and centrifuge for 10 min at $10,000 \times g$ at 4 °C.

6. Transfer the aqueous phase (upper, clear) into a new 1.5 mL tube and add an equal volume of isopropanol (e.g., ~500 μL isopropanol needed per 1 mL TRI Reagent).

7. Mix thoroughly and precipitate RNA overnight at -20 °C (*see* **Note 12**).

8. Centrifuge RNA at 4 °C at $12,000 \times g$ for 60 min. Implement a wash step with 70% ethanol.

9. Resuspend pellet in 30–100 μL ultrapure water. Resuspended RNA can be stored at -20 °C for several years (*see* **Note 13**).

3.5 RNA Purification by Size-Exclusion Chromatography (SEC)

1. For purification of tRNA and bulk rRNA, size-exclusion chromatography (SEC) on an HPLC system is employed using SEC buffer as the mobile phase [12]. An AdvanceBio SEC 300 Å, 2.7 μm, 7.8 × 300 mm column allows fast separation of tRNA from rRNAs using an isocratic elution at 1 mL/min with a column temperature of 40 °C [13]. After equilibration of the column for at least 30 min, up to 100 μg of total RNA can be injected. The large rRNA subunits co-elute from 3.5 to 4.8 min and the pure tRNA elutes from 6.9 to 7.9 min (*see* **Note 14**).

2. For purification of 18S rRNA and 28S rRNA an AdvanceBio SEC 1000 Å, 2.7 μm, 7.8 × 300 mm column is used.
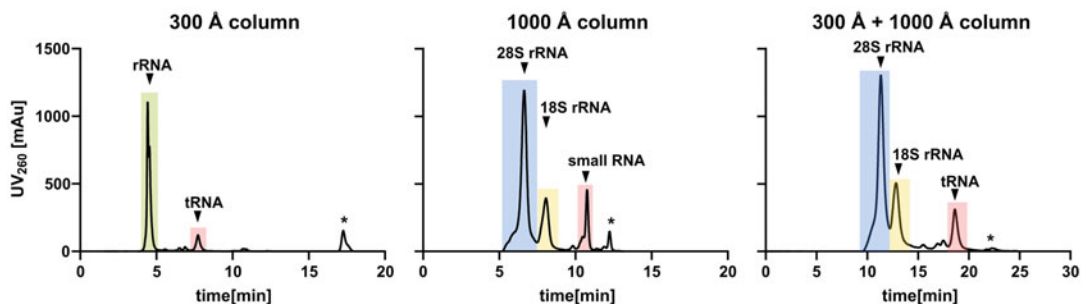
Fig. 5 Elution profiles of total RNA separated by size-exclusion chromatography. For purification, a 300 Å column, a 1000 Å column, or both in combination are used. Small RNA may consist of 5S rRNA, 5.8S rRNA, and tRNA. (*) indicates common contaminants (small molecules) in RNA samples isolated with TRI Reagent

Parameters are the same as for the 300 Å column. 28S rRNA elutes from 5.0 to 7.2 min and 18S rRNA from 7.5 to 8.5 min (*see* **Note 15**).

3. If parallel purification of 28S rRNA, 18S rRNA, and tRNA is desired, the two columns can be installed in tandem. The 1000 Å column can be directly connected behind the 300 Å column using a short piece of 0.15 mm inner diameter capillary. One run takes 30 min with all other chromatographic parameters remaining identical to single column use. 28S rRNA then elutes from 9.5 to 11.9 min, 18S rRNA from 12.6 to 14.8 min, and tRNA from 18.0 to 20.0 min (Fig. 5).

4. Collect the desired fractions and concentrate them to ~50 μL using a lyophilizer or vacuum concentrator (*see* **Note 16**).

5. Add $\frac{1}{10}$ of the volume of 5 M NH_4OAc, and then add 2.5× of the volume of ice-cold 100% ethanol (*see* **Note 17**). Precipitate and resuspend RNA as described in Subheading 3.4.

3.6 RNA Purification by Oligonucleotide Hybridization Assay

Purification of specific RNA types can be done by oligonucleotide hybridization. Our protocol is a variation of a published protocol by the Helm lab [14]. Please *see* Fig. 6 for an overview of the procedure.

1. A DNA oligonucleotide (ON) of ~30 nucleotides in length with an additional AAA-tail and a biotin tag is designed complementary to the sequence of the target RNA (*see* **Note 18**). For example, the biotinylated ON used to purify tRNA$_{Phe}$ is shown here: [Btn] AAATGGTGCCGAAACCCGGGATC GAACCAGGGT.

2. Equilibrate the beads by transferring 25 μL of streptavidin beads T1 (*see* **Note 19**) for each sample into a 1.5 mL tube (e.g., 200 μL for eight samples).

3. Place the tube on a magnetic rack and let the beads attach to the wall of the tube. Then, carefully aspirate and discard the liquid.
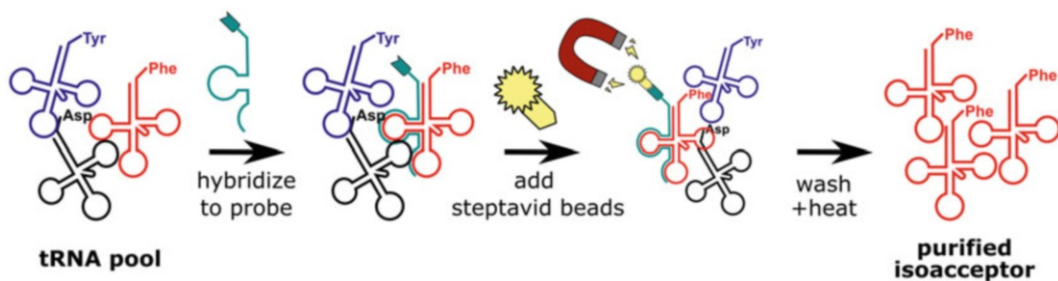
Fig. 6 RNA purification using oligonucleotide hybridization. SEC-purified tRNA is incubated with a biotinylated DNA probe (green) complementary to the tRNA of interest (red). Using streptavidin-coated magnetic beads, the tRNA of interest is purified from total tRNA. This procedure can be applied to other RNA molecules in addition to tRNA

4. Resuspend the remaining beads in 25 μL (here: 200 μL for eight samples) of B&W buffer and repeat **step 3**. Repeat the wash twice with B&W buffer and then once with SSC buffer (5×). Finally, resuspend the beads in 25 μL (here: 200 μL for eight samples) SSC buffer (5×).

5. For hybridization of RNA, mix up to 100 pmol of, ideally, size-prepurified RNA (*see* **Note 20**) with 100 pmol of ON in SSC buffer (5×) in a final volume of 100 μL. For example, mix 20 μL total tRNA (or RNA of interest) (75 ng/μL), 1 μL ON (100 μM), 25 μL SSC buffer (20×), and 54 μL water. Heat the mixture to 90 °C for 3 min, and then instantaneously incubate for 10 min at 65 °C. Finally, allow the mixture to cool down to room temperature.

6. Transfer 25 μL of equilibrated beads into each of the hybridized RNA samples at room temperature. Mix thoroughly and incubate the samples on a thermomixer for 30–60 min shaking at 600 rpm at room temperature.

7. To remove unbound RNA, place the sample back on the magnetic rack and repeat **step 3** above. Resuspend the beads in 50 μL SSC buffer (1×). Repeat **step 7** with SSC buffer (0.1×) three times.

8. Finally, resuspend the beads in 10–30 μL water and incubate for 2 min at 75 °C. Subsequently, put the sample on the magnetic rack and transfer the RNA-containing liquid into a new tube.

3.7 RNA Digestion and Filtration

1. Dilute up to 1 μg of purified RNA in 20 μL ultrapure water.

2. Freshly prepare a master mix for digestion according to Table 2.

3. Add 15 μL of the master mix to each sample and mix by pipetting up and down.

4. Incubate the samples for 2 h at 37 °C.

5. Add 15 µL of LC-MS buffer (*see* **Note 21**).

6. Transfer the whole sample volume to a 96-well filter plate (10 kDa MWCO) mounted on a skirted PCR plate (alternatively use single-filter tubes with 10 kDa MWCO; see the manufacturer's manual) and centrifuge for 30 min at 3000 × *g* and 4 °C (*see* **Note 22**).

7. Transfer 2.5 µL of SILIS (10×) into LC-MS vials and add 22.5 µL of the digested centrifuged RNA sample (from **step 6**). Mix by pipetting up and down (*see* **Note 23**).

3.8 Calibration

1. For calibration, weigh and dissolve all synthetic nucleosides (Table 1) in water to a stock concentration of 10 mM. Exceptions are G, m^2G, m^{22}G, and Q which are, due to low solubility in water, dissolved to a stock concentration of 1 mM.

2. To prepare the calibration solutions, first mix and dilute the desired nucleosides in a final concentration of 100 µM for the canonical nucleosides and 5 µM for the modified nucleosides (*see* **Note 24**). Aliquot the resulting "nucleoside start mix" in 25 µL and store at -20 °C.

3. Prior to usage, thaw one aliquot and dilute 10 µL with 90 µL water. This solution is used for the highest concentration of calibration (=L12).

4. We suggest serial dilution by mixing 50 µL with 50 µL water. Repeat until 12 solutions with descending concentration are prepared (L1–L12, 12×, 1:2 dilution).

5. Transfer 2.5 µL of SILIS (10×) into LC-MS vials and add 22.5 µL of calibration solution. Mix by pipetting up and down. An overview of preparation of calibration solutions and the resulting chromatogram is shown in Fig. 7.
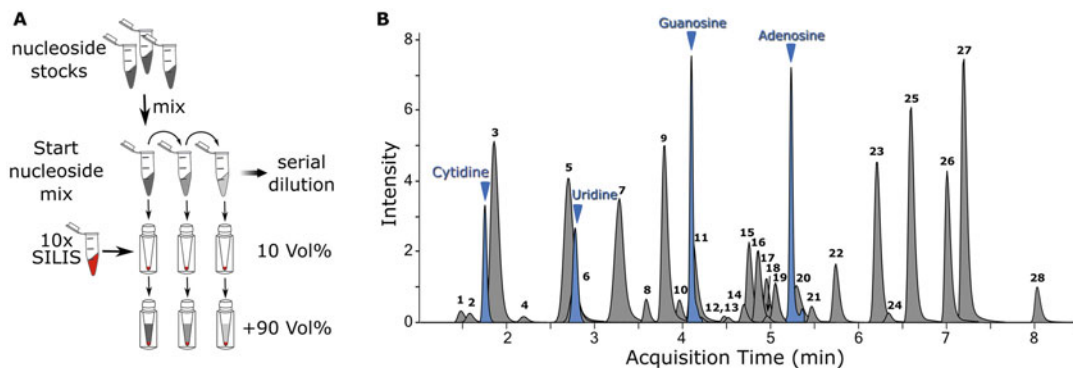
Fig. 7 LC-MS/MS calibration measurement. (**a**) Overview of calibration solution preparation. (**b**) Chromatogram with UV (blue) and MS (black) trace indicating the identity of all peaks. 1: D, 2: Ψ, 3: C, 4: ncm^5U, 5: m^3C, 6: U, 7: m^1A, 8: m^5C, 9: m^7G, 10: I, 11: G, 12: s^2U, 13: Um, 14: m^3U, 15: m^1I, 16: m^1G, 17: Gm, 18: mcm^5U, 19: m^2G, 20: A, 21: t^6A, 22: m^{22}G, 23: Am, 24: mcm^5s^2U, 25: m^6A, 26: m^6Am, 27: m^{66}A, 28: i^6A

3.9 LC-MS/MS Measurement

1. The chromatographic separation of the analytes is implemented by high-performance liquid chromatography (HPLC). The separation is performed using a Synergi Fusion-RP 100 Å 2.5 μm, 150 × 2.0 mm column from Phenomenex with a gradient elution.

2. Regarding the separation of nucleosides on the HPLC we suggest the following setup:

Parameter	Setting
Column oven temperature	35 °C
Flow rate	0.35 mL/min
Aqueous solvent (A)	LC-MS buffer
Organic solvent (B)	Pure acetonitrile

3. The liquid chromatography is run with the following gradient:

Time	Aqueous solvent (A) LC-MS buffer	Organic solvent (B) Acetonitrile
0–1 min	100%	0%
1–4 min	Decreasing to 90%	Increasing to 10%
4–7 min	Decreasing to 60%	Increasing to 40%
7–8 min	60%	40%
8–11 min	100%	0%

4. We suggest optimization of source parameters using a mix of the four canonical nucleosides. Optimized parameters determined by our lab using Agilent's "Source Optimizer" software are as follows:

Parameter	Setting
Ionization	ESI
Ion mode	Positive
Skimmer voltage	15 V
Cell accelerator voltage	5 V
N2 gas temperature	230 °C
N2 gas flow rate	6 L/min
Sheath gas (N2) temperature	400 °C
Sheath gas (N2) flow rate	12 L/min

(continued)

Parameter	Setting
Capillary voltage	2500 V
Nozzle voltage	0 V
Nebulizer	40 psi

5. Continue with the optimization of individual nucleosides. Implement a product ion scan in order to determine the fragmentation of the respective nucleoside. Then continue with the optimization of fragmentor voltage (50–250 V) and collision energy (5–25 eV) and the determination of retention times for every nucleoside of interest with the goal of receiving the highest sensitivity possible. This can be conducted manually or with Agilent's "Optimizer" software. Optimized parameters for each nucleoside used in our settings are given in Table 3 (*see* **Note 25**).

6. Design a dynamic multiple reaction monitoring (dMRM) method for sample measurement. As the sample contains a mixture of nucleosides and their isotopologues, determine the precursor ion and product ion for every isotopologue that might occur in your mixture. Then, fill in the optimized parameters for each nucleoside. If you decide to use a retention time window (ΔRT), determine the ΔRT for each compound. For most modified nucleosides, we suggest a 0.5–1 min ΔRT. Slight changes of pH in the samples might lead to shifted retention times, especially for m^3C, m^1A, and m^7G. Thus, we recommend a ΔRT of 2–3 min for those.

7. For later analysis using Agilent's "Quantitative Mass Hunter" software it is important to specify every isotopologue by a unique compound name. Additionally, assign every analyte, including SILIS and the isotopologues, to their respective compound group and tick the SILIS box only for the SILIS isotopologue. An example is given in Table 4.

3.10 Data Analysis   We analyze data with the quantitative and qualitative MassHunter Software from Agilent. Detailed instruction on how to handle Agilent's MassHunter Workstation Software for Quantitative Analysis is given in their Familiarization Guide which can be found online (https://www.agilent.com/cs/library/usermanuals/Public/G3335_90061_Quant_Familiarization-EN.pdf) or in a video on the Kellner lab homepage (https://www.cup.lmu.de/oc/kellner/).

1. The areas of the MS signals are integrated for each nucleoside. For the calibration, the values of integrated MS signals from target nucleosides are set in relation to the MS signals of the

Table 3
Optimized QQQ parameters for nucleosides

Compound name	Precursor ion	MS1 Res	Product ion	MS2 Res	Ret time (min)	Delta Ret time	Fragmentor	Collision energy
A	268.1	Wide	136	Unit	5.2	1	200	20
ac^4C	286.1	Wide	154	Unit	5	1	85	9
acp^3U	346.1	Wide	214	Unit	2.3	1	95	15
Am	282.1	Wide	136	Unit	6	1	130	17
C	244.1	Wide	112	Unit	2.1	1	200	20
Cm	258.1	Wide	112	Unit	4.1	1	180	9
D	247.1	Wide	115	Unit	1.6	1	70	5
G	284.1	Wide	152	Unit	4.3	1	200	20
Gm	298.1	Wide	152	Unit	5	1	100	9
I	269.1	Wide	137	Unit	4.1	1	100	10
i^6A	336.3	Wide	204	Unit	8	1	140	17
m^1A	282.1	Wide	150	Unit	2.2	1.5	150	25
m^1G	298.1	Wide	166	Unit	4.9	1	105	13
m^1I	283.1	Wide	151	Unit	4.8	1	80	12
$m^1\Psi$	259.0	Wide	223	Unit	3.1	1	85	5
$m^{22}G$	312.1	Wide	180	Unit	5.7	1	105	13
m^2G	298.1	Wide	166	Unit	5.1	1	95	17
m^3C	258.1	Wide	126	Unit	2.3	1.5	88	14
m^3U	259.1	Wide	127	Unit	4.8	1	75	9
m^5C	258.1	Wide	126	Unit	3.8	1	185	13
m^5U	259.1	Wide	127	Unit	4.4	1	95	9
$m^{66}A$	296.0	Wide	164	Unit	7.1	1	130	21
$m^{66}Am$	310.0	Wide	164	Unit	7.5	1	120	15
m^6A	282.1	Wide	150	Unit	6.5	1	125	17
m^6Am	296.0	Wide	150	Unit	7	1	125	17
m^7G	298.1	Wide	166	Unit	3.6	1	100	13
mcm^5s^2U	333.1	Wide	201	Unit	6.2	1	92	8
ncm^5s^2U	318.1	Wide	186	Unit	4.2	1	95	7
mcm^5U	317.1	Wide	185	Unit	5	1	95	5
ncm^5U	302.0	Wide	170	Unit	2.5	1	85	8
cm^5U	303.1	Wide	171	Unit	2	1	100	7

(continued)

Table 3
(continued)

Compound name	Precursor ion	MS1 Res	Product ion	MS2 Res	Ret time (min)	Delta Ret time	Fragmentor	Collision energy
Q	410.2	Wide	295	Unit	4.3	1	115	12
s^2U	261.1	Wide	129	Unit	4.3	1	80	6
t^6A	413.1	Wide	281	Unit	5.8	1	130	9
U	245.1	Wide	113	Unit	3	1	95	5
Um	259.2	Wide	113	Unit	4.6	1	96	8
Ψ	245.1	Wide	209	Unit	1.7	1	90	5
ManQ	572.3	Wide	295.5	Unit	3.9	1	120	20
GalQ	572.3	Wide	295.5	Unit	4.1	1	115	20

Determined by using unlabeled synthetic nucleosides. The same parameters can be applied to labeled nucleosides (with separate compound name) when m/z of precursor and product ions are increased accordingly

Table 4
Setting of compound groups during dMRM method setup of nucleosides (here shown for Am)

Compound name	Compound group	ISTD
Am	Am	False
Am D_3 lab	Am	False
Am core lab	Am	False
Am fully lab	Am	False
Am SILIS	Am SILIS	True

respective SILIS (*see* **Note 26**) to receive the nucleoside isotope factor (NIF, Eq. (1)):

$$NIF_{nucleoside} = \frac{signal\ area_{nucleoside}}{signal\ area_{respective\ SILIS}} \qquad (1)$$

2. Results from Eq. (1) are plotted against the expected molar amount of nucleosides and regression curves are plotted through the data points. The slopes represent the respective relative response factors for the nucleosides (rRFN) and enable an absolute quantification. The plotting of these calibration curves is done automatically by the quantitative MassHunter software and should be checked manually for linearity (*see* **Note 27**). This principle is outlined in Fig. 8a. The calibration
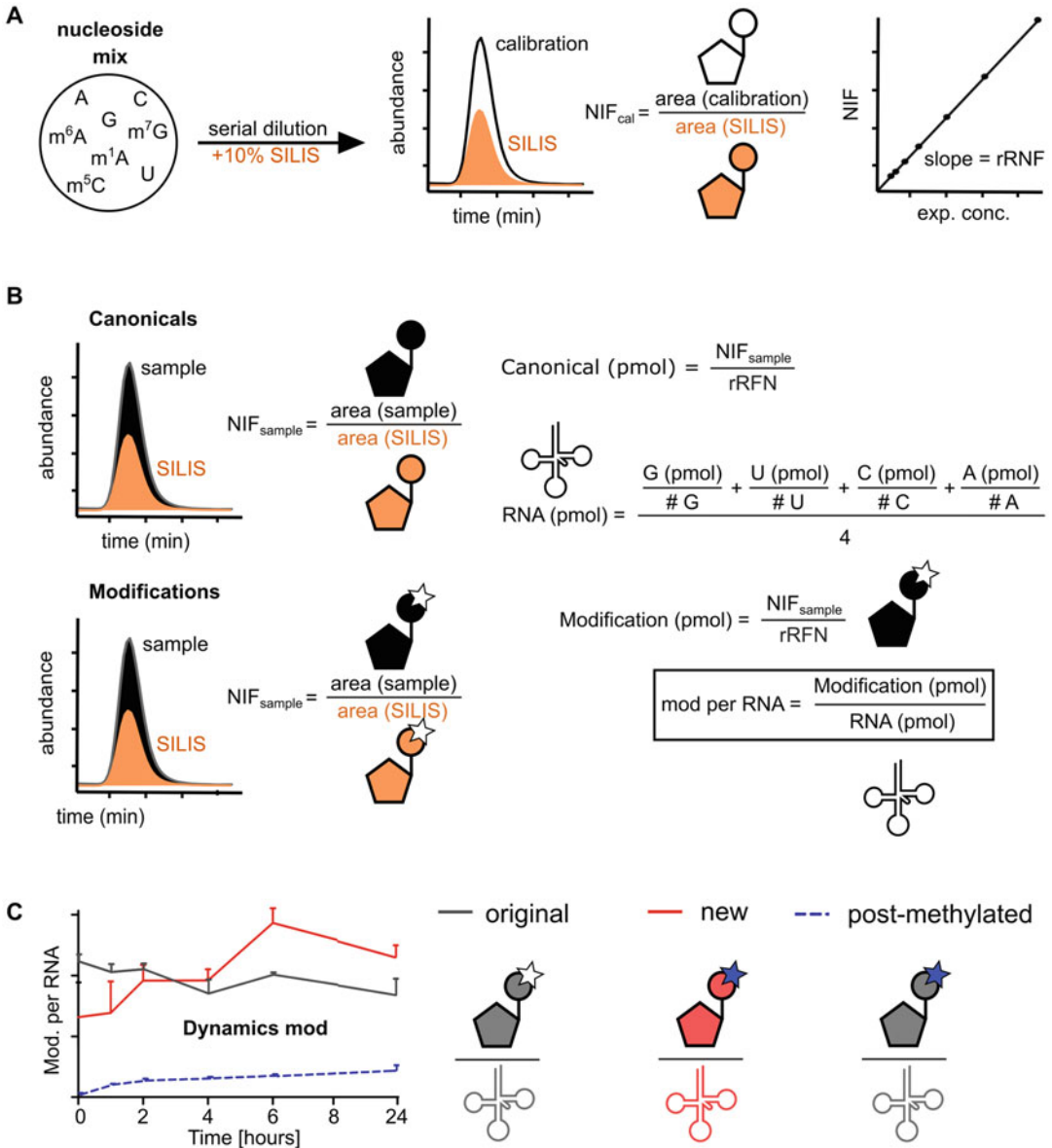
Fig. 8 Concepts of absolute quantification by LC-MS/MS. (**a**) Calibration curves are plotted by calculating the NIF of each nucleoside. (**b**) Absolute amounts of canonicals and modifications are calculated by dividing the sample area by the respective SILIS area and by applying the previously determined calibration curves. Modification per tRNA can be calculated by referencing to the injected amount of RNA molecules (based on expected numbers of canonicals). (**c**) Absolute amounts of modifications have to be referenced to the respective labeled canonicals. Thereby original, new, and post-methylated transcripts can be investigated in parallel

curves measured with the synthetic standards (unlabeled) are used for each isotopologue of the respective nucleoside (*see* **Note 28**).

3. Molar amounts of nucleosides in samples are then calculated according to Eq. (2) using the signal areas of target compounds and SILIS in the samples and the respective rRFN, determined by calibration measurements. This step is done automatically by the quantitative MassHunter software:

$$n_{\text{sample nucleoside}} = \frac{\text{signal area}_{\text{sample nucleoside}}}{\text{rRFN}_{\text{nucleoside}} \times \text{signal area}_{\text{respective SILIS}}} \quad (2)$$

4. To make different samples quantitatively comparable, the molar amount of each modified nucleoside has to be normalized by the molar amount of canonical nucleosides. This can be done by normalizing the molar amount of a single or the sum of canonical nucleosides or by normalizing the molar amount of injected RNA to receive the number of modifications per RNA molecule. Therefore, the calculated amount of injected canonical nucleotides must be divided by their expected occurrence in the respective RNAs and averaged afterwards (Eq. (3)). The numbers for each canonical nucleoside are either taken from known sequences or determined empirically. Figure 8b summarizes the theory of MS quantification. The different isotopologues have to be referenced to their corresponding labeled canonicals (e.g., unlabeled modifications have to be referenced to unlabeled canonicals). An example for the labeled nucleoside m^7G is shown in Table 5. A quick overview of the principles of absolute quantification in a NAIL-MS experiment is shown in Fig. 8c:

$$n_{t\text{RNA}} = \frac{\frac{n_C}{\#C} + \frac{n_U}{\#U} + \frac{n_G}{\#G} + \frac{n_A}{\#A}}{4} \quad (3)$$

Table 5
Quantification of m^7G per tRNA (based on G)

m^7G (pmol)	G (pmol)	m^7G per tRNA
$\dfrac{\text{area } m^7G \left({}^{15}N, CD_3 \right)}{\text{rRFN } m^7G \text{ area } m^7G \text{ (SILIS)}}$	$\dfrac{\text{area G} \left({}^{15}N \right)}{\text{rRFN G area G (SILIS)}}$	$\dfrac{m^7G \text{ (pmol)}}{\frac{G \text{ (pmol)}}{\#of \ G \ in \ sequence}}$

First the molar amount of injected nucleosides is calculated based on the signal areas of target nucleosides and SILIS and the respective calibration curves (here for m^7G and G). These steps are done automatically by the quantitative MassHunter software. Then the molar amount of modification is divided by the molar amount of respective RNA molecules calculated by dividing the molar amount of canonical nucleosides by the expected number (#) of the respective canonical (here based on G)

4   Notes

1. The first generation of SILIS is suited for quantification of unlabeled RNA samples from, for example, tissue samples. It is produced in the presence of $^{12}CD_3$-methionine using ^{13}C-rich growth medium. Due to the presence of $^{13}CH_3$-methionine in the ^{13}C-rich growth medium only 80% of the methyl marks were $^{12}CD_3$-labeled while 20% were $^{13}CH_3$-labeled and thus 2 Da lighter than required for our newly developed cell culture NAIL-MS experiments (Fig. 9). Our second-generation SILIS does not rely on methionine labeling and the successful monoisotopic labeling is shown in Fig. 10.
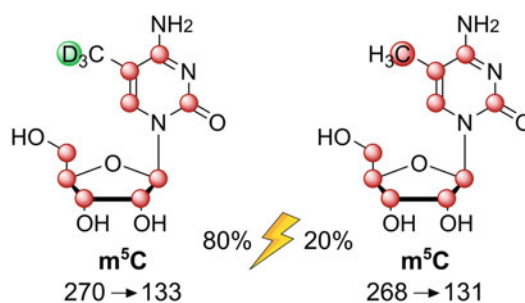
Fig. 9 Incomplete labeling of SILIS[Gen1]. Desired labeling of SILIS[Gen1] (80%) (left) and undesired labeling of SILIS[Gen1], where the methyl groups are $^{13}CH_3$- instead of $^{12}CD_3$-labeled (20%) (right)
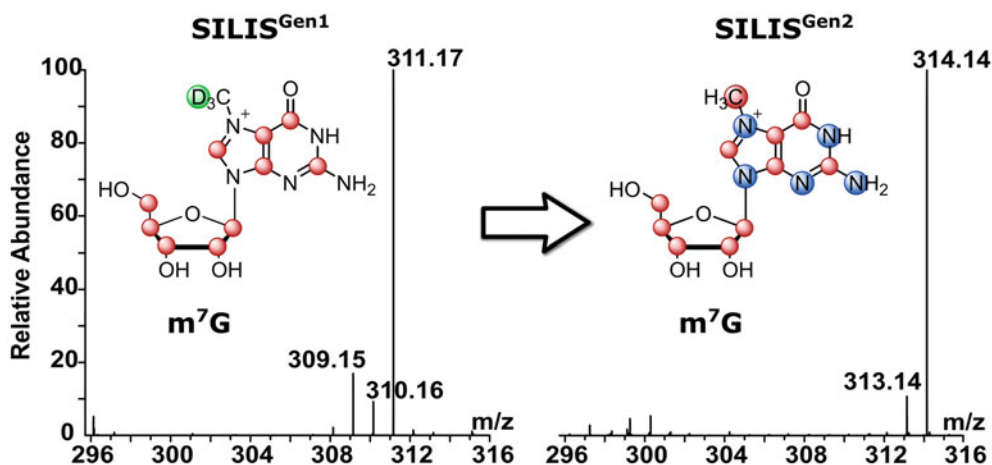
Fig. 10 Improvement of defined labeling of methylated nucleosides in SILIS[Gen2]. SILIS[Gen1] was labeled using ^{13}C Silantes-rich growth medium supplemented with ^{13}C-glucose and CD_3-methionine which led to incomplete labeling of methylated nucleosides (left). SILIS[Gen2] was labeled using ^{13}C, ^{15}N Silantes-rich growth medium supplemented with ^{13}C-glucose where all respective atoms are ^{13}C- or ^{15}N-labeled (right)

2. For example, labeling of purines and the respective modifications in medium A (+5 for adenosines, +4 for guanosines) with a CD_3-methionine-labeled culture in medium B (+3 for all methylated nucleosides) could potentially lead to difficulties in data acquisition as the resulting isotopologues differ by only 1–2 Da. We recommend one unlabeled culture while the other culture is ^{15}N-, ^{13}C-, and D_3-labeled.

3. Compared to unlabeled nucleosides, we observe very low MS background signals for labeled nucleosides. Therefore, we recommend starting with unlabeled cells and switch to labeled medium upon initiation of the experiment. We refer to this approach as "forward" experiment. This is important as the abundance of new nucleosides (=labeled in forward) is naturally very low upon experiment initiation.

4. Collart et al. [15] suggest storing acidic phenol at 4 °C. We observed diminished extraction efficiency over time, when acidic phenol was stored at 4 °C. Instead we store acidic phenol in aliquots at −20 °C and thaw it directly before use to ensure uniform extraction efficacy.

5. If a NAIL-MS sample shows low signal intensity of the SILIS, the UV detection of theophylline at 260 nm acts as an external standard to rule out potential errors.

6. Instead of DMEM medium, it is possible to use RPMI R0883 or IMDM I3390 if supplemented with the labeled compounds using the same concentrations as for DMEM. As these media are already supplemented with methionine, the complete labeling of methyl groups may not work. For this purpose, choose media that lack methionine. HeLa (in DMEM or RPMI) and HAP (in DMEM or IMDM) cells were also successfully labeled in these media.

7. Labeled adenine is used for the labeling of all purines whereas labeled uridine is used for the labeling of all pyrimidines. Labeled methionine is supplemented if the labeling of SAM-dependent methyl groups is desired. If available, different labeled isotopologues can be used. Keep in mind that all mass transitions potentially differentiate from the ones reported here. For proper evaluation by MS, all resulting masses should be more than 2 Da higher than the respective other isotopologues to ensure that there is no overlap with nucleosides already carrying naturally occurring ^{13}C-atoms.

8. Instead of quenching medium, the respective growth medium (supplemented with the labeled compounds as desired) can be used. For economic reasons, we suggest using the cheaper quenching medium.

9. PBS wash is only required if a portion of cells are to be harvested for protein/non-RNA analysis or if a large portion of the cells die during the experiment.

10. Lysed cells can be stored at −20 °C for up to 1 month.

11. The harvesting procedure might differ for other cell types, e.g., for yeast.

12. Precipitation for 1 h is possible but may result in lower quantities. Incubating at −80 °C should be considered for very short precipitation times.

13. From our experience, thiolated nucleosides and especially wobble uridines are susceptible to degradation upon long-term storage.

14. The retention time of rRNA subunits from different organisms may be different as a consequence of the subunit's sizes. Also, retention times might shift depending on the operating life of the column. Always inject a test sample for proper evaluation of retention times.

15. tiRNA purification is possible with a 130 Å column [13].

16. Take care that the RNA is not concentrated to dryness as this might lead to RNA degradation.

17. Alternatively, co-precipitants (e.g., 1 μL of GlycoBlue™) can be added to each sample to facilitate RNA precipitation and simplify washing steps.

18. CG- or AU-rich sequences and a high number of modifications can alter the necessary melting temperature and may require optimization.

19. In our experience, all four types of streptavidin-coated Dynabeads$^{®}$ can be used. M-270 and M-280 beads require the use of 50 μL instead of 25 μL (as used for T1 and C1 beads) per sample.

20. The use of pre-purified RNA is crucial to avoid unspecific binding of other RNAs with similar sequence motifs which are highly abundant in total RNA preparations.

21. The addition of LC-MS buffer (pH 5.3) to digested samples (pH 8.0) lowers the pH and improves nucleoside stability, detection efficiency, and retention time reproducibility. However, if the concentration of RNA is very low, it may be considered to skip this step.

22. Filtration of digested samples is important to avoid contamination of the HPLC with digestion enzymes and particles. However, some filter materials might interact with the nucleosides (e.g., Na^{+} adducts) and perturb analysis.

23. Samples should be measured directly after digestion to avoid degradation of unstable modifications such as wobble uridines.

24. These concentrations apply for tRNA analysis. Here, the modified nucleosides D and Ψ might be added in higher concentrations (e.g., 20 μM). For the analysis of different RNA molecules, it may be necessary to change the concentration ratio in order to adapt to the natural-occurring RNA modification density.

25. Most types of RNA have a substantial excess of canonical nucleosides, which would saturate the MS detector at the injection amounts needed for detection of modified nucleosides. Thus, we use nonoptimal MS parameters for all canonical nucleoside isotopologues (e.g., fragmentor voltage or collision energy substantially higher) to artificially impair sensitivity. For analysis of purified RNAs such as tRNA isoacceptors, we recommend using optimal parameters for canonical nucleosides.

26. Assign each isotopologue to its respective SILIS compound. For example, all m^{5}C compounds (unlabeled and labeled) should be referenced to the m^{5}C SILIS compound.

27. The range of calibration must be wide enough to include the measured concentrations of each isotopologue. Special care should be taken that the calibration curve is linear in each segment that is used for sample analysis. Therefore, calibration points that fall below the lower limit of quantification and higher concentrations that result in detector saturation should be excluded.

28. In Agilent's quantitative MassHunter software navigate to "Tools" > "Actions" > "Copy Calibration Level" in the method adjustment window. This will activate a script which transfers the respective peak areas of the unlabeled calibration compound to each isotopologue which is in the respective compound group (also *see* Table 4). From this timepoint on it is crucial to only quantify the batch if needed. Activating "Analyze batch" results in reversing of the used script to copy calibration curves to each isotopologue.

Acknowledgments

References

1. Heiss M, Kellner S (2016) Detection of nucleic acid modifications by chemical reagents. RNA Biol 0. https://doi.org/10.1080/15476286.2016.1261788

2. Motorin Y, Helm M (2019) Methods for RNA modification mapping using deep sequencing: established and new emerging technologies. Genes (Basel) 10(1). https://doi.org/10.3390/genes10010035

3. Helm M, Lyko F, Motorin Y (2019) Limited antibody specificity compromises epitranscriptomic analyses. Nat Commun 10(1):5669. https://doi.org/10.1038/s41467-019-13684-3

4. Brandmayr C, Wagner M, Bruckl T, Globisch D, Pearson D, Kneuttinger AC, Reiter V, Hienzsch A, Koch S, Thoma I, Thumbs P, Michalakis S, Muller M, Biel M, Carell T (2012) Isotope-based analysis of modified tRNA nucleosides correlates modification density with translational efficiency. Angew Chem Int Ed Engl 51(44):11162–11165. https://doi.org/10.1002/anie.201203769

5. Kellner S, Ochel A, Thuring K, Spenkuch F, Neumann J, Sharma S, Entian KD, Schneider D, Helm M (2014) Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. Nucleic Acids Res 42(18):e142. https://doi.org/10.1093/nar/gku733

6. Thuring K, Schmid K, Keller P, Helm M (2017) LC-MS analysis of methylated RNA. Methods Mol Biol 1562:3–18. https://doi.org/10.1007/978-1-4939-6807-7_1

7. Borland K, Diesend J, Ito-Kureha T, Heissmeyer V, Hammann C, Buck AH, Michalakis S, Kellner S (2019) Production and application of stable isotope-labeled internal standards for RNA modification analysis. Genes (Basel) 10(1). https://doi.org/10.3390/genes10010026

8. Reichle VF, Petrov DP, Weber V, Jung K, Kellner S (2019) NAIL-MS reveals the repair of 2-methylthiocytidine by AlkB in E. coli. Nat Commun 10(1):5600. https://doi.org/10.1038/s41467-019-13565-9

9. Heiss M, Reichle VF, Kellner S (2017) Observing the fate of tRNA and its modifications by nucleic acid isotope labeling mass spectrometry: NAIL-MS. RNA Biol 14(9):1260–1268. https://doi.org/10.1080/15476286.2017.1325063

10. Heiss M, Hagelskamp F, Kellner S (2020) Cell culture NAIL-MS allows insight into human RNA modification dynamics in vivo. bioRxiv. https://doi.org/10.1101/2020.04.28.067314

11. Thumbs P, Ensfelder TT, Hillmeier M, Wagner M, Heiss M, Scheel C, Schon A, Muller M, Michalakis S, Kellner S, Carell T (2020) Synthesis of galactosyl-queuosine and distribution of hypermodified Q-nucleosides in mouse tissues. Angew Chem Int Ed Engl. https://doi.org/10.1002/anie.202002295

12. Chionh YH, Ho CH, Pruksakorn D, Ramesh Babu I, Ng CS, Hia F, McBee ME, Su D, Pang YL, Gu C, Dong H, Prestwich EG, Shi PY, Preiser PR, Alonso S, Dedon PC (2013) A multidimensional platform for the purification of non-coding RNA species. Nucleic Acids Res 41(17):e168. https://doi.org/10.1093/nar/gkt668

13. Hagelskamp F, Borland K, Ramos J, Hendrick AG, Fu D, Kellner S (2020) Broadly applicable oligonucleotide mass spectrometry for the analysis of RNA writers and erasers in vitro. Nucleic Acids Res 48(7):e41. https://doi.org/10.1093/nar/gkaa091

14. Hauenschild R, Tserovski L, Schmid K, Thuring K, Winz ML, Sharma S, Entian KD, Wacheul L, Lafontaine DL, Anderson J, Alfonzo J, Hildebrandt A, Jaschke A, Motorin Y, Helm M (2015) The reverse transcription signature of N-1-methyladenosine in RNA-Seq is sequence dependent. Nucleic Acids Res 43(20):9950–9964. https://doi.org/10.1093/nar/gkv895

15. Collart MA, Oliviero S (2001) Preparation of yeast RNA. Curr Protoc Mol Biol Chapter 13: Unit13 12. https://doi.org/10.1002/0471142727.mb1312s23

The Stress-Dependent Dynamics of *Saccharomyces cerevisiae* tRNA and rRNA Modification Profiles

Yasemin Yoluç [1] , **Erik van de Logt** [2] **and Stefanie Kellner-Kaiser** [1],*

[1] Department of Pharmaceutical Chemistry, Goethe University Frankfurt, 60438 Frankfurt, Germany; yoluc@pharmchem.uni-frankfurt.de

[2] Department of Chemistry, Ludwig-Maximilians University Munich, 81377 Munich, Germany; vandelogt@genzentrum.lmu.de

* Correspondence: Kellner@pharmchem.uni-frankfurt.de; Tel.: +49-069-798-29919

Abstract: RNAs are key players in the cell, and to fulfil their functions, they are enzymatically modified. These modifications have been found to be dynamic and dependent on internal and external factors, such as stress. In this study we used nucleic acid isotope labeling coupled mass spectrometry (NAIL-MS) to address the question of which mechanisms allow the dynamic adaptation of RNA modifications during stress in the model organism *S. cerevisiae*. We found that both tRNA and rRNA transcription is stalled in yeast exposed to stressors such as H_2O_2, $NaAsO_2$ or methyl methanesulfonate (MMS). From the absence of new transcripts, we concluded that most RNA modification profile changes observed to date are linked to changes happening on the pre-existing RNAs. We confirmed these changes, and we followed the fate of the pre-existing tRNAs and rRNAs during stress recovery. For MMS, we found previously described damage products in tRNA, and in addition, we found evidence for direct base methylation damage of 2′O-ribose methylated nucleosides in rRNA. While we found no evidence for increased RNA degradation after MMS exposure, we observed rapid loss of all methylation damages in all studied RNAs. With NAIL-MS we further established the modification speed in new tRNA and 18S and 25S rRNA from unstressed *S. cerevisiae*. During stress exposure, the placement of modifications was delayed overall. Only the tRNA modifications 1-methyladenosine and pseudouridine were incorporated as fast in stressed cells as in control cells. Similarly, 2′-O-methyladenosine in both 18S and 25S rRNA was unaffected by the stressor, but all other rRNA modifications were incorporated after a delay. In summary, we present mechanistic insights into stress-dependent RNA modification profiling in *S. cerevisiae* tRNA and rRNA.

Keywords: stress dependent RNA modification dynamics; absolute quantification of RNA modifications; isotope labeling; mass spectrometry; *Saccharomyces cerevisiae*

1. Introduction

The central dogma of molecular biology states that DNA is the storage of the genetic code, which is transcribed into messenger RNA (mRNA) and translated into proteins with the help of transfer RNA (tRNA) and ribosomal RNA (rRNA). This fundamental life process is dominated by nucleic acids, which are composed of the canonical nucleosides adenosine, guanosine, cytosine and uridine (and thymidine in DNA). The sequence of these building blocks defines the genetic code of an organism. Additional chemical groups on these building blocks, commonly methylations, form a second layer of information on top of the code. In DNA, methylation was found to be dynamic. The addition or removal of a methylation on carbon C5 of cytosine can switch genes off or on [1,2]. Since this chemical code is additional information on top of the sequence, it is referred to as the epigenetic code. While epigenetics is an intensively studied area, the analogous process in RNA, termed epitranscriptomics, is far less studied [3]. This is mainly due to limited number of tools that can be used to study the dynamics of RNA modifications, and in addition, the

complex process of finding biological consequences of RNA modifications. While DNA modifications must be removed by enzymatic or chemical processes to maintain the genetic sequence and its function, RNA has the option of simple degradation to disband unwanted RNA strands and the subsequent transcription of new RNA. This dynamic degradation and dilution, by new transcripts, constitutes a fundamental difficulty in the accurate assessment and quantification of RNA modifications.

A fundamental study was presented by Chan et al. They provided insight into the changes to modifications in small RNA (<200 nts) as a cause of chemical stress exposure. This study coined the term "stress-dependent RNA modification reprogramming," and there is clear evidence that RNA modifications are regulated by stress [4,5]. The methodological foundation of this and later studies was quantitative mass spectrometry, which allows one to assess changes in RNA modification abundance and compare, e.g., stressed samples with controls. A problem in the interpretation of the underlying data consists in the simultaneous analysis of RNA subspecies within the cell. For example, a higher modification density can be explained by additional modification events or by degradation of non-modified RNAs. A lower modification density is even more challenging to interpret. It can be caused by (a) enzymatic demodification processes, such as m^6A for human mRNA [6,7] or ms^2C in bacterial tRNA [8]; (b) by increased degradation of modified RNA; or c) by increased transcription of the RNA, without it being modified at all. As there are numerous RNAs within each cell and over two dozen abundant RNA modifications, it is very likely that a combination of all those processes is happening. To unravel the different mechanisms used for RNA modification adaptation, we developed stable isotope labeled pulse-chase studies in yeast [9], bacteria [10] and human cell culture [11]. Nucleic acid isotope labeling coupled mass spectrometry (NAIL-MS) relies on metabolic labeling of RNA to distinguish RNA modifications from different RNA subspecies, e.g., RNAs which existed during stress-exposure and RNAs which are transcribed during the stress recovery phase. For this purpose, we utilized isotope dilution mass spectrometry, which not only allows qualitative assessments of RNA modification changes, but furthermore, absolute quantification of RNA modifications [12].

In this work, we focused on the mechanisms which lead to the stress-dependent adaptation of tRNA and rRNA modifications during and up to 20 h after stress exposure in *S. cerevisisae*. We utilized stressors previously described by Chan et al., such as H_2O_2, MMS and $NaAsO_2$. In addition, we studied the oxidant TBH and determined its impact on tRNA modifications. We applied our unique NAIL-MS technology to follow the fate of original RNAs exposed to stressors and how their modification profiles were impacted. For both tRNA and rRNA, we only observed minor changes. Only methyl-methanesulfonate which directly damages RNA [10], led to a substantial increase in methylated nucleosides. With a methylome discrimination assay, we proved direct methylation of the RNA, and we observed two undescribed RNA damage products which emerged from base methylation of 2′-O-ribose methylated nucleosides in rRNA. Furthermore, we closely looked at the original RNAs, but also new transcripts, during the time after stress exposure, when cells were striving to recover from the stressors. We found that cells exposed to stress barely showed signs of transcription. In addition, the speed of tRNA and rRNA modification during maturation slowed or even stalled after stress. Only some modifications, such as m^1A and Ψ in tRNA and Am in rRNA, are more quickly incorporated into the new transcripts upon arsenite stress. These modifications are good candidates for future studies focusing on the role of RNA modifications in the stress response.

2. Materials and Methods

2.1. Chemicals and Reagents

All salts were obtained from Sigma Aldrich (Munich, Germany) at molecular biology grade, unless stated otherwise. Isotopically labeled compounds: $^{15}N_2$- uracil (≥98% atom, Eurisotope), $^{13}C_6$-glucose (≥99% atom, Eurisotope) and L-methionine-[$^2H_3$]-methyl (98 atom % D, Sigma-Aldrich). All solutions and buffers were made with ultrapure

water (Milli-Q, Merck, Kenilworth, NJ, USA). Supplementary Materials Table S1 shows all synthetic standards of modified nucleosides and their respective vendors.

2.2. Growth Media for S. cerevisiae

Yeast-nitrogen-base (YNB) (Carl Roth, Karlsruhe, Germany) minimal medium was prepared by mixing a $10\times$ YNB stock solution supplemented with a mix of unlabeled aminoacids (final concentration in $1\times$ growth medium: 0.02 g/L arginine, 0.02 g/L histidine, 0.06 g/L leucine, 0.03 g/L lysine, 0.05 g/L, phenylalanine, 0.4 g/L serine, 0.2 g/L threonine, 0.04 g/L tryptophane, 0.03 g/L tyrosine and 0.15 g/L valine). Depending on the desired stable isotope labeling, 0.02 g/L $^{15}N_2$-uracil, 0.01 g/L $^{13}C_6$-glucose, 0.72 g/L-methionine-[$^2H_3$]-methyl or their unlabeled isotopomers were used.

2.3. Yeast Cultivation

A single colony of *S. cerevisiae* BY4741 was picked from a YPD-agar plate and used for inoculation of 5 mL YNB. The cells were grown at 30 °C at 250 rpm. The next day, the cell density was assessed by OD600 measurement (Eppendorf, Biophotometer plus) and the cell suspension diluted to OD 1. The cells were allowed to grow for 3 h to reach mid-log phase and the stressor was added. After 1 h of stress exposure the medium was exchanged by centrifugation (5 min, $3000\times g$, 24 °C). The resulting pellet was resuspended in fresh medium to initiate the recovery phase. Afterwards, 2 mL of cell suspension was harvested at set time points by centrifugation (5 min, $12,000\times g$, 4 °C). The RNA was isolated as described in Section 2.11.

2.4. LD$_{50}$ Assay

For determination of LD$_{50}$ values, yeast was cultivated as described in 2.3 and exposed to various concentrations of H_2O_2, MMS, NaAsO$_2$, TBH or HOCl. After 1 h, 100 μL of each culture was diluted to $1/10^5$ with sterile water. From this dilution 70 μL was plated on a pre-warmed YPD agar plate. The YPD plates were incubated at 30 °C for 48 h and the colonies were counted.

2.5. Stress Study

For assessment of tRNA and rRNA modification profiles under stress, yeast was cultivated as described in Section 2.3, split into a control and stress groups and exposed to the determined LD$_{50}$ concentrations of H_2O_2 (2 mM), NaAsO$_2$ (40 mM), HOCl (3 mM), MMS (12 mM) and TBH (10 mM). Then, 2 mL samples were drawn as indicated in the Figure 1, and the RNA was isolated and purified as described in Sections 2.11–2.13.

2.6. Comparative NAIL-MS for Validation

For validation of NAIL-MS conditions, *S. cerevisiae* BY4741 was grown overnight in unlabeled YNB medium or in $^{13}C_6$-glucose/$^{15}N_2$-uracil labeled YNB medium, as described in Section 2.3. Before RNA isolation, 1 mL of labeled culture and 1 mL of unlabeled culture were mixed, and the RNA was immediately harvested (Section 2.11) before mass spectrometric analysis (Sections 2.12 and 2.13, but without SILIS).

2.7. Pulse-Chase NAIL-MS Experiment

A 5 mL overnight culture of *S. cerevisiae* BY4741 was grown in $^{13}C_6$-glucose and $^{15}N_2$-uracil labeled YNB medium. The next day, the culture was diluted with $^{13}C_6$-glucose and $^{15}N_2$-uracil labeled YNB medium to OD 1 with a total volume of 32 mL. After 3 h of growth, the first sample was harvested. Afterwards the culture was split, and one half was exposed to the LD$_{50}$ concentration of the respective stressor. After one hour of stress exposure, the next sample was harvested and afterwards the medium was exchanged by centrifugation (5 min, $3000\times g$, 24 °C). The resulting pellet was resuspended L-methionine-[$^2H_3$]-methyl labeled YNB medium. More samples were harvested (2 mL) at set timepoints after the

initiation of the recovery phase. The RNA was extracted and purified as described in Sections 2.11–2.13.
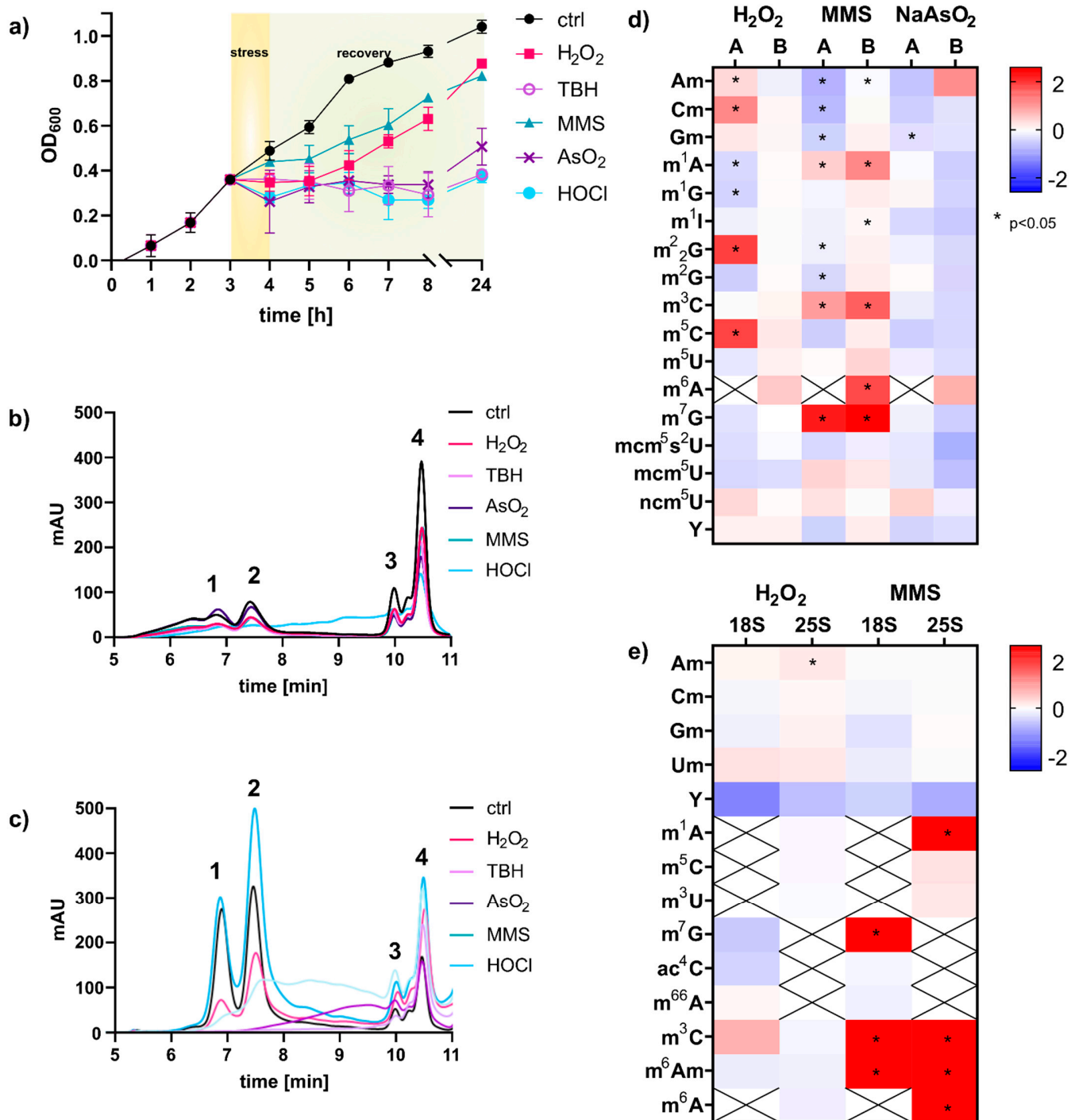
Figure 1. Analysis of total RNA and RNA modification changes upon chemical stress exposure of *S. cerevisiae* cultures. (**a**) Logarithmic growth curve in YNB medium using LD_{50} concentrations of stressors: H_2O_2 (2 mM), $NaAsO_2$ (40 mM), HOCl (3 mM), MMS (12 mM), TBH (10 mM). (**b**) Total RNA of *S. cerevisiae* isolated with TRI reagent or hot phenol (**c**) assessed with size exclusion chromatography, 1: 25S, 2: 18S, 3: 5.8S rRNA, 4: tRNA (**d**) Fold changes in modification density in small RNA < 200 nts: (A) Chan et al. and (B) total tRNA from this work. An increase is shown in red and a decrease in blue. Student's *t*-test results: $p < 0.05$ indicated with *. (**e**) Fold changes in modification density in 18S and 25S rRNA; Student's *t*-test results: $p < 0.05$ indicated with *. All experiments were done in biological triplicates; error bars reflect standard deviations.

2.8. Methylome Discrimination Assay

L-Methionine-[^{2}H$_3$]-methyl labeled YNB medium was used, following the culturing method described in Section 2.3. After stress exposure, the stressor containing medium was removed and fresh L-methionine-[^{2}H$_3$]-methyl labeled YNB medium was used.

2.9. Pulse-Chase NAIL-MS Experiment MMS Damage Repair

To monitor MMS induced RNA damage repair, cells were grown as described in Section 2.3 using L-methionine-[^{2}H$_3$]-methyl labeled YNB medium in the overnight culture. The next day the culture was diluted with L-methionine-[^{2}H$_3$]-methyl labeled medium to OD 1, grown for 3 h and exposed to 12 mM MMS. After 1 h of exposure, the medium was exchanged by centrifugation (5 min, $3000\times g$, 24 °C), and the resulting pellet was resuspended in ^{13}C$_6$-glucose, ^{15}N$_2$-uracil and L-methionine-[^{2}H$_3$]-methyl labeled YNB medium.

2.10. Knock-Out Screening

An overnight culture of each knock-out strain was grown in unlabeled YNB medium as described in Section 2.3, and 12 mM MMS was used as a stressor. After 1 h of stress exposure, samples were taken from the control and from the stress-exposed culture. Afterwards, the RNA was extracted and further processed for analysis.

2.11. RNA Extraction—Hot Phenol

Total RNA was isolated according to the hot-phenol extraction protocol of Collart et al. [13]. The extraction was followed by ethanol precipitation. Therefore, 0.1xV 3 M NH$_4$OAc and 2.5xV 100% ice cold ethanol were added to the aqueous phase; the mixture was stored at −20 °C overnight. The next day, samples were centrifuged (40 min, $12,000\times g$, 4 °C), the supernatant was discarded and the reaction tube was rinsed with 200 μL of 70% ice-cold ethanol. After another step of centrifugation (10 min, $12,000\times g$, 4 °C), the supernatant was discarded, and the ethanol was air dried for 10 min. Afterwards, the total RNA was suspended in 50 μL H$_2$O.

2.12. rRNA and tRNA Purification

18S and 25S rRNA and tRNA were purified by size exclusion chromatography (SEC) (AdvanceBio SEC 300 Å, 2.7 μm, 7.8 × 300 mm for tRNA combined with BioSEC 1000 Å, 2.7 μm, 7.8 × 300 mm for 18S and 25 S rRNA, Agilent Technologies) according to our published protocol [14]. After purification, the RNA was precipitated and dissolved in 30 μL H$_2$O.

2.13. RNA Digestion for Mass Spectrometry

RNA (300–500 ng) in aqueous digestion mix (30 μL) was digested to single nucleosides by using 2 U alkaline phosphatase, 0.2 U phosphodiesterase I (VWR, Radnor, Pennsylvania, USA) and 2 U benzonase in Tris (pH 8, 5 mM) and MgCl$_2$ (1 mM) containing buffer. Furthermore, 5 μg tetrahydrouridine (Merck, Darmstadt, Germany), 10 μM butylated hydroxytoluene and 1 μg pentostatin were added to avoid deamination and oxidation of the nucleosides. The mixture was incubated 2 h at 37 °C and then filtered through 96-well 10 kDa molecular-weight cut-off plates (AcroPrep Advance 350 10 K Omega, PALL Corporation, New York, NY, USA) at $3000\times g$ and 4 °C for 30 min. Then, 1/10 Vol. of SILIS (stable isotope labeled internal standard) was added to each filtrate before analysis by QQQ mass spectrometry.

2.14. Preparation of rRNA and tRNA SILIS

S. cerevisiae BY4741 was grown in 5 mL of ^{13}C, ^{15}N Silantes rich growth medium (Silantes, Munich, Germany Product no.: 111601402) supplemented with 1% (*w/w*) ^{13}C$_6$-glucose. The culture was incubated overnight, and the next day it was diluted to OD 0.1 with fresh ^{13}C, ^{15}N Silantes rich growth medium supplemented with 1% (*w/w*) ^{13}C$_6$-glucose. The culture was incubated for another 2 days at 30 °C. The cells were harvested and

RNA was extracted according to Collart et al. [13]. rRNA and tRNA were purified by size exclusion chromatography (SEC) (AdvanceBio SEC 300 Å, 2.7 µm, 7.8 × 300 mm, Agilent Technologies), as described in [14]. Subsequently, the RNA was hydrolyzed to single nucleosides, as described in Section 2.13. As an external standard 10 mM theophylline was added to a final concentration of 1 mM in the digestion solution. The resulting digest/theophylline mixture is referred to as 10 × SILIS, which was added to a final concentration of 1 x to samples and calibration solutions. The labeling efficiency was confirmed by high resolution mass spectrometry (HRMS). Spectra of precursor and product ions were recorded by a ThermoFinnigan LTQ Orbitrap XL operated in positive ionization mode after LC separation of ribonucleosides.

2.15. QQQ Mass Spectrometry

For quantitative mass spectrometry, an Agilent 1290 Infinity II equipped with a diode-array detector (DAD) combined with an Agilent Technologies G6470A Triple Quadrupole system and electrospray ionization (ESI-MS, Agilent Jetstream) was used. Operating parameters: positive-ion mode, skimmer voltage of 15 V, cell accelerator voltage of 5 V, N_2 gas temperature of 230 °C and N_2 gas flow of 6 L/min, sheath gas (N_2) temperature of 400 °C with a flow of 12 L/min, capillary voltage of 2500 V, nozzle voltage of 0 V and nebulizer at 40 psi. The instrument was operated in dynamic MRM mode (multiple reaction monitoring, MRM). Mass transitions for all monitored analytes and their isotopologues are found in Table S5. For separation a Core-Shell Technology column (Synergi, 2.5 µm Fusion-RP, 100 Å, 100 × 2 mm column, Phenomenex, Torrance, CA, USA) at 35 °C and a flow rate of 0.35 mL/min were used in combination with a binary mobile phase of 5 mM NH_4OAc aqueous buffer A, brought to pH 5.6 with glacial acetic acid (65 µL), and an organic buffer B of pure acetonitrile (Roth, LC-MS grade, purity ≥.99.95). The gradient started at 100% solvent A for 1 min, followed by an increase to 10% over 3 min. From 4 to 7 min, solvent B was increased to 40% and was maintained for 1 min before returning to 100% solvent A and a 3 min re-equilibration period.

2.16. Calibration

For calibration, synthetic nucleosides were weighed and dissolved in water to a stock concentration of 1–10 mM. The calibration solutions ranged from 0.3 to 500 pmol for each canonical nucleoside and from 0.3 to 500 fmol for each modified nucleoside and were spiked with 1/10 volume of SILIS. The sample data were analyzed by MassHunter Quantitative Software from Agilent. The areas of the MRM signals were integrated for each modification and their isotopologues. The absolute amounts of the modifications were referenced to the absolute amounts of the respective canonical. In the case of the pulse-chase experiment, the different isotopomers were referenced to their respective labeled canonicals, so that original modifications were referenced to original canonicals and new modifications were referenced to new canonicals.

2.17. Statistics

All experiments were performed at least three times (biological replicates) to allow student *t*-test analysis. The *p*-values of the Student's *t*-test (unpaired, two-tailed, equal distribution) were calculated using Excel or Graphpad Prism.

3. Results

3.1. S. cerevisiae's Total RNA Composition Is Changed by Chemical Stress Exposure

Intrigued by the concept of stress-dependent RNA modification reprogramming [4], we set out to study the reaction of *S. cerevisiae* on the transcriptome level in more detail. For this purpose, we wanted to use our established NAIL-MS methodology which is based on controlled stable isotope nutrient's addition to minimal medium [9]. NAIL-MS relies on yeast nitrogen based medium (YNB), which differs largely from the commonly used yeast extract peptone dextrose medium (YPD). Thus, we first determined the 50% lethal dose of

every stressor for a *S. cerevisisae* BY4741 culture grown in YNB medium (Figure S1). For this purpose, an overnight yeast culture was diluted in YNB medium and grown for 3 h until mid-log growth phase and stressed with either methyl-methanesulfonate (MMS) or one of the oxidants: hydrogen peroxide (H_2O_2), arsenite ($NaAsO_2$), tert-butyl hydroperoxide (TBH) or hypochloric acid (HOCl). After one hour of stress exposure, the cells were pelleted and resuspended in fresh YNB medium for recovery. The growth curve for the non-exposed control cells shows the unaltered growth of the cells, whereas the cells exposed to MMS and H_2O_2 showed a delay in growth after stress exposure (Figure 1). The cells exposed to the oxidants $NaAsO_2$, TBH and HOCl did not recover within 24 h and showed no growth within this timespan. We next extracted the total RNA from cells after one hour of exposure using either the commercial TRI reagent and glass bead approach or hot phenol [13]. The total RNA was loaded onto a size exclusion chromatography column of 1000 Å, and the eluting RNA was detected using UV absorption at 254 nm. As shown in Figure 1b,c, the TRI based method yielded mainly RNAs smaller than 200 nts. 18S and 25S rRNA were of low abundance and undefined size. In contrast, the hot-phenol method yielded high amounts for 18S, 25S and tRNA. Judging from the elution profile in Figure 1c, the integrity of rRNAs remained under MMS and H_2O_2 exposure, whereas all rRNA was lost in cells exposed to the oxidants $NaAsO_2$, TBH and HOCl. Overall, the profile of total RNA was bizarre in these cells, and the fate of the rRNA is unclear. Therefore, the stressor HOCl was not further pursued for RNA modification analysis. Only $NaAsO_2$ and TBH showed acceptable integrity of tRNA, and thus tRNA modification profiles can be analyzed.

3.2. tRNA Modification Reprogramming in S. cerevisiae Is Stress Dependent; rRNA Modifications Are Unaltered

For RNA modification analysis we used our established stable isotope dilution LC-MS/MS protocol [12]. In the acute phase, 60 min after stress exposure, we found changes in tRNA modification density in dependence of the chemical used, as previously suggested by Chan et al. [4]. A comparison of the published data and our fold-change data is given in Figure 1d. The direct comparison revealed several differences between our and the published data. For H_2O_2, we found less tRNA modification reprogramming, while similar trends are found for MMS and $NaAsO_2$ exposure. We have identified three major experimental differences which contributed to the observed differences: (1) The published experiments were performed in rich YPD growth medium, whereas we used minimal YNB medium. (2) Chan et al. used a column affinity-based protocol for purification of RNA smaller than 200 nts, whereas we used size exclusion chromatography for tRNA purification. (3) Our mass spectrometric data was acquired using stable isotope dilution, which allows absolute quantification of RNA modifications. Therefore, we are confident that our data reflect the changes in tRNA modification profiles accurately. With our study we confirm the findings by Chan et al. that tRNA modifications are reprogrammed in the acute moment of chemical stress exposure and that the changes are dependent on the chemical stressor. However, especially for H_2O_2 exposure, the observed changes were minimal and were not statistically significant. For MMS, we found substantial formation of 1-methyladenosine (m^1A), 3-methylcytidine (m^3C), 6-methyladenosine (m^6A) and 7-methylguanosine (m^7G), as recently described as RNA main damage products [10,15]. From the same experiments, we purified the 18S and 25S rRNA and subjected them to RNA modification quantification by LC-MS/MS. After 60 min of stress exposure, we found only minor changes in the natural epitranscriptomes of both rRNAs. This is in accordance with a recent study from the Novoa laboratory [5]. However, in rRNA from MMS exposed yeast, we found high numbers of the potential damage products m^1A, m^7G, m^3C and m^6A and a damage-methylated 2′-O-methyladenosine (m^xAm).

3.3. MMS Directly Methylates tRNA and rRNA in S. cerevisiae

After MMS exposure, we detected high abundances of those RNA modifications, which are RNA damage products that were described in *E. coli* studies [10,15]. With the goal of elucidating the origins of these RNA modifications in *S. cerevisiae*, we envisioned a

methylation discrimination assay which distinguishes enzymatic RNA methylation from direct methylation damage. S-Adenosylmethionine (SAM) is the natural methyl-donor for yeast RNA methyltransferases, and by feeding L-methionine-[^{2}H$_3$]-methyl enzymatic methylations, they receive a +3 mass increase. As shown in Figure 2a, optimal labeling of native RNA modifications m^{7}G, m^{1}A and m^{3}C was achieved with 18 mM L-methionine-[^{2}H$_3$]-methyl. Cells were exposed to 12 mM MMS in the continuous presence of 18 mM L-methionine-[^{2}H$_3$]-methyl. After one hour of MMS exposure, tRNA, 18S and 25S rRNA were extracted, and mass spectrometry analysis revealed the ratio of enzymatically placed methylations (m/z +3) to damage-derived methylations ($m/z \pm 0$). As shown in Figure 2b, up to 60% of all m^{7}G marks were caused by direct methylation with MMS. To a lower extent, m^{3}C, m^{1}A and m^{6}A were caused by direct methylation of canonical nucleosides in tRNA. For rRNA, we found the same damage products (Figure 2c,d), and in addition, two base-methylated 2′-O-methyladenosine species designated as m^{x}Am. A comparison to our synthetic standards of m^{1}Am and m^{6}Am indicates that the early eluting damage product was m^{1}Am and the later one was m^{6}Am (Figure S2). Both were caused by direct base methylation of the highly abundant Am of both rRNAs during MMS exposure. With the power of our methylome discrimination assay, we could clearly identify the origin of the base methylation from MMS and the enzymatic origin of the ribose methylation (Figure 2e). Intrigued by this finding, we searched for a methylation damage product of Gm in rRNA, and we observed a clear signal of m^{7}Gm in the MMS exposed yeast samples. Further identification through the comparison with a synthetic standard has not yet been possible (Figure S2). A detailed analysis of the observed absolute quantities is given Table S2.
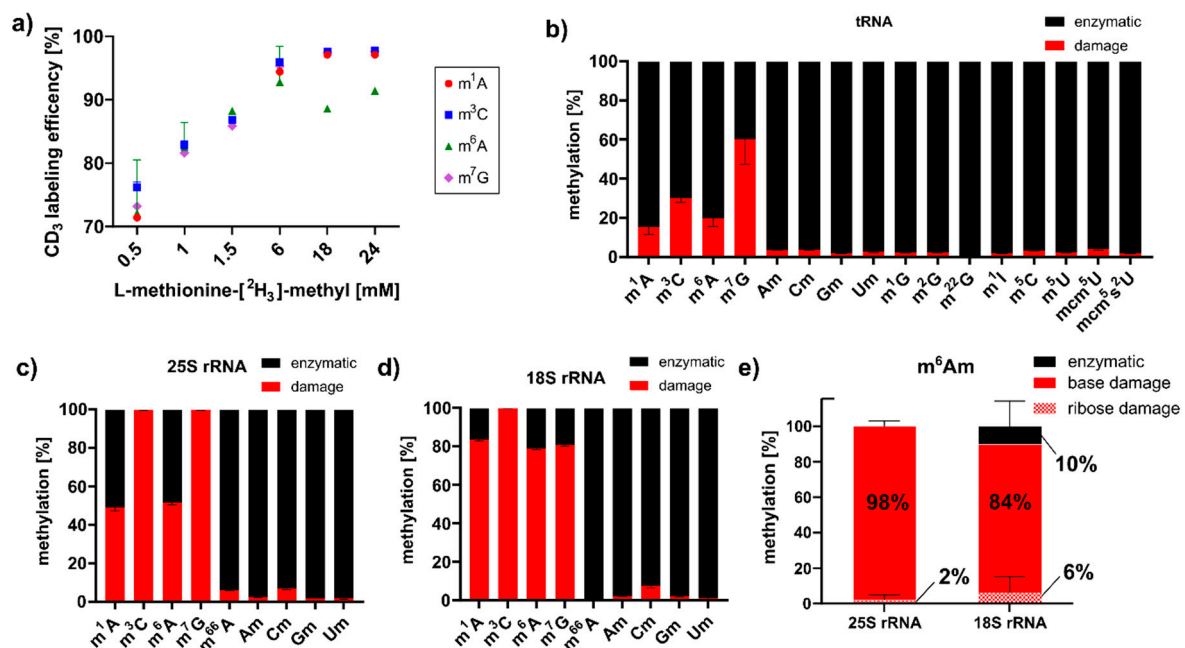
Figure 2. Methylome discrimination assay in *S. cerevisiae*. (**a**) Titration of optimal L-methionine-[^{2}H$_3$]-methyl supplementation to receive the highest abundance of CD$_3$-labeled RNA modifications. (**b**–**d**) Ratios of damaged (red) to enzymatically (black) methylated nucleosides in *S. cerevisiae* tRNA (**b**), 25S rRNA (**c**) and 18S rRNA (**d**) after 60 min of MMS exposure. (**e**) The methylome discrimination assay reveals the origin of the methylation in the newly identified RNA damage product m^{6}Am.

3.4. Modification Density in Existing tRNAs Rises upon S. cerevisiae Stress Exposure

Although the changes in tRNA modification density were small, we were curious to find out how they emerged mechanistically. With a pulse chase NAIL-MS experiment, we aimed to answer the questions: How does transcription change due to stress exposure? Are existing tRNAs degraded? Is it the original tRNAs which are modified or even demodified? When do new transcripts emerge, and how quickly are they modified? To answer these

questions, we required a robust NAIL-MS method which allows accurate and precise analysis of mainly methylated nucleosides. In our previous method published in 2017 [9], we established the necessary medium for such an experiment; however, the stable isotope labeled internal standard (SILIS) was problematic. For methylated cytidine derivatives especially, the same m/z was found in the SILIS, along with the original modification of the tRNA. Therefore, a new SILIS, without m/z overlap, with the analytes had to be produced. For this purpose, we utilized a commercially available yeast medium which was enriched with carbon-13 and nitrogen-15 instead of carbon-12 and nitrogen-14 atoms. For *S. cerevisiae*, glucose is the carbon source and an ideal energy source, and by addition of 0.1 g/L of $^{13}C_6$-glucose, we received well growing cultures and high numbers of fully ^{13}C-labeled yeast cells. The total ^{13}C- and ^{15}N-labeled RNA was isolated, the rRNA and tRNA were purified and the SILIS was prepared following our established protocol [12]. A comparison of our new SILIS and the previous SILIS is found in Figure S4 and our recently published protocols [14].

With the new SILIS in hand, we followed our published yeast NAIL-MS protocol (Figure 3a). Briefly summarized, yeast is grown overnight in YNB medium supplemented with $^{15}N_2$-uracil and $^{13}C_6$-glucose. The next day, cells were brought to OD 1 in the same medium, left for 3 h to enter mid-log phase and then exposed to the chemical stressor. After one hour of exposure, the stressor was removed by medium exchange. For the chase phase, medium with L-methionine-[$^2H_3$]-methyl was used. Due to the mass spectrometric detection, we followed the abundances of modified nucleosides in RNA existing during exposure to the chemical, determined the abundance of new canonical nucleosides forming after stress and determined the modification incoperation in new transcripts.
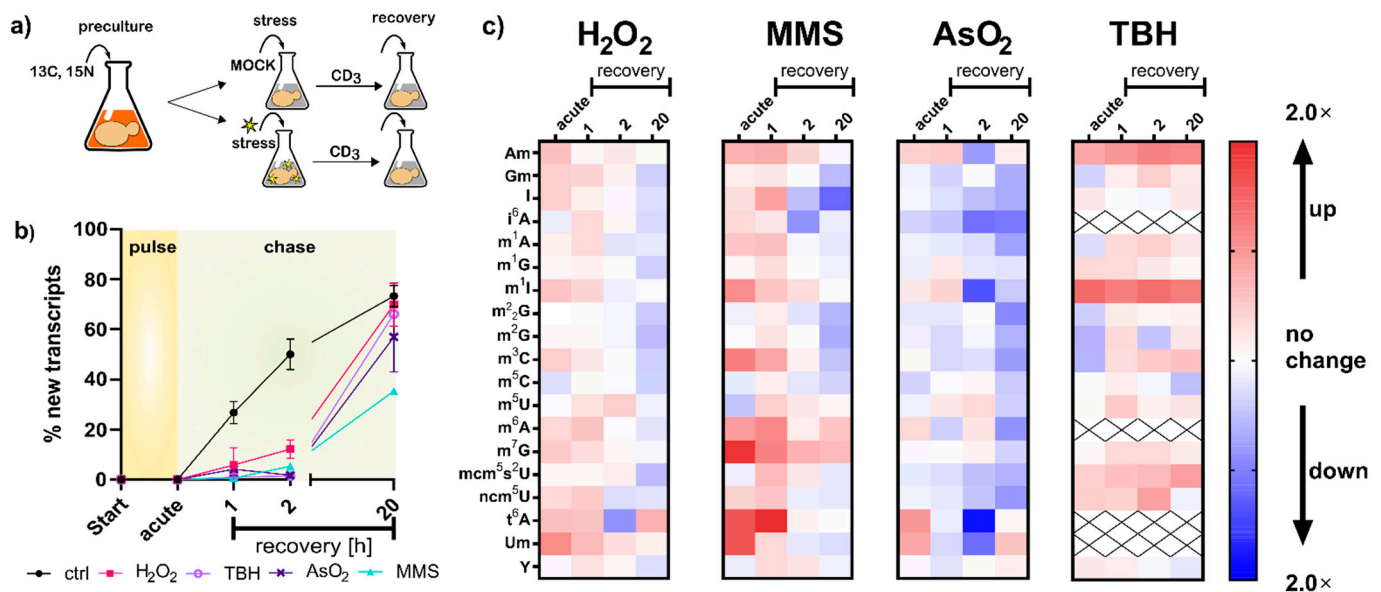
Figure 3. NAIL-MS pulse chase experiment with different stressors. (**a**) A concept sketch of the pulse-chase NAIL-MS assay. (**b**) Ratios of new and original tRNA transcripts displayed as percentages of new transcripts with the LD$_{50}$ dose; data from n = 3 biological replicates; error bars reflect standard deviations. (**c**) Relative abundances of modified nucleosides in total tRNA compared to the time-matched control. The resulting fold changes in red indicate higher modification densities, and blue indicates lower abundances. Data are averages from three biological replicates.

In the first step, we compared the number of new canonical nucleosides to the number of original canonical nucleosides, which is an indicator of cellular metabolism. After two hours of growth in the new but stable isotope labeled medium, 50% of all tRNAs contained new canonical nucleosides. In contrast, all stressed cells contained less than 10% of new canonical nucleosides, which indicates that transcription of tRNA is substantially repressed during the stress recovery phase (Figure 3b). Thus, all RNA modification density changes

observed in Figure 1d must have been derived from from changed modification patterns in the original tRNAs. To further investigate the impacts of stress on the tRNA modification profiles, we determined the number of modifications per original tRNA and compared the quantities of stressed cells to the numbers found in the respective control cells. Under acute stress, we mainly observed higher numbers of tRNA modifications for H_2O_2, MMS and TBH stress and lower numbers for $NaAsO_2$ stress (Figure 3c). These observations are in good agreement with our findings in Figure 1d.

Regarding the overarching hypothesis of stress-dependent RNA modification reprogramming, our findings concerning transcription rates (Figure 3b) indicate a non-active adaptation scenario for increases in tRNA modification abundance. The increase in tRNA modification density might have been caused by a combination of (a) halted transcription in stressed cells, which left only existing transcripts as substrates for RNA writer enzymes, and thus higher modification numbers were observed; and (b) ongoing transcription and slow maturation in the control cells, and thus we found lower modification numbers in the control cells. Lower numbers of modifications, as observed for $NaAsO_2$, might indicate active removal of modifications; however, given the large number of different modifications that are affected, a global effect on mature tRNA might be causative. In theory, mature and thus modified tRNAs might be targeted for degradation. Yet, we see no evidence of increased degradation of original tRNAs (Figure 3b). The trend of decreased tRNA modification density in arsenite-exposed *S. cerevisiae* remained visible even after two hours of recovery. Our methodology is currently not able to determine the biological mechanism behind this intriguing observation. A more detailed overview and absolute numbers of all native modifications per tRNA, 18S and 25S rRNA can be found in Figure S8.

One of the most striking findings of Figure 3c is the direct methylation of nucleobases, which appears to have been reverted during the recovery phase. We designed a pulse chase assay based on the methylome discrimination methodology to follow the fate of enzymatically placed modified nucleosides and damage-derived nucleosides (Figure 4a). On average, both rRNAs receive more than a dozen damage methylations during the experiment, and at least one methylation damage is found per tRNA. We found for both tRNA and rRNA, unexpectedly fast loss of these damages within one hour of recovery. The abundance of native methylations remained unchanged (Figures 4b and S6). We tested several knockout strains of known nucleic acid damage repair enzymes for their potential involvement in the demethylation process in *S. cerevisiae*. However, except for met18, which showed a decent involvement in total RNA demethylation, no enzyme was found to be part of an active demethylation machinery (Figure S7 and Table S4). Loss of damaged RNAs through targeted degradation of the damaged RNA subpopulation is another valid hypothesis, but our data on original-to-new transcript ratios in Figure 3B do not favor this hypothesis. In summary, rRNA and tRNA receive substantial methylation damage through MMS exposure of *S. cerevisiae*, but we do not know by which mechanism the damaged nucleosides were lost within 60 min of recovery.
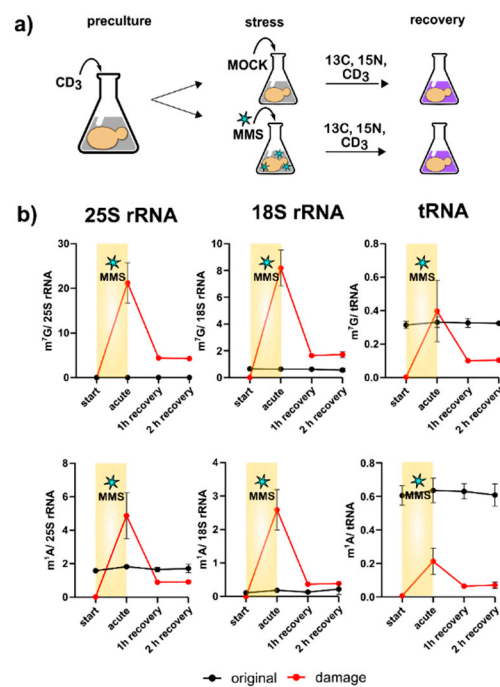
Figure 4. Methylome discrimination-based pulse chase experiment. (**a**) A concept sketch of the assay. Initially cells are grown in a medium containing L-methionine-[^{2}H$_3$]-methyl; after stress exposure the cells are cultivated in medium with L-methionine-[^{2}H$_3$]-methyl and ^{15}N$_2$-uracil, ^{13}C$_6$-glucose. (**b**) Numbers of m^{7}G (upper panel) and m^{1}A (lower panel) per respective RNA. Red line shows damage-induced methylation; black line shows enzymatic methylation. Data for tRNA from three biological replicates; error bars represent standard deviations. Data for 25S and 18S rRNA from two biological replicates; error bars represent standard deviations.

3.5. tRNA Modification Placement in New Transcripts Is Stress Dependent

With NAIL-MS, we have the unique opportunity to observe the speed of modifications in RNAs transcribed after a stress event. Figure 5a shows the early formation of highly abundant and less abundant modifications in total tRNA from *S. cerevisiae*. In accordance with our results from 2017 [9], we found immediately high amounts for modified nucleosides, such as pseudouridine (Ψ) and 2′-O-methylguanosine (Gm). Interestingly, other modifications of the anticodon-stem-loop mcm^5(s^2)U, t^{6}A and i^{6}A also appeared early on in total tRNA (Figures 5 and S9). For Ψ, an immediate placement was also observed in human total tRNA, human tRNA^{Phe} [11] and yeast tRNA^{Phe} [16]. However, for Gm, which is incorporated fast in yeast tRNA, we found slow incorporation in human tRNA. Other modified nucleosides such as m^{1}A, m^{5}C, Cm and m^{7}G showed similar incorporation speeds to human tRNAs [11].

For ribosomal RNAs, we found an immediate steady-state abundance of ribose methylated modifications, as expected (Figure 5b) [17]. In 18S rRNA, m^{7}G and in 25S rRNA, m^{3}U, m^{5}C and m^{1}A, were also placed early on, which is unsurprising, given their later inaccessibility to modification enzymes in the mature ribosome. Previous studies proposed rRNA methylation as a co-transcriptional process in the early phase of ribosome assembly [17,18]. To our great surprise, the isomerization of uridine to Ψ was substantially slower, and it took more than two hours to reach the final modification density in both 18S and 25S rRNA. This observation is in stark contrast to our findings in human rRNA, where we observed a fast pseudouridinylation within minutes [11]. Our most puzzling result was observed for Ψ in the latest 20 h timepoint. Its abundance exceeded the steady state level observed in the unlabeled experiment (Table S2). We excluded a methodological bias, due to the biologically valid data received for tRNA. Thus, this observation deserves continued research to discover the underlying mechanism.
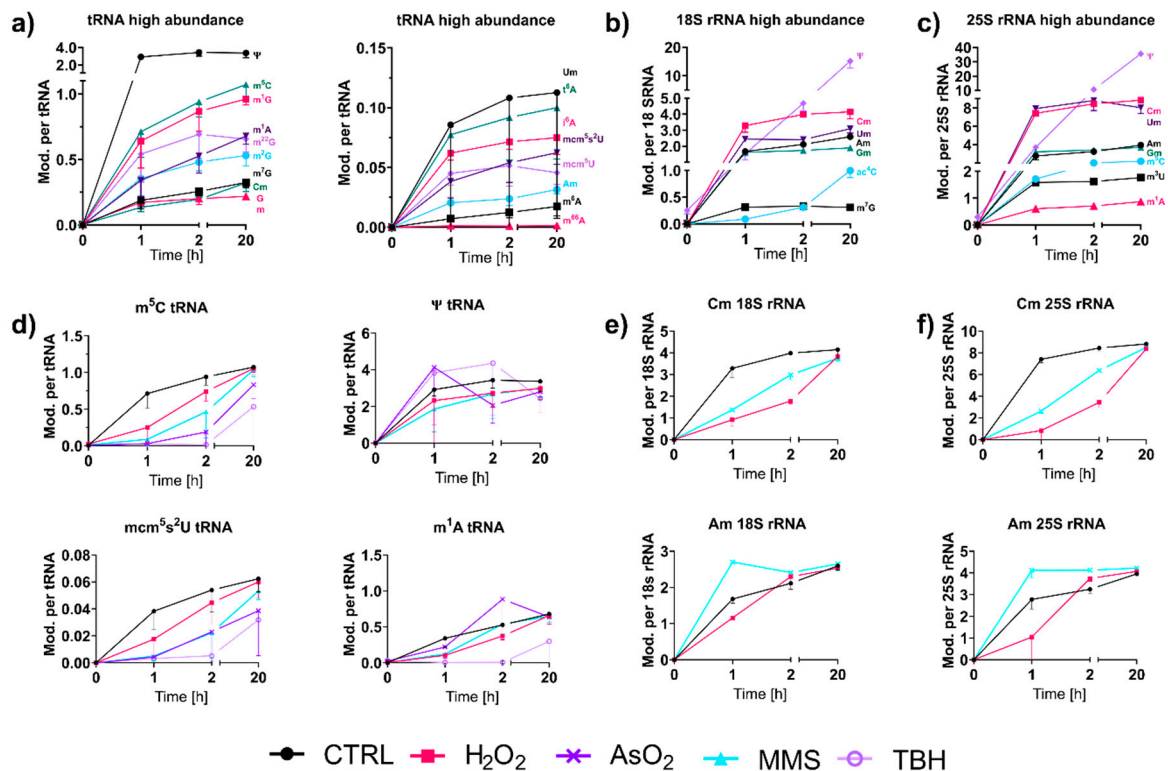
Figure 5. Incorporation of RNA modifications into new transcripts over time. Cells were initially grown in medium containing $^{15}N_2$-uracil, $^{13}C_6$-glucose; after stress exposure, the cells were cultivated in medium with L-methionine-[$^2H_3$]-methyl. (**a**) Abundances of modifications in new tRNA sorted by high and low abundances of modifications. (means of $n = 3$ and error bars reflect standard deviation). (**b**,**c**) Abundance of modifications in new 18S (**b**) and 25S (**c**) rRNA. (means of $n = 2$, and error bars reflect standard deviation). (**d**) A comparison of modification abundances in new total tRNA with dependence on stress. (means of $n = 3$, and error bars reflect standard deviations). (**e**,**f**) A comparison of modification abundances in new 18S (**e**) and 25S (**f**) rRNA in dependence on stress. (means of $n = 2$, and error bars reflect standard deviations).

For most tRNA and rRNA modifications, we observed immediate or extremely fast incorporation within one hour of transcription (Figure 5a–c). From our NAIL-MS stress exposure studies in Figure 3b, we now know that transcription is differentially impacted by the chosen stressors. However, how about the subsequent RNA modification processes? To address this yet unsolved question, we analyzed the emergence of modified nucleosides within new transcripts in the recovery phase using the NAIL-MS experiment from Figure 3c. For new tRNAs, we observed delayed incorporation of most modified nucleosides. This was exemplarily shown for m^5C and mcm^5s^2U in Figure 5d, but also for Cm, m^3C, Gm, m^1G, m^2G, $m^{22}G$ and m^7G. For most modified nucleosides, the incorporation delay was strongest for the oxidants TBH and $NaAsO_2$, and H_2O_2 and MMS had more modest delays. For rRNA, a similar delay in modification speed was observed. Cm formation was especially slower under stress in comparison to unstressed cells (Figure 5e,f). In contrast, m^1A and Ψ in tRNA and Am in both rRNAs were incorporated as fast or even faster after stress and to a higher degree compared to the unstressed controls. In yeast $tRNA^{Phe}$, m^1A and Ψ were recently shown to be the starting point during maturation, which indicates their important role in this tRNA's modification network [16]. $NaAsO_2$ appeared to have increased m^1A and Ψ abundances in newly transcribed RNAs, which might indicate involvement of these modified nucleosides in the arsenite stress response.

Oxidative stress has a substantial impact on the cell, and thiolated biomolecules suffer especially from exposure to reactive oxidant species. For example, bacterial DNA can be naturally thiolated at a non-bridging oxygen of the phosphodiester bond. During exposure to hypochloric acid, the sulfur can be replaced by oxygen, which either causes lethal strand

breaks or a regular phosphodiester bond [19]. Our NAIL-MS data from $NaAsO_2$ and TBH stressed cells showed a potentially similar effect on thiolated tRNA modifications. In eukaryotes, methylthiolation of adenine at position 2 has been reported, and the resulting modifications ms^2i^6A and ms^2t^6A are known to reside in the anticodon stem-loop of tRNAs [3]. As shown in Figure S9A, the abundance of original t^6A increased during TBH exposure. In addition, both i^6A and t^6A were substantially more abundant in new tRNA transcripts compared to tRNAs from unstressed controls (Figure S9). Our NAIL-MS data indicate that thiolated nucleosides might be either direct substrates to reactive oxygen species or are impacted by the disturbed sulfur homeostasis.

4. Discussion

RNA and its modifications have gained renewed interest, and we are at the start of understanding their dynamic nature and their underlying mechanisms. In this study, we followed the fate of the model organism *S. cerevisiae* during stress exposure and stress recovery. As previously described by the pioneers in the field [4], we observed a stress-dependent change in tRNA modification abundance. Thanks to our study, we now have the mechanistic insight to hypothesize on the mechanism of tRNA modification profiling. Our data suggest that primarily RNA transcription, and especially the rates of activity of RNA modification enzymes, were causative of the reported changes. For 18S and 25S rRNA, we observed no or only minimal changes in RNA modification profiles due to stress in *S. cerevisiae*. However, due to the complete hydrolysis of RNA to nucleosides, it is possible that site-specific effects on RNA modifications found at multiple positions were lost in our analysis.

Exposure to H_2O_2, $NaAsO_2$ and TBH all resulted in oxidative stress, following different mechanisms. While H_2O_2 generates hydroxyl radicals by Fenton chemistry [20–22], arsenite stress induces indirect oxidative stress, which results in downregulation of RNA synthesis [23] and the overproduction of reactive oxygen species. Exposure to TBH leads to the generation of butoxyl radicals via a Fenton-type reaction [24]. Our analysis of both total RNA integrity (Figure 1b,c) and RNA modification adaptation substantiates the differential cellular effects of the oxidants. We think it is noteworthy that the common RNA composition of *S. cerevisiae* (85% rRNA, 10% tRNA and 5% mRNA) was scrambled during and after the stress exposure, which must be taken into account if RNA modification densities are quantified.

In the context of methylation stress, we have previously reported on RNA damage products in *E. coli* and human cells [10,11]. Until now, we have observed the damages of canonical nucleosides and recently of thiolated RNA modifications. In this work, we described the methylation damage found on 2′-O-ribose methylated nucleosides Am and Gm, which are highly abundant in rRNA. We suggest that m^7Gm is a new type of RNA damage; however, m^7Gm requires further structural validation by comparison with synthetic standards, which are currently not available. In addition, we observed both m^1Am and m^6Am. However, similarly to m^6A, we are unsure about the mechanistic origin of m^6Am. Recently, the Dedon laboratory presented convincing evidence that m^6A might emerge from m^1A through dimroth rearrangement due to the RNA hydrolysis protocol required for LC-MS analysis [25]. Thus, it is possible that m^6Am is in fact a secondary damage product of m^1Am which undergoes dimroth rearrangement.

Regarding the increases and sudden disappearance of all tRNA and rRNA methylation damage products, we can only speculate. In *E. coli*, the repair of m^3C takes two hours through AlkB. m^1A takes with more than four hours, and m^7G is not removed at all. The slow speed of repair in combination with a clear substrate specificity observed in *E. coli* makes us wonder how all methylation damage types are repaired within one hour in *S. cerevisiae*. Maybe, there are no demethylases involved, but a more global mechanism. Interestingly, we studied methylation damage in human RNA after MMS exposure, and to our surprise, we observed only minor quantities of RNA damage products which might argue towards instantaneous repair of the damage or a more sophisticated detoxification

processes [11]. Thus, enzymatic involvement in the repair cannot be excluded. However, regarding the various enzyme knockouts screened, we did not find a satisfying candidate for an involvement in RNA repair through demethylation. Thus, the question remains: how do eukaryotic cells remove RNA methylation damage? We hope to uncover the biology behind this question in future studies.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/genes12091344/s1, Figure S1: LD$_{50}$ for *S. cerevisiae* BY4741; Figure S2: detected damage products in total RNA after MMS treatment; Figure S3: comparative NAIL-MS for labeling validation of $^{13}C_6$-glucose and $^{15}N_2$-uracil labeling; Figure S4: comparison of SILIS^{Gen1} and SILIS^{Gen2}; Figure S5: High-resolution mass spectra of canonical nucleosides in SILIS^{Gen2}; Figure S6: Methylome discrimination based pulse chase experiment, modification levels of m$_3$C and m$_7$G; Figure S7: examination of knock-out strains; Figure S8: Absolute abundance of modified nucleosides per original tRNA from NAIL-MS pulse chase experiment.; Figure S9: NAIL-MS pulse chase experiment, levels of i$_6$A and t$_6$A after AsO$_2$ and TBH treatment.; Table S1: Overview of used synthetic standards; Table S2: Comparison of *S. cerevisiae* 18S and 25S rRNA modification levels determined in this work and literature; Table S3: absolute number of enzymatically and damage methylated nucleosides in 25S, 18S rRNA and tRNA; Table S4: absolute modifications levels in total RNA after MMS treatment in knock-out strains; Table S5: mass transitions of nucleoside isotopologues.

Author Contributions: Conceptualization, Y.Y. and S.K.-K.; methodology, Y.Y.; software, Y.Y. and E.v.d.L.; validation, Y.Y.; investigation, Y.Y. and E.v.d.L.; resources, S.K.-K.; data curation, Y.Y. and S.K.-K.; writing—original draft preparation, Y.Y. and S.K.-K.; writing—review and editing, Y.Y. and S.K.-K.; visualization, Y.Y.; supervision, S.K.-K.; project administration, S.K.-K.; funding acquisition, S.K.-K. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Holliday, R.; Pugh, J.E. DNA modification mechanisms and gene activity during development. *Science* **1975**, *187*, 226–232. [CrossRef] [PubMed]
2. Compere, S.J.; Palmiter, R.D. DNA methylation controls the inducibility of the mouse metallothionein-I gene lymphoid cells. *Cell* **1981**, *25*, 233–240. [CrossRef]
3. Boccaletto, P.; Machnicka, M.A.; Purta, E.; Piątkowski, P.; Bagiński, B.; Wirecki, T.K.; de Crécy-Lagard, V.; Ross, R.; Limbach, P.A.; Kotter, A.; et al. MODOMICS: A database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* **2017**, *46*, D303–D307. [CrossRef]
4. Chan, C.T.Y.; Dyavaiah, M.; DeMott, M.S.; Taghizadeh, K.; Dedon, P.C.; Begley, T.J. A Quantitative Systems Approach Reveals Dynamic Control of tRNA Modifications during Cellular Stress. *PLoS Genet.* **2010**, *6*, e1001247. [CrossRef] [PubMed]
5. Begik, O.; Lucas, M.C.; Pryszcz, L.P.; Ramirez, J.M.; Medina, R.; Milenkovic, I.; Cruciani, S.; Liu, H.; Vieira, H.G.S.; Sas-Chen, A.; et al. Quantitative profiling of native RNA modifications and their dynamics using nanopore sequencing. *bioRxiv* **2021**. bioRxiv:2020.07.06.189969.
6. Mauer, J.; Luo, X.; Blanjoie, A.; Jiao, X.; Grozhik, A.V.; Patil, D.P.; Linder, B.; Pickering, B.F.; Vasseur, J.J.; Chen, Q.; et al. Reversible methylation of m(6)A(m) in the 5′ cap controls mRNA stability. *Nature* **2017**, *541*, 371–375. [CrossRef]
7. Zheng, G.; Dahl, J.A.; Niu, Y.; Fedorcsak, P.; Huang, C.M.; Li, C.J.; Vågbø, C.B.; Shi, Y.; Wang, W.L.; Song, S.H.; et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **2013**, *49*, 18–29. [CrossRef]
8. Reichle, V.F.; Petrov, D.P.; Weber, V.; Jung, K.; Kellner, S. NAIL-MS reveals the repair of 2-methylthiocytidine by AlkB in *E. coli*. *Nat. Commun.* **2019**, *10*, 5600. [CrossRef]
9. Heiss, M.; Reichle, V.F.; Kellner, S. Observing the fate of tRNA and its modifications by nucleic acid isotope labeling mass spectrometry: NAIL-MS. *RNA Biol.* **2017**, *14*, 1260–1268. [CrossRef]
10. Reichle, V.F.; Weber, V.; Kellner, S. NAIL-MS in *E. coli* determines the source and fate of methylation in tRNA. *ChemBioChem* **2018**, *19*, 2575. [CrossRef] [PubMed]

11. Heiss, M.; Hagelskamp, F.; Marchand, V.; Motorin, Y.; Kellner, S. Cell culture NAIL-MS allows insight into human tRNA and rRNA modification dynamics in vivo. *Nat. Commun.* **2021**, *12*, 389. [CrossRef] [PubMed]

12. Borland, K.; Diesend, J.; Ito-Kureha, T.; Heissmeyer, V.; Hammann, C.; Buck, A.H.; Michalakis, S.; Kellner, S. Production and Application of Stable Isotope-Labeled Internal Standards for RNA Modification Analysis. *Genes* **2019**, *10*, 26. [CrossRef] [PubMed]

13. Collart, M.A.; Oliviero, S. Preparation of Yeast RNA. *Curr. Methods Mol. Biol.* **1993**, *23*, 13.12.1–13.12.5. [CrossRef] [PubMed]

14. Heiss, M.; Borland, K.; Yoluç, Y.; Kellner, S. Quantification of Modified Nucleosides in the Context of NAIL-MS. In *RNA Modifications: Methods and Protocols*; McMahon, M., Ed.; Springer: New York, NY, USA, 2021; pp. 279–306.

15. Reichle, V.F.; Kaiser, S.; Heiss, M.; Hagelskamp, F.; Borland, K.; Kellner, S. Surpassing limits of static RNA modification analysis with dynamic NAIL-MS. *Methods* **2018**, *156*, 91–101. [CrossRef]

16. Barraud, P.; Gato, A.; Heiss, M.; Catala, M.; Kellner, S.; Tisne, C. Time-resolved NMR monitoring of tRNA maturation. *Nat. Commun.* **2019**, *10*, 3373. [CrossRef] [PubMed]

17. Koš, M.; Tollervey, D. Yeast Pre-rRNA Processing and Modification Occur Cotranscriptionally. *Mol. Cell* **2010**, *37*, 809–820. [CrossRef]

18. Osheim, Y.N.; French, S.L.; Keck, K.M.; Champion, E.A.; Spasov, K.; Dragon, F.; Baserga, S.J.; Beyer, A.L. Pre-18S Ribosomal RNA Is Structurally Compacted into the SSU Processome Prior to Being Cleaved from Nascent Transcripts in *Saccharomyces cerevisiae*. *Mol. Cell* **2004**, *16*, 943–954. [CrossRef]

19. Kellner, S.; DeMott, M.S.; Cheng, C.P.; Russell, B.S.; Cao, B.; You, D.; Dedon, P.C. Oxidation of phosphorothioate DNA modifications leads to lethal genomic instability. *Nat. Chem. Biol.* **2017**, *13*, 888–894. [CrossRef]

20. Ohshima, H.; Tatemichi, M.; Sawa, T. Chemical basis of inflammation-induced carcinogenesis. *Arch. Biochem. Biophys.* **2003**, *417*, 3–11. [CrossRef]

21. Hampton, M.B.; Kettle, A.J.; Winterbourn, C.C. Inside the Neutrophil Phagosome: Oxidants, Myeloperoxidase, and Bacterial Killing. *Blood* **1998**, *92*, 3007–3017. [CrossRef]

22. Dat, J.; Vandenabeele, S.; Vranová, E.; Van Montagu, M.; Inzé, D.; Van Breusegem, F. Dual action of the active oxygen species during plant stress responses. *Cell. Mol. Life Sci. CMLS* **2000**, *57*, 779–795. [CrossRef] [PubMed]

23. Haugen, A.C.; Kelley, R.; Collins, J.B.; Tucker, C.J.; Deng, C.; Afshari, C.A.; Brown, J.M.; Ideker, T.; Van Houten, B. Integrating phenotypic and expression profiles to map arsenic-response networks. *Genome Biol.* **2004**, *5*, R95. [CrossRef]

24. Gazdag, Z.; Máté, G.; Čertik, M.; Türmer, K.; Virág, E.; Pócsi, I.; Pesti, M. tert-Butyl hydroperoxide-induced differing plasma membrane and oxidative stress processes in yeast strains BY4741 and erg5Δ. *J. Basic Microbiol.* **2014**, *54*, S50–S62. [CrossRef] [PubMed]

25. Wang, J.; Alvin Chew, B.L.; Lai, Y.; Dong, H.; Xu, L.; Balamkundu, S.; Cai, W.M.; Cui, L.; Liu, C.F.; Fu, X.Y.; et al. Quantifying the RNA cap epitranscriptome reveals novel caps in cellular and viral RNA. *Nucleic Acids Res.* **2019**, *47*, e130. [CrossRef] [PubMed]

RESEARCH ARTICLE

SYSTEMATIC ASSESEMENT OF METHYLATION DAMAGE IN NUCLEIC ACIDS

Yasemin Yoluç,[a] Hurina Hu,[b] Matthias Heiss,[b] Stefanie Kaiser, * [a,b]

[a] Y. Yoluç, Prof. Stefanie Kaiser
Institute of Pharmaceutical Chemistry
Goethe University Frankfurt
Max-von-Laue Str.9, Frankfurt, Germany
E-mail: stefanie.kaiser@pharmchem.uni-frankfurt.de

[b] H. Hu and M. Heiss
Faculty of Chemistry and Pharmacy
Ludwig-Maximilians University Munich
Butenandtstr. 5, Munich, Germany

Abstract: Enzymatic modification of nucleic acid is the basis of gene regulation and RNA activity. In addition, nucleic acids are nucleophiles and target of direct and potentially harmful methylation. The discrimination of enzymatic and damage methylation in RNA is possible using nucleic acid isotope labeling coupled mass spectrometry (short NAIL-MS). Here, we expand NAIL-MS to DNA for assessment of methylation damage abundance in DNA. In both *E. coli* and *S. cerevisiae*, we find a 10-20 fold higher degree of damaged nucleic acids compared to HEK cells. In a systematic comparison, we find equal abundance of the main damage product 7-methylguanine in DNA and RNA, while damage to the N-1 of adenine is mainly found in RNA but not DNA. In a pulse chase NAIL-MS study, we follow the repair of DNA *in vivo* in dependence of known and potential erasers in *E. coli* and *S. cerevisiae*.

Introduction

Living cells are a mixture of various biomolecules and their carefully orchestrated interaction is crucial for cellular homeostasis. From a chemist's perspective, the biomolecules of life are either electrophiles or nucleophiles which readily react with external chemicals, *e.g.* excreted from competing organisms or administered during drug therapy. Especially nucleic acids, the storage of the genetic code in case of DNA and the main regulator of the cell in case of RNA, exhibit various nucleophilic sites. These sites are prone to be modified in response to stress, *e.g.* UV radiation[2], alcohol[3], infections[4]. Stress exposure can lead to changes in modification densities, but also to addition of chemical groups on the nucleophilic sites. The stress induced modifications can range from methylation to more divers oxidative damage products like hydroxylated, carboxylated and even glycosylated nucleosides.[5-7] It is thus not surprising, that microbes use little electrophiles, e.g. alkylating agents such as streptozotocin, azaserine or methylenchloride to harm their competitor's biomolecules.[8-10] Similarly, we use alkylating agents in modern cancer therapy. Goodman and Gilman first described the effect of chloroethylamines, better known as nitrogen mustard (Figure 1), on Hodgkin's disease, lymphosarcoma and leukemia.[11] Exposition to chloroethylamines lead to cross-linking of guanine to guanine or adenosine and results in blocked replication.[12] Today we still benefit from these findings, as alkylating agents are still part of first-line therapies. The most nucleophilic positions of nucleic acids are found on the nucleobases (Figure 1) and react with the various electrophiles

either in an S_N1 or S_N2 type reaction. S_N1 reactions usually result in methylation of oxygens and nitrogens, whereas S_N2 reactions yield methylated nitrogens.[13-14] As the nucleobases of DNA and RNA are identical or in case of uridine and thymidine electronically similar, past *in vitro* studies did not discriminate between the two molecules.[14] In 1983, Singer and Grunberger provided an overview of various electrophiles and the resulting alkylation products from reaction with single and double-stranded nucleic acids *in vitro* and *in vivo*.[14] One of them, methyl methanesulfonate (MMS) is an S_N2 type reagent and is the most commonly used agent to study methylation damage in cells. In both single and double stranded nucleic acids, position N-7 of guanine was found to be the most reactive substrate. In addition, positions N-3 and N-1 of adenine are methylated in double stranded nucleic acids. In single-stranded nucleic acids position N-1 of adenine is more reactive than in double-stranded nucleic acids where the N-1 position is occupied by Watson-Crick base pairing.

The least reactive position is the N-3 of cytosine, which is methylated to a lower extent than the other reported positions.[14] *In vivo* studies commonly focused on the methylation damage inflicted on genomic DNA for two reasons. The first is the biological relevance of DNA damage as unwanted methylation is potentially mutagenic and thus carcinogenic. The second is that the major product of direct methylation agents is 7-methylguanine, a modification which is commonly found as a natural modification in all major RNA types such as mRNA, tRNA and rRNA.[15] Until recently, it was not easily possible to distinguish RNA methylation marks which emerge either from enzymatic reaction and/or direct methylation. Thus, it was not possible to assess the absolute number
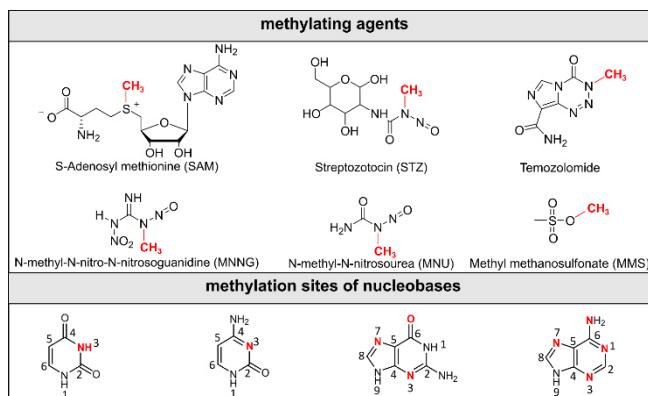

Figure 1. Major methylation agents (transferred methyl group is accentuated in bold red) and their targets in nucleobases of RNA and DNA indicated in red

of damage sites in RNA and compare it to those found in DNA. In 2017 our lab developed metabolic nucleic acid isotope labeling coupled mass spectrometry (NAIL-MS), a key technique to absolutely quantify nucleosides and to trace their origin and fate over time.[16] With NAIL-MS, we provided data for absolute numbers of enzymatic and damage methylation in tRNA of $E.\ coli$[1, 17-18], rRNA and tRNA of $S.\ cerevisiae$[19] as well as 18S rRNA and tRNA of HEK cells[20].

All organisms have developed various pathways to maintain the integrity of their chemical reactive nucleic acids and thus avoid the fulminant consequences. Over the last decades, various repair pathways have been postulated, such as direct reverse base excision repair, mismatch repair, nucleotide excision repair, homologues and non-homologues end joining pathway.[21] Repair mechanisms in $E.\ coli$ are well studied and the key player for direct repair is Ada. Ada has an active demethylation activity which triggers its actual role as initiator for adaptive response and provokes the expression of the repair enzymes AlkA and AlkB.[22] The glycosylase AlkA removes methylations on position N-3 of adenine glycosylase AlkA removes methylations on position N-3 of adenine and positions N-3 and N-7 of guanine in bacterial DNA through glycosidic bond hydrolysis.[23] Subsequently, the apurinic/apyrimidic site is repaired by the base excision pathway.[22] AlkB, an alpha-ketoglutarate-dependent dioxygenase, can repair methylation damage on DNA.[24-25] It was shown that AlkB oxides the methyl groups of m^1A and m^3C to hydroxymethyl derivatives, which dissociate to form formaldehyde and recover adenine and cytosine in DNA and RNA.[24-25] Later it was reported that AlkB additionally has the ability to demethylate m^1G[26], m^3T[26-27], m^6A[28] $in\ vitro$ and to reverse ms^2C to s^2C by direct sulfur demethylation in tRNA of $E.\ coli\ in\ vivo$.[1] Surprisingly, only little is known about demethylation in the well-studied model organism $S.\ cerevisiae$. Mag1, the potential homologue of bacterial AlkA, removes methylation damage on m^7dG and m^3dA in DNA $in\ vivo$[29] and m^1dA and m^3dC $in\ vitro$[30]. Besides Mag1, the apurinic endonuclease/3'-diesterase Apn1 is reported to repair DNA methylation through the base excision repair pathway[31-33]. In 2002, various $S.\ cerevisiae$ knockout strains were exposed to MMS and 40 strains showed increased sensitivity towards the treatment.[33] Among them were the knockouts of Mag1, Apn1 and in addition Met18. Met18 is a subunit of the cytosolic iron sulfur protein assembly machinery[34-35], which is also found in dioxygenases such as AlkB. The above mentioned repair mechanisms and identification of enzymes involved in the process have been done either $in\ vitro$ or in radio-isotope-labeled $in\ vivo$ studies[36]. With the development of NAIL-MS, we introduced pulse-chase assays independent of radio-isotopes and tracing of nucleic acid damages over time is now possible. We have used NAIL-MS in $E.\ coli$, $S.\ cerevisiae$ and HEK cells to follow the fate and repair of methylation damage in both tRNA and rRNA but mechanisms in DNA were not investigated yet.

Here, we expand our NAIL-MS technology to DNA and compare the extent of nucleic acid damage in the model organisms $E.\ coli$, $S.\ cerevisiae$ and HEK cells. The combination of NAIL-MS with a pulse chase assay, allowed us to trace the fate of the endogenous and damage caused methylations. In $E.\ coli$ we observed m^6dA and m^5dC to be endogenous modifications and determined m^7dG as the main damage product of MMS. In $S.\ cerevisiae$ we did not detect any endogenous DNA methylations but could show that m^7dG and m^1dA emerge upon MMS exposure. Damage induced

methylations decreased substantially during recovery. By examination of knockouts of known repair enzymes, we could confirm the contribution of AlkB to m^1dA m^3dC DNA repair in $E.\ coli$. In yeast we could attest the involvement of Mag1 in m^7dG and m^1dA repair and of Met18 in m^1dA repair. In HEK cells we could identify m^5dC in the genomic endogenous methylome. Further, we could confirm the formation of m^7dG in DNA but we see no evidence of active repair. In a final step we compare the abundance of DNA damage and RNA damages in a dose and organism dependent manner.

Results and Discussion

Methylation damage of DNA and RNA of $E.\ coli$

$E.\ coli$ was cultivated in unlabeled medium and after reaching the mid log phase cells were exposed to MMS doses of 5 mM and 20 mM (LD_{20}-LD_{50} of used strains Figure S1 and Reichle $et\ al.$[1]). In addition to the wild type strain (wt), $\Delta alkA$ and $\Delta alkB$ strains were examined. Samples were taken right before stress exposure to quantify native methylation marks and just after stress exposure to examine the acute phase[33]. MMS was removed by pelleting $E.\ coli$ and resuspension in fresh, but stable isotope labeled medium containing $^{15}NH_4Cl$ and L-methionine-[$^2H_3$]-methyl. S-Adenosylmethionine (SAM) is the natural methyl-donor for methyltransferases, and by feeding L-methionine-[$^2H_3$]-methyl enzymatic methylations result in a mass increase of +3 m/z whereas methylations resulting from MMS comply the unlabeled mass transition. The metabolic labeling allows to distinguish between nucleic acids present during stress exposure (unlabeled) and newly replicated/transcribed (labeled) nucleic acid molecules. After 5 hours, the recovery sample was drawn to trace the fate of methylated nucleosides of nucleic acids present during the MMS exposure. The concept is graphically displayed in Figure 2A.

We find that the levels of the bacterial epigenetic marks m^6dA and m^5dC are not affected by MMS, while m^7dG, m^1dA and m^3dC are solely formed upon MMS exposure (Figure 2B). Hence, there was no need to differentiate the source of methylation during MMS exposure and the experiment could be initiated with unlabeled medium (Figure 2A). m^7dG is the most prominent damage product in DNA with an abundance of 2.2 per 1000 canonical nucleotides (nts, Figure 2B). We confirm the previous report of AlkA involvement in m^7dG repair[23], as we observe a slower loss of m^7dG in the $\Delta alkA$ strain in the NAIL-MS context (Figure S2). The less abundant DNA damage products m^1dA and m^3dC (0.015 and 0.005 per 1000 nts, respectively) appear to accumulate in the $\Delta alkB$ strain, which confirms the reported activity of AlkB towards these damages[24-25]. As expected, the bacterial epigenetic marks m^6dA and m^5dC are unchanged by the MMS stress and tested genetic backgrounds as expected.

At the LD_{50} dose (20 mM), we find 17.9 damaged nucleosides per 1000 canonical nucleosides, which are distributed 1/3 on RNA and 2/3 on DNA (Table 1). In $E.\ coli$, m^7dG accounts for 66% of all nucleic acid damages while the RNA damages m^1A, m^3C and ms^2C account for a total of 27% of all MMS induced damages, m^1A 16%, ms^2C 10%, m^3C 1.5% respectively.
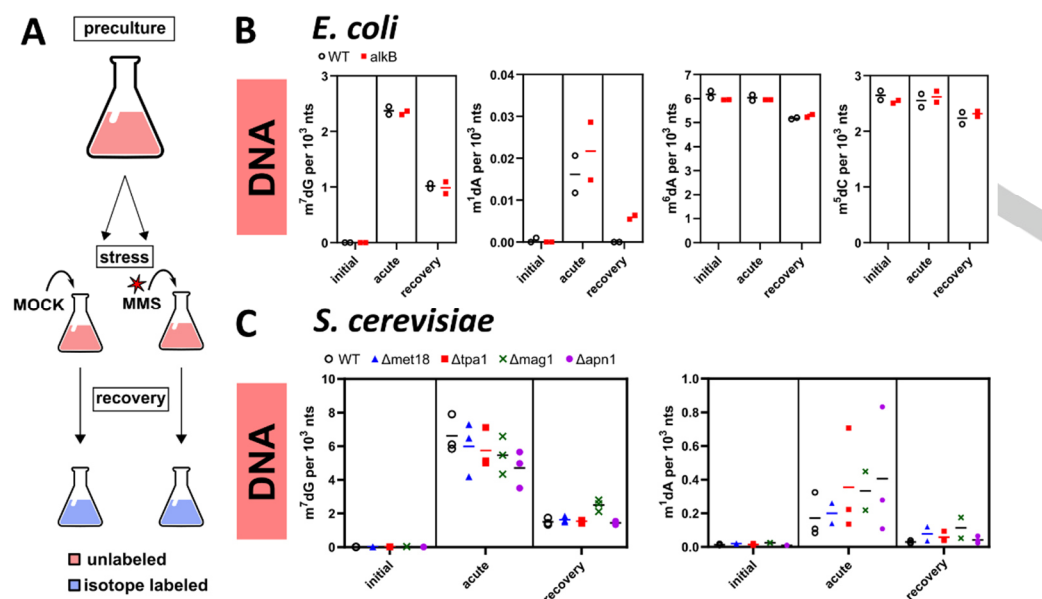
Figure 2. A) concept sketch of NAIL-MS pulse chase assay for *E. coli* and *S. cerevisiae* B) absolute number of methylated nucleosides per 1000 canonical nucleosides in DNA of wt (black) and AlkB (red) knock out *E. coli* in initial phase, after 1 h of 5 mM MMS exposure and after 5 h of recovery, n=2, mean as line and individual values C) absolute number of methylated nucleosides per 1000 canonical nucleosides in DNA of WT (black), Met18 (blue) and Tpa1 (red) Mag1 (green) and Apn1 (pink) knock out *S. cerevisiae* in initial phase, after 1 h of 12 mM MMS exposure and after 5 h of recovery, n=3, mean as line and individual values.

Methylation damage of DNA and RNA of *S. cerevisiae*

In *E. coli* DNA, we identified the N7 of guanine to be the most prominent target of DNA damage. We were wondering whether organisms with their DNA safely enclosed by the nuclear membrane have a similar high abundance of potentially lethal DNA damages and exposed *S. cerevisiae* to MMS. In an initial experiment we cultivated *S. cerevisiae* in unlabeled medium without MMS exposure and took samples when cells reached the mid-log phase (initial), after one hour (acute) and after two hours (recovery) to assess the endogenous methylome in DNA. No endogenous methylome was detected in the DNA of *S. cerevisiae* (Figure S3A), thus there was no need to differentiate the source of methylation during MMS exposure. Therefore, *S. cerevisiae* was cultivated in unlabeled medium, until mid-log phase. Subsequently, cells were exposed to the LD$_{50}$ concentration of 12 mM MMS for one hour[19]. After that, the recovery phase was initiated by medium exchange into stable isotope labeled medium. Again samples were taken right before stress exposure to quantify native methylation marks, just after stress exposure to examine the acute phase and after two hours of recovery to investigate potential repair mechanisms (Figure 2C). Knock out strains of Met18[37-38], Tpa1[39], Mag1[29, 40] and Apn1[31] were investigated with the intention of examining their role in nucleic acid methylation damage repair.

Before MMS exposure, we found no evidence for the presence of m^{5}dC[41] nor any other methylated nucleosides (m^{7}dG, m^{6}dA, m^{1}dA, m^{3}dC) in *S. cerevisiae* DNA (Figure S3A) which is in accordance with literature.[41-42] After MMS exposure, m^{7}dG (6.6 per 1000 nts) and m^{1}dA (0.17 per 1000 nts) appeared as the most abundant DNA methylation damages (Figure 2C). The abundance is the same order of magnitude as in *E. coli* (for comparison see Table 1). Both damages decrease over time, which may be explained by 3 scenarios. The first scenario, dilution of damaged nucleosides through replication, can be excluded as replication happens in stable isotope nutrient medium, and new DNA is distinguished by the mass spectrometric analysis. The second scenario is active repair through dealkylation, as found to be the case

in *E. coli* DNA (Figure 2B). Indeed, for Δ*mag1* cells, we observe an attenuated repair of m^{7}dG of 23% (p-value 0.013) compared to the wt, which is in line with literature.[29] In addition, m^{1}dA repair is reduced by 17% in Δ*mag1* cells but not significant according to student t-test (p-value 0.16). Our NAIL-MS based *in vivo* observation confirms an *in vitro* study from Admiraal *et al.*, which reported that Mag1 recognizes substrates of AlkB (m^{1}dA and m^{3}dC) *in vitro*. Further, we see a reduced repair of m^{1}dA in the absence of Met18 (21%), but we did not achieve statistical significance (p-value 0.23) due to the low number of analyzed samples for Δ*mag1* and Δ*met18* (n=2). We did not observe the involvement of Tpa1 or Apn1 in DNA damage repair of m^{7}dG and m^{1}dA (n=3).[31] Yet, our data also shows that DNA damage is partly lost independently of the repair enzymes Mag1 and Met18. Thus, a third scenario requires consideration. The third scenario is death of the most strongly damaged cells, which are lost at the step of cell harvesting through centrifugation. In this scenario, DNA with high abundance of damage is lost from the analysis over time and the number of damages per 1000 nts drops as only the DNA of the less damaged survivor cells remains upon extraction.

The impact of MMS exposure on the native tRNA modification profile has been recently reported elsewhere.[16, 19] In the acute phase 6.41 × m^{7}G, 3.44 × m^{1}A, 0.66 × m^{3}C and 0.18 × m^{1}A per 1000 nts are observed. At the LD$_{50}$ dose (12 mM), we find 17.5 damaged nucleosides per 1000 canonical nucleosides, which are distributed 2/3 on RNA and 1/3 on DNA (Table 1).

Methylation damage of DNA of HEK cells

In higher eukaryotes, gene expression is regulated through m^{5}dC and its oxidized derivatives.[43-44] In terms of alkylation damage, m^{7}dG is reported as the main damage[13] after MMS exposure but little is known about its repair in human cells. Using m^{5}dC as a stable point of reference, we wanted to investigate the extent of MMS damage to DNA and the subsequent repair of damaged sites in human cells (Figure S4).
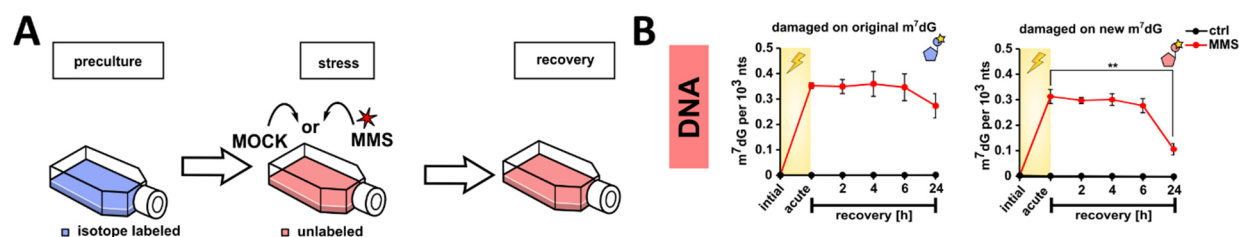
Figure 3. A) concept sketch of NAIL-MS pulse chase assay for HEK DNA analysis B) absolute number of enzymatic (black) and damage (red) m^{7}dG per 1000 canonical nucleosides in DNA of HEK in initial phase, after 1h of 1 mM MMS exposure and after 2 h, 4 h,6 h and 24 h of recovery ; n=3, error bars reflect standard deviation C) concept sketch of NAIL-MS pulse chase assay for HEK RNA D) absolute number of enzymatic (black) and damage (red) methylated nucleosides per 1000 canonical nucleosides in tRNA of HEK in initial phase, after 1h of 1 mM MMS exposure and after 2 h, 4 h and 6 h of recovery ; n=3, error bars reflect standard deviation. Statistical analysis was performed using student t-test, two sided. ** = p-value < 0.01.

Therefore, cells were cultured in stable isotope-nutrient containing medium for 7 days.[20] For the addition of sublethal doses of MMS (1 mM[20]), we switched to unlabeled medium, which in turn was replaced with fresh unlabeled medium after 1 hour of MMS exposure (Figure 3A). Samples were drawn before addition of MMS, after 1 hour of exposure and at several time points during the recovery phase. As expected, the abundance of m^{5}dC (Figure S6) does not change significantly due to the MMS exposure. In both, the existing and the newly replicated DNA, m^{7}dG forms up to 0.34 damages per 1000 nts. The number of damage sites is substantially lower in human cells compared to *E. coli* (11.8 (LD$_{50}$) and 2.2 (LD$_{20}$) per 1000 nts) and *S. cerevisiae* (6.62 per 1000 nts). This is in agreement with the sophisticated detoxification mechanism and secure packaging of the human genome, which makes it less accessible to the MMS electrophile. Other damage products such as m^{1}dA, m^{3}dA and/or m^{6}dA could also be detected, but at very low abundance (< 0.001 modification per 1000 nts). During the early recovery phase (1-6 hours after MMS removal) the abundance of m^{7}dG stays elevated. It takes 24 hours of recovery for a statistically significant decrease of m^{7}dG damage products from the newly replicated DNA. For the original DNA a decrease after 24 hours is also observable but of no statistical significance. Here, the potential loss of m^{7}dG from existing DNA would have to be attributed to either DNA repair or death of damaged cells, as described for *S. cerevisiae* in scenario 3. The more substantial decrease in new DNA might be a result of dilution through DNA replication after stress removal. Based on the collected data, it is not possible to fully determine which of the two mechanisms is responsible but an active removal of m^{7}dG from original DNA seems unlikely. Based on the collected data, it is not possible to fully determine which of the two mechanisms is responsible but most probably both contribute to some extent.

As observed for *E. coli* and *S. cerevisiae*, the same RNA damage products after MMS exposure in HEK cells were reported (~0.15 × m^{7}G and ~0.05 × m^{1}A per 1000 nts at 1 mM MMS[20] (Table 1). But (i) the abundance is at least 10-fold lower and (ii) we do not observe loss and/or repair of these tRNA damages within the experimental recovery timeframe of 6 hours.

Conclusion

Damage to nucleic acids harms the storage and translation of the genetic code on both DNA and RNA level. In this work we raise our attention to methylation damages, as native methylation of nucleic acids and damage could not be distinguished in the past. With NAIL-MS, we can now study damage methylation in both DNA and RNA in parallel to native nucleic acid modifications inside living cells. In this systematic *in vivo* comparison, we find an organism dependence. For all assessed organisms, the methylation at position N-7 of guanine in DNA is the main nucleic acid damage accounting for 99% of all damages in *E. coli*, 97% in *S. cerevisiae* and 71% in HEK cells at the respective organisms LD$_{50}$ dose. Damage to the N-1 of adenine is mostly low abundant for DNA, but in RNA it is of similar abundance as the RNA N7-guanine damage. Damages to purine nucleosides is very low abundant in all organisms for DNA and RNA. Interestingly, damage does not only occur to canonical ribonucleosides but in addition to native RNA modifications such as 2-thiocytidine. The respective damage product, ms^{2}C, is of similar abundance in *E. coli* tRNA as is m^{1}A or m^{7}G.[1]

Nucleic acid methylation damage is fatal to organisms and with its DNA unprotected, *E. coli* has developed enzymatic strategies to remove these damages efficiently. Here, we confirm the role of AlkA for removal of m^{7}dG and AlkB for repair of m^{1}dA and m^{3}dC *in vivo* using our NAIL-MS technique.

S. cerevisiae has its DNA safely packaged in the nucleus and should be less accessible for electrophiles such as MMS. Yet, *S. cerevisiae* DNA damage is of similar abundance as in *E. coli*. This argues for an efficient diffusion of the MMS molecule inside the cell and potentially inefficient or absent electrophile detoxification and damage repair. Concerning the latter, our NAIL-MS based experiment hints towards the involvement of Mag1 as the m^{1}dA and m^{3}dC dealkylation enzyme *in vivo*. Our finding confirms an *in vitro* study, which reported Mag1 to be similar to *E. coli* AlkB and defined m^{1}dA and m^{3}dC as substrates *in vitro*.

Both *S. cerevisiae* and human HEK cells have their DNA wrapped safely in the nucleus. Yet, the extent of DNA damage is 20-fold lower in HEK cells compared to yeast. In addition, RNA damage

Table 1. Structural formula of formed damage nucleosides m^{7}G, m^{1}A, m^{6}A, m^{3}C and ms^{2}C, damage caused by methylation is displayed in red; number of formed damage product per 1000 canonical nucleosides in *E. coli*, *S. cerevisiae* and HEK. Values for tRNA damage at LD$_{50}$ for *E. coli* were taken from Reichle *et al*.[17], for *S. cerevisiae* from Yoluç *et al*.[19] and for HEK from Heiss *et al*.[20]

methylation products		m^{7}G		m^{1}A		m^{6}A		m^{3}C		ms^{2}C	
c (MMS)	per 1000 nts	DNA	tRNA	DNA	tRNA	DNA	tRNA	DNA	tRNA	DNA	tRNA
LD$_{50}$	*E. coli*	11.8	0.71[17]	0.03	2.96[17]	-	0.38[17]	0.03	0.26[17]	n.a.	1.67[17]
	S. cerevisiae	6.62	6.41[19]	0.17	3.44[19]	-	0.18[19]	-	0.66[19]	n.a.	n.a.
	HEK	0.34	0.15[20]	-	0.14[20]	-	-	-	-	n.a.	n.a.

is 40-fold lower in HEK cells which argues for a more sophisticated detoxification mechanism in HEK cells compared to yeast. In addition, more than 9 nucleic acid demethylases are reported and/or predicted in human cells, which might be further involved in the efficient and fast repair of nucleic acid methylation damage. Modern mass spectrometry and its unmatched sensitivity in combination with stable isotope labeling allowed a systematic assessment of nucleic acid methylation damage *in vivo* in the most prominent model organisms. On one side, our data confirms and strengthens knowledge for *E. coli* nucleic acid repair which is of paramount importance in the context of antibiotic development. On the other side, we realize how little is known about RNA damage and actual repair mechanisms in eukaryotes, especially humans. Given the rising number of drugs acting on DNA as anticancer drugs or even the new epigenome targeting drugs, a knowledge gap has been uncovered. We suggest NAIL-MS as a fast and reliable tool for studying human DNA and RNA modification dynamics for closing it.

Experimental Section

Chemicals and Reagents

All salts, reagents and media were obtained from Sigma-Aldrich (Munich, Germany) at molecular biology grade unless stated otherwise. The isotopically labeled compounds $^{13}C_5$,$^{15}N_2$-Uridine (Ribose-$^{13}C_5$, 98%; $^{15}N_2$, 96-98%) and $^{15}N_5$-Adenine ($^{15}N_5$, 98%) were obtained from Cambridge Isotope Laboratories (Tewksbury, MA, USA). Unlabeled glutamine, isotopically labeled L-glutamine-amide-^{15}N (98 atom% ^{15}N), L-aspartic-^{15}N acid (98 atom% ^{15}N) and L-methionine-[$^2H_3$]-methyl (98 atom% D) were obtained from Sigma-Aldrich. Isotopically labeled $^{13}C_6$-glucose (≥99 atom% ^{13}C) was obtained from Eurisotope (Saarbruecken, Germany). All solutions and buffers were made with water from a Sartorious arium® pro ultrapure water system (Goettingen, Germany). The nucleosides (dA) deoxyadenosine, (dC) deoxycytidine, (dG) deoxyguanosine, (T) thymidine adenosine (A), cytidine (C), guanosine (G) and uridine (U), were obtained from Sigma-Aldrich. (m^{6}dA) 6-methyldeoxyadenosine, 1-methyladenosine (m^{1}A), N3-methylcytidine (m^{3}C), N6-methyladenosine (m^{6}A), 7-methylguanosine (m^{7}G), 5-methylcytidine (m^{5}C), 5-methyluridine (m^{5}U) were obtained from Carbosynth (Newbury, UK). (m^{1}dA) 1-methyldeoxyadenosine, (m^{3}dC) 3-methyldeoxycytidine were purchased from Jena BioScience (Jena, Germany). (m^{7}dG) 7-methyldeoxyguanosine was provided by Dr. Christoph Borek.

Growth media for *E. coli*

Minimal medium M9 was used with and without the indicated isotopes. Unlabeled M9 was prepared by mixing a 10 × M9 stock solution with glucose, $MgCl_2$, Na_2SO_4 and $CaCl_2$. For unlabeled 10 × M9 stock solution, Na_2HPO4 (68 g/L), KH_2PO_4 (30 g/L), NaCl (2.5 g/L), and NH_4Cl (10 g g/L) were mixed and autoclaved. For 15N-labeled 10 x M9 stock solution Na_2HPO4 (68 g g/L), KH_2PO_4 (30 g g/L), NaCl (2.5 g g/L), and $^{15}NH_4Cl$ (10 g g/L) were mixed and autoclaved. $MgCl_2$ (0.1 M), $CalCl_2$ (0.1 M) Na_2SO_4 (0.1 M), and 20 % (w/w) glucose were prepared by sterile filtration. For a 5 mL M9 preculture 500 μL M9 stock solution was mixed with 100 μL glucose, 100 μL $MgCl_2$, 100 μL Na_2SO_4, and 5 μL $CaCl_2$. For 15N, CD3-labeled cultures, the ^{15}N-10 × M9 stock solution was used and 200 μL L-methionine-[$^2H_3$]-methyl (stock 5 g/L) was added to 5 mL of culture volume.

Growth media for *S. cerevisiae*

YNB minimal medium was used with and without the indicated isotopes. YNB medium was prepared by mixing a 10 × YNB stock solution supplemented with a mix of needed unlabeled aminoacids (final concentration:

0.02 g/L arginine, 0.02 g/L histidine, 0.06 g/L leucine, 0.03 g/L lysine, 0.05 g/L, phenylalanine, 0.4 g/L serine, 0.2 g/L threonine, 0.04 g/L tryptophane, 0.03 g/L tyrosine and 0.15 g/L valine). Depending on the desired labeling either 0.02 g/L $^{15}N_2$-uracil, 0.01 g/L $^{13}C_6$-glucose or their unlabeled isotopomers and 0.02 g/L L-methionine or 0.72 g/ L-methionine-[$^2H_3$]-methyl were used. All unlabeled additives were supplied by Sigma-Aldrich, Munich, Germany

Growth Medium for HEK cells

All cell culture media and supplements were obtained from Sigma-Aldrich (Munich, Germany) unless stated otherwise. Standard Basal medium for HEK 293 culture was DMEM D6546 high glucose supplemented with 10% FBS and 0.584 g/L L-glutamine. For all experiments where labeling of nucleosides was involved DMEM D0422 without methionine and cysteine was used. DMEM D0422 was supplemented with 10% dialyzed FBS (Biowest, Nuaillé, France), 0.584 g/L L-glutamine, 0.063 g/L cystine (stock concentration 78.75 g/L dissolved in 1M HCl), 0.03 g/L methionine, 0.05 g/L uridine and 0.015 g/L adenine. Uridine, adenine and methionine were either added as unlabeled or labeled compounds depending on the desired labeling.

E. coli cultivation

Bacterial cells were grown at 37 °C, 250 rpm, all media were prewarmed to 37 °C. A single colony of *E. coli* was inoculation in 5 mL M9 minimal medium. The next day, the cell suspension was diluted to OD 1 and grown to mid-log phase. MMS [DNA: 5 mM] was added to the culture. After 1h of stress exposure the medium was exchanged, by centrifugation [5 min, 1200 × g, 24 °C]. The resulting pellet was resuspended in fresh medium to initiate the recovery phase. Cells were harvested (7 mL) at set time points by centrifugation [5 min, 12000 × g, 4°C].

S. cerevisiae cultivation

Yeast cells were grown at 30 °C, 250 rpm, all media were prewarmed to 30 °C. A single colony of *S. cerevisiae* BY4741 was inoculation in 5 mL YNB minimal medium. The next day, the cell suspension was diluted to OD 1 and grown for 3 h to reach the mid-log phase. MMS [12 mM] was added to the culture. After 1h of stress exposure the medium was exchanged, by centrifugation [5 min, 3000 × g, 24 °C]. The resulting pellet was resuspended in fresh medium to initiate the recovery phase. Cells were harvested (2 mL) at set time points by centrifugation [5 min, 12000 × g, 4°C].

Cell culture

Cells were split 1:7 using standard procedures every 2-3 days to counter overgrowth. Cells cultured in DMEM medium were kept at 10% CO_2 for proper pH adjustment.

DNA extraction – *E. coli*

The bacterial total DNA was extracted using the E.Z.N.A® Bacterial DNA Kit from Omega Bio-tek (Norcross, USA). Isolation was performed according to the manufacturer's protocol.

DNA extraction – *S. cerevisiae*

Isolation of yeast DNA was achieved by a combination of two isolation methods. Cell lysis was performed according to the Bust'n Grab method developed by Harju et al.[45] The collected cell pellet was resuspended in 200 μL lysis buffer. Afterwards, the suspension was put on dry-ice until frozen completely (> 2 min). The suspension was quickly thawed by incubation in a hot water bath (80 °C) for 1 minute. The dry-ice/hot water bath cycle was repeated twice. Subsequently, the sample was vortexed vigorously for 40 seconds. Purification of the DNA was performed according to

a modified protocol of the E.Z.N.A® Bacterial DNA Kit from Omega Bio-tek (Norcross, USA). The following centrifugation steps were all carried out at room temperature. 200 µL TL buffer and 20 µL proteinase K solution were added to the suspension. The mixture was incubated at 55 °C and vortexed every 15 min. After one hour, 8 µL of a prepared RNase A solution (100 mg/mL) was added. The sample was incubated at room temperature for 60 min. Subsequently, the sample was centrifuged (5 min, 12 000 ×rcf) and 420 µL BL buffer was added to the supernatant. After incubation at 65 °C for 10 minutes 420 µL absolute ethanol (ice-cold) was added. The solution was vortexed thoroughly. Afterwards, the sample was loaded on to a HiBind® DNA Mini Column. The solution was transferred through the column material by centrifugation at 10 000 ×rcf for 1 minute. The column was washed with 500 µL HBC buffer and two times with 700 µL DNA wash buffer via centrifugation (1 min, 10 000 ×rcf). Subsequently, the empty column was centrifuged (5 min, 12 000 ×rcf) to completely dry the column. For the elution of the DNA 55 µL elution buffer, heated to 65 °C, was added to the column. After incubation at 65 °C for 5 min the DNA was eluted by centrifugation (1 min, 10 000 ×rcf). The elution step was repeated once. The eluate was concentrated using a centrifugal evaporator to a total volume ≤ 20 µL. The concentrated sample was stored at -20 °C.

DNA extraction – HEK

Isolation of HEK DNA was done by the Monarch® Genomic DNA Purification Kit following the manufacturer's instructions.

Nucleoside hydrolysis for mass spectrometry

300 ng DNA or RNA in aqueous digestion mix (30 µL) was digested to single nucleosides by using 0.2 U alkaline phosphatase, 0.02 U phosphodiesterase I (VWR, Radnor, Pennsylvania, USA), and 0.2 U benzonase in Tris (pH 8, 5 mM) and MgCl2 (1 mM) containing buffer. Furthermore, 0.5 µg tetrahydrouridine (Merck, Darmstadt, Germany), 1 µM butylated hydroxytoluene, and 0.1 µg pentostatin were added to avoid deamination and oxidation of the nucleosides. The mixture was incubated for 2 h at 37 °C and then filtered through 96-well 10 kDa molecular-weight cut-off plates (AcroPrep Advance 350 10 K Omega, PALL Corporation, New York, USA) at 3000 × g and 4 °C for 30 min. 1/10 Vol. of SILIS (stable isotope labeled internal standard) was added to each filtrate before analysis by QQQ mass spectrometry.

Preparation of DNA SILIS

Metabolically isotope labeled total DNA used for the production of the SILIS (stable isotope labeled internal standard) was prepared using bacterial DNA. Therefore, *E. coli* wild-type (BW25113) was grown in 15N- and $^{13}C_6$-labeled M9 minimal medium (overnight, 37 °C, 250 rpm). After dilution to a OD600 of 1 with M9 minimal medium [^{15}N- and ^{13}C-labeled] the culture was incubated (60 min, 37 °C, 250 rpm) to reach early mid-log phase. Subsequently, 18.7 mM MMS was added and the culture was incubated for 60 min. The cells were harvested by centrifugation (5 min, RT, 3000 rpm) and the total DNA was isolated using the Qiagen Genomic tip 500/G from Qiagen (Hilden, Germany) specialized for the isolation of large amounts of DNA. For the preparation of DNA SILIS, total DNA was enzymatically digested to single nucleosides according to chapter 3.12. Theophylline functioning as an external standard was added to the digested solution to a final concentration of 10 µM. Finally, 5 mM ammonium acetate (pH 5.3) was added to the filtrate to obtain a final concentration of 20 ng/µL of DNA SILIS. The solution prepared in the above mentioned way is referred to as 10×SILIS.

QQQ mass spectrometry

For quantitative mass spectrometry an Agilent 1290 Infinity II equipped with a diode-array detector (DAD) combined with an Agilent Technologies G6470A Triple Quad system and electrospray ionization (ESI-MS, Agilent Jetstream) was used. Operating parameters: positive-ion mode, skimmer voltage of 15 V, cell accelerator voltage of 5 V, N2 gas temperature of 230 °C and N2 gas flow of 6 L/min, sheath gas (N2) temperature of 400 °C with a flow of 12 L/min, capillary voltage of 2500 V, nozzle voltage of 0 V, and nebulizer at 40 psi. The instrument was operated in dynamic MRM mode. For separation a Core-Shell Technology column (Synergi, 2.5 µm Fusion-RP, 100 Å, 100 × 2 mm column, Phenomenex, Torrance, CA, USA) at 35 °C and a flow rate of 0.35 mL/min were used in combination with a binary mobile phase of 5 mM NH4OAc aqueous buffer A, brought to pH 5.6 with glacial acetic acid (65 µL), and an organic buffer B of pure acetonitrile (Roth, LC-MS grade, purity ≥.99.95). The gradient started at 100% solvent A for 1 min, followed by an increase to 10 % over 3 min. From 4 to 7 min, solvent B was increased to 40 % and was maintained for 1 min before returning to 100 % solvent A and a 3 min re-equilibration period.

Calibration

For calibration, synthetic nucleosides were weighed and dissolved in water to a stock concentration of 1–10 mM. The calibration solutions range from 0.3 to 500 pmol for each canonical nucleoside and from 0.3 to 500 fmol for each modified nucleoside and were spiked with 10 % SILIS. The sample data were analyzed by the Quantitative Software from Agilent. The areas of the MRM signals were integrated for each modification and their isotope derivatives. The absolute amounts of the modifications were referenced to the absolute amounts of the respective canonical. In the case of the pulse-chase experiment, the different isotopomers were referenced to their respective labeled.

Statistics

All experiments were performed at least three times (biological replicates) to allow student t-test analysis. p-values of student t-test (unpaired, two-tailed, equal distribution) were calculated using Excel or Graphpad Prism.

Acknowledgements

[1] V. F. Reichle, D. P. Petrov, V. Weber, K. Jung, S. Kellner, *Nature communications* **2019**, *10*, 5600.
[2] T. L. de Jager, A. E. Cockrell, S. S. Du Plessis, *Adv Exp Med Biol* **2017**, *996*, 15-23.
[3] M.-Y. Chen, Z. Gui, K.-K. Chen, J.-H. Ding, J.-G. He, J. Xiong, J.-L. Li, J. Wang, B.-F. Yuan, Y.-Q. Feng, *Chinese Chemical Letters* **2022**, *33*, 2086-2090.
[4] M. Liu, F. Chen, T. Liu, F. Chen, S. Liu, J. Yang, *Microbes Infect* **2017**, *19*, 580-586.
[5] M. Guo, X. Li, L. Zhang, D. Liu, W. Du, D. Yin, N. Lyu, G. Zhao, C. Guo, D. Tang, *Oncotarget* **2017**, *8*, 91248-91257.
[6] F. Tang, F. Yuan, B.-F. Yuan, Y. Wang, *Journal of the American Chemical Society* **2022**, *144*, 454-462.
[7] Y. Tang, S. J. Zheng, C. B. Qi, Y. Q. Feng, B. F. Yuan, *Analytical chemistry* **2015**, *87*, 3445-3452.
[8] P. Taverna, B. Sedgwick, *J Bacteriol* **1996**, *178*, 5105-5111.
[9] P. Vaughan, B. Sedgwick, J. Hall, J. Gannon, T. Lindahl, *Carcinogenesis* **1991**, *12*, 263-268.
[10] A. M. Wuosmaa, L. P. Hager, *Science* **1990**, *249*, 160-162.
[11] L. S. Goodman, M. M. Wintrobe, W. Dameshek, M. J. Goodman, A. Gilman, M. T. McLennan, *Jama* **1984**, *251*, 2255-2261.

[12]   J. Hansson, R. Lewensohn, U. Ringborg, B. Nilsson, *Cancer Res*
         1987, *47*, 2631-2637.
[13]   D. T. Beranek, *Mutation Research/Fundamental and Molecular
         Mechanisms of Mutagenesis* **1990**, *231*, 11-30.
[14]   B. Singer, D. Grunberger, *Molecular Biology of Mutagens and
         Carcinogens*, 1 ed., Springer USA, Boston, MA, **1983**.
[15]   P. Boccaletto, F. Stefaniak, A. Ray, A. Cappannini, S. Mukherjee,
         E. Purta, M. Kurkowska, N. Shirvanizadeh, E. Destefanis, P.
         Groza, G. Avşar, A. Romitelli, P. Pir, E. Dassi, S. G. Conticello, F.
         Aguilo, J. M. Bujnicki, *Nucleic Acids Research* **2021**, *50*, D231-
         D235.
[16]   M. Heiss, V. F. Reichle, S. Kellner, *RNA Biology* **2017**, *14*, 1260-
         1268.
[17]   C. Borek, V. F. Reichle, S. Kellner, *Chembiochem* **2020**, *21*, 2768-
         2771.
[18]   V. F. Reichle, V. Weber, S. Kellner, *ChemBioChem* **2018**.
[19]   Y. Yoluç, E. van de Logt, S. Kellner-Kaiser, *Genes (Basel)* **2021**,
         12.
[20]   M. Heiss, F. Hagelskamp, V. Marchand, Y. Motorin, S. Kellner,
         Nature communications **2021**, *12*, 389.
[21]   R. Hakem, *Embo j* **2008**, *27*, 589-605.
[22]   D. Mielecki, E. Grzesiuk, *FEMS Microbiology Letters* **2014**, *355*, 1-
         11.
[23]   O. D. Schaerer, J.-Y. Ortholand, A. Ganesan, K. Ezaz-Nikpay, G.
         L. Verdine, *Journal of the American Chemical Society* **1995**, *117*,
         6623-6624.
[24]   S. C. Trewick, T. F. Henshaw, R. P. Hausinger, T. Lindahl, B.
         Sedgwick, *Nature* **2002**, *419*, 174-178.
[25]   P. Falnes, R. F. Johansen, E. Seeberg, *Nature* **2002**, *419*, 178-
         182.
[26]   P. Falnes, *Nucleic Acids Res* **2004**, *32*, 6260-6267.
[27]   J. C. Delaney, J. M. Essigmann, *Proceedings of the National
         Academy of Sciences* **2004**, *101*, 14051-14056.
[28]   D. Li, J. C. Delaney, C. M. Page, X. Yang, A. S. Chen, C. Wong,
         C. L. Drennan, J. M. Essigmann, *Journal of the American
         Chemical Society* **2012**, *134*, 8896-8901.
[29]   K. G. Berdal, M. Bjørås, S. Bjelland, E. Seeberg, *Embo j* **1990**, *9*,
         4563-4568.
[30]   S. J. Admiraal, D. E. Eyler, M. R. Baldwin, E. M. Brines, C. T.
         Lohans, C. J. Schofield, P. J. O'Brien, *The Journal of biological
         chemistry* **2019**.
[31]   S. C. Popoff, A. I. Spira, A. W. Johnson, B. Demple, *Proc Natl
         Acad Sci U S A* **1990**, *87*, 4193-4197.
[32]   D. Ramotar, S. C. Popoff, E. B. Gralla, B. Demple, *Molecular and
         Cellular Biology* **1991**, *11*, 4537-4544.
[33]   M. Chang, M. Bellaoui, C. Boone, G. W. Brown, **2002**, *99*, 16934-
         16939.
[34]   D. J. Netz, J. Mascarenhas, O. Stehling, A. J. Pierik, R. Lill, *Trends
         Cell Biol* **2014**, *24*, 303-312.
[35]   J. Balk, M. Pilon, *Trends Plant Sci* **2011**, *16*, 218-226.
[36]   R. Ougland, C. M. Zhang, A. Liiv, R. F. Johansen, E. Seeberg, Y.
         M. Hou, J. Remme, P. Falnes, *Molecular cell* **2004**, *16*, 107-116.
[37]   O. Stehling, A. A. Vashisht, J. Mascarenhas, Z. O. Jonsson, T.
         Sharma, D. J. Netz, A. J. Pierik, J. A. Wohlschlegel, R. Lill,
         Science (New York, N.Y.) **2012**, *337*, 195-199.
[38]   K. Gari, A. M. León Ortiz, V. Borel, H. Flynn, J. M. Skehel, S. J.
         Boulton, *Science (New York, N.Y.)* **2012**, *337*, 243-245.
[39]   L. Aravind, E. V. Koonin, *Genome Biol* **2001**, *2*, Research0007.
[40]   J. Chen, B. Derfler, L. Samson, *Embo j* **1990**, *9*, 4569-4575.
[41]   J. H. Proffitt, J. R. Davie, D. Swinton, S. Hattman, *Mol Cell Biol*
         1984, *4*, 985-988.
[42]   S. Hattman, C. Kenny, L. Berger, K. Pratt, *J Bacteriol* **1978**, *135*,
         1156-1157.
[43]   A. P. Bird, A. P. Wolffe, *Cell* **1999**, *99*, 451-454.
[44]   V. Colot, J. L. Rossignol, *Bioessays* **1999**, *21*, 402-411.
[45]   S. Harju, H. Fedosyuk, K. R. Peterson, *BMC Biotechnol* **2004**, *4*, 8-
         8.

Entry for the Table of Contents
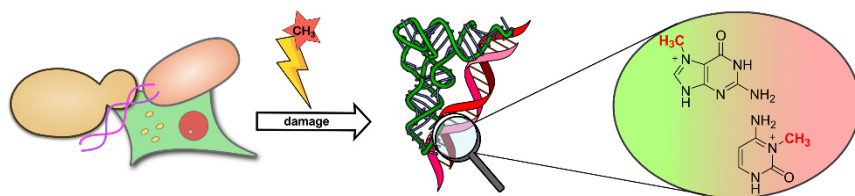
Methylation is a common modification of nucleic acids. They can be introduced enzymatically to regulate cellular homeostasis or can be damage induced which leads to malfunction in replication and translation. Here we provide a systematic study to identify and quantify damage besides the endogenous methylome covering DNA and RNA of *E. coli, S. cerevisiae* and HEK 293T cells. Nucleic acid isotope labeling couples mass spectrometry (NAIL-MS) is used to discriminate between enzymatic and damage methylation of nucleic acids.

Institute and/or researcher Twitter usernames: @KellnerLab

3.2 Challenges in viral RNA modification profiling

The first case of COVID-19 caused by the severe acute respiratory syndrome corona-virus-2 (SARS-CoV-2) was reported by the end of 2019. Since then, the virus has spread quickly and caused infections worldwide leading to a global pandemic with high mortality.[27] Despite much attention and numerous investigations, the nucleic acid modifications of this RNA virus have not been fully elucidated yet. An early work in 2020 identified 41 modification sites using nanopore sequencing, suggesting that most of the modifications consist of methylations.[245] Another study in 2021 determined 130 2'-O-methylations of canonical nucleosides in the genome of SARS-CoV-2 with Nm-Seq.[246] Intrigued by the variances in quantity and unknown identity of these modifications, my goal was to identify and quantify the modifications in the mutants D614 and G614 as well as in the variants of concern (VOCs) alpha, beta and delta of SARS-CoV-2 with LC-MS/MS. The biological features of the mutants and variants have been reviewed intensively elsewhere.[247, 248] With our robust LC-MS/MS method for nucleoside analysis, we proved nucleoside quantities in various domains of life previously.[158, 178, 240, 242, 249] It is necessary to understand the characteristics of the viral genome and the differences in mutants and variants in order to identify novel drug targets and optimize the application of available therapeutics and vaccines. The sample purity and integrity are fundamental for a valid analysis and are the requirements for an accurate quantification. In the following, I am going to point out the obstacles in evaluation of sample identity, purity and integrity of the genome of SARS-CoV-2, which might give a hint for the persisting lack of availability of the modification profile of SARS-CoV-2.

3.2.1 Validation of sample integrity

When the sequence of the viral RNA of SARS-CoV-2 was examined, it was determined that the RNA consists of 5493 × C (18%), 9595 × U (32%), 5863 × G (20%) and 8955 × A (30%)[30], this distribution is displayed in Figure 12A. The samples in this study were kindly provided by Prof. Denisa Bojkova from the Institute for Medical Virology, University Hospital, Goethe University, Frankfurt am Main. Caco-2 cells were infected with SARS-CoV-2, and the viral particles were concentrated by sucrose density gradient centrifugation of the cell culture supernatant 72 h post infection (hpi). After the RNA was extracted with the QIAamp Viral RNA Kit, the samples were handed over to me for quantitative analysis with LC-MS/MS.

The first quantification of hydrolyzed viral RNA resulted in a different distribution of canonical nucleosides in contrast the published sequence (Figure 12B) Here, adenosine constitutes 92% of all canonical nucleosides.
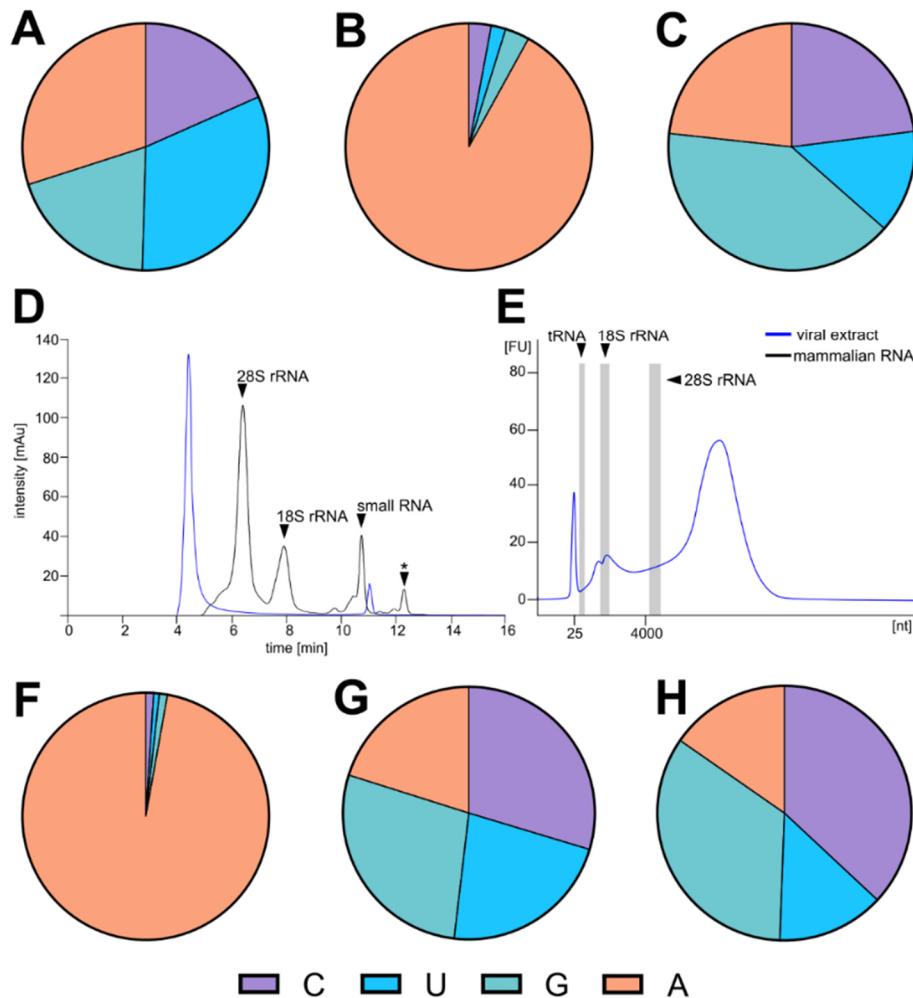

Figure 12: Pie chart of canonical ratios in (A) published sequence of RNA of SARS-CoV-2[30] (B) sample extracted with QIAmp viral RNA kit (C) sample extracted with QIAmp viral RNA kit and hydrolyzed with CIP only (D) UV trace from size exclusion chromatography of sample extracted with QIAmp viral RNA kit (blue) and mammalian total RNA (black), (*) indicates small degradated RNA (E) electropherogram of fraction collected at 4.4 min with SEC (blue) and size of mammalian tRNA, 18S and 28S rRNA indicated with grey bars (F) pie chart of canonical ratios in sample extracted with QIAmp viral RNA kit and purified with size exclusion chromatograph (G) pie chart of canonical ratios in sample extracted with QIAmp viral RNA kit without poly A carrier RNA (H) host cell samples

In order to identify the source of the excessive amount of adenosine, the sample was hydrolyzed with calf intestinal phosphatase (CIP) only. It was investigated if free nucleotides in the sample may have distorted the quantification. CIP only dephosphorylates the 5' and 3' ends of RNA phosphomonoesters, meaning that the RNA strand remains intact while single nucleotides are dephosphorylated. This approach showed no major abundance of a single species of nucleotides in the sample, even if the sample contained a minority of single nucleotides, those were evenly distributed with shares of 13% to 40%. No excess of adenosine was observed (Figure 12C).

Next, the sample integrity was examined. Therefore, the provided RNA was separated by size exclusion chromatography on a 1000 Å column, as described by Heiss *et al.*[243] The UV trace of the separation of mammalian total RNA is visualized in black for comparison (Figure 12D). The chromatographic separation of the viral extract resulted in a major peak at 4.4 min and a minor peak at 11 min (Figure 12D). As reported in Heiss *et al.* 28S rRNA elutes after 5 min, all RNA subspecies smaller than this (18S, tRNA) elute later.[243] Based on this knowledge the first peak eluting at 4.4 min must consist of RNA larger than 5000 nts (human 28S rRNA) and therefore was assigned to the viral RNA, which is consisting of nearly 30000 nucleotides, and the later peak was assigned to small RNA, in accordance with Heiss *et al.*[243] The size of the RNA isolated at 4.4 min was further characterized with the Agilent Bioanalyzer. The electropherogram revealed that the purified RNA consists of oligonucleotides larger than 4000 nucleotides (Figure 12E). From this, it was concluded that no ribosomal RNA from the host, which might have co-eluted during SEC, was interfering with the analysis and might be the reason for distorted abundances. Nevertheless, the quantification of this purified large RNA still resulted in 97% A (Figure 12F).

The source of the excess adenosine was later identified to be derived from the QIAmp viral RNA kit. This kit purifies viral RNA with the aid of poly A carrier RNA. From here on the viral RNA was purified without poly A carrier RNA. The canonical composition of RNA resulting from this adapted procedure is displayed in Figure 12G. Here, the distribution is similar to the expected ratio shown in Figure 12A. However, the abundance of C and G is higher than expected. The determination of the canonical distribution in the host RNA reveals the prevalence of C and G over A and U (Figure 12H). Hence remaining host RNA in the viral extracts may contributed to the slightly higher abundances of C and G in comparison to the published sequence.

In order to further examine the sample quality and scrutinize for contaminant host RNA, the extracted RNA was separated with a 0.7% agarose gel. A DNA ladder ranging until 20000 nts was utilized, instead of a RNA ladder, as the latter only ranges until 6000 nts and the viral RNA is expected to be 30000 nts long. In order to circumvent misinterpretation of sizes due to different migration of DNA and RNA, total RNA of *S. cerevisiae* was loaded as a control, with its known sizes of 25S, 18S rRNA and tRNA (Figure 13). The RNA extract from the host (Caco-2) and the viral extracts were loaded in triplicates.
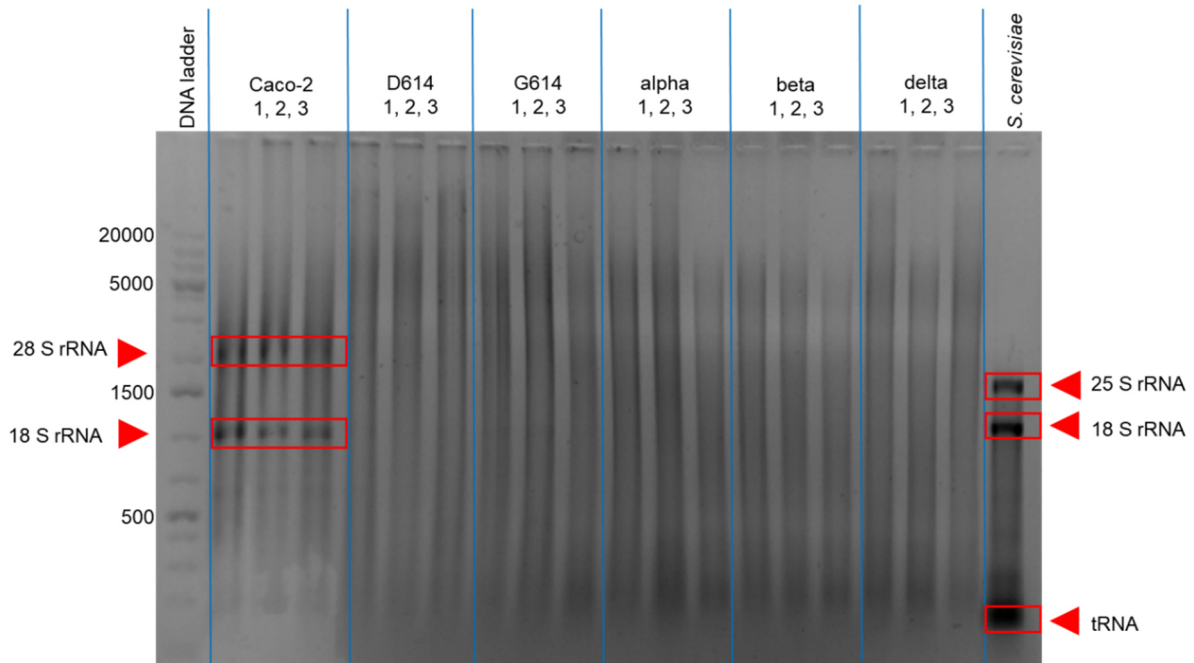
Figure 13: 0.7% agarose gel electrophoresis of RNA extracts from the host cell, the mutants and variant of SARS-CoV-2 and total RNA of *S. cerevisiae.* Lane 1: DNA ladder, lane 2-4: Caco-2, lane 5-7: D614, lane 8-10: G614, lane 11-13: alpha, lane 14-16: beta, lane 17-19: delta, lane 20: total RNA *S. cerevisiae*, 1 µg of each sample was loaded onto the gel, the gel was stained with GelRed and imaged with a UV transilluminator (λ=312 nm)

The 25S, 18S rRNA and tRNA of *S. cerevisiae* can be clearly identified in the last lane (Figure 13). The large ribosomal subunits of the extract from the mammalian host Caco-2 were also clearly visualized in the first to third lane. As anticipated, here the 28S rRNA band is slightly higher than the 25S rRNA from *S. cerevisiae* (3400 nts) due to the larger size of the mammalian rRNA (5000 nts). In lanes of the viral extracts no distinct bands were observed. The orthogonal approach of the agarose gel electrophoreses confirmed the absence of intact host ribosomal and tRNA in the viral extracts. However, the abundance of degraded host RNA in the viral extracts cannot be excluded. In Richter *et al.* and my own work on *The Stress-Dependent Dynamics of Saccharomyces cerevisiae tRNA and rRNA Modification Profiles* it was shown that rRNA degrades over time or upon stress, and is no further detectable, but co-elutes with RNA of smaller size.[99, 250] Even though most degradation fragments cannot be visualized in distinct bands, due to low abundance, they were shown to dilute small RNA and/or enrich small RNA fractions with *e.g.* rRNA specific modifications like m^{66}A.[250] The quantification of modifications is going to be discussed in the following chapter: 3.2.3 *Absolute quantification of modifications*.

Collectively, the presence of viral SARS-CoV-2 RNA could not be certified by this method, but the presence of potentially distorting intact rRNA or tRNA from host could

be excluded by size exclusion chromatography and subsequent analysis with the Agilent Bioanalyzer as well as with the agarose gel (Figure 12D, E, Figure 13). However, mammalian cells possess lncRNA which range from 200 to 10000 nts.[251] With the utilized methods, the presence of RNA subspecies smaller than 5000 nts was excluded, whereas the presence of lncRNA larger than 5000 nts could not be precluded. In general, the identity of RNA can be determined by sequencing, however this was beyond the scope of this study.

3.2.2 Validation of sample integrity on nucleosides level

The sample integrity was further examined on the nucleoside level by chromatographic separation of the sample after hydrolysis. The UV chromatogram of the hydrolyzed sample shows deoxynucleosides (dN, T) from DNA, besides the ribonucleosides from RNA (Figure 14A). The deoxynucleosides in the chromatogram were assigned as reported in Yoluç *et al.*[244] As a consequence, in the subsequent experiments the samples were treated with DNase I prior to RNA hydrolysis, in order to deplete remaining DNA. The successful DNA depletion is shown in Figure 14B. This UV chromatogram shows ribonucleosides only and theophylline, which is added to the SILIS as a control.[243]
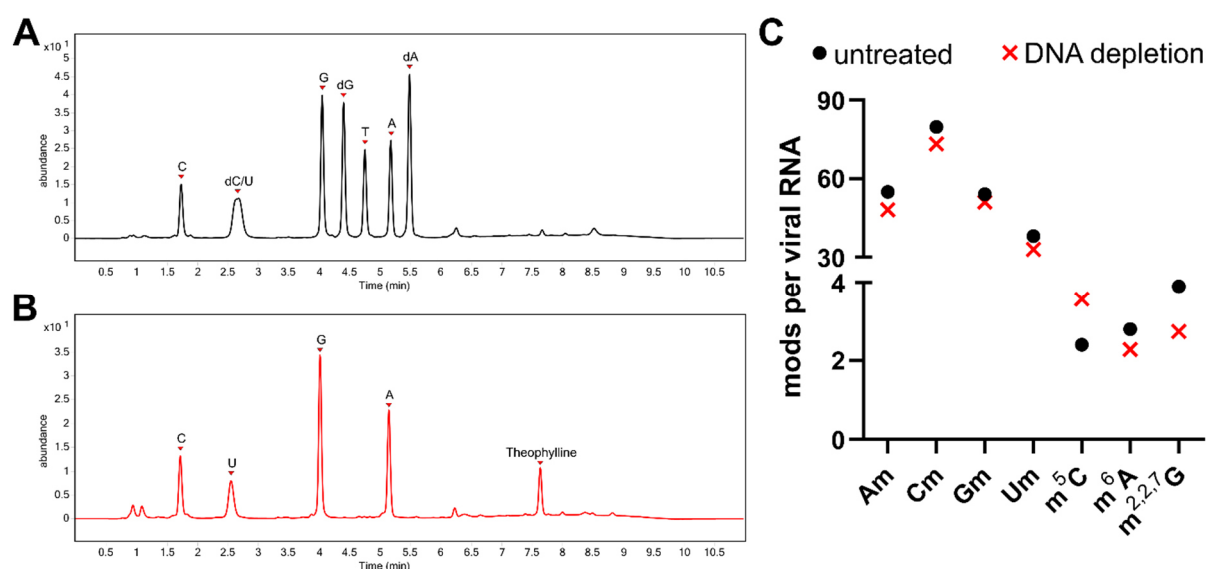
Figure 14: Effect of DNase I treatment on quantification of modifications (A) UV chromatogram of untreated RNA extract from SARS-CoV-2, peak identity labeled above (λ=260 nm) (B) UV chromatogram of DNase I treated RNA extract from SARS-CoV-2, peak identity labeled above (λ=260 nm) (C) comparison of modifications per viral RNA strand of untreated (black) and DNase I treated (red) RNA extract from SARS-CoV-2 (D614)

In an additional approach, it was examined if remaining DNA may interfere with the quantification of ribonucleosides. Therefore, the same sample was split prior to RNA hydrolysis and one aliquot was treated with DNase I. As depicted in Figure 14C, the modification densities are comparable regardless of DNase I treatment. From this, it is

concluded that residual DNA in the sample has no effect on quantification of RNA modification abundance. Hence, the DNA depletion can be considered optional.

3.2.3 Absolute quantification of modifications

After the source of excessive A has been identified, the presence of intact host RNA smaller than 5000 nts and a distortion by deoxynucleosides was excluded, the viral RNA profiling was conducted. The obtained quantities are summarized in table 1. Slight differences in modification density among the variants were observed. The determined values of this quantification clearly exceeds the reported 41 modification sites.[245] Here, solely the sum of base methylations accounted for 22 to 43 modifications per viral RNA strand, adding the ribose methylated modifications as well as Ψ and I, quantities up to 248 modifications per viral RNA strand were reached (Table 1). From the current literature it can be discerned that there are differences in modification density based on (a) the host cell line, (b) the SARS-CoV-2 variant and (c) the infection time prior to RNA extraction. In addition, it has been reported that the modifications on the viral RNA are highly dynamic.[252] The most studied modifications in the context of SARS-CoV-2 are m^6A, Ψ and the 2'-O-methylations.

The extent of variances, based on the above mentioned factors, can be clearly seen by the comparison of determined quantities for m^6A. In my work, the quantification of m^6A revealed an abundance of 1.7 - 2.8 per viral RNA, depending on the variant (Table 1). In contrast to that, 14 and 15 m^6A sites were determined with direct RNA sequencing (DRS) from SARS-CoV-2 (IVCAS 6.7512) or the gamma variant infected Vero E6 cells.[253, 254] Other studies using MeRIP determined 4 (24 hpi) and 13 (56 hpi) m^6A sites in Vero E6 cells infected with the BetaCoV/Wuhan/Ime-BJ01/2020 variant. While 7 sites were detected in Huh7 cells,120 hpi with the BetaCoV/Wuhan/Ime-BJ01/2020 variant[255], and 9 m^6A sites in Huh7 cells infected with the IVCAS 6.7512 variant by DRS[254].

Li *et al.* analyzed the variant USA-WA1/2020, from infected Vero cells 24 hpi.[256] The authors found in sum 208 ribose methylated modifications, 288 Ψ, 8 m^6A and 7 m^5C sites by LC-MS analysis, while I determined 180 – 227 ribose methylated modifications, up to 140 Ψ, 2.8 m^6A and 7 m^5C per viral RNA, depending on the variant. Another

study found 42 m⁵C sites with DRS in extracts from Vero cells infected with the Australia/VIC01/2020 variant [257] and Fleming *et al.* propose 5 conserved sites at the 3' end of the viral RNA with DRS[258].

As previously outlined, Yang *et al.*, proposed the presence of 130 2'-O-methylated nucleosides by Nm-seq of viral RNA from SARS-CoV-2 WT and the deletion variant Δ382, 48 hpi in Vero E6 cells.[246] However, the abundances of the different ribose methylated nucleosides was not assessed. Li *et al.*, who found 208 2'-O-methylated nucleosides in sum, reported 89 Am, 49 Cm, 64 Gm and only 5 Um per viral RNA. In my work, the sum of ribose methylated modifications ranges from 181 to 227 depending on the variant, which is in conformity with the reported dimensions.[246, 256] However, I quantified about 40 - 50 Am and Gm, respectively as well as 65 – 80 Cm and 29 – 38 Um, which is significantly more as reported by Li *et al.*[256] Nevertheless, the largest alignment in modification densities with my data was observed with the results of Li *et al.*[256], which is the sole other study utilizing LC-MS/MS for quantification. Indications for deviating abundances, especially for 2'-O-methylated nucleosides should be discussed in the following by comparing the procedures of sample preparation.

The first difference is the selection of the host cell line, while in my study Caco-2 cells were infected with the various mutants and variants, in Li *et al.* Vero cells were utilized[256]. Further, here the supernatant of infected Caco-2 cells was harvested after 72 hpi, while Li *et al.* analyzed cellular RNA 24 hpi. Also, the RNA extraction procedure varies, here viral particles were concentrated by sucrose gradient centrifugation and subsequent RNA extraction with the QIAmp viral RNA Kit and finally the large RNA was separated from small fragments by SEC. In contrast to that, Li *et al.* lysed the Vero cells with Trizol and extracted the RNA with the Direct-zol RNA Kit (Zymo).[256] Additionally, Li *et al.* treated the samples with DNase I prior to hydrolysis, while I could show that deoxynucleosides had no impact on quantification of RNA modifications and thus I omitted the DNA depletion. Also, the RNA hydrolysis protocol varies. I applied our usual one-pot hydrolysis for 1h at 37 °C, containing the enzymes calf intestinal phosphatase (CIP), benzonase, phosphodiesterase (PDE1), the deaminase inhibitor pentostatin and the antioxidant butylated hydroxytoluene (BHT), herein after the samples were directly proceeded for mass spectrometric analysis. In contrast to that, Li *et al.* carried out a two-step hydrolysis, beginning with hydrolysis by nuclease P1 (NP1) for 2h at 37 °C, followed by incubation with alkaline phosphatase for 2h at 37 °C and thus

a total incubation time of 4h. Subsequently, the hydrolysates were concentrated by centrifugal evaporation before analyzing them by mass spectrometry.[256] Apparent from this comparison, the sample preparation is significantly different and only the analysis technique is congruent. However, the comparison of different hydrolysis protocols in *Strategies to Avoid Artifacts in Mass Spectrometry-Based Epitranscriptome Analyses* by my colleague Dr. Steffen Kaiser and me, revealed that canonicals are released in comparable abundances whereas 2'-O-methylated nucleosides, which are reported to be high abundant in viral RNA, are released to a significantly lower extent. It was shown that PDE1 is indispensable for hydrolysis of 2'-O-methylated nucleosides. Especially, the release of Cm and Um was shown to be dependent on PDE1.[259] Interestingly, these two modifications were quantified to a considerably lower extent by Li *et al.*, who spared PDE1 from the hydrolysis, in contrast to my quantification resulting from PDE1 containing enzyme mixture. This may explain the variance in quantification of ribose methylated modifications.

Comparing the quantities in table 1, it appears that the modification abundances are comparable in the mutants D614 and G614. The similarities in the genome of D614 and G614 are plausible, as these mutants only differ in one amino acid of the spike protein. While the mutants D614 and G614 contain more 2'-O-methylated nucleosides in comparison to the variants alpha, beta and delta. The variants possess more mutations for spike proteins. The variants alpha and delta have even mutations in the nucleocapsid, an overview of all mutations is given by Tao *et al.*[247] Therefore, it is conclusive that their genome might differ more from the mutants D614 and G614. The variants beta and delta possess more base methylated modifications than the other examined mutants and variants. Especially, the number of methylated guanosines is higher in beta and delta variants.

The results given in table 1 should be interpreted with caution, as the analysis is limited by the pending confirmation of sample integrity, albeit the presence of potential distorting intact host RNA smaller than 5000 nts was excluded with multiple approaches. Yet, none of the methods were suitable for verification of sample identity and this should be taken into account as a possible source of error.

Table 1: Quantities of modifications in RNA extracts from SARS-CoV-2 variants. Values for certain modifications can be found in the rows. Values for the different mutants and variants can be taken from the columns. Fields with a blue background highlight the 2'-O-methylated nucleosides. Fields with a green background highlight the base methylated modifications. At the bottom of the table the sum of all ribose methylated and base methylated modifications as well as the sum of all modifications is given. The mean of n=3 is given, SD can be found in Table S1 in the appendix

#per RNA	D614	G614	alpha	beta	delta
Y	140.0	134.0	100.2	112.4	86.2
I	81.0	80.3	34.0	79.0	34.4
Am	54.9	51.7	41.4	47.0	42.7
Cm	79.7	75.3	65.8	69.3	72.9
Gm	54.0	51.6	41.4	46.1	42.9
Um	38.0	38.2	32.2	32.6	28.9
m^5C	2.4	4.1	4.8	5.3	7.0
m^1A	3.2	3.2	4.2	4.6	6.4
m^6A	2.8	2.7	2.0	2.2	1.7
m^1G	3.2	3.2	4.2	4.6	6.4
m^2G	1.8	2.1	3.5	3.9	6.9
m^7G	3.7	3.5	3.3	3.7	3.6
$m^{2,2}G$	0.6	0.7	1.3	1.3	2.5
$m^{2,2,7}G$	3.9	3.3	2.3	5.1	8.4
sum ribose methyl.	226.6	216.8	180.9	195.0	187.5
sum base methyl.	21.7	22.8	25.6	30.7	43.0
total	248.3	239.6	206.6	225.7	230.5

On the basis of marker modifications, as described by Richter *et al.*[250], the abundances of certain modifications from host and viral extracts should be compared. Richter *et al.* declared m^6_6A and m^3U as rRNA specific modifications and reported their abundance in RNAs of smaller size upon rRNA degradation, thus these modifications can be used as tool for rRNA identification.[250] Here, D614 was selected exemplarily for the viral extracts and its modification abundance was compared to control sample from the host (Caco-2) RNA, consisting of rRNA (see agarose gel, Figure 13). The number of modifications was referenced to the length of the whole viral genome. The abundance of m^6_6A and m^3U in viral extracts was less or equal 1 per whole viral genome, whereas double the amount was detected in RNA of Caco-2 cells (Figure 15). Same applies for ribose methylated modifications (Nm), also here twice as much modifications were detected in the eukaryotic RNA in comparison to the RNA extract from D614. The sum of Nm from Caco-2 cells equals 429 per viral RNA (30000 nts), which equals 102 Nm per human rRNA and thus is in alignment with literature.[260] Interestingly, the proportion of Nm abundance from viral extracts is comparable to the proportion from Caco-2. It ap-

pears that the ribosomal modification pattern of the host prevails in the viral RNA extracts, albeit the absolute levels are lower in viral RNA extracts. This might be attributed to the dilution of host RNA in viral extracts by the less modified viral RNA itself. These samples were purified by sucrose density gradient centrifugation, which is based on size dependent separation, thus the viral RNA, which is 6 × larger than
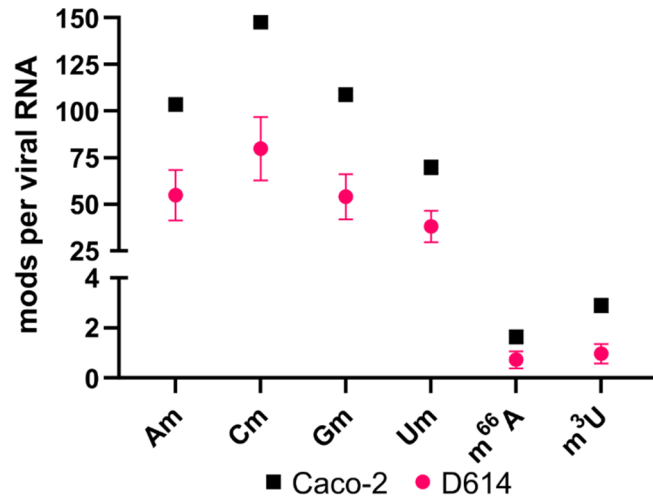

Figure 15: Comparison of modifications per viral RNA strand of Caco-2 rRNA (black) and RNA extracts from D614 (pink), the mean of n=3, error bars reflect standard deviation

e.g. 28S rRNA (5000 nts), should be distinctly separable from host RNA. However, the abundance of rRNA markers ($m^{66}A$, m^3U) in the viral extract as well as the similarity in proportion of Nm abundance between Caco-2 and viral extracts points to the likelihood of host rRNA contamination in the viral extracts, which distorts the sample integrity and consequently the quantification of viral modifications. A further indicator for inconsistency in sample integrity of the viral RNA extracts may be the large error of viral RNA extracts in contrast to the host. Even though RNA from both sources was processed the same way, the error of replicates is more pronounced in viral RNA extracts than in host RNA (Figure 15, Table S1 for more strains and modifications).

From this it was assumed that the purification method might account for inconsistency in sample integrity and contamination by host RNA. As a result, it was decided to enrich the viral RNA with hybridization probes specific to the sequence of SARS-CoV-2. Vero E6 cells were infected SARS-CoV-2 (D614) and RNA was extracted 48 hpi with Trizol. Subsequently, viral RNA was enriched with MagIC (Element Zero, Berlin, Germany) from cellular RNA and from cell culture supernatant of infected cells as well as from uninfected cells to determine the extent of unspecific binding. While the enrichment from cellular RNA was conducted with three individual biological replicates, the RNA enriched from the supernatant was obtained by pooling 3 biological replicates. The probes utilized for hybridization consist of 45 different probes along the genomic sequence of SARS-CoV-2 with the length of 55 nucleotides. The enriched RNA was provided by Dr. Ryan Bennett and Dr. Harald Smith from OyaGen (Rochester, NY, USA).

Figure 16 shows the canonical ratios for the reported sequence of SARS-CoV-2[30] (Figure 16A), for enriched samples from cellular RNA of infected cells (Figure 16B) and viral particles from the cell culture supernatant of infected cells (Figure 16C). In enriched viral RNA from cellular RNA of infected cells the abundance of the nucleosides C and G are in good alignment with the published abundances, while A is slightly more and U less abundant (Figure 16B), it can be concluded that the canonical ratio for enriched RNA from cellular source is comparable to literature. From Figure 16C, it is apparent that the abundance of A, from enriched RNA from cell culture supernatant of infected cells is higher than reported, while the abundance of the other canonical nucleosides is lower, which might be attributed to low sample amount or insufficient purity and is going to be discussed later.

Even though, the hybridization probes were reported to be specific to the genome of SARS-CoV-2, the extent of unspecific binding and the resulting abundance of modifications in the control samples should be examined. First the total amounts of canonical nucleosides were compared. From figure 16D, it is evident that the canonicals from cellular RNA of infected samples are up to 110 times more abundant than in the cellular control sample and thus this marginal unspecific binding can be neglected. The abundances of canonical nucleosides from supernatant of control and infected samples was below 2 pmol, in similarity to the unspecific binding observed in cellular control samples. Even though, these samples from cell culture supernatant were pooled from 3 biological replicates, the yield appears to be quite low and thus it needs to be taken into account, that the abundance of modifications in these samples might be below the limit of quantification.

Next, the modification abundances for ribose methylated modifications and the rRNA marker $m^{66}A$ in the samples were compared. Firstly, no $m^{66}A$ was detected in enriched RNA from cellular RNA of infected cells, which is an indicator for the absence of host rRNA (Figure 16E). The abundance of 2'-O-methylated nucleosides is about 10 times lower, as determined for the initial viral RNA extracts and their proportions distinctly vary from the proportion of eukaryotes (compare to Figure 15). Also the abundance of 2'-O-methylated nucleosides from the cellular control sample is distinctly lower, this might be attributed to the composition of the sample. While the previous host sample consisted of rRNA (agarose gel, Figure 13), which contains ~100 ribose methylated nucleosides, here the control sample was yielded from total cellular RNA and thus

might consist of a mixture of RNA. As in mammals only 0.52% of mRNA and 0.8% of tRNA is ribose methylated, the overall density of 2'-O-methylated nucleosides is diluted in total RNA, and thus is lower than in rRNA. The abundances of modifications in samples from cell culture supernatant of uninfected cells is similar to the abundances from RNA enriched from uninfected cells, hence the unspecific binding in supernatant and cellular RNA is similar and low. The quantification of modifications from cell culture supernatant of infected cells revealed the abundance of 4.9 $m^{66}A$ per viral strand, which is an rRNA marker. Further, the abundance of ribose methylated modifications is also comparable to the control samples. Two scenarios might explain this observation (a) this might be an indicator for the presence of released host RNA in the cell culture supernatant and its co-enrichment, due to unspecific binding, which distorts the abundance of viral modifications (b) under consideration of the comparability of canonical amounts in the sample enriched from cell culture supernatant of infected cells to the control samples, the results might rather be attributed to unspecific binding than representation of viral modification density.

In Figure 16F, the quantities of modifications per viral RNA from infected samples are displayed. Besides $m^{66}A$, the modifications m^1A, m^1G, $m^{22}G$ and m^3C could be identified to an extent of up to 38 modifications per viral strand. These modifications are reported to be present in tRNA and rRNA[23], which again might be an indicator for host RNA in the enriched sample. However, this line of argument may be countered by the absence of Ψ, which is a highly abundant modification in eukaryotes[64, 261, 262] (Figure 16F).
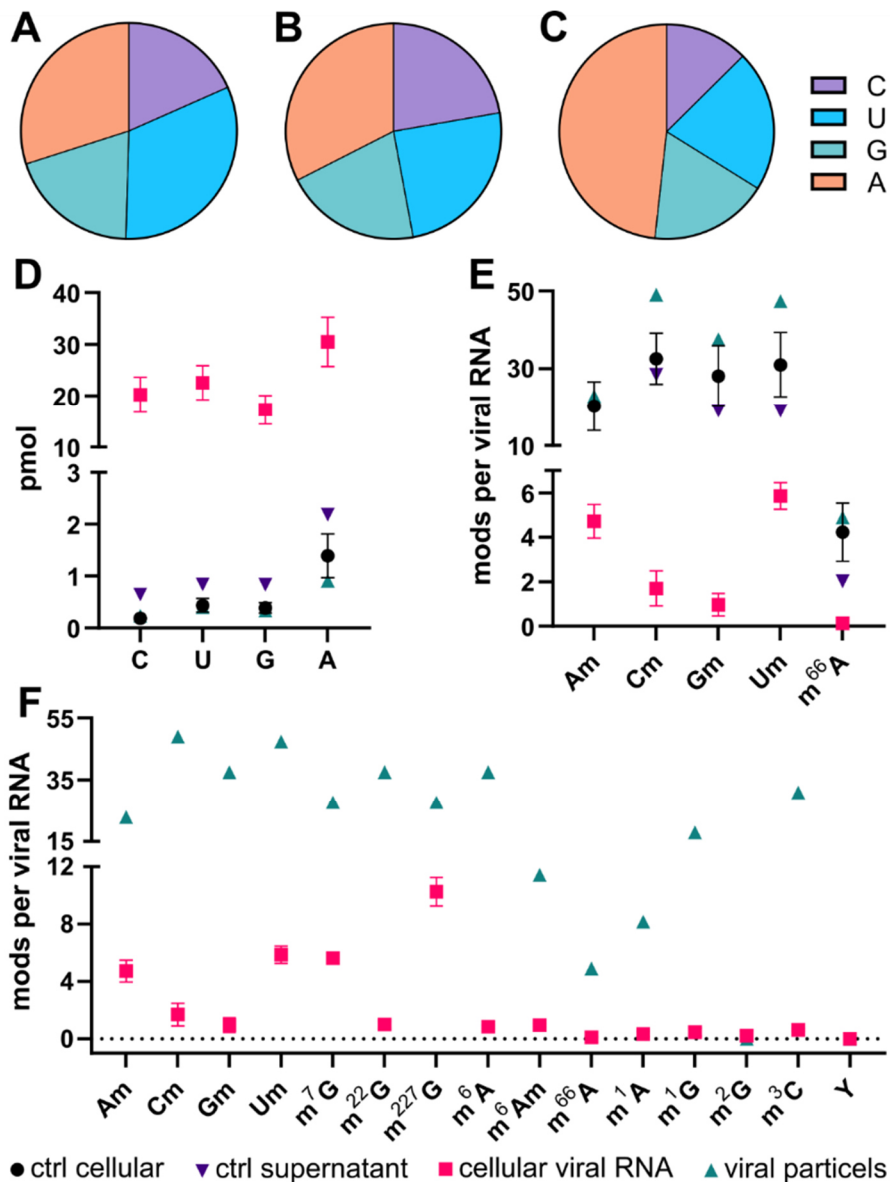
Figure 16: Canonical nucleoside and modification abundance in samples enriched for sequence of SARS-CoV-2 (A) pie chart of canonical ratios of published SARS-CoV-2 sequence[30] (B) pie chart of canonical ratios in enriched samples from cellular RNA (C) pie chart of canonical ratios in enriched samples from cell culture supernatant (D) total amounts of canonical nucleosides enriched from cellular RNA of uninfected cells (ctrl cellular, black), enriched from cell culture supernatant of uninfected cells (ctrl supernatant, purple), enriched from cellular RNA of infected cells (cellular viral RNA, pink) and enriched from viral particles from cell culture supernatant (viral particles, green) (E) number of modifications per viral RNA in samples enriched from cellular RNA of uninfected cells (ctrl cellular, black), enriched from cell culture supernatant of uninfected cells (ctrl supernatant, purple), enriched from cellular RNA of infected cells (cellular viral RNA, pink) and enriched from viral particles from cell culture supernatant (viral particles, green) (F) number of modifications per viral RNA in samples enriched from cellular RNA of infected cells (cellular viral RNA, pink) and enriched from viral particles from cell culture supernatant (viral particles, green); enrichment from supernatant n=1, enrichment from cells n=3, error bars reflect standard deviation

The aforementioned typical rRNA and tRNA modifications were not, or just to a really low extent, detected in the samples enriched from infected cells. This might be interpreted as the absence of host RNA. The modification abundance in these samples, is remarkably lower than the initially determined values depicted in table 1. Most of the

base methylated modifications occur 1 – 5 times per viral RNA. Interestingly, the cap modification $m^{227}G$ was quantified to an extent of 10 times per viral RNA, however each viral RNA is decorated with only one cap per strand and the incorporation as internal RNA modification was not reported yet. The hybridization probes consist of 55-mers along the genomic sequence of SARS-CoV-2. Hence there might be hypotheses to explain the obtained results. On the one hand, the short hybridization probes along the sequence, might not enrich only full length genomic RNA, but also capped subgenomic RNAs. These subgenomic RNAs are shorter and less structured than the full length genomic RNA, and thus might be more accessible to the hybridization probes, and this might even favor the enrichment of subgenomic RNA over the genomic viral RNA. On the other hand, especially fragments from the 5' end of the viral RNA might be enriched and are giving rise to the high quantities of this cap modification in relation to the whole viral RNA. An additional indication for this assumption is the absence of Ψ in the analyzed samples (Figure 16F). While Fleming *et al*. found at least 5 modification sites of Ψ at the 3'UTR[258], there was no evidence for the presence of Ψ in these enriched samples (Figure 16F).

In sum, it appears that the unspecific binding of the enrichment is low and can be neglected. Further, it is suggested that the RNA enriched from viral particles containing cell culture supernatant, is either not concentrated enough, so also here unspecific binding is prevailing over viral RNA enrichment or the sample integrity is distorted by released host RNA that is attributing to the quantification due to unspecific binding. The highest amounts of RNA could be enriched from cellular RNA of infected cells, however here the sample integrity or rather the length of enriched RNA needs to be inspected for further conclusions. Future work should consist of identifying the length of enriched RNA or even the quantity of certain sections of the sequence by RT-qPCR. Moreover, the purity of enriched samples should be validated with RNA-Seq to exclude the presence of host RNA populations.

From the compilation of modification densities in this study and the comparison to published data, it is evident that distinct conclusions on the modification abundance of SARS-CoV-2 are difficult to draw. The absolute amounts are highly dependent on the cultivation and infection conditions and as a result of my work it is demonstrated how crucial sample purity and integrity is for valid analysis. Nevertheless, this study is the

first broad trial to profile and compare simultaneously various modifications in the mutants and variants of SARS-CoV-2. The analysis with LC-MS/MS allows the consideration of a large number of different modifications at once, which is a strong advantage in contrast to *e.g.* sequencing, where either the identity of the modification remains elusive or solely some limited number of modifications can be detected. Further, this approach benefits from the high detection sensitivity, which even allows the quantification of low abundant modifications. Moreover, the quantitative approach with LC-MS/MS revealed the distorted composition of canonical nucleosides with the initial extraction method and thereby identified this common extraction method as source of error for future investigations. Wu *et al.* who first reported the full length sequence of SARS-CoV-2, purified the viral RNA from patient samples with a commercially available kit (RNeasy Plus Universal Mini kit, Qiagen), followed by rRNA depletion and thereby reached homology of >99.9%.[30] Indeed, the rRNA marker modification m^{66}A was determine in the samples, purified by sucrose density gradient centrifugation, hence for future work it is advice to implement a rRNA depletion. The enrichment of viral RNA with hybridization probes, appeared to be promising, however the sample integrity has to be validated by RNA-Seq to exclude the presence of host RNA population and the length and abundance of enriched sections should to be determined with RT-qPCR.

In conclusion, additional work is required before a complete understanding of the viral genome can be reached. Future studies should focus on optimization of RNA extraction from SARS-CoV-2 and establish a reliable and easily applicable method for verification of sample purity and identity.

3.2.4 Enrichment of viral RNA sequences with biotin enriched mung bean assay

In SARS-CoV-2 the 3'-UTR stem loop 1 (SL1) (nts 29548-29613) is reported to be highly structured.[263] Attributed to the contribution of modifications on structure stabilization, it was anticipated that this section might have a high modification density. The SL1 is a large and bulged SL with a U rich loop. The stem of SL1 can form a pseudoknot with the loop of the adjacent SL2.[263] It is suggested that the SL1 and SL2 element can act as a putative RNA switch involved in the regulation of replication.[264, 265] It is summarized that this section is building the most important structural element at the 3'-UTR.[263] The identification of modifications within this specific region is of particular interest, as these modifications may have an impact on structure and

function of this RNA section. This certain section of the viral RNA should be enriched with a reverse complementary DNA oligonucleotide. This procedure does not only introduce an additional assurance of specificity but also allows the determination of modifications in a sequence dependent manner. In brief, a biotin tagged DNA oligonucleotide reverse complementary to the sequence of interest is hybridized to the target RNA. Subsequently this DNA-RNA hybrid was incubated with mung bean nuclease, hydrolyzing all single stranded RNA and leaving intact the hybridized section only. This remaining section was purified using streptavidin coated magnetic beads. After dissociation of the hybrid, the RNA section of interest is released and hydrolyzed to single nucleosides for quantification with LC-MS/MS.

Prior to analysis of viral RNA, the method was established and validated and its accuracy was confirmed by the analysis of the section nts 76 - 126 of 18S rRNA from *S. cerevisiae*. As it can be seen in figure 17A, the calculated quantities are in accordance with literature.[266] For analysis of the SL1 in RNA of SARS-CoV-2, a reverse complementary DNA oligonucleotide, ranging from nts 29548 to 29631, was designed. This oligonucleotide was hybridized to RNA isolated from the mutants D614 and G614, provided by Prof. Denisa Bojkova. The section of interest was successfully enriched and the number of nucleosides in this section was quantified (Figure 17B). The number of canonical nucleosides are comparable to literature[30]. Nevertheless, there was no evidence for modifications in this section.

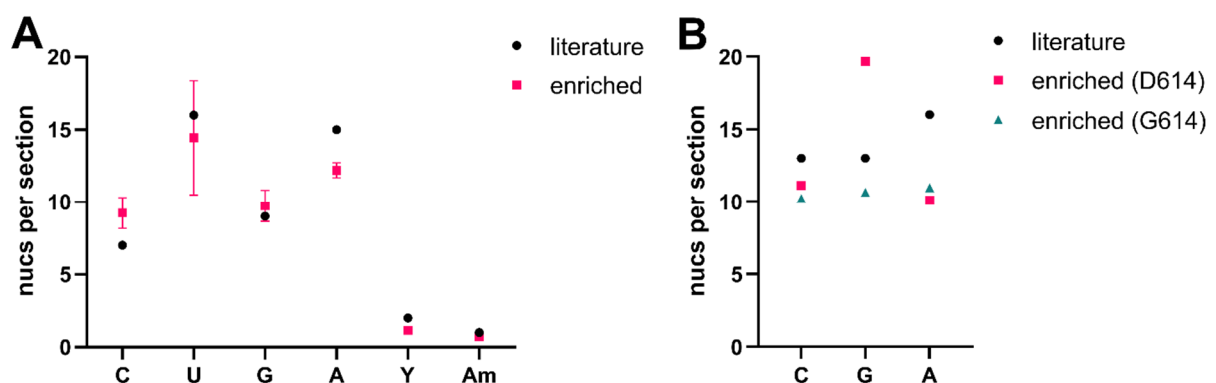Figure 17: Quantification of nucleosides from specific RNA section (A) nts 76-126 from 18S rRNA of *S. cerevisiae*, literature[266] (black), enriched (pink), (B) nts 29548-29613 of SARS-CoV-2, literature[30] (black) of variant D614 (pink) and G614 (green), Uridine was excluded from the quantification due to bad peak shape, *S. cerevisiae* n=3, error bars reflect standard deviation, SARS-CoV-2 n=1

This chapter shed light onto several hindrances for viral RNA modification profiling and might offer an explanation for continuing incomplete modification profiling of the SARS-CoV-2 genome, despite numerous investigations worldwide. A comparison of modifications in the mutants and variants of SARS-CoV-2 is presented, with the limitation of unsatisfactory validation of sample purity. In order to eliminate distortions in modification abundances by presence of other RNA populations, two enrichment procedures and their outcome were discussed. However, the integrity of samples enriched with MagIC remains to be validated. Whereas, the biotin enriched mung bean assay enables target specific enrichment besides identification and quantification of modifications in a sequence dependent manner with LC-MS/MS.

3.3 LC-MS/MS method development for absolute quantification of RNA cap modifications

3.3.1 Chromatographic method development

3.3.1.1 Chromatographic separation on system for nucleosides analysis

To start with, the analytes' retention behavior was investigated on a known system. Therefore, the nucleoside monophosphates (NMPs) were separated with our chromatographic method for nucleoside analysis. In contrast to nucleosides, NMPs possess an additional phosphate ester group and thus a negative charge at pH 5.3. It was taken into account that the additional negative charge might impact the interaction with the stationary phase and thus the retention may differ from the nucleosides' retention.[243] However, the retention of NMPs should be assessed in comparison to nucleosides, whose retention behavior is well studied on this system. Briefly, the system was equipped with a RP C18 column (Synergi Fusion-RP, Phenomenex) in combination with a binary mobile phase of 5 mM NH_4OAc aqueous buffer (pH 5.3) and acetonitrile as organic component. A gradient elution was conducted beginning from 0% organic phase for 1 min and increasing to 40% over time, followed by a re-equilibration to 0% organic phase (Figure 18).

Indeed, the elution order of the NMPs was the same as for nucleosides in this set up. Nevertheless, the separation of CMP, UMP, GMP was insufficient with selectivites of 1.05 and 1.02, while this method usually yields selectivities of at least 1.5 for canonical nucleosides. As aforementioned, the negative charge may influence the interaction with the non-polar RP stationary phase. The hydrophobic stationary phase of RP columns has a strong affinity towards hydrophobic or less polar compounds, while polar compounds are retained insufficiently or not at all. This effect can be clearly seen here (Figure 18), while the canonical nucleosides elute between 1.72 min and 5.17 min, three of four NMPs elute within 1.2 min. NMPs are less retained, attributed to the additional phosphate group and its negative charge and thus to the higher polarity in comparison to the nucleosides. The cap analogue GpppA, with a triphosphate and thus three negative charges at pH 5.3, could not be retained at all with this system (data not shown here). These highly polar analytes were not sufficiently retained on this column, as a consequence this approach was discarded. A more suitable stationary phase had to be selected.
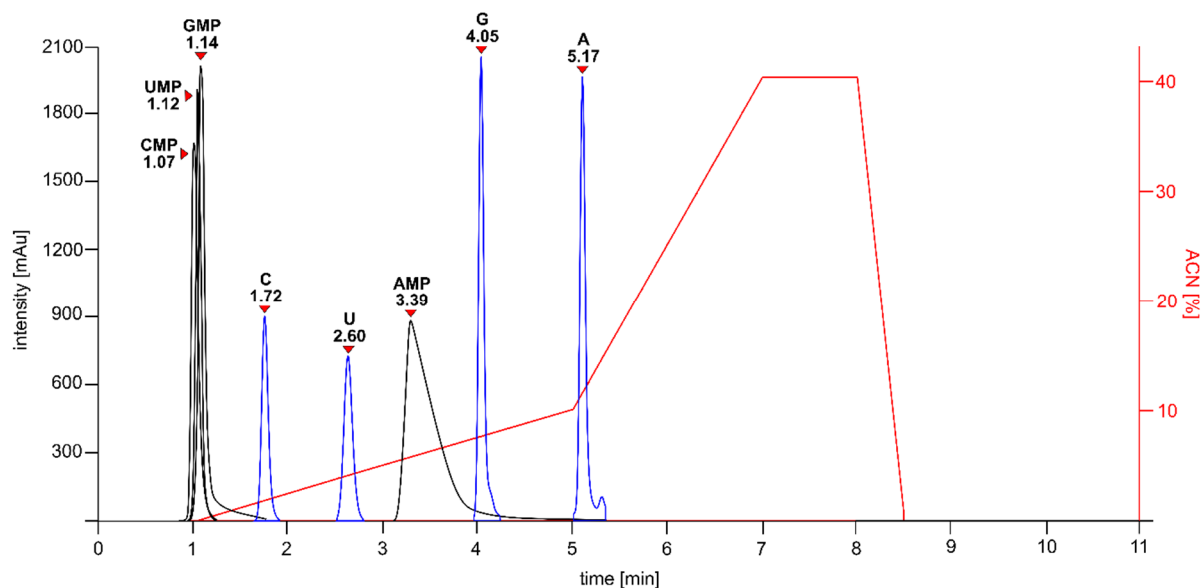
Figure 18: UV chromatogram of nucleosides (blue) and nucleoside monoposphates (black), system equipped with RP C18 column (Synergi Fusion-RP, Phenomenex), 5 mM NH₄OAc aqueous buffer (pH 5.3) and acetonitrile, peak identity and t_R labeled above, λ=260 nm

3.3.1.2 Selection of stationary phase

Usually, nucleotides are separated on RP systems with the aid of ion pairing reagents. [267-270] However, this approach was initially developed for UV detection and is disadvantageous for electrospray ionization (ESI) and mass spectrometric analysis. Most ion pairing reagents are not volatile and thus cause ion suppression during ESI.[271] Moreover, the device is persistently contaminated with the ion pairing reagent and the method's reproducibility is limited by often occurring retention time shifts.[272] Another possibility might be the separation by hydrophilic interaction liquid chromatography (HILIC) with a polar stationary phase, which is highly suitable for separation of polar analytes and subsequent mass spectrometric analysis. HILIC excels in lower back pressure and higher sensitivity for ESI-MS in opposition to RP chromatography. For HILIC organic solvent rich mobile phases are used, while mobile phases for RP chromatography consist of aqueous buffers for the most part. Organic mobile phases are less viscous than aqueous mobile phases, and result in lower back pressure, which allows for higher flow rates. Moreover, the organic solvent rich mobile phase is more volatile than aqueous mobile phases and hence favors the formation of gas phase ions during ESI and thereby ameliorates ionization efficiency. Despite its good performance in separation of polar analytes and the high ionization efficiency, it was decided against the usage of a HILIC column as it would not match the desired features of the method. Previously, it was shown that highly concentrated buffers have to be used to elute

nucleotides from a HILIC column[273], while high salt concentrations are not recommended for ESI. It leads to increased conductivity, which results in impaired ionization.[271] Moreover, separation on HILIC columns are known to be less reproducible, little changes in mobile phase composition or the analytes solvent can have major impact on retention time and peak shape. In the method to be developed, an external calibration, prepared in water and samples hydrolyzed in reaction buffer should be analyzed. Thus the developed method needs to be robust enough to attenuate the arising fluctuations from differences in calibration and sample solvent. Further, the goal consisted of developing a short method, but the equilibration of HILIC columns usually takes two times longer than of RP columns.[274] Additionally, HILIC columns are known to have a rather low loading capacity[274], while it was planned to inject larger sample amounts in order to reach the limits of quantification for cap modifications, as each RNA molecule only contains a maximum of one cap modification.

Next, the trifunctional RP C18 column (ACQUITY HSS T3, Waters, Milford, USA) was taken into consideration. This RP column is reported to separate nucleotides, despite their structural similarity.[275] According to the vendor, this column enables maximum retention of polar analytes due to reduced C18 occupancy. This allows an extended access for polar analytes to silanol groups and thereby ameliorates the interaction capacity to enhance the retention.[274] The column is suggested for separation of polar analytes over a pH range from 2 to 8. This column is offered with a pore size of 100 Å, particle size of 1.8 µm and inner diameter of 2.1 mm, which are common dimensions of a column for LC-MS application, the column length is available in 50, 75, 100 and 150 mm. It was decided to utilize a long column, thus the 150 mm, in order to aim for a high theoretical plate number and thus a better resolution.[276]

3.3.1.3 Investigation of mobile phase

After the column was decided on, the parameters of buffer pH, temperature, flow rate and ion strength should be investigated in order to develop a method, that is optimized for short analysis time and high detection sensitivity.

In first instance, the pH of the mobile phase was selected, as it has a major impact on the chromatographic separation, especially for polar analytes. The retention of polar analytes on an RP system is influenced by their degree of ionization, ionized analytes are less retained than their neutral forms.[276] If the pH is selected in proximity (±1.5) to

the analyte's pk_a, the analyte is going to be ionized only partly, which accounts for instable retention times and broad peaks.[276, 277] It is advised to select the pH at least ±2 from the analytes' pk_a, in order to obtain analytes of a uniform charge state.[276] Additionally, it should be considered that also the state of the stationary phase can vary at different pH and this also influences the analytes' retention.[277] Further, it was planned to couple the chromatographic separation with ESI-MS. The ionization of the analyte in the aqueous phase is relevant for the ionization during ESI.[271] Hence, the selected pH will impact the retention as well as the ion mode in the subsequent mass analysis.

Aqueous buffers of different pH were tested (pH 2.7, 4, 5, 6, 6.9). The buffer with pH 2.7 was prepared by adding 0.1% formic acid to 1 L water. The buffers with pH 4 to 6 were prepared with 5 mM NH_4OAc and the addition of acidic acid for pH adjustment. The buffer with pH 6.9 was prepared with 5 mM NH_4OAc and the addition of ammonia for pH adjustment. This pH range was selected on the basis of the given pH limits for the column and the suitability of buffer systems for LC-MS analysis. Commonly, LC buffers with low pH are prepared with trifluoracidic acid (TFA) resulting in a pH of 2, but TFA acts as ion pairing reagent and therefore does not fulfill the requirements for this method. Apart from TFA, citric acid can be used to produce buffers at low pH, but it was refrained from this approach because of its high UV-cutoff.[277] Even though, the analytes should be detected by MS, the UV detection should be acquirable for method development and troubleshooting. Buffers with pH higher than 6, are usually phosphate buffers, which are not volatile and thus incompatible with MS analysis. Further, it was decided not to exceed the pH of 6.9, under consideration of the buffer capacity of NH_4OAc and the pH limit of the column.

Acetonitrile was utilized as organic eluent. It is an aprotic solvent, highly suitable for RP systems. Its lone electron pair can build hydrogen bonds with silanol groups of the stationary phase and thus withdraw the possibility for the analyte to interact with silanol groups and consequently reduces their retention.[274] Moreover, its eluting strength was shown to be practical for separation of NMPs and cap analogues on RP systems.[275, 278] Further, acetonitrile is a highly suitable organic solvent for ESI ionization, attributing to good ionization efficiency, due to its low vaporization enthalpy.[279] Additionally, acetonitrile reduces the surface tension and thereby eases solvent evaporation.[280] The

improved solvent evaporation brings about noise reduction and thus increases the signal-to-noise ratio.[203]

The gradient was adapted from He *et al.* with slight adjustments to obtain a suitable separation of NMPs and cap analogues, under consideration of short analysis time.[275] The gradient is provided in table 2. 50 pmol of each analyte was injected. The selected parameters were the same for all analyses, and are given in table 3. The resulting UV chromatograms are displayed in figure 19.

Table 2: Gradient, time is given in minutes, ratio of buffer B (acetonitrile) is given in [%]

time [min]	buffer B [%]
0-3	0
3-6	3
6-8	15
8-10	15
10-13	60
13-14	0
14-15	0

Table 3: selected parameters for investigation of suitable pH

Column	Acquity HSS T3, 100 Å, 1.8 μm, 2.1 mm x 150 mm
Temperature	20 °C
Flowrate	0.25 ml/min
Detection	UV (λ=260 nm)
Organic Phase	acetonitrile
Aqueous Phase	0.1% FA, pH 2.7/ 5 mM NH_4OAc pH 4/ 5/ 6/ 6.9

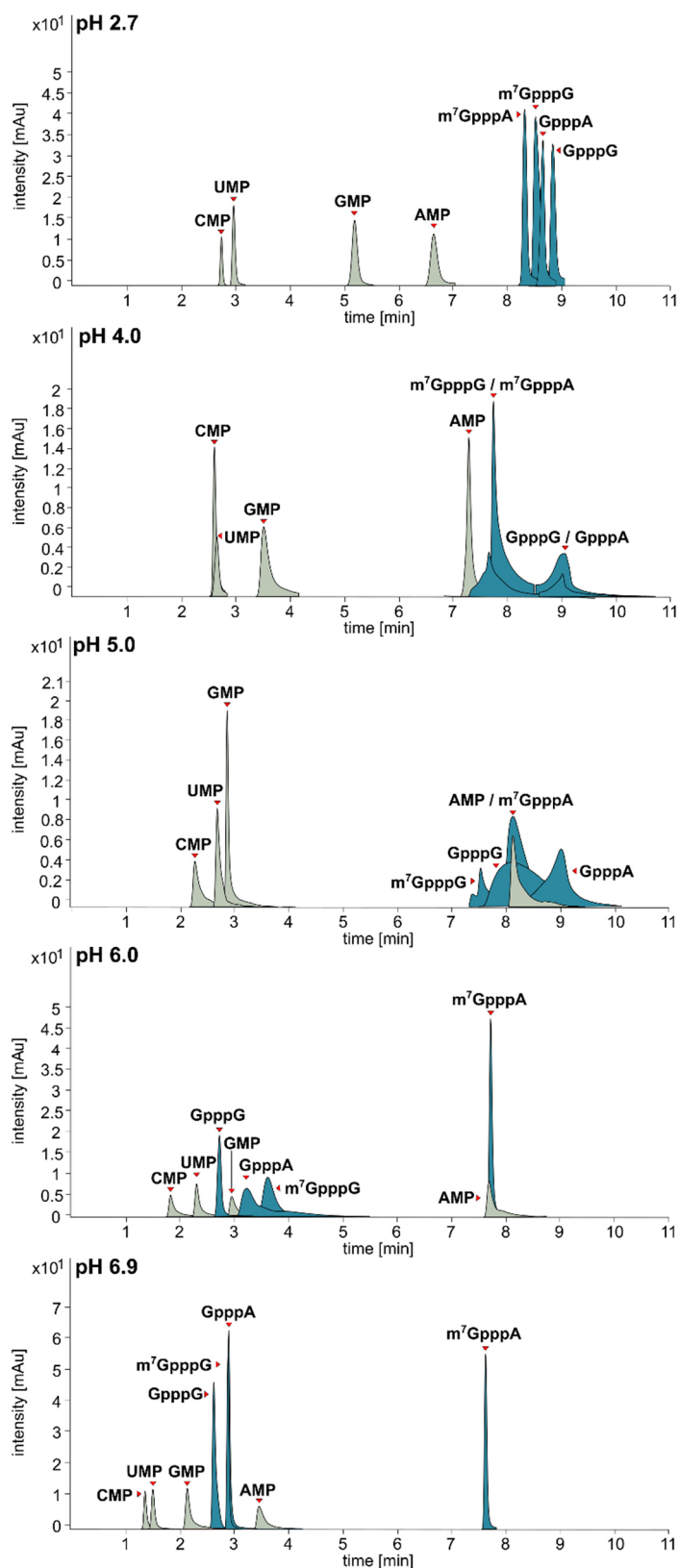Figure 19: UV chromatograms at different pH, conditions can be taken from upper left corner of chromatograms, peak identity is labeled above, λ=260 nm

In contrast to the initial approach (Figure 18), all analytes were sufficiently retained on the new column (Figure 19). As anticipated, the pH had a great impact on the retention of the analytes (Figure 19). In order to evaluate the results and assess the retention

behavior of the analytes, the retention factor (k) in dependence of the pH was plotted. Ideally, retention factors should be between 2 and 8 and should be divers amongst the analytes.[198] The retention factor (k) was calculated according to equation (1), where k is the retention factor, t_R the retention time and t_0 the void time, which was previously determined to be 1.4 min. The resulting retention factors for every pH are illustrated in figure (20).

(1)
$$k = \frac{t_R - t_0}{t_0}$$
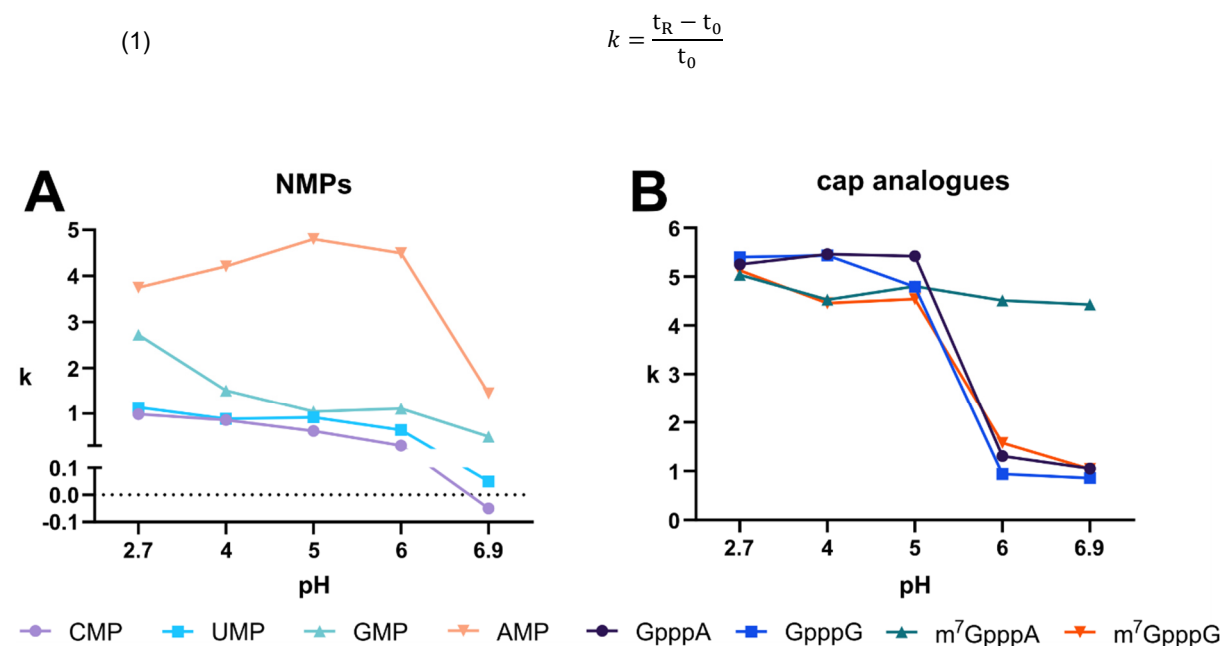
Figure 20: Plot of retention factor as function of pH (A) for NMPs CMP (purple), UMP (blue), GMP (green), AMP (orange), (B) for cap analogues GpppA (purple), GpppG (blue), m⁷GpppA (green), m⁷GpppG (orange)

The retention of CMP, UMP and GMP is slightly declining with increasing pH. As expected, the retention factors of CMP and UMP are similar, potentially attributed to similar retention mechanisms due to their structural resemblances. Interestingly, the retention of AMP is increasing until pH 5 and drops at pH 6 (Figure 20A). The retention of the cap analogues seems to be comparable at the different pH values, attributed to their similarity in number and type of functional groups. The retention factors of the cap analogues change barely from pH 2.7 to 5, but starting at pH 6 the retention factors are reduced significantly from 5 to 1, for GpppA, GpppG and m⁷GpppG. For m⁷GpppA, the reduction of the retention factor is less pronounced from 5 to 4.4 over the whole pH range.

As previously outlined and also evident from the results of this study (Figure 19, 20) the pH of the mobile phase has a great impact on the retention of ionic analytes. Nev-

ertheless, conclusion on effects of hydrophobicity, polarity and thus the retention behavior of multiply charged analytes as function of the pH are complex.[276, 277] The retention is dependent on the number and type of functional groups. Various mechanisms, classified in primary and secondary interactions contribute to the retention of an analyte. These mechanisms can be more or less pronounced based on the analyte's degree of ionization. In order to assess the charge state at a certain pH, one needs to know the pk_a of the functional group. pk_a values are usually determined in water, but the addition of organic solvents during the chromatography can change the pk_a of the analyte as well as the pH of the aqueous phase.[276] Despite the complexity, an estimation of ionization and selection of a suitable pH should be done in the following.

The analytes consist of nucleobases, riboses and phosphate esters. The structures and their pk_a values are displayed in Figure 21. The functionality of riboses with a pk_a of 12 can be neglected, as this is far from the investigated pH range. The pk_a of the phosphate ester groups is known to be 1, as a result the optimum for good chromatography should be pH = pk_a ± 2 = 3. The pk_a values of the nucleobases range from 3.5 to 9.8. On the basis of pH = pk_a ± 2 for all moieties, the optimal pH should be between 6.2 and 7.2.
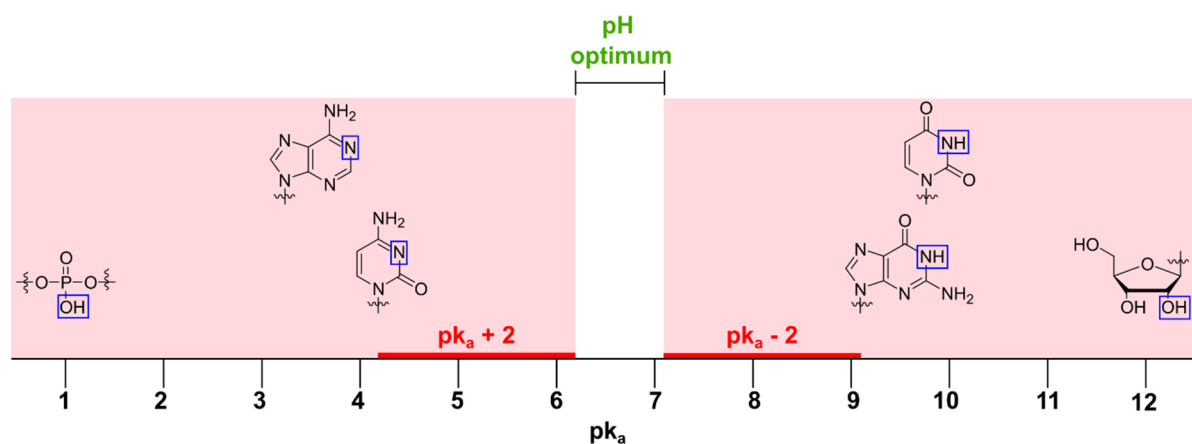
Figure 21: pk_a values of moieties and optimum pH range, respective functional groups are framed in blue, unsuitable pH range is highlighted in red, optimum pH range is indicated by line on top

In summary, it is difficult to estimate the retention behavior of the analytes as multiple ionizable groups are present and the organic solvent can change the pH of the mobile phase as well as the pk_a of the analyte. In such cases the attribute of the molecule as entity, can be inferred from the inflection point of retention factor as function of the

pH.[277] The inflection point for GpppA, GpppG and m^{7}GpppG can be clearly seen between pH 5 and 6 in Figure 20B. The aforementioned peak broadening, as a cause of partly ionized analyte molecules is also evident at pH 4 and 5 for the cap analogues, which are in the predicted range of ±1.5 of the inflection point (Figure 20B). In order to further validate the peak shape, the asymmetry factor was investigated. The asymmetry factor is described by the ratio of the distance between the leading and trailing widths at 10% maximum to the center line at the peak maximum (Equation 8). An asymmetry factor of 1 is describing an ideal gaussian-shaped and symmetric peak, whereas if the value is <1 fronting is faced and if this value is >1 tailing is observed.

$$(8) \qquad\qquad\qquad As = \frac{B}{A}$$

In Figure 22A, it can be seen that the asymmetry factor is generally increasing with the pH and peaks of NMPs are tailing. The fronting of peaks and hence asymmetry factors of <1 were not determined for NMPs. In Figure 22B, the asymmetry factors of cap analogues are displayed. At pH 2.7 tailing was observed for all cap analogues. Whereas, fronting was determined at pH 4 for GpppA and GpppG, as well as for m^{7}GpppG at pH 5, while tailing was determined for the remaining cap analogues with a maximum factor of 9.6 for m^{7}GpppA at pH 4. Also, at pH 6 strong tailing was determined for GpppA with a factor of 7.5, while the tailing for the other cap analogues was less pronounced. At pH 6.9 the asymmetry factors deviate the least from the ideal value of 1, slight tailing was observed here. In combination with the illustration of the peaks in the chromatograms (Figure 19), it evinces that with pH at 4, 5 and 6 symmetries strongly deviate from the ideal gaussian-shaped peak, while the largest compliance to symmetric peak shape was observed for pH 6.9.
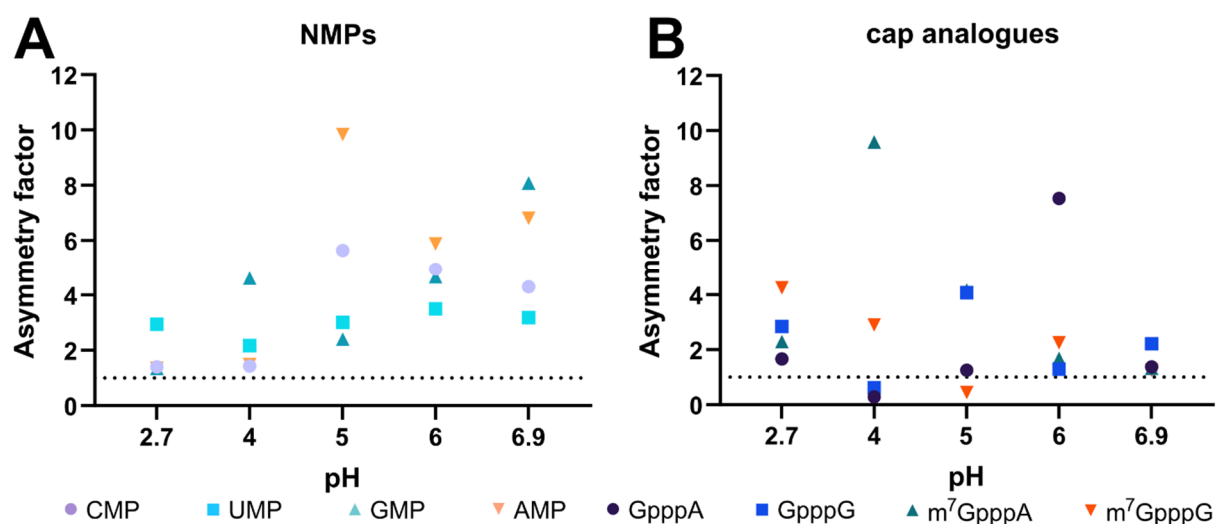
Figure 22: Asymmetry factors at different pH for (A) NMPs and (B) cap analogues, asymmetry factor of 1 for ideal gaussian-shaped peaks is indicated by dotted line

After the examination of the retention factor and peak symmetry at different pH, the resolution should be investigated to find a suitable pH for the aqueous phase. Therefore, it should be noted that the elution order of the cap analogues is changing over the investigated pH range (Figure 19). Hence, peaks were defined as peak pair 1 to 7 in the following examination of resolution (Figure 23). The elution order of ionic analytes on RP system can change due to alterations in mobile phase composition.[277] The applied chromatography is a gradient method, the ratio of organic solvent is changing over time. As previously shown, the retention factors of the analytes vary at different pH values. Along with the changes in retention, the solvent composition varies at the different retention times, which may explain the changing elution order. A resolution of at least 1.5 yields in base line separation of two adjacent peaks and is assumed to be ideal, if the resolution is lower the separation is insufficient. The resolution was determined according to equation 6, where R_s is the resolution, t_R is the retention time in min and w_h is the peak width at the base line in min.

$$(6) \qquad R_s = \frac{1.18 * (t_R - t_R)}{w_{h1} + w_{h2}}$$

From Figure 23, it is apparent that the resolution does not meet the requirement of ≥1.5 for every peak at the investigated pH. The largest compliance was observed at pH 2.7, where the majority of NMPs are separated adequately amongst each other (peak pair 2 and 3) and from the cap analogues (peak pair 4). However, this method is not depended on the chromatographic resolution and thus the full base line separation to

identify peaks, as the chromatographic separation is followed by an MS analysis, where each analyte can be clearly identified by its specific mass transition.
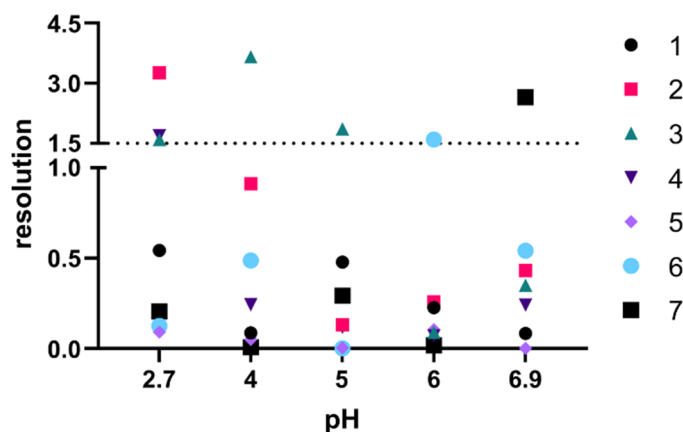
Figure 23: resolution of adjacent peak pairs 1-7 at different pH, the ideal resolution of 1.5 is indicated by the dotted line

In summary, pH 4, 5 and 6 were excluded, as pH ±1.5 around the pk_a or in case of multiply ionizable molecules, around the inflection point of k as a function of the pH, are not considered to be suitable for robust retention times and symmetric peak shape and the inflection point for most cap analogues was observed between pH 5 and 6. Strong deviations from ideal peak shape were determined at pH 4 and 5. Overall, the resolution was below 1.5 for the majority of peak pairs over the tested pH range. Especially, at pH 6 the resolution of peak pairs was lower than at other pH. At pH 2.7 the resolution for 3 peak pairs was ≥1.5 and acceptable peak shapes were obtained. At pH 6.9 the peak shapes of cap modifications were the closest to ideal. Even though, pH of 2.7 yielded retention factors in the recommended range of 2-8[198] for the majority of analytes, it was not considered further because of the high retention factors for cap analogues (5-6) and resulting late retention times. Here, pH of 6.9 was favored, which is within the optimal pH range (Figure 21), results in good peak shape (Figure 22) and matches the desired reduced retention times for short analysis (Figure 19). The retention time of m⁷GpppA was not appreciably altered by changing pH and is going to be adjusted via temperature and flow rate in the following sections.

3.3.1.4 Investigation of column temperature

As outlined in the introduction, one can differentiate between partition and adsorption chromatography. As partition chromatography is based on the variances in solubility of analytes in the two immiscible liquid phases and partitioning of analytes between these phases, the temperature influences the solubility and thus the separation. According

to the adsorption theory, the adsorption decreases with increasing temperature. Further changes in temperature can have an effect on (a) the pH of the mobile phase (b) the pk_a of the analyte, and thus the degree of ionization and (c) the interaction of the ionized analyte with silanol.[277] In conclusion the impact of temperature is more distinct for ionized samples than for neutral samples. The temperature can impact the retention, the selectivity and it can improve the resolution of peaks, by reducing peak width.[277] According to the principle of Le Chatelier, temperature increase shifts the equilibrium towards the mobile phase.[274] In conclusion, temperature increase results in reduced retention, whereas the reduction of temperature can be a tool for adjustment of peak selectivity. However, optimizing the selectivity via temperature is limited to analytes that differ in their chemical attributes and the therewith associated retention mechanisms.[274] In this analysis the retention mechanism of NMPs among each other and cap analogues among each other are assumed to be similar, due to their structural similarity. Here, the goal consisted of reducing retention times while avoiding co-elution. The retention of the analytes at common column temperatures for LC-MS analysis (20 °C, 30 °C, 40 °C) were compared. The selected parameters were the same for all analyses, and are given in table 4. The resulting UV chromatograms are given in Figure 24.

Table 4: selected parameters for investigation of suitable temperature

Column	Acquity HSS T3, 100 Å, 1.8 µm, 2.1 mm x 150 mm
Temperature	20 / 30 / 40 °C
Flowrate	0.25 ml/min
Detection	UV (λ=260 nm)
Organic Phase	acetonitrile
Aqueous Phase	5 mM NH_4OAc, pH 6.9
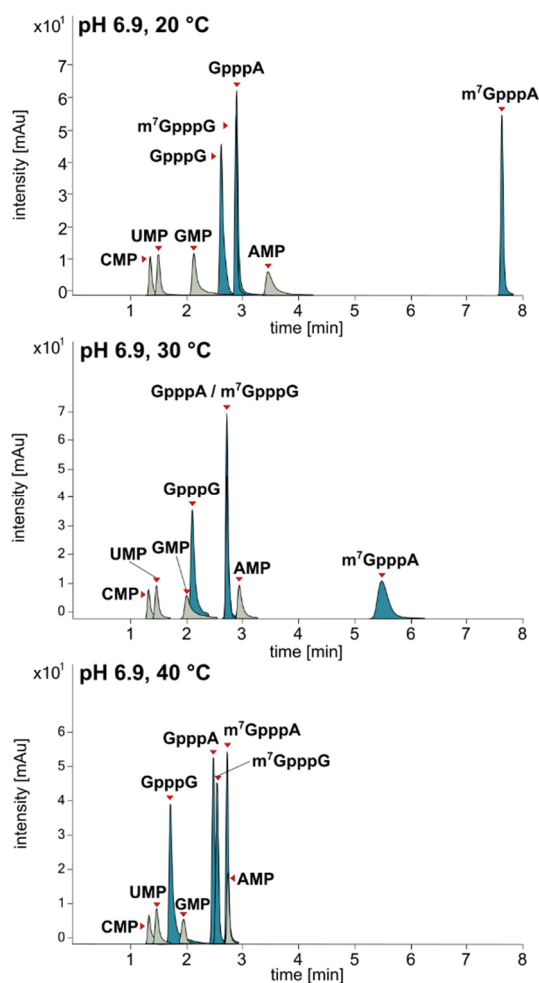
Figure 24: UV chromatograms varying column temperature, conditions can be taken from upper left corner of chromatograms, peak identity is labeled above, λ=260 nm

In line with the expectations, the retention times decreased with increasing temperature (Figure 24). The retentions dependency on temperature should be evaluated by visualizing the data in a Van't Hoff plot, with the logarithmic values of the retention factors and the reciprocal temperature in Kelvin (K) (Figure 25). The Van't Hoff plot allows to draw conclusions on similarities in retention mechanisms, by parallelism of slopes and changing selectivities, which can be discerned by intersecting slopes.[271]

CMP was excluded from this analysis, as its retention time was <1.4 min and therefore, eluting with the void volume. The slopes for NMPs are almost parallel to the x axis (Figure 25A), indicating that the temperature has no impact on retention. Further, the slopes are close to be parallel (Figure 25A), which is a sign of equal retention mechanisms[274], and might be attributed to the structural similarity of NMPs. The elution order of GMP and GpppG changes over the examined temperatures. A change in elution order, as a consequence of temperature change is commonly observed for adjacent peaks.[274] The slopes of GpppG and m^{7}GpppA are negative, meaning that the retention

is decreasing with higher temperature (Figure 25B). Especially, the retention factor of the last eluting analyte m[7]GpppA is notably reduced at 40 °C. The slopes of GpppG and m[7]GpppA are running in line, again indicating similarity of underlying retention mechanisms.[274] The slopes of the critical peak pair GpppA and m[7]GpppG seem to be congruent. But if contemplated closer, the slopes are intersecting at 30 °C (Figure 25B, right site), where the analytes co-elute (Figure 24) and show a changed elution order at 20 °C and 40 °C. While this critical peak pair was initially hardly separated, the largest variance in retention factor was determined at 40 °C (Figure 25B, right side).
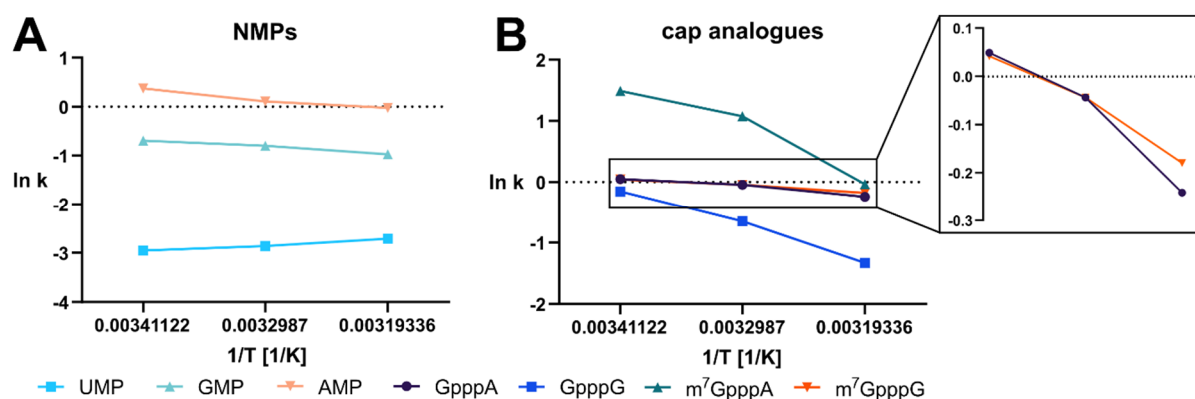
Figure 25: Van't Hoff Plot at 20 °C (0.00341122 1/K), 30 °C (0.0032987 1/K), 40 °C (0.00319336 1/K) (A) for NMPs, UMP (blue), GMP (green), AMP (orange) (C) for cap analogues, GpppA (purple), GpppG (blue), m[7]GpppA (green), m[7]GpppG (orange)

In order to investigate the variances in peak separation attributed to temperature, the resolution of adjacent peak pairs at different temperatures were determined according to equation 6, where R_s is the resolution, t_R is the retention time in min and w_h is the peak width at the base line in min.

$$(6) \qquad R_s = \frac{1.18 * (t_R - t_R)}{w_{h1} + w_{h2}}$$

As the elution order varies, peaks were designated as peak pair 1 to 7. It was assumed that resolutions might deteriorate with increasing temperature as retention times were expected to be reduced. In general, this assumption was confirmed, the ideal resolution of ≥1.5 was achieved solely for the last peak pair at 20 °C and 30 °C (Figure 26). However, only at 40 °C separation of GpppA and m[7]GpppA (peak pair 5) was achieved with a resolution of 0.17, in contrast to 30 °C and 20 °C, where the resolution is 0 and 0.017 respectively. This result was already indicated by the comparison in the Van't Hoff plot (Figure 25, right side). The retention times of NMPs are affected marginally by increasing temperature, whereas the retention times of the cap analogues could be

distinctly reduced with increasing temperature. Especially, the retention time for the last eluting analyte m^7GpppA was decreased from 7.60 min to 2.75 min.
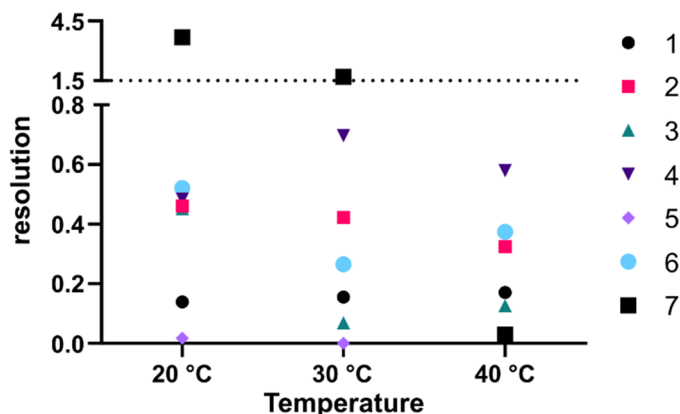
Figure 26: resolution of adjacent peak pairs 1-7 at 20 °C, 30 °C and 40 °C (red), the ideal resolution of 1.5 is indicated by the dotted line

From Figure 24, it is evident that the retention factors of cap analogues converge with increasing temperature. Concomitantly, the resolution is being reduced (Figure 26). Nevertheless, it was decided to continue with 40 °C, in favor of reduced retention times, since the resolution is not meeting the requirements at any temperature for the majority of the peaks. Moreover, this method is not depended on complete separation and hence on resolutions ≥1.5. The analytes can be identified by their specific mass transitions in the subsequent mass spectrometric analysis.

3.3.1.5 Investigation of flow rate

Next, it was aimed to optimize the plate number (N). This is an analyte specific measure for the separation efficiency of the set up (column, elution method, system) under consideration of the peak dispersion. High plate numbers are considered to yield high efficient separations. The plate number can be influenced by column length, particle size, temperature and flow rate.[274] As the dimension of the column is fixed and the temperature was already optimized in favor of reduced retention times, it was aimed to further reduce retention times by adjusting the flow rate. In general, the flow rate should be selected on the basis of good peak symmetries while maintaining their separation, under consideration of short retention and low backpressure.[274] The selected parameters were the same for all analyses, and are given in table 5. The resulting UV chromatograms are displayed in Figure 27.

Table 5: selected parameters for investigation of suitable flow rate

Column	Acquity HSS T3, 100 Å, 1.8 µm, 2.1 mm x 150 mm
Temperature	40 °C
Flowrate	0.15/ 0.25/ 0.35/ 0.45 ml/min
Detection	UV (λ=260 nm)
Organic Phase	acetonitrile
Aqueous Phase	5 mM NH$_4$OAc, pH 6.9

From Figure 27, it is evident that the retention times and peak widths are distinctly reduced with increasing flow rate but also the peak separation is impaired. The van Deemter equation was taken into account to judge the separation efficiency in dependency of the flow rate. This equation evaluates the plate height (H) as function of the flow rate, under consideration of peak width. Prior to determination of the plate height, the plate number needs to be determined according to equation 3, where N is the plate number, t_R is the retention time in min and $w_{0.5}$ is the peak width at half maximum in min. The values can be taken from table 6 and 7.

$$(3) \qquad N = 5.54 * \left(\frac{t_R}{w_{0.5}} \right)^2$$

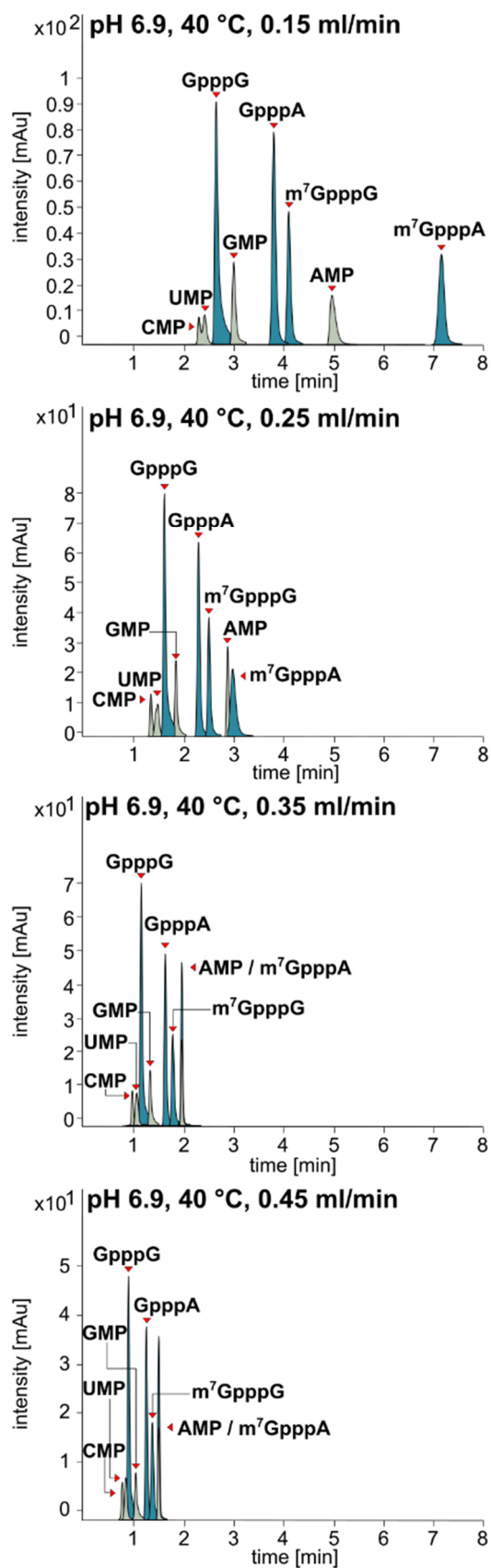Figure 27: UV chromatograms varying flow rate, conditions can be taken from upper left corner of chromatograms, peak identity is labeled above, λ=260 nm

Table 6: Values for determination of plate number for NMPs, flow rate [ml/min], (t_R) retention time [min], ($w_{0.5}$) peak width at half maximum [min], (N) plate number

flow rate	CMP			UMP			GMP			AMP		
	t_R	$w_{0.5}$	N	t_R	$w_{0.5}$	N	t_R	$w_{0.5}$	N	t_R	$w_{0.5}$	N
0.15	2.18	0.062	6849	2.40	0.066	7326	2.99	0.073	9313	4.96	0.129	8190
0.25	1.34	0.049	4143	1.47	0.093	1384	1.83	0.051	7133	2.87	0.050	18253
0.35	0.97	0.045	2558	1.05	0.069	1290	1.32	0.048	4190	1.95	0.042	11979
0.45	0.77	0.045	1609	0.83	0.057	1183	1.03	0.050	2365	1.49	0.041	7346

Table 7: Values for determination of plate number for cap analogous, flow rate [ml/min], (t_R) retention time [min], ($w_{0.5}$) peak width at half maximum [min], (N) plate number

flow rate	GpppA			GpppG			m^7GpppA			m^7GpppG		
	t_R	$w_{0.5}$	N	t_R	$w_{0.5}$	N	t_R	$w_{0.5}$	N	t_R	$w_{0.5}$	N
0.15	3.79	0.082	11854	2.64	0.073	7246	7.15	0.132	16268	4.09	0.078	15255
0.25	2.29	0.058	8636	1.61	0.053	5112	2.97	0.091	5901	2.49	0.059	9867
0.35	1.62	0.05	5816	1.14	0.047	3259	1.95	0.042	11979	1.77	0.059	5003
0.45	1.25	0.047	3900	0.89	0.046	2060	1.49	0.042	7001	1.37	0.058	3077

The plate height (H) is the product of column length and the reciprocal plate number (N) and described by the equation (4), where L is the column length in mm (150 mm) and N is the plate number that was previously calculated.

$$(4) \qquad\qquad H = \frac{L}{N}$$

The optimal flow rate is determined at the lowest point of the van Deemter curve. Hence, a good efficiency is obtained if the plate height (H) is low, thus the theoretical plate (N) is high, as the column length is a static value here.[274] The resulting graphs are displayed in figure 28.

In general, the plate heights appear to be smaller with lower flow rates (Figure 28). As a result, flow rates of 0.35 ml/min and 0.45 ml/min were excluded from further analysis. The plate height for UMP is lower at 0.15 ml/min in comparison to 0.25 ml/min, whereas for AMP the plate height is the lowest at 0.25 ml/min. For CMP, GMP, GpppA, GpppG and m^7GpppG the plate heights are just slightly higher at 0.25 ml/min in comparison to 0.15 ml/min. An unexpected curve is observed for m^7GpppA. While the plate height of other cap analogues is increasing with the flow rate, here the plate height peaks at 0.25 ml/min. This can be explained by the peak width at this flow rate. Usually, the peak width is decreasing with rising flow rate. But here the peak widths are comparable at 0.15 ml/min and 0.25 ml/min (see Figure 27 and table 7), while the retention time is reduced more than half. Inserted in equation 3, this means that the plate number is

reduced to one fourth from 0.15 ml/min to 0.25 ml/min and thus the plate height increases. Just with 0.35 ml/min the peak width for m⁷GpppA was reducing and the curve declining again.
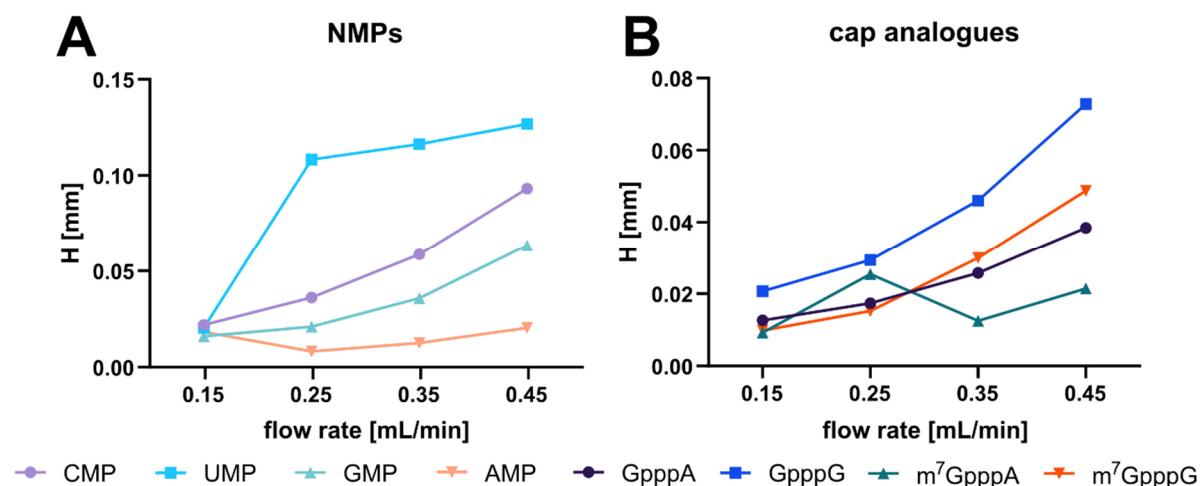
Figure 28: Van Deemter plot (A) for NMPs at different flow rates CMP (purple), UMP (blue), GMP (green), AMP (orange) (B) for cap analogues at different flow rates GpppA (purple), GpppG (blue), m⁷GpppA (green), m⁷GpppG (orange)

Further, it was already outlined that one needs to find a compromise between minimal plate height, sufficient separation and short analysis time. As evident from the chromatograms (Figure 27), the last analyte elutes at 7.15 min at 0.15 ml/min, whereas at 0.25 ml/min the last analyte elutes at 2.97 min, which reduces the analysis time by more than 4 minutes. Additionally, the peak width of NMPs was reduced up to 61% and of cap analogues up to 31% (Figure 27, Table 6, 7). Narrow peaks contribute improved S/N ratios and thus to increased sensitivities[281], which is on behalf of requirements for this method development. In order to examine the separation, the resolution at 0.15 ml/min and 0.25 ml/min were determined according to equation 6, where R_s is the resolution, t_R is the retention time in min and w_h is the peak width at the base line in min.

$$(6) \qquad R_s = \frac{1.18 * (t_R - t_R)}{w_{h1} + w_{h2}}$$

Again, as the elution order is changing, peak pairs were annotated with numbers. Only peak pair 7 at 0.15 ml/min is meeting the requirement of resolution ≥ 1.5 (Figure 29). While the resolution of the remaining peaks at 0.15 ml/min varies between 0.13 and 1.28. the resolution for all peak pairs at 0.25 ml/min is ≤ 0.5.
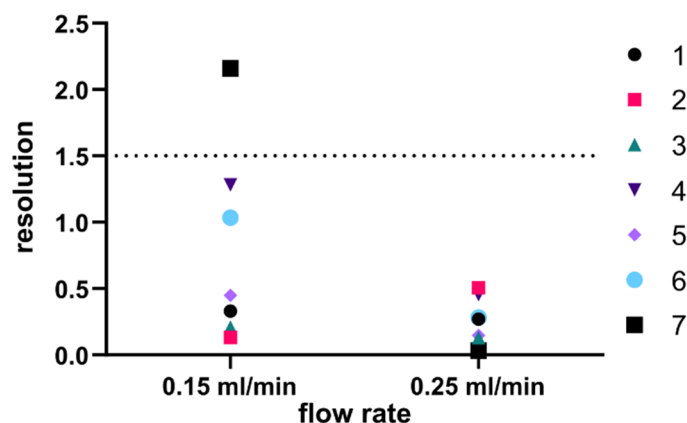
Figure 29: resolution of adjacent peak pairs 1-7 at 0.15 ml/min and 0.25 ml/min, the ideal resolution of 1.5 is indicated by the dotted line

Taken together, it was decided to continue the analysis with the flow rate of 0.25 ml/min, as the plate height for the majority of the analytes is comparable at 0.15 ml/min and 0.25 ml/min, but the analysis time is reduced considerably by 4 minutes. Further, the peak widths are distinctly reduced, which is advisable to ameliorate detection sensitivity. The impaired resolution was accepted, as specificity of this method was not dependent on base line separation, but peaks can be further identified by their specific mass transition.

3.3.1.6 Gradient adjustment

In summary, it was decided to utilize an ammonium acetate buffer at pH 6.9, the temperature of 40 °C and a flow rate of 0.25 ml/min. As evident from figure 27 all analytes elute within 3 min. Regarding the gradient that was introduced previously and used for the chromatographic separation (Figure 30, Gradient 1), it is evident that the analytes elute without the support of organic phase. In order to reduce the organic solvent consumption, for economic and ecologic reasons the gradient was adjusted. Still, the organic phase was not omitted completely, in order to flush off less polar residuals. Further, the re-equilibration time at the end of the method was adjusted, which is crucial for reproducibility of retention times. For re-equilibration to starting conditions it is suggested to apply 10 × column volume. The column volume was calculated according to equation 9, where V is the column volume in ml, L the column length in cm and d the inner diameter in cm.

(9) $$V = 0.5 \text{Ld}$$

The column length is 15 cm and the inner diameter is 2.1 mm (0.21 cm), resulting in a column volume of 1.575 ml. Consequently, 10 column volumes equal 15.75 ml, divided by the flow rate of 0.25 ml/min this corresponds to a re-equilibration time of 6.3 minutes, which was reduced to 6 minutes here.

For the new gradient 2, it was decided to start with 100% aqueous buffer for the first 5 minutes. The organic solvent was raised to 70% from 5 to 8 minutes and was then reduced to 0% until minute 9. Subsequently the column was re-equilibrated with 100% aqueous buffer until 15 minutes (Figure 30, Gradient 2).
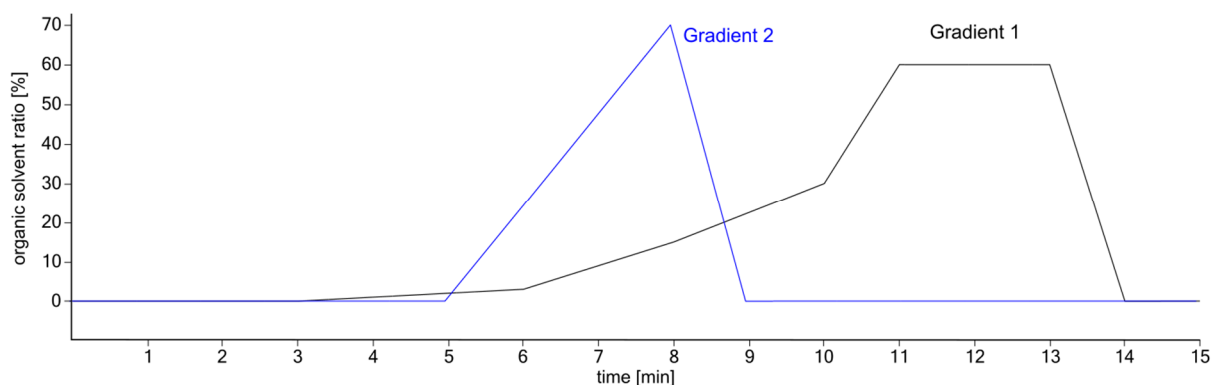
Figure 30: Overlay of organic solvent ratio of Gradient 1 and Gradient 2

Even though, the re-equilibration time was increased by the factor of 6, the analysis time of 15 min could be maintained. Only one published method for cap analysis is slightly shorter (12 min) [282], whereas this method does not include the simultaneous separation of NMPs and cap analogues but the separation of cap analogues only. Moreover, the acetonitrile consumption was reduced successfully by 47% with gradient 2, which was determined by the comparison of the area under curve (AUC) of the organic solvent ratios (Figure 30). The method could be further optimized, by reduction of re-equilibration time under surveillance of reproducible retention times. However, this investigation was not within the framework of this study.

3.3.1.7 Investigation of ion strength

The last parameter to be optimized was the ion strength. High ion strength improves the buffer capacity and thereby strengthens the prevention of retention time shifts.[271] The increase of ion strength can lead to altered retention as well as improved resolution and peak shape, as secondary interactions of polar and/or charged analytes with the stationary phase are suppressed.[271, 276] But increased salt load impairs the ESI efficiency by generation of less stable sprays due to enhanced conductivity[271, 281]. In order

to evaluate the impact of ion strength on retention and ionization, LC-MS/MS data was
adduced (Figure 31).
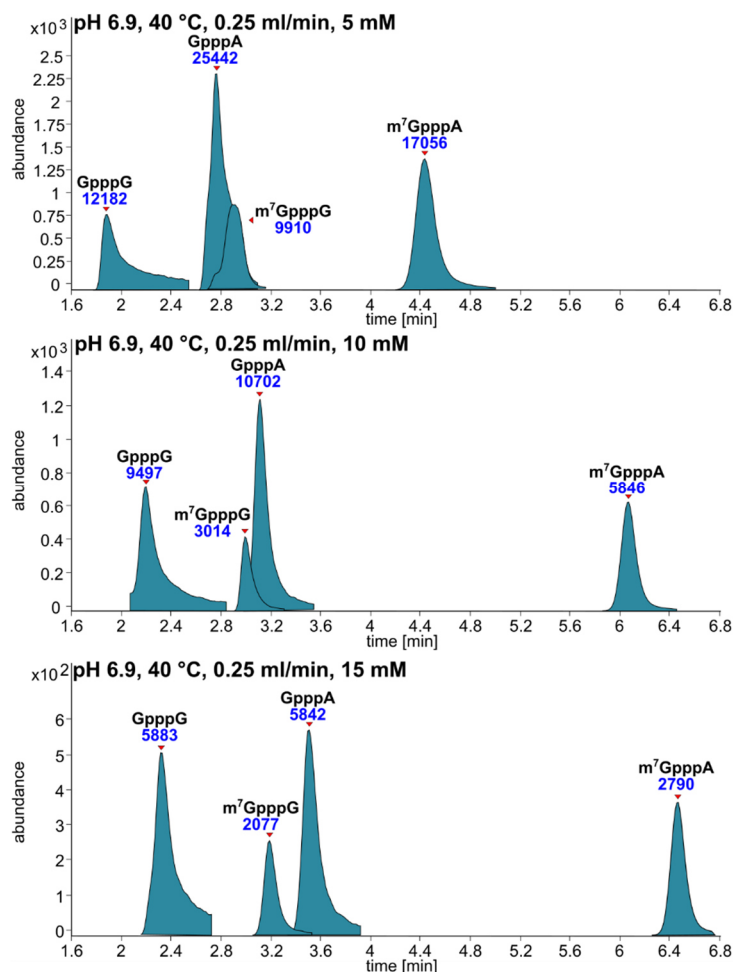
Figure 31: LC-MS/MS chromatograms from aqueous buffer pH 6.9 with varying concentrations of NH_4OAc,
conditions can be taken from upper left corner of chromatograms, peak identity and area is labeled above

Before the ion strength was investigated, the column was replaced by a new one of

the same model, due to column aging. This explains the slight retention time shifts in

the upper panel of figure 31, in contrast to the chromatogram in figure 27, despite the

same conditions. From Figure 31 it is clearly evident that the retention was changing

with increasing ion strength, especially m^7GpppA was retained stronger with increasing

salt concentration. An improvement in peak shape is also evident, especially for

GpppG, which showed a shoulder at 5 mM NH_4OAc, but not anymore at higher salt

concentrations. Additionally, a slight decrease in peak width was observed. This might

result from reduced electrostatic repulsion of the negatively charged analyte from the

stationary phase with increasing salt concentration. At pH > 5 silanol groups are neg-

atively charged, if the analyte is also negatively charged and the buffer concentration

is low, the sample can be repelled from the stationary phase, which results in less

retention and poor resolution. This effect can be reduced with higher salt concentration[283], which might be the explanation for increased retention and improved resolution with rising buffer concentration.

In order to judge if the increase in buffer concentration improved the separation of adjacent peaks, as well as their peak shape, the resolution of the critical peak pair GpppA and m⁷GpppG was determined according to the equation 6, where R_s is the resolution, t_R is the retention time in min and w_h is the peak width at the base line in min.

$$(6) \qquad R_s = \frac{1.18 * (t_R - t_R)}{w_{h1} + w_{h2}}$$

Usually, a resolution larger than 1.5 is desired for symmetric peaks of same height or a minimal resolution of 1.8-2 for non-symmetrical peaks of different height. The resolution of two peaks is relevantly dependent on the retention factor (k), the selectivity (α) and the theoretical plate (N), which were already introduced and examined in the previous sections. The values can be taken from table 8.

Table 8: Retention times (t_R) [min], peak width at base line (w_h) [min] and resolution (R_s) for GpppA and m⁷GpppG at different aqueous phase concentrations

	GpppA		m⁷GpppG		R_s
	t_R	w_h	t_R	w_h	
5 mM	2.76	0.46	2.89	0.47	0.16
10 mM	3.11	0.57	2.99	0.41	0.15
15 mM	3.51	0.50	3.19	0.48	0.38

Evident from table 8 the resolution increases with rising buffer concentration from 0.16 to 0.38. Still the resolution is far from the desirable value of 2 for peaks of different height. Here, it needs to be considered that the chromatographic parameters were optimized for short analysis time and high detection sensitivity. The low resolution is brought about these facts. Again, it should be underlined that this methods specificity is not dependent on the chromatographic base line separation, as the analytes can identified by their specific mass transitions.

More interesting for these purposes is the y axis of Figure 31, hence the peak height. The peak height and area, thus the signal intensity was decreasing consistently with higher ion strength. As outlined previously, the goal was to develop a highly sensitive method for absolute quantification of cap modifications, from which we know that they

only occur once per analyzed RNA molecule. One could argue that the sensitivity could be further improved by utilizing even lower amounts of NH_4OAc, but it was refrained from this as a certain amount of charge carrier is essential for ESI, in order to generate a reproducible electro spray with high ionization efficiency.[274] As the resolution was not significantly improved by higher ion strength but the signal intensity was reduced, it was decided to keep the concentration of 5 mM NH_4OAc.

3.3.2 Optimization of mass spectrometric parameters

After the chromatographic method development was completed, the parameters for mass analysis had to be optimized. It is important to develop the chromatographic method first as the flow rate, solvent composition and the degree of ionization of the compound have an impact on the parameters necessary for a sensitive detection. One needs to distinguish between the optimization of analyte specific parameters, that are selected individually for each compound and the optimization of the ion source parameters, which are selected on the basis of general high signal intensity for all analytes of interest. Our mass spectrometer is equipped with an electrospray ionization (ESI) source, here the capillary voltage for droplet charging as well as gas flow rates and temperatures need to be optimized for effective droplet formation and ion desolvation. Before selecting ideal parameters for the ion source the optimal analyte specific parameters were determined. The goal was to develop a highly sensitive method for quantification of cap modifications, therefore it was decided to develop a method with dynamic multiple reaction monitoring (dMRM). Here analytes are specified by retention time and their mass transition. In this monitoring only a certain mass transition at a certain time window is measured, and therefore allows detections of high sensitivity. Our mass spectrometer is equipped with a triple quadrupole. After analytes were ionized in the ion source, they move into the first quadrupole (Q1), which acts as mass filter and selects the precursor ion. Subsequently, the selected precursor ion enters the hexapole collision cell. In here the precursor ion is fragmented under collision induced dissociation (CID), generating product ions. From here the product ions migrate to the third quadrupole (Q3), where the previously determined product ions are selected and subsequently detected in the detector. In the following, the optimization of the mass spectrometric parameters is outlined.

3.3.2.1 Optimization of analyte specific parameters

The analyte specific parameters to be optimized are the mass to charge (*m/z)* of precursor and product ion, the fragmentor voltage and collision energy. The cell accelerator voltage (CAV) determines the speed of the product ion leaving the collision cell and can be varied from 1 V to 8 V. For the analysis of isotopologues, yielding product ions with the same *m/z*, the CAV has to be high enough to quickly accelerate the product ions into the second mass analyzer before the entrance of the following ions into the collision cell. Slow transition of product ions might result in signal interferences. According to our intensive studies of nucleoside isotopologues[243] and the recommendation of 4 V to 5 V for dMRM[284], the CAV was kept at 5 V and was not further optimized, as it has no significant impact on the signal intensity.[284] Before optimization of parameters, it needs to be decided on the ion mode. Under physiological conditions the phosphate diester backbone of DNA and RNA is deprotonated while the nucleobases are not ionized. With the selected mobile phase of pH 6.9, it was assumed that the analytes possess at least one negative charge. Conclusively, the negative ion mode was chosen. The optimization for GpppA is outlined exemplarily in the following. The other analytes were optimized analogously, the optimized parameters can be found in table S3. The optimization was implemented individually with synthetic standards of every analyte, therefore no chromatographic separation was needed. For the sake of time the optimization was performed on a short C18 RP column (Agilent, Zorbax Eclipse Plus, 95Å, 2.1 x 50 mm, 1.8 µm), which reduces the analysis time to 1 min. In first instance the *m/z* of the precursor ion was determined. Therefore, the mass spectrometer was operated in full scan mode, where all ions entering the mass spectrometer are detected. The precursor ion with *m/z* 771 was detected in the resulting mass spectrum. Minor signals for *m/z* 408 and 424 were also detected, which are the product ions of this precursor ion (Figure 32A). In this scan mode no collision energy was applied, nevertheless product ions were generated, which is attributed to in source fragmentation. Therefore, it is of great importance to optimize the fragmentor voltage. The fragmentor voltage is applied to accelerate the analytes from the electrospray chamber, which is under atmospheric pressure, towards the mass analyzer, which is set under vacuum. The fragmentor voltage should be high enough to give maximal abundance of precursor ions without fragmenting them. In order to optimize the fragmentor voltage, a selected ion monitoring (SIM) was implemented, where the Q1 and Q3 were fixed to the *m/z* of the precursor ion. Sequential varying fragmentor voltages

ranging from 50 V to 270 V were applied and the optimum was selected on the basis of highest signal-to-noise ratio (S/N). Here 220 V was determined to give the highest signal (Figure 32B).

Next, a product ion scan was implemented. Here the precursor ion was fixed, the previously determined fragmentor voltage was applied and a suitable collision energy of 60 V was selected. In Figure 32C the mass spectrum of the product ion scan is given. Several product ions were generated. Usually the most abundant product ion is selected, in this case it would be the product ion with *m/z* 159, but here it was refrained from this criterion. For explanation the attention should be drawn to figure 32E, the detected *m/z* values for precursor and product ions were assigned to structures, based on reported fragmentation[285, 286]. This illustration shows the precursor ion on top, in the middle the product ions resulting from phosphoanhydride cleavage, consisting of a base and two phosphate ester groups. These product ions can be further fragmented to the ions on the base panel, which are yielded from all cap analogues. It is evident that the product ion with *m/z* 159 is generated from every cap analogue and therefore is not suitable as tool of specificity. As a consequence, it was decided to select the product ion illustrated in the middle panel (*m/z* 408 or 424) as they are specific to the analytes.

Subsequently, the collision energy was optimized to achieve the highest abundance for the selected mass transition. A SRM with fixed *m/z* for precursor and product ion as well as the previously determined fragmentor voltage was conducted with sequentially varying collision energies ranging from 10 eV to 50 eV. As displayed in Figure 32D, 35 V was found to yield the highest signal.
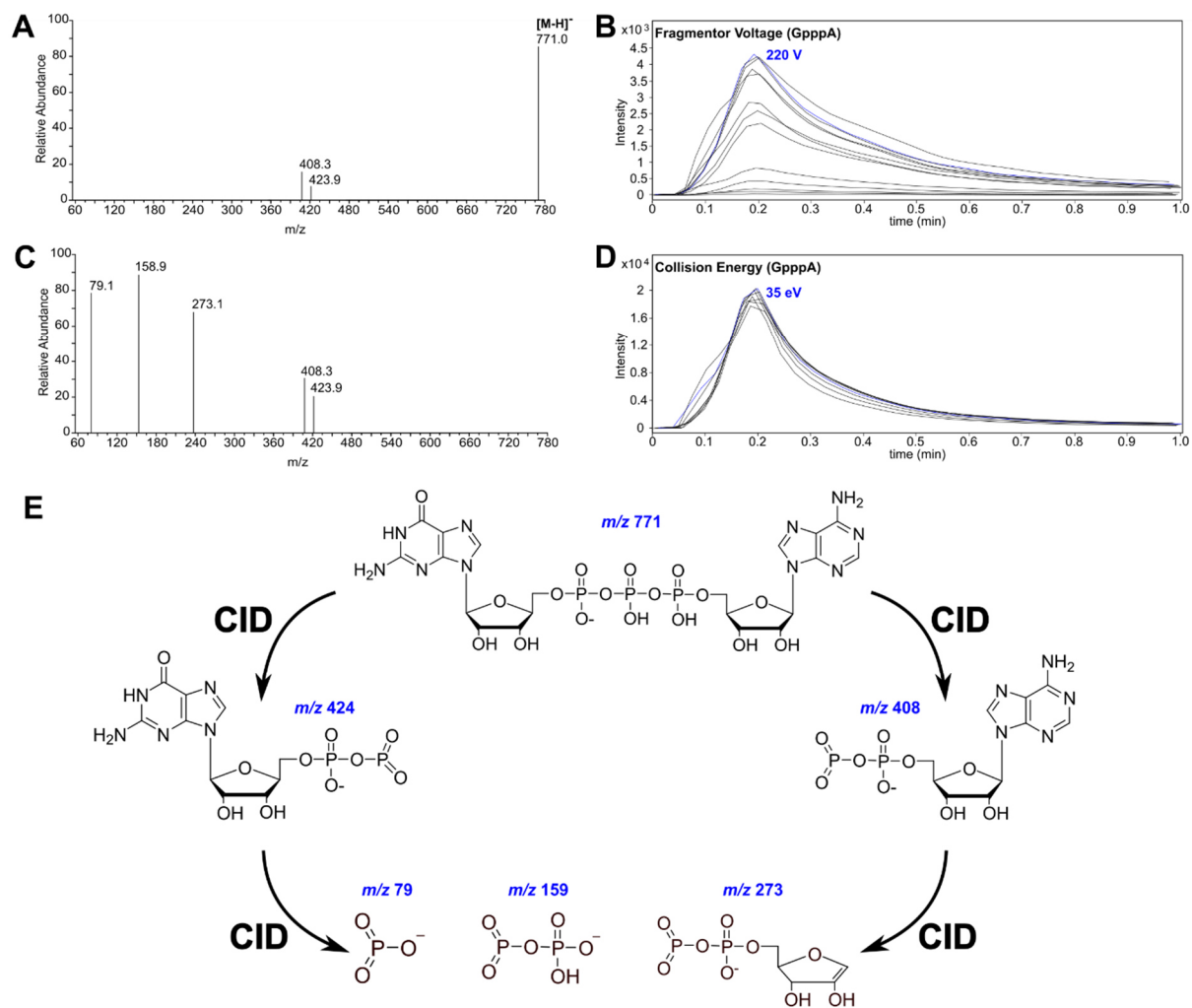
Figure 32: Optimization of analyte specific parameters for GpppA (A) mass spectrum of full scan (B) overlay of LC-MS/MS chromatograms of different fragmentor voltages, peak yielding the highest signal and results from the optimal value is highlighted in blue (C) mass spectrum of product ion scan (D) overlay of LC-MS/MS chromatograms of different collision energies, resulting in formation of product ion with *m/z* 408, peak yielding the highest signal and results from the optimal value is highlighted in blue (E) proposed fragmentation

From this a dMRM method was designed, with the retention times, previously determined in the optimization of the chromatographic method. The optimal parameters for every analyte can be found in table S3.

3.3.2.2 Optimization of ion source specific parameters

The detection sensitivity can be further improved by selection of optimized parameters for the ion source. Here the parameters were selected on the basis of optimal values for the cap analogues, as those will be the less abundant species in the sample, in comparison to the NMPs. The ion source parameters were optimized according to the one variable at a time approach (OVAT), meaning that only one parameter is varied, while all other parameters are kept stable. Even though it is known that the parameters influence each other and a multidimensional approach would be more accurate, it is

not common in the laboratory routine. The optimization of the source parameters was implemented under consideration of previously optimized parameters of chromatography and analyte specific parameters. This is of great importance as the mobile phase composition and flow rate have great impact on the ionization of the analyte. The pH and ion strength affect the ionization efficiency and the aqueous and organic phase composition influences the desolvation of the ions in terms of surface tension.[203, 209-211] The parameters to be optimized, were the capillary voltage which is responsible for droplet charging, the flow rate and temperature of the drying gas, which are crucial for droplet formation, control of droplet size and desolvation. Especially, in this method the drying gas is of major significance, as the negative ions are already formed in solution and have to be transferred to gas phase. The nebulizer pressure impacts the nebulization process efficiency of the spray and in combination with the drying gas, it is responsible for ion formation. A further plane to improve ionization efficiency and thus the signal intensity is the application of the sheath gas for pneumatic assistance during ESI. The sheath gas is applied concentrically around the spray needle. Its purpose is to narrow the spray, which results in more effective desolvation.[214] The additional thermal focusing concentrates the ions spatial.[215] Further the nozzle voltage can be applied, which improves the ionization of apolar analytes.

In the following the optimization is outlined exemplarily for GpppA and was conducted analogously for the other cap analogues. The optimum values were determined based on the highest signal intensity. The optimization of the capillary voltage was implemented between 2500 V and 5000 V, with steps of 500 V. The capillary voltage of 2500 V was found to yield the highest signal intensity (Figure 33A). The drying gas flow rate was examined from 4 L/min to 12 L/min with steps of 2. The drying gas temperature was investigated from 230 °C to 340 °C, in steps of 30. The values yielding the highest signal intensities were 12 L/min and 340 °C, respectively (Figure 33B, C). The impact of the nebulizer pressure was monitored from 20 psi to 40 psi by steps of 5 and 40 psi was found to yield the highest signal intensity (Figure 33D). The sheath gas flow and temperature were examined in the ranges from 10 L/min to 12 L/min by steps of 1 and from 200 °C to 400 °C by steps of 50. It was found that the highest signals were obtained for 11 L/min and 400 °C, respectively (Figure 33E, F). The nozzle voltage was examined from 0 V to 2000 V, by steps of 500 and found to impair the signal intensity and therefore was kept at 0 V (Figure 33G). It was recognized that the highest sheath gas temperature tested gave the highest signal intensity, one could argue that an even

higher temperature could result in a more intense signal. Nevertheless, it was refrained from testing higher temperatures, as the tested range is according to manufacturer's recommendation and it was not desired to exceed this.
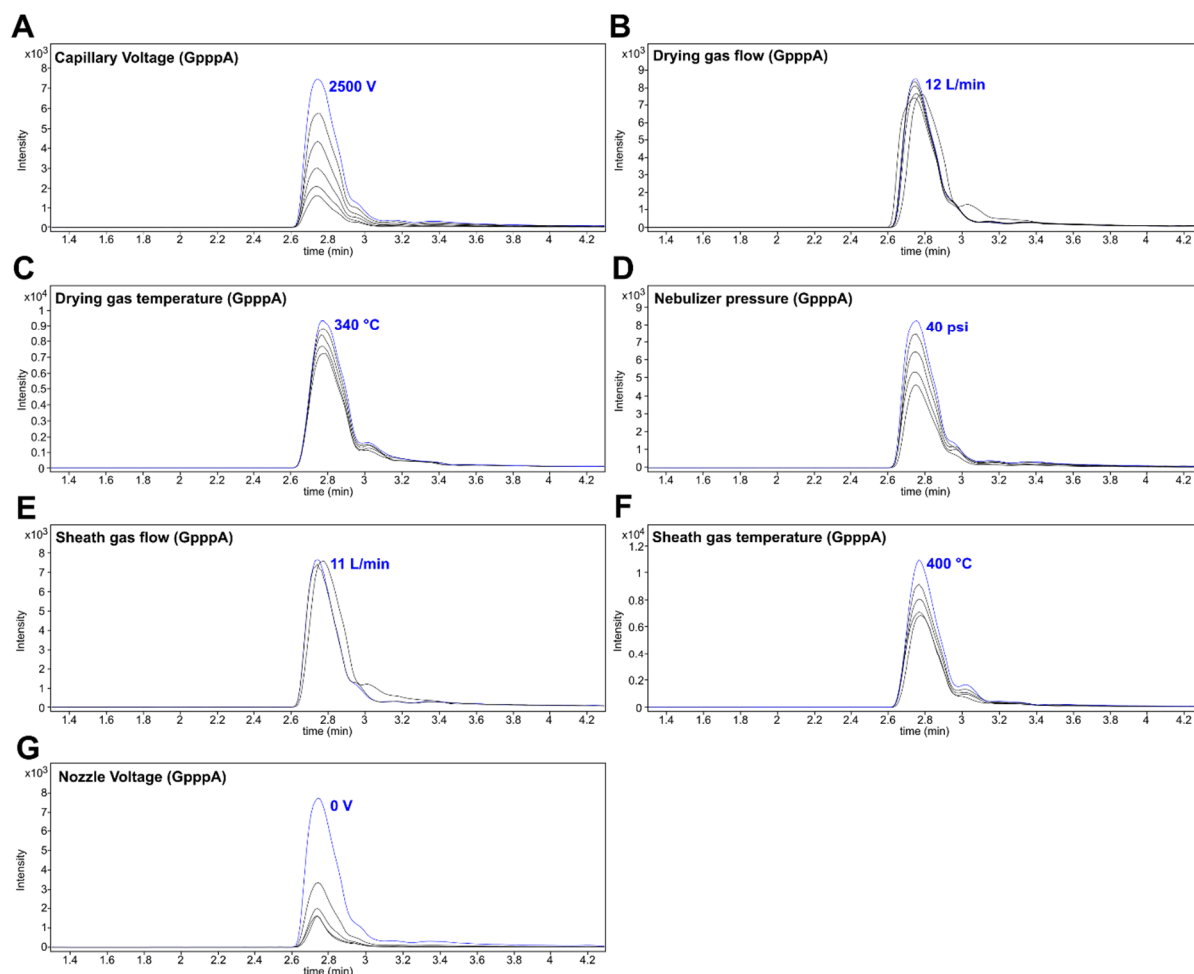
Figure 33: Optimization of ion source specific parameters. Overlay of LC-MS/MS chromatograms of GpppA with different parameters of the ion source to identify the optimal value for (A) capillary voltage (B) drying gas flow (C) drying gas temperature (D) nebulizer pressure (E) sheath gas flow (F) sheath gas temperature and (G) nozzle voltage. The peak yielding the highest signal and results from the optimal value is highlighted in blue

Finally, the optimized parameters for all cap analogues were collated and the terminal parameters were selected on the basis of accordance for all cap analogues and are presented in the methods section. The final LC-MS/MS chromatogram after optimization of all parameters is displayed in Figure 34.
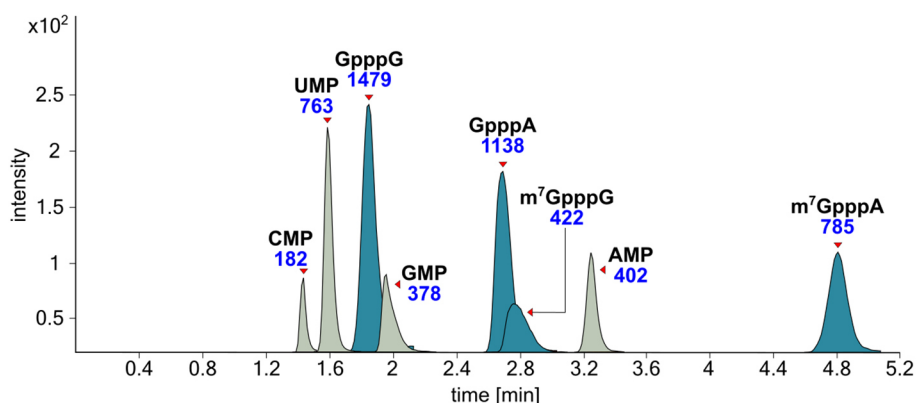
Figure 34: LC-MS/MS chromatogram from final parameters, peak identity and area are labeled above, injected amount of each NMP equals 2.38 pmol and of each cap analogue 0.32 pmol

3.3.2.3 Determination of LOD and LOQ for cap analogues

After the optimization the limit of detection (LOD), limit of quantification (LOQ) and the linear range were determined. The LOD is the lowest amount of analyte that is detectable and is defined as 3 × S/N. The LOQ is the lowest amount of analyte that can be quantified reliably and is defined as 10 × S/N. The S/N is the ratio of signal-to-noise and is specific for every analyte and defined by the quotient of the analyte's signal height (S) and the noise (N). The determination of LOD and LOQ for m7GpppA is illustrated exemplarily in figure 35 by the overlay of LC-MS/MS chromatograms.
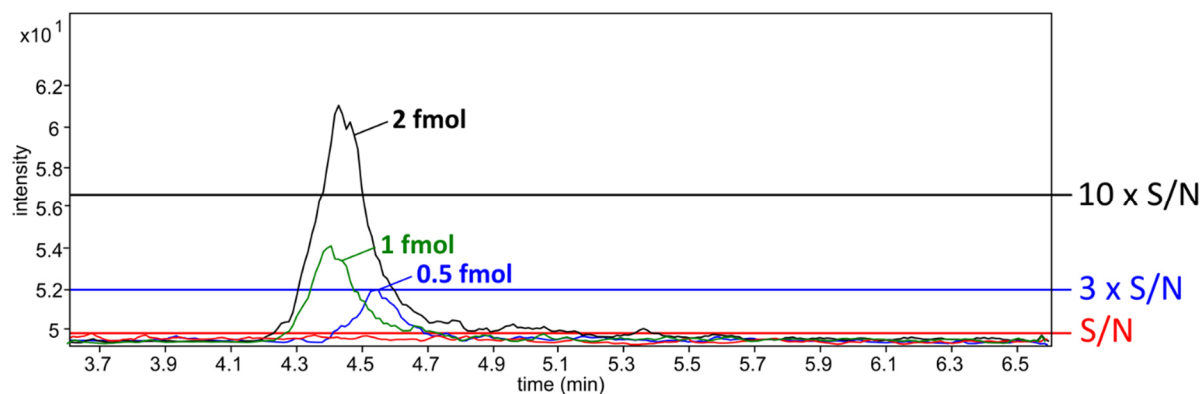
Figure 35: Overlay of LC-MS/MS chromatograms of m7GpppA for determination of LOD and LOQ, blank injection (red), 0.5 fmol (blue), 1 fmol (green), 2 fmol (black), vertical lines indicate the multiples of the signal-to-noise (S/N), S/N (red), 3 × SN = LOD (blue), 10 × SN = LOQ (black)

The trace in red results from a blank inject and represents the S/N. The LOD (3 × S/N) is indicated by the blue line and the LOQ (10 × S/N) by the black line. Further, the signals for 0.5 fmol, 1 fmol and 2 fmol m7GpppA are displayed. As it can be observed in figure 35, the LOD for m7GpppA is 0.5 fmol, while the LOQ is not reached at 1 fmol but at 2 fmol. The results for all cap analogues are given in figure 35 and table 9.

Further, the linear range should be introduced, which is the optimal working range. The linearity of the calibration is given within this range. Its lower limit is the LOQ and the upper limit is the limit of linearity (LOL). After exceeding the LOL the response is saturated and the curve is flattening. It is suggested that the saturation of response is a result of maximum concentration of analyte that can be ionized.[216, 287] The graphs for determination of the aforementioned values are illustrated in Figure 36.
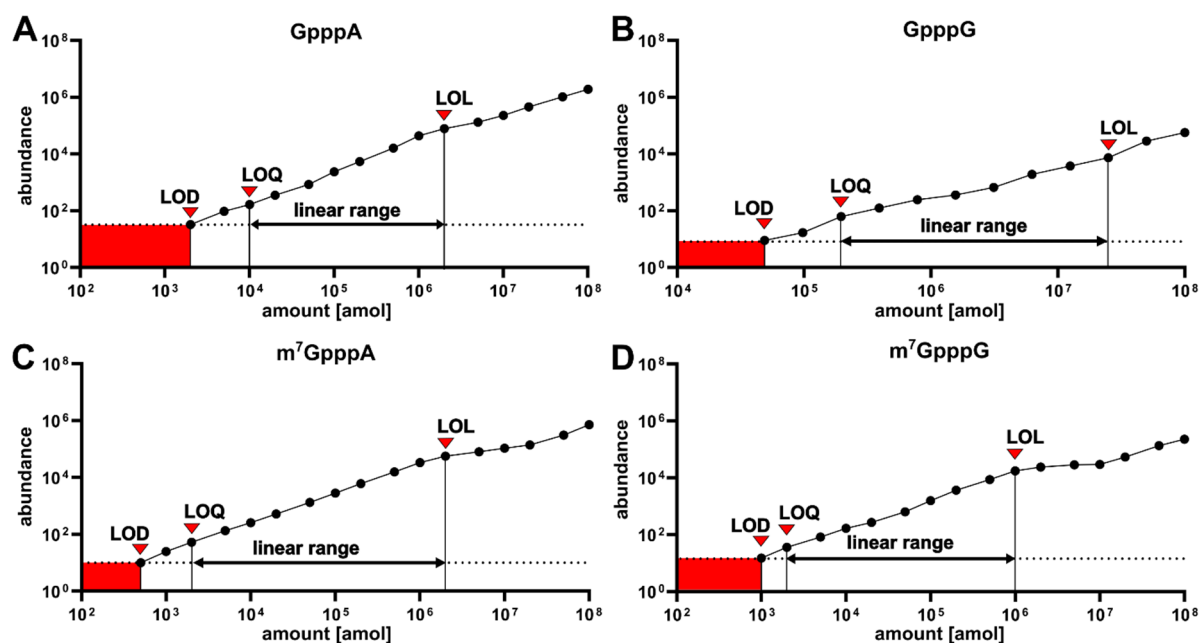
Figure 36: Calibration curves for cap analogues (A) GpppA (B) GpppG (C) m[7]GpppG and (D) m[7]GpppA, limit of detection (LOD), limit of quantification (LOQ), limit of linearity (LOL) and the linear range are marked in the plots, the red squares highlight the noise region

The coefficient of determination for confirmation of linearity was yielding R^2(GpppA) = 0.9976, R^2(GpppG) = 0.9913, R^2(m[7]GpppA) = 1.0 and R^2(m[7]GpppG) = 0.9966. The determined LOD and LOQ values are given in table 9.

Table 9: List of limit of detection (LOD) and limit of quantification (LOQ) for cap analogues with developed method

analyte	LOD [fmol]	LOQ [fmol]
GpppA	2	10
GpppG	49	195
m[7]GpppA	0.5	2
m[7]GpppG	1	2

To our knowledge this is the first time a cap analogue could be detected in attomole amounts. Moreover, these are the lowest LOQ values for m[7]GpppA and m[7]GpppG that were determined with a LC-MS/MS method. With this method the LOQ for m[7]GpppA and m[7]GpppG was reduced up to 12 times in comparison to previous reports[278, 282, 288].

It appears that the LOD and LOQ for GpppG were higher than for the other cap analogues. Also other studies obtained impaired detection and quantification limits for GpppG in comparison to the other analogues with factors of 5 to 6.[278, 282] In general, the S/N ratio could be raised by noise reduction. The noise can be reduced by preferring a dMRM over a full scan, which was realized here, or by reduction of the chemical noise, which can be achieved by removal of solvent clusters, through increased heat during vaporization. However, the drying gas temperature was selected at the upper limit of the manufacturers recommendation, therefore it was refrained from reducing noise by further temperature increase.

3.3.3 Determination of dynamic range

Besides the determination of limits in detection and quantification a valid method should also include the determination of the dynamic range of biological sample composition. To address this question, different amounts and compositions of possible molecules of interest were mimicked. The compositions of a 30, 300, 3000 and 30000 mer were rebuild in order to imitate cap modification containing macromolecules, like IVTs, mRNA or the gRNA of SARS-CoV-2. This approach should allow the determination of detectability of a cap modification besides the nucleotides of the N-mer. Samples ranging from 50 ng to 5 µg were prepared, consisting of 1/N-mer cap analogue and NMPs in equimolar amounts.

Figure 37 illustrates the detectability of a cap analogue besides multiple NMPs for different sample amounts. Fields highlighted in blue represent compositions, where the detection of the cap analogue was successful, whereas purple indicated that it was not. Cap analogues from molecules within a two-digit length were successfully detectable even at low amounts of 50 ng. For the detection of cap modifications of a 300 mer at least 1 µg of sample needed to be injected. The injection of larger amounts is not recommended, as it can result in column overload and thus impaired peak shape and resolution. Moreover, the injection of vast amounts can lead to saturated response, as only a certain amount of molecules can get ionized in the source. In conclusion, the analysis of cap modifications from full length RNA (mRNA, viral RNA) would not possible.
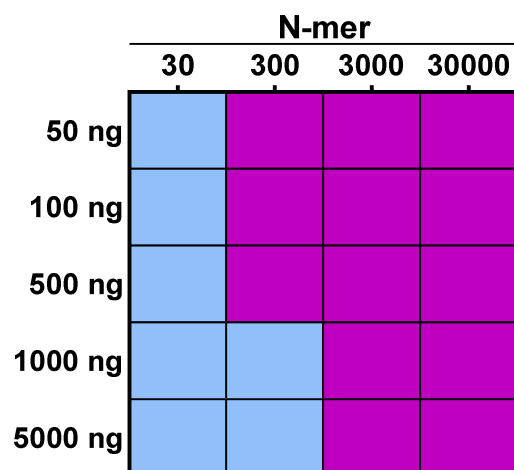
Figure 37: Illustration of method's dynamic range. Columns reflect the composition of a N-mer, rows refelct the injected sample amount in ng, fields highlighted in blue display detectable conditions for cap analogues, fields highlighted in purple display the limits of the detectability of cap analogues.

Also in other studies intact long RNA from biological material could not be examined. Either short length RNA was transcribed *in vitro* with 30 nucleotides length[278] or sRNA from *E. coli* was analyzed[286], which is reported to be smaller than 300 nts[289]. One study did analyze mRNA from biological source, but there the RNA was hydrolyzed and cap modifications were purified with HPLC. The fractions containing solely cap modifications were further pursued to LC-MS analyses and thereby excluded NMPs as potential interfering factor.[282] This additional enrichment step can result in substantial sample loss.

For the analysis of long RNA, it is advised to implement the previously introduced *biotin enriched mung bean assay*. The feasibility of this approach was already shown in the chapter 3.2.4 *Enrichment of viral RNA sequences with biotin enriched mung bean assay*. In contrast to the application mentioned earlier, here the goal consists in shortening the target RNA rather than specifically enriching it. The RNA should be shortened to a sequence of the 5' end with a specific length within the dynamic range.

3.3.4 Effect of injection volume on peak shape

Ideally a peak has a Gaussian shape, which can be validated by the asymmetry factor. For symmetric peaks this value is 1, if the value is <1 fronting is faced, if this value is > 1 tailing is observed. Fronting and tailing can be both a result of column overloading. Overloading can be differentiated in a mass overloading, as a result of excessive sample amount and volume overloading, as a result of large injection volumes. Here the

limits of overloading should be investigated. 500 ng of a 30-mer with an injection volume of 20 µL (Figure 38, black) and 1000 ng a 30-mer but with an injection volume of 4 µL (Figure 38, blue) were separated. An overlay of the resulting peaks from the two conditions is illustrated in Figure 38.
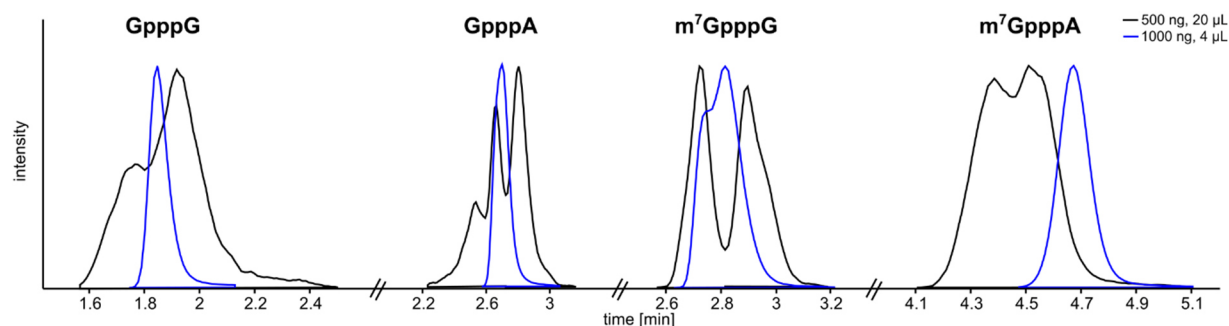
Figure 38: Extract from LC-MS/MS chromatogram of the cap analogues GpppG, GpppA, m[7]GpppG and m[7]GpppA from an injection with 500 ng and 20 µL (black) and 1000 ng and 4 µL (blue)

The peaks were scaled for better comparability of peak dimension. Peaks illustrated in black (500 ng, 20 µL) have an impaired peak shape in contrast to peaks in blue (1000 ng, 4 µL). Here the source of impaired peak shape must be the attributed to volume overloading, as the 20 µL inject yields bad peak shape despite the lower sample amount. Further, peak splitting was observed for the 20 µL injection. Usually, peak splitting is attributed to poorly packed columns, which can be excluded here as the column was purchased from a well-known company producing high quality columns. Another source for peak splitting and peak broadening, emerges if the analytes solvent is stronger than the eluent. This is not the case here, as solvent and eluent were both aqueous. From this, it is suggested to analyze samples of high concentration and to avoid large injection volumes for symmetric peaks.

3.3.5 Preparation of cap-SILIS

As outlined in the introduction, the utilization of SILIS besides a calibration allows for absolute quantification. Our nucleoside SILIS is produced biosynthetically.[243] For a biosynthetical cap-SILIS the enrichment of mRNA would be necessary. However, the production of a mRNA SILIS is not purposive as (a) the purification of mRNA is labor intensive and remaining of rRNA and tRNA cannot be excluded, (b) the yield after mRNA purification is rather low[242] and (c) each mRNA molecule has only one cap modification, thus the composition does not match the method's dynamic range.

Therefore, it was decided to produce the cap-SILIS by *in vitro* transcription. The IVT was transcribed according to Hagelskamp *et al.*[290] with a slight modification in the

rNTPs utilized for transcription. Here, ^{13}C labeled rNTPs were used with the intention of generating heavy isotopologues of the naturally occurring nucleotides. The ^{13}C labeled pyrimidines are 9 Da and purines 10 Da heavier than their naturally occurring isotopologues, as every carbon atom in the base and ribose consist of the heavy ^{13}C isotope. Additionally, the 5' end of the IVT was capped with m[7]G and optionally a 2'-O-methylation of the first nucleotide of the sequence was implemented, yielding either cap 0 or cap 1. For NMPs this results in a mass increase of +9 Da or +10 Da. For cap modifications, consisting of two purine bases, this results in a mass increase of +20 Da for the precursor ion and +10 Da for the product ion.
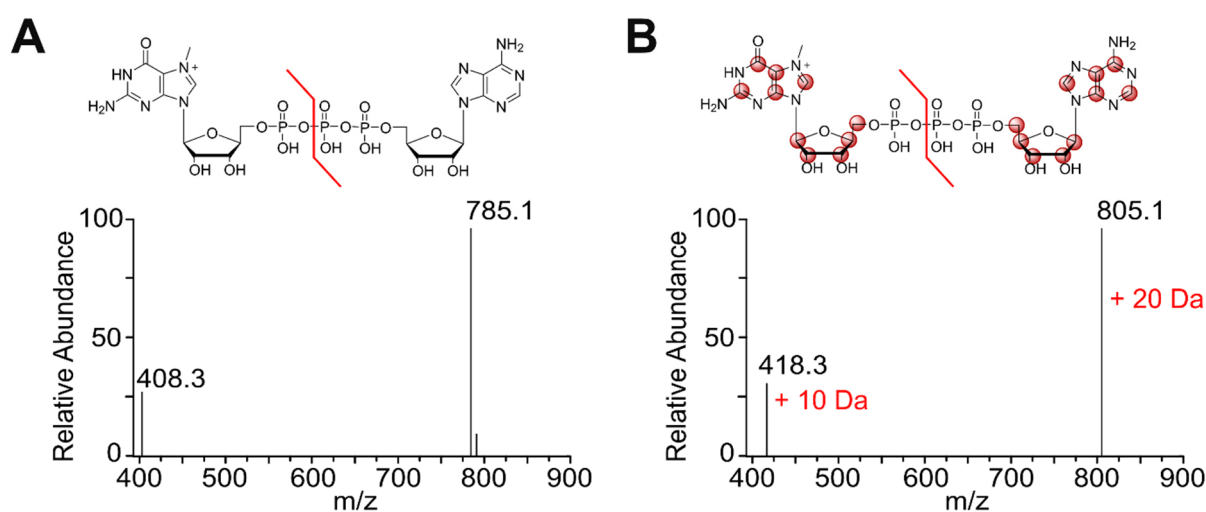
Figure 39: Structure of m[7]GpppA and the resulting mass spectra in negative ion mode (A) structure of m[7]GpppA the fragmentation is indicated by the red line, the respective mass spectrum in negative ion mode with the precursor ion (*m/z* 785.1) and the product ion (*m/z* 408.3) (B) structure of m[7]GpppA SILIS, ^{13}C atoms are highlighted in red, the fragmentation is indicated by the red line, the respective mass spectrum in negative ion mode with the precursor ion (*m/z* 805.1) and the product ion (*m/z* 418.3), the mass increase in contrast to the unlabeled isotopologue is displayed in red next to the peaks

In Figure 39 the structures of the isotopologues of m[7]GpppA are illustrated in the upper panel and the lower panel shows the mass spectra of the resulting ions in negative ion mode with precursor and product ion. All mass transitions for the isotopologues used as SILIS are given table S3 in the appendix.

4   Conclusion and Outlook

The combination of pulse-chase experiments and NAIL-MS offered a better understanding of the stress-induced RNA modification dynamics and the underlying mechanisms. The metabolic labeling strategy in *S. cerevisiae* was improved to generate a novel SILIS, which is clearly distinguishable from the isotopologues generated in the NAIL-MS experiments and to yield efficient deuteration of methyl groups, enabling the differentiation of endogenous and damage-induced methylation. The application of NAIL-MS allowed to identify original tRNA, 18S and 25S rRNA transcripts as target of damage, while new transcripts remained unaffected. Moreover, methylation damage by MMS was found to occur by direct methylation on nucleobases. In addition, novel damage products of MMS could be identified, which occur by direct base methylation of 2'-O-methylated nucleotides in rRNA. The trace of methylated nucleosides revealed, that the level of endogenous modifications remained stable, while the abundance of damage-induced methylations was decreasing quickly after stress exposure.

This technique was successfully expanded to study the methylation dynamics on the genomic level. The quantities of endogenous and damage-induced methylations in the genome and transcriptome of *E. coli*, *S. cerevisiae* were assessed by the combination of pulse-chase experiments and NAIL-MS. In line with the previous findings, the fast decrease of damage-induced methylation after MMS exposure could be determined. Moreover, m^7G and m^7dG were identified as the major damage products in the genome and transcriptome of the examined organisms.

The combination of pulse-chase experiments and NAIL-MS can be utilized to study many more dynamics of nucleic acids in divers organisms. The knowledge about absolute modification quantities and their impact on organismal function can help to advance the application of RNA therapeutics and personalized medicine.

In parallel to the aforementioned organisms, modifications in the viral genome are highly dynamic. But in contrast to prokaryotes and eukaryotes the underlying mechanisms are less understood. SARS-CoV-2, which was first discovered in 2019, already developed several mutants and variants within three years and the differences in the modification profile of the genome remain to be assessed. Absolute modification quantities in the mutants D614 and G614 as well as in the variants alpha, beta and delta were determined. However, it appeared that the found quantities are dependent on the

cultivation and purification methods and hence the validation of sample purity and integrity prior to analysis is suggested. Two approaches for viral RNA enrichment were provided but the validity of resulting quantities should be reinforced by confirming the integrity of enriched samples, *e.g.* through RT-qPCR and RNA-Seq.

Apart from internal modifications, the genome of SARS-CoV-2 possess a cap modification at the 5' end, which the virus adapts from the host mRNA to escape the innate immune response. Further insight into capping mechanisms might offer a way to limit the evasion from the host immune system and hence restrict the replication and the contagiousness. Here, a successful LC-MS/MS method development for time efficient and highly sensitive detection of cap modifications was presented. In addition, the cap-SILIS was generated which allows for absolute quantification of cap modifications per molecule. This method excels by its high detection sensitivity, short analysis time and the minor use of organic solvent, which is desirable for ecological and economic reasons. Therefore, it might be feasible for industrial application and quality control of mRNA therapeutics, besides its use to study capping mechanisms.

Taken together, the described methods offer a way to gain a better understanding of epigenomics and epitranscriptomics. The analysis of modification abundances and dynamics can drive the production and application of mRNA therapeutics and personalized medicine ahead and thus is of great importance for medical advances.

5   Material and Methods

The described materials and methods in this section refer to the unpublished data. Protocols of published data can be found in the materials and methods section of the respective publication. Only internal or more detailed descriptions of the procedures already described in the publications are supplemented in this section.

5.1   Materials

Chemicals, nucleosides and cap analogues

All salts were obtained from Sigma-Aldrich (Munich, Germany) in molecular biology grade, unless stated otherwise. Chemicals such as MMS (99%, SKU: 129925-25G), sodium arsenite (≥90%, SKU: S7400-100G), H_2O_2 (30%, 1072980250:), TBH (70%, SKU: 458139), HOCl (380 mM, SKU: A1727) were also purchased from Sigma-Aldrich. The isotopically labeled compounds $^{15}N_2$- uracil (≥98%) and $^{13}C_6$-glucose (≥99%) were purchased from Euroisotope and L-methionine-[$^2H_3$]-methyl (98 atom % D) was obtained from Sigma-Aldrich. LC-MS grade solvents (acetonitrile, isopropanol, methanol) and HPLC grade isopropanol were purchased from Roth (Karlsruhe, Germany). All solutions and buffers were made with water from a Millipore device (Milli-Q, Merck, Kenilworth, NJ, USA).

The nucleosides adenosine (A), cytidine (C), guanosine (G) and uridine (U) were purchased from Sigma-Aldrich. Dihydrouridine (D) was acquired from Apollo Scientific (Stockport, UK). N6-threonylcarbamoyladenosine (t^6A) was acquired from TRC (North York, Canada). N6-dimethyladenosine ($m^{66}A$) was acquired from Alfa Chemistry (New York, USA). N3-methyluridine (m^3U), N6-isopentenyladenosine (i^6A), 2-methylguanosine (m^2G), and 5-methoxycarbonylmethyluridine (mcm^5U) were generous gifts from the Dedon laboratory. 1-Methylinosine (m^1I) was a generous gift from STORM Therapeutics LTD (Cambridge, UK). All other RNA nucleosides were purchased from Carbosynth (Newbury, UK). The deoxynucleosides 1-methyldesoxyadenosine (m^1dA) and 3-methyldesoxycytidine (m^3dC) were purchased from Jena BioScience (Jena, Germany). 6-Methyldesoxyadenosine (m^6dA) was purchased from Carbosynth. 7-Methyldeoxyguanosine (m^7dG) was provided by Dr. Christoph Borek. The nucleoside monophosphates, cytidine 5′-monophosphate (CMP), uridine 5′-monophosphate (UMP), guanosine 5′-monophosphate (GMP), adenosine 5′-monophosphate (AMP) were purchased from Sigma-Aldrich (Munich, Germany). The cap analogues GpppA, GpppG,

m^{7}GpppA and m^{7}GpppG were purchased from New England Biolabs (Ipswitch, USA). For all materials not shown here, the sources are given in the respective chapter.

SEC buffer

7.7 g NH$_4$OAc [0.1 M] (molecular biology grade, SUK: A1542-250G) was dissolved in 1000 ml ultrapure water (final concentration 0.1 M) in a Duran ® Schott bottle.

MOPS buffer

For the preparation of 10 × MOPS buffer, 16.94 g 3-(N-morpholino)propane sulfonic acid (MOPS), 2.72 g NaOAc and 1.17 g ethylenediaminetetraacetic acid (EDTA) were dissolved in 400 ml ultrapure water, yielding the final concentrations of 200 mM MOPS, 50 mM NaOAc and 10 mM EDTA. The pH was adjusted 7 with 10 M NaOH.

20 × SSC buffer

1.75 g NaCl was added to a 2 ml solution of 1.5 M trisodium citrate solution (pH 7.0). The solution was diluted with ultrapure water to a volume of 10 ml, yielding the final concentrations of 300 mM trisodium citrate and 3 M NaCl. Other concentrations of the SSC buffer were prepared by appropriate dilution of the 20 × SSC buffer with ultrapure water.

1× B&W buffer

548 mg NaCl was dissolved in a mixture of 38.5 μL 1 M EDTA solution and 50 μL 1 M Tris-HCl (pH 7.5). The solution was filled up to 10 ml with ultrapure water, yielding the final concentrations of 1 M NaCl, 0.5 mM EDTA and 5 mM Tris-HCl.

Hybridization Buffer (MBN)

0.596 g HEPES and 0.373 g KCl were dissolved in 9.5 ml ultrapure water. Subsequently, 500 μL DMSO was added and the pH was adjusted to 7 by addition of 2 M NaOH. The final concentrations were 250 mM HEPES, 500 mM KCl and 5% (v/v) DMSO.

LC-MS Buffer for nucleoside analysis

For nucleoside analysis, 0.385 g NH$_4$OAc (LC-MS grade, ≥99%, VWR, SUK: 84885.180) was dissolved in 1000 ml ultrapure water (final concentration 5 mM) in a

Duran ® Schott bottle. The pH was adjusted to 5.3 with 65 µL acetic acid (HiPerSolv CHROMANORM for LC/MS, acetic acid 99%, VWR Chemicals, SUK: 84874.180).

LC-MS Buffer for analysis of cap modifications

For nucleoside analysis, 0.385 g NH_4OAc (LC-MS grade, ≥99%, VWR, SUK: 84885.180) was dissolved in 1000 ml ultrapure water (final concentration 5 mM) in a Duran ® Schott bottle. The pH was adjusted to 6.9 with 7 µL ammonia (LC/MS grade, ammonia 25%, VWR Chemicals, SUK: 5.33003.0050)

Technical equipment

Table 10: List of technical equipment with device name, model and vendor

Device	Model	Vendor
HPLC	Agilent 1100 HPLC System	Agilent, Santa Clara, USA
LC-MS/MS	Agilent 1260 Infinity II LC System, Agilent 6470A QQQ with Jet Stream ESI source	Agilent, Santa Clara, USA
SpeedVac	Genevac EZ-2 Plus Evaporating System	Genevac, Ipswich, UK
Nanophotometer	Implen NanoPhotometer® N60	Implen GmbH, Munich, Germany
Bioanalyzer	Agilent 1100 HPLC System	Agilent, Santa Clara, USA
Thermocycler	SensoQuest Labcycler	SensoQuest, Göttingen, Germany
electrophorese system	ROTIPHORESE® PROfessional runVIEW	Carl Roth, Karlsruhe, Germany
Imager	Universal Hood II	BioRad, Hercules, USA

5.2 Biochemical methods

Determination of RNA concentration

Concentration of RNA was determined spectrophotometrically using a NanoPhotometer® (Implen GmbH, Munich, Germany).

RNA purification by SEC

The large RNA from viral extracts was purified by size exclusion chromatography. The system was equipped with an AdvanceBio SEC 1000 Å, 2.7 µm, 7.8 × 300 mm column (Agilent, Santa Clara, USA) and 0.1 M NH_4OAc as mobile phase. The separation was conducted with an isocratic elution at 1 ml/min and a column temperature of 40 °C. A maximum of 50 µg of RNA was injected per run. The large RNA fraction was collected from 4 min to 6.2 min. Subsequently, the fraction was concentrated under vacuum to a final volume of 50 µL. The RNA was precipitated by the addition of 0.1 × V NH_4OAc [5 M] and 2.5 × V ice-cold ethanol (100%). After, RNA precipitation over night at -20 °C, the sample was centrifuged (4 °C, 12000 × g, 40 min). The reaction tube was washed with 180 µL ice-cold ethanol (70%) and centrifuged once again (4 °C, 12000 × g, 10 min). The supernatant was discarded and the pellet was left to dry at room temperature for 5 min. Then, the RNA was resuspended in 30 µL ultrapure water.

Bioanalyzer

After preparation of the Agilent 6000 Pico RNA Chip according to the manufacturer's instructions, 1 µL of sample was loaded onto the chip and analyzed in the Agilent 2100 Bioanalyzer using the pre-installed "Eukaryote Total RNA Pico Series II.xsy" method.

Agarose Gel

First, 0.7 g Agarose (Carl Roth, Karlsruhe, Germany) and 0.25 g Synergel (Carl Roth, Karlsruhe, Germany) were suspended in 1 ml ethanol. Next, 84.5 ml ultrapure water was added to the suspension and it was heated in the microwave until the agarose was completely dissolved. The solution was allowed to cool to ~ 50 °C, then 10 ml 10 × MOPS buffer, 5.5 ml formaldehyde and 5 µL GelRed stain were added to the solution. The gel was poured and allowed to solidify. Loading dye in the ratio 1/1 (v/v) was added to the sample and was subsequently heated 10 minutes at 70 °C and then kept on ice for 1 minute. 1 µg of RNA was loaded onto the gel. The gel was run for 90 minutes at 75 V in 1 × MOPS buffer. Finally, the gel was imaged by UV transillumination at 312

nm (Universal Hood II, BioRad, Hercules, USA). The exposure time was varied to optimize the signal-to-background ratio of the image.

DNase I digest

For DNA depletion 2 units of DNase I (NEB, Ipswich, USA) were added to the sample and subsequently incubated 10 minutes at 37 °C. In order to extract the intact RNA, a Trizol (Sigma-Aldrich, Munich, Germany) extraction was performed, according to the manufacturers protocol.

RNA hydrolysis for nucleoside analysis

A master mix for RNA hydrolysis was prepared according to table 11, the given amounts have to be multiplied by the number of samples. 14 µL of master mix is added to 21 µL sample (150-500 ng). The reaction was incubated 2 h at 37 °C. After hydrolysis the sample was diluted with 0.2 × V LC-MS buffer (*e.g.* 35 µL digest + 7 µL LC-MS buffer). For analysis, 1 µL SILIS (10 ×) was co-injected with the sample (*e.g.* 9 µL sample + 1 µL SILIS).

Table 11: Composition of master mix for nucleoside hydrolysis

Compound	Stock conc.		Goal conc.	1x
MgCl$_2$	10 mM	→	1 mM	3.5 µl
Tris (pH 8)	50 mM	→	5 mM	3.5 µl
Benzonase	1 U/µL	→	2 U	2 µl
CIP	1 U/µL	→	2 U	2 µl
PDE1	0.1 U/µL	→	0.2 U	2 µl
Pentostatin	1 mg/mL	→	1 µg	1 µl
BHT	10 mM	→	10 uM	1 µl

RNA hydrolysis for analysis of cap modifications and NMPs

A master mix for RNA hydrolysis was prepared according to table 12, the given amounts have to be multiplied by the number of samples. 5 µL of master mix is added to 15 µL sample (1 µg). The reaction was incubated 1 h at 37 °C. For analysis, 1 µL cap-SILIS was co-injected with the sample (*e.g.* 9 µL sample + 1 µL SILIS).

Table 12: Composition of master mix for nucleoside monophosphate and cap hydrolysis

Compound	Stock conc.		Goal conc.	1x
NH$_4$OAc (pH 5.3)	330 mM	→	50 mM	3 µl
ZnCl$_2$	20 mM	→	1 mM	1 µl
NP1	1 U/µL	→	0.05 U	1 µl

In vitro transcription for cap-SILIS

First, a polymerase chain reaction (PCR) was performed in a total volume of 50 µl with a final concentration of 1 × Phusion Buffer HF (New England Biolabs, Ipswich, USA) and 0.8 µM forward (fwd) and reverse (rev) primer. The sequences of template and primers are given in table 13. Additionally, 1 µl dNTPs, 0.5 µl Phusion polymerase and 100 ng of the DNA template were added. The sample was amplified according to the following PCR program: 95 °C for 2 min, 95 °C for 30 s for 20 amplification cycles, 57 °C for 30 s for 20 times and 68 °C for 1 min for 20 times. At the end of the program, the PCR reaction was incubated at 68° C for 1 min and was cooled down to 4 °C. For each *in vitro* transcription two PCR reactions were pooled. A total of 100 µl PCR product was added to T7 buffer mix and T7 enzyme (Transcript Aid T7 High Yield Transcription Kit, Thermo Fisher Scientific, Waltham, USA) and 1.6 µl of each ^{13}C-rNTP (100 mM) (Silantes,Munich,Germany, SUK: 121206100). The mixture was incubated 2 h at 37 °C and 600 rpm. Then, the sample was treated with 2 µl T7 enzyme mix and 5 µl 50 mM $MgCl_2$ and incubated for additional 2 h. Afterwards, the sample was treated again with 2 µl T7 enzyme mix and 5 µl 50 mM $MgCl_2$ and incubated for additional 2 h. After the transcription, the DNA template was removed by addition of 4 µl DNase I [2U/µL] and incubation for 1 h at 37 °C. Subsequently, the sample was centrifuged (RT, 5000 × g, 5 min) to remove the insoluble pyrophosphate of the transcription reaction. The *in vitro* transcript (IVT) containing supernatant was precipitated by addition of 0.1 × V of 5 M NH_4OAc and 2.5 × V of ice-cold ethanol (100%) followed by overnight incubation at −20 °C. The IVT was pelleted by centrifugation (4 °C, 12000 × g, 40 min). The reaction tube was washed with 180 µL ice-cold ethanol (70%) and centrifuged once again (4 °C, 12000 × g, 10 min). The supernatant was discarded and the pellet was left to dry at room temperature for 5 min. Finally, the IVT was resuspended in 30 µL ultrapure water. Subsequently, the IVT was purified by SEC. The system was equipped with an AdvanceBio SEC 130 Å, 2.7 µm, 7.8 × 300 mm column (Agilent, Santa Clara, USA) and 0.1 M NH_4OAc as mobile phase. The separation was conducted with an isocratic elution at 0.5 ml/min and a column temperature of 60 °C. A maximum of 50 µg of IVT was injected per run. The IVT fraction was collected from 9 min to 10.2 min. Subsequently, the fraction was concentrated under vacuum to a final volume of 50 µL. The IVT was precipitated by the addition of 0.1 × V NH_4OAc [5 M] and 2.5 × V ice-cold ethanol (100%). After, precipitation over night at -20 °C, the sample was centrifuged (4 °C, 12000 × g, 40 min). The reaction tube was washed with 180

µL ice-cold ethanol (70%) and centrifuged once again (4 °C, 12000 × g, 10 min). The supernatant was discarded and the pellet was left to dry at room temperature for 5 min. Then, the IVT was resuspended in 30 µL ultrapure water and hydrolyzed as described under RNA hydrolysis for analysis of cap modifications and NMPs for the generation of the cap-SILIS.

Table 13: List of primers and oligonucleotide used for PCR, the code is an internal identification number from the oligonucleotide database of the Kaiser group (as of July 2022)

target	code	sequence 5' - 3'
fw primer	SK11	TGG CGT AGT CGG C
rev primer	SK10	CGC GCG AAG CTT AAT ACG ACT CAC TAT A
DNA template	SK167	TGGCGTAGTCGGCCTGTGAGTGGCTGTGAG- TGGCTGTGAGTGGCTGTGAGTGGCTGTGAG- TGGCTGTGAGTGGCTGTGAGTGGCTGTGAG- TGGCTGTGAGTGGCTGTGAGTGGTCTCCTTATAGTGAG-

Biotin enriched mung bean assay

For the enrichment of specific RNA sections, the biotin enriched mung bean assay was developed and established. This method consists of a modified combination of the protocols for isoacceptor purification[291] and mung bean assay[266]. For hybridization, 600 pmol of biotin tagged oligonucleotide was incubated with of 60 pmol of yeast 18 S rRNA or 0.9 pmol of viral RNA and 0.3 × V hybridization buffer (MBN) for 5 minutes at 95 °C. Then, the temperature was decreased step-wise over 120 min to 45 °C. After hybridization, 8.75 units of mung bean nuclease, 0.1 × V mung bean reaction buffer and RNase A with a final concentration of 2 ng/µL was added. The mixture was pipetted up and down prior to incubation at 1 h at 35 °C. The hybridized nucleic acids were extracted with Trizol (Sigma-Aldrich, Munich, Germany) according to the manufacturers protocol. After extraction the hybrid was precipitated by addition of 0.1 × V of 5 M NH$_4$OAc and 2.5 × V of ice-cold ethanol (100%) followed by overnight incubation at −20 °C. The hybrid was pelleted by centrifugation (4 °C, 12000 × g, 40 min). The reaction tube was washed with 180 µL ice-cold ethanol (70%) and centrifuged once again (4 °C, 12000 × g, 10 min). The supernatant was discarded and the pellet was left to dry at room temperature for 5 min. Afterwards, the hybrid was resuspended in 75 µL ultrapure water and was enriched by the biotin tag. Therefore, 50 µL streptavidin coated dynabeads (type M280) (Thermo Fisher, Waltham, USA) were washed once with 25 µL B&W buffer and three times with 25 µL 5 × SSC buffer. Afterwards, 25 µL

of 20 × SSC buffer was added to the sample and this solution was given to the strep-tavidin coated magnetic beads and was incubated 30 min at 25 °C. Then, the liquid was discarded and the beads were washed once with 50 µL 1 × SSC buffer and three times with 25 µL 0.1 × SSC buffer. Finally, the beads were resuspended in 20 µL ultrapure water and incubated 3 min at 75 °C to dissociate the RNA from the oligonucleotide. Subsequently, the RNA containing supernatant was further pursued for RNA hydrolysis.

Table 14: Oligonucleotides for biotin enriched mung bean assay, the code is an internal identification number from the oligonucleotide database of the Kaiser group (as of July 2022)

target	code	sequence 5' - 3'
S.c. 18S rRNA	SK171	[Btn]ATCAAATAAACGATAACTGAT-TTAATGAGCCATTCGCAGTTTCACTGTAT
3'UTR SARS-CoV-2	SK192	[Btn]CACAAGAGTAGACTATATATCGTAAACGGAAAA-GCGAAAACGTTTATATAGCCCATCTGCCTTGTG

5.3   Analytical methods

Nucleoside analysis

For quantitative mass spectrometry an Agilent 1290 Infinity II equipped with a diode-array detector (DAD) combined with an Agilent Technologies G6470A Triple Quad system and electrospray ionization (ESI-MS, Agilent Jetstream) was used. Operating parameters: positive ion mode, cell accelerator voltage of 5 V, N_2 gas temperature of 230 °C and N_2 with a flow of 6 L/min, sheath gas (N_2) temperature of 400 °C with a flow of 12 L/min, capillary voltage of 2500 V, nozzle voltage of 0 V, and nebulizer at 40 psi. The instrument was operated in dynamic MRM mode. For separation a Core-Shell Technology column (Synergi, 2.5 µm Fusion-RP, 100 Å, 100 × 2 mm column, Phenomenex, Torrance, USA) at 35 °C and a flow rate of 0.35 ml/min were used in combination with a binary mobile phase of 5 mM NH_4OAc aqueous buffer A, brought to pH 5.6 with glacial acetic acid (65 µL), and an organic buffer B of pure acetonitrile (Roth, LC-MS grade, purity ≥.99.95). The gradient started at 100% solvent A for 1 min, followed by an increase to 10% solvent B over 4 min. From 5 to 7 min, solvent B was increased to 40% and maintained for 1 min before returning to 100% solvent A in 0.5 min and a 2.5 min re-equilibration period. The analyte specific parameters can be taken from table S2.

Calibration for nucleoside analysis

For calibration, synthetic nucleosides were weighed and dissolved in water to a stock concentration of 1–10 mM. The calibration solutions range from 0.3 to 500 pmol for each canonical nucleoside and from 0.3 to 500 fmol for each modified nucleoside and were spiked with 10% SILIS. The sample data were analyzed by the Quantitative Software from Agilent. The areas of the MRM signals were integrated for each modification and their isotope derivatives. The absolute amounts of the modifications were referenced to the absolute amounts of the respective canonical.

Normalization / Quantification of nucleosides per RNA molecule

In order to compare quantities of modifications, the amount was referenced to the amount of canonical nucleosides of the analyzed sequence. Therefore, the calculated molar amounts of canonical nucleosides were divided by their known frequency, taken from the published sequence[30], according to equation 10. Subsequently, the mean value was built to give the molar amount of the whole RNA. In order to determine the

number of modifications per viral RNA, the molar amount of the modification was divided by the injected molar amount of viral RNA.

$$(10) \qquad n(RNA_{SARS-CoV-2}) = \frac{\frac{n(C)}{5493} + \frac{n(U)}{9595} + \frac{n(G)}{5863} + \frac{n(A)}{8955}}{4}$$

Cap modification analysis

For quantitative mass spectrometry an Agilent 1290 Infinity II equipped with a diode-array detector (DAD) combined with an Agilent Technologies G6470A Triple Quad system and electrospray ionization (ESI-MS, Agilent Jetstream) was used. Operating parameters: negative-ion mode, cell accelerator voltage of 5 V, N_2 gas temperature of 340 °C and N_2 with a flow of 12 L/min, sheath gas (N_2) temperature of 400 °C with a flow of 10 L/min, capillary voltage of 2500 V, nozzle voltage of 0 V, and nebulizer at 40 psi. The instrument was operated in dynamic MRM mode. For separation a RP 18 column (Acquity HSS T3, 100 Å, 1.8 μm, 2.1 mm x 150 mm, Waters, Milford, USA) at 40 °C and a flow rate of 0.25 ml/min were used in combination with a binary mobile phase of 5 mM NH_4OAc aqueous buffer A, brought to pH 6.9 with ammonia (7 μL), and an organic buffer B of pure acetonitrile (Roth, LC-MS grade, purity ≥.99.95). The gradient started at 100% solvent A for 5 min, followed by an increase to 70% over 3 min. Then, solvent B was decreased to 0% over 1 min and a re-equilibration of 6 min was carried out. The analyte specific parameters can be taken from table S3.

Calibration for quantification of nucleoside monophosphates and cap modifications

For calibration, synthetic 5' NMPs were weighed and dissolved in water to a stock concentration of 1 mM. The cap analogues were dissolved in 5 mM NH_4OAc (pH 5.3) to a final concentration of 1 mM. The calibration solutions range from 30 fmol to 500 pmol for each 5' NMP and from 3 fmol to 100 pmol for each cap analogue and were spiked with 10% cap-SILIS. The sample data were analyzed by the Quantitative Software from Agilent. The areas of the MRM signals were integrated for each modification. The absolute amounts of the cap modifications were referenced to the absolute amounts of the respective nucleoside monophosphates.

Normalization / Quantification of cap modifications per RNA molecule

In order to compare quantities of cap modifications, the amount was referenced to amount of nucleoside monophosphates in the analyzed sequence. Therefore, the calculated molar amounts of nucleoside monophosphates were divided by their known frequency, according to equation 11. Subsequently, the mean value was built to give the molar amount of the sequence. In order to determine the number of cap modification per sequence, the molar amount of the cap modification was divided by the mean of the nucleoside monophosphates.

$$(11) \qquad n(RNA) = \frac{\frac{n(CMP)}{\#CMP} + \frac{n(UMP)}{\#UMP} + \frac{n(GMP)}{\#GMP} + \frac{n(AMP)}{\#AMP}}{4}$$

6   Appendix

Table S 1: Mean and standard deviation (n=3) of modifications in RNA extracts from SARS-CoV-2 variants. Values for certain modifications can be found in the rows. Values for the different mutants and variants can be taken from the columns. Fields with a blue background highlight the 2'-O-methylated nucleosides. Fields with a green background highlight the base methylated modifications.

#per RNA	D614		G614		alpha		beta		delta		Caco-2	
	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
Y	140.0	±25.9	134.0	±26.4	100.2	±21.4	112.4	±21.3	86.2	±14.3	297.9	±2.46
I	81.0	±4.0	80.3	±16.0	34.0	±12.2	79.0	±20.1	34.4	±8.7	20.6	±14.1
Am	54.9	±11.1	51.7	±6.01	41.4	±6.3	47.0	±9.2	42.7	±2.6	103.5	±1.7
Cm	79.7	±13.9	75.3	±9.4	65.8	±11.9	69.3	±14.0	72.9	±5.9	147.5	±1.6
Gm	54.0	±9.9	51.6	±6.0	41.4	±7.1	46.1	±8.4	42.9	±8.4	108.4	±1.7
Um	38.0	±6.9	38.2	±6.2	32.2	±9.3	32.6	±5.7	28.9	±5.6	69.8	±3.0
m^5C	2.4	±0.5	4.1	±0.3	4.8	±0.3	5.3	±1.6	7.0	±0.4	7.3	±0.4
m^3U	1.0	±0.4	0.8	±0.2	0.3	±0.2	0.4	±0.04	0.2	±0.05	2.9	±0.2
m^1A	3.2	±0.4	3.2	±0.01	4.2	±0.1	4.6	±1.5	6.4	±0.1	4.5	±0.2
m^6A	2.8	±0.7	2.7	±0.4	2.0	±0.4	2.2	±0.4	1.7	±0.1	6.3	±0.07
$m^{66}A$	0.7	±0.2	0.9	±0.2	0.4	±0.2	0.6	±0.2	0.4	±0.2	1.6	±0.2
m^1G	0.6	±0.1	0.7	±0.01	1.2	±0.04	0.9	±0.6	1.4	±0.3	0.5	±0.06
m^2G	1.8	±0.3	2.1	±0.08	3.5	±0.05	3.9	±1.5	6.9	±0.5	1.8	±0.3
m^7G	3.7	±0.4	3.5	±0.4	3.3	±0.2	3.7	±0.6	3.6	±0.004	4.7	±0.2
$m^{2,2}G$	0.6	±0.1	0.7	±0.1	1.3	±0.02	1.3	±0.5	2.5	±0.2	0.5	±0.05
$m^{2,2,7}G$	3.9	±1.0	3.3	±0.8	2.3	±0.1	5.1	±0.6	8.4	±2.5	3.6	±0.8

Table S 2: Analyte specific parameters for nucleoside analysis, *m/z* precursor ion, *m/z* product ion, retention time (t_R), fragmentor voltage (Frag.), collision energy (CE)

Compound Name	Precursor Ion	Product Ion	t_R [min]	Frag. [V]	CE [eV]
A	268	136	5.20	200	21
A SILIS	283	146	5.19	200	21
ac^4C	286	153	5.04	85	9
ac^4C SILIS	301	163	5.04	85	9
ac^4Cm	300	154	6.11	80	5
Am	282	136	6.04	110	17
Am SILIS	298	146	6.04	110	17
C	244	112	1.74	175	13
C SILIS	256	119	1.74	175	13
Cm	258	112	3.80	180	9
Cm SILIS	271	119	3.80	180	9
D	247	115	1.56	70	5
D SILIS	258	121	1.58	70	5
G	284	152	4.08	130	17
G SILIS	299	162	4.09	130	17

Compound Name	Precursor Ion	Product Ion	t_R [min]	Frag. [V]	CE [eV]
Gm	298	152	4.88	100	9
Gm SILIS	314	162	4.88	100	9
i^6A	336	204	7.94	140	17
i^6A SILIS	356	219	7.94	140	17
I	269	137	3.94	100	10
I SILIS	283	146	3.94	100	10
Im	283	137	4.79	75	9
m^1A	282	150	1.68	110	21
m^1A SILIS	298	161	1.68	110	21
m^1Am	296	150	3.64	100	20
m^1G	298	166	4.99	105	13
m^1G SILIS	314	177	4.99	105	13
m^1I	283	151	4.71	80	12
m^1I SILIS	298	161	4.71	80	12
m^2A	282	150	6.28	110	21
m^2G	298	166	4.99	95	17
m^2G SILIS	314	177	4.99	95	17
m^3C	258	126	1.66	88	14
m^3C SILIS	271	134	1.66	88	14
m^3U	259	127	4.66	75	9
m^4Cm	272	126	4.63	80	12
m^5C	258	126	3.54	185	13
m^5C SILIS	271	134	3.55	185	13
m^5Cm	272	126	4.63	190	13
m^5U	259	127	4.19	95	9
m^5U SILIS	271	134	4.19	95	9
m^5Um	273	127	5.49	90	13
m^6A	282	150	6.46	125	17
m^6A SILIS	298	161	6.47	125	17
m^6Am	296	150	6.92	125	17
m^7G	298	166	3.10	100	13
m^7G SILIS	314	177	3.11	100	13
$m^{22}G$	312	180	5.64	105	13
$m^{22}G$ SILIS	329	192	5.64	105	13
$m^{66}A$	296	164	7.10	130	21
$m^{227}G$	326	194	4.92	110	15
$m^{227}G$ SILIS	344	207	4.92	110	15
mcm^5s^2U	333	201	6.26	95	8

Compound Name	Precursor Ion	Product Ion	t_R [min]	Frag. [V]	CE [eV]
mcm^{5}s^{2}U SILIS	347	210	6.26	95	8
t^{6}A	413	281	5.80	130	9
t^{6}A SILIS	434	297	5.80	130	9
U	245	113	2.65	95	5
U SILIS	256	119	2.66	95	5
Um	259	113	4.50	96	8
Um SILIS	271	119	4.50	96	8
Y	245	209	1.64	80	5
Y SILIS	256	220	1.65	80	5

Table S 3: Analyte specific parameters for NMP and cap analysis, *m/z* precursor ion, *m/z* product ion, retention time (t_R), fragmentor voltage (Frag.), collision energy (CE)

Compound Name	Precursor Ion	Product Ion	t_R [min]	Frag. [V]	CE [eV]
AMP	346	79	3.25	250	37
AMP SILIS	356	79	3.25	250	37
CMP	322	79	1.44	250	37
CMP SILIS	332	79	1.44	250	37
GMP	362	79	1.96	250	37
GMP SILIS	372	79	1.96	250	37
UMP	323	79	1.59	250	37
UMP SILIS	333	79	1.59	250	37
GpppA	771	424	2.69	220	35
GpppG	787	424	1.85	220	35
m^{7}GpppA	785	408	4.80	220	35
m^{7}GpppA SILIS	805	418	4.80	220	35
m^{7}GpppAm	799	422	9.00	220	35
m^{7}GpppAm SILIS	819	432	9.00	220	35
m^{7}GpppG	801	424	2.76	250	37
m^{7}GmpppG	815	424	4.90	220	40

Table S 4: List of abbreviations

Abbreviation	Definition
°C	degree Celsius
µg	microgram
µL	microliter
µM	micromolar
Å	Ångström
AMP	Adenosine monophosphate
AUC	area under curve

Abbreviation	Definition
B&W	binding and wash
BHT	butylated hydroxytoluene
Caco-2 cells	caucasian colon adenocarcinoma cells
CE	capillary electrophoresis
CID	collision induced dissociation
CIP	calf intestinal phosphatase
CMP	Cytidine monophosphate
Cryo EM	cryogenic electron microscopy
dMRM	dynamic multiple reaction monitoring
DNA	deoxyribonucleic acid
Dnase I	deoxyribonuclease I
dNTP	2'-deoxyribonucleoside triphosphate
DRS	direct RNA sequencing
E. coli	*Escheria coli*
EDTA	ethylenediaminetetraacetic acid
fmol	femtomol
g	gram
GMP	Guanosine monophosphate
gRNA	genomic RNA
GTP	Guanosine triphosphate
H	plate height
h	hours
HEK cells	human embryonic kidney cells
hpi	hours post transfection
HPLC	high performance liquid chromatography
Huh7 cells	human hepatoma cells
IUPAC	International Union of Pure and Applied Chemistry
IVT	*in vitro* transcript
k	retention factor
kDa	kilodalton
kV	kilovolt
LC-MS/MS	liquid chromatography coupled tandem mass
LOD	lower limit of detection
LOL	limit of linearity
LOQ	lower limit of quantification
m/z	mass-to-charge
MALDI	matrix assisted laser desorption ionization
MBN	mung bean nuclease
MeRIP	methylated RNA immunoprecipitation sequencing
mg	milligram
min	minutes
mL	milliliter

Abbreviation	Definition
mM	millimolar
MMS	methyl methanesulfonate
MOPS	3-(N-morpholino)propanesulfonic acid
MRM	multiple reaction monitoring
mRNA	messenger RNA
MS	mass spectrometer
N	theoretical plate number
NAIL-MS	nucleic acid isotope labeling coupled mass spectrometry
ng	nanogram
NIF	nucleoside isotope factor
nm	nanometer
nM	nanomolar
NMP	nucleoside monophosphate
NMR	nuclear magnetic resonance spectroscopy
NP1	nuclease P1
nts	nucleosides
PDE1	phosphodiesterase 1
Phe	Phenylalanine
pmol	picomol
QQQ	Triple-Quadrupol
RNA	ribonucleic acid
RNase A	bovine pancreatic ribonuclease A
rNTP	ribonucleoside triphosphate
ROS	reactive oxygen species
RP	reversed phase
rpm	revolutions per minute
rRNA	ribosomal RNA
rRNF	relative response factor for nucleoside
R_s	resolution
RT	room temperature
S. cerevisiae	*Saccharomyces cerevisiae*
S/N	signal-to-noise
SAM	S-Adenosyl methionine
SEC	size exclusion chromatography
SILAC	isotope labeling with amino acids in cell culture
SILIS	isotope labeled internal standard
SIM	selected ion monitoring
SL	stem loop
SRM	selected reaction monitoring
SSC	saline sodium citrate
t_0	void time

Abbreviation	Definition
THU	tetrahydrouridine
TIC	total ion current chromatogram
TOF	time of flight
t_R	retention time
t'_R	netto retention time
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
UMP	uridine monophosphate
UV	ultra violet
V	volt
Vero cells	kidney epithelial cells from african green monkey
VOC	variant of concern
w	peak width at base line
$w_{0.5}$	peak width at half maximum
α	selectivity

7   References

[1.]   Griffith, F., The Significance of Pneumococcal Types. *J Hyg (Lond),* **1928,** *27* (2), 113-59.

[2.]   Watson, J. D.; Crick, F. H. C., Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature,* **1953,** *171* (4356), 737-738.

[3.]   Crick, F., Central Dogma of Molecular Biology. *Nature,* **1970,** *227* (5258), 561-563.

[4.]   Jordan, A.; Reichard, P., RIBONUCLEOTIDE REDUCTASES. *Annual Review of Biochemistry,* **1998,** *67* (1), 71-98.

[5.]   Trun, N. J.; Marko, J. F., Architecture of a bacterial chromosome. *Asm News,* **1998,** *64* (5), 276-283.

[6.]   Holmes, V. F.; Cozzarelli, N. R., Closing the ring: Links between SMC proteins and chromosome partitioning, condensation, and supercoiling. *Proceedings of the National Academy of Sciences,* **2000,** *97* (4), 1322-1324.

[7.]   Ochman, H.; Caro-Quintero, A., Genome Size and Structure, Bacterial. In *Encyclopedia of Evolutionary Biology*, Kliman, R. M., Ed. Academic Press: Oxford, **2016**; pp 179-185.

[8.]   Goffeau, A.*, et al.*, Life with 6000 genes. *Science,* **1996,** *274* (5287), 546, 563-7.

[9.]   Lander, E. S.*, et al.*, Initial sequencing and analysis of the human genome. *Nature,* **2001,** *409* (6822), 860-921.

[10.]   Palazzo, A. F.; Lee, E. S., Non-coding RNA: what is functional and what is junk? *Front Genet,* **2015,** *6*, 2.

[11.]   Iserentant, D.; Fiers, W., Secondary structure of mRNA and efficiency of translation initiation. *Gene,* **1980,** *9* (1-2), 1-12.

[12.]   Hall, M. N.*, et al.*, A role for mRNA secondary structure in the control of translation initiation. *Nature,* **1982,** *295* (5850), 616-618.

[13.]   Renciuk, D.*, et al.*, Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine. *Nucleic Acids Res,* **2013,** *41* (21), 9891-900.

[14.]   Chittori, S.*, et al.*, Structure of the primed state of the ATPase domain of chromatin remodeling factor ISWI bound to the nucleosome. *Nucleic Acids Res,* **2019,** *47* (17), 9400-9409.

[15.]   Shi, H.; Moore, P. B., The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. *Rna,* **2000,** *6* (8), 1091-105.

[16.]   Thaplyal, P.; Bevilacqua, P. C., Chapter Nine - Experimental Approaches for Measuring pKa's in RNA and DNA. In *Methods in Enzymology*, Burke-Aguero, D. H., Ed. Academic Press: **2014**; Vol. 549, pp 189-219.

[17.]   Glotz, C.*, et al.*, Secondary structure of the large subunit ribosomal RNA from Escherichia coli, Zea mays chloroplast, and human and mouse mitochondrial ribosomes. *Nucleic Acids Res,* **1981,** *9* (14), 3287-306.

[18.]   Maly, P.; Brimacombe, R., Refined secondary structure models for the 16S and 23S ribosomal RNA of Escherichia coli. *Nucleic Acids Res,* **1983,** *11* (21), 7263-86.

[19.]   Darlix, J. L.; Rochaix, J. D., Nucleotide sequence and structure of cytoplasmic 5S RNA and 5.8S RNA of Chlamydomonas reinhardii. *Nucleic Acids Res,* **1981,** *9* (6), 1291-9.

[20.]   Noller, H. F.*, et al.*, Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Res,* **1981,** *9* (22), 6167-89.

[21.] Neefs, J.-M.; De Wachter, R., A proposal for the secondary structure of a variable area of eukaryotic small ribosomal subunit RNA involving the existence of a pseudoknot. *Nucleic Acids Research,* **1990,** *18* (19), 5695-5704.

[22.] Nissen, P.*, et al.*, The structural basis of ribosome activity in peptide bond synthesis. *Science,* **2000,** *289* (5481), 920-30.

[23.] Boccaletto, P.*, et al.*, MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Research,* **2017,** *46* (D1), D303-D307.

[24.] Sharp, S. J.*, et al.*, Structure and Transcription of Eukaryotic tRNA Gene. *Critical Reviews in Biochemistry,* **1985,** *19* (2), 107-144.

[25.] Roundtree, I. A.*, et al.*, Dynamic RNA Modifications in Gene Expression Regulation. *Cell,* **2017,** *169* (7), 1187-1200.

[26.] Gelderblom, H., Structure and Classification of Viruses. In *Medical Microbiology*, S, B., Ed. University of Texas Medical Branch at Galveston, **1996**.

[27.] Callaway, E.*, et al.*, Six months of coronavirus: the mysteries scientists are still racing to solve. *Nature,* **2020**, 178-179.

[28.] Duffy, S.*, et al.*, Rates of evolutionary change in viruses: patterns and determinants. *Nature Reviews Genetics,* **2008,** *9* (4), 267-276.

[29.] Huang, C. R.; Lo, S. J., Evolution and diversity of the human hepatitis d virus genome. *Adv Bioinformatics,* **2010,** *2010*, 323654.

[30.] Wu, F.*, et al.*, A new coronavirus associated with human respiratory disease in China. *Nature,* **2020,** *579* (7798), 265-269.

[31.] Poltronieri, P.*, et al.*, RNA Viruses: RNA Roles in Pathogenesis, Coreplication and Viral Load. *Curr Genomics,* **2015,** *16* (5), 327-35.

[32.] Hallgrímsson, B.; Hall, B. K., Epigenetics: linking genotype and phenotype in development and evolution. Univ of California Press: **2011**.

[33.] Raiber, E.-A.*, et al.*, Mapping and elucidating the function of modified bases in DNA. *Nature Reviews Chemistry,* **2017,** *1* (9), 0069.

[34.] Furuichi, Y.; Shatkin, A. J., Characterization of cap structures. *Methods Enzymol,* **1989,** *180*, 164-76.

[35.] Holliday, R.; Pugh, J. E., DNA modification mechanisms and gene activity during development. *Science,* **1975,** *187* (4173), 226-32.

[36.] Auclair, G.; Weber, M., Mechanisms of DNA methylation and demethylation in mammals. *Biochimie,* **2012,** *94* (11), 2202-11.

[37.] Ndlovu, M. N.*, et al.*, Exposing the DNA methylome iceberg. *Trends Biochem Sci,* **2011,** *36* (7), 381-7.

[38.] Laurent, L.*, et al.*, Dynamic changes in the human methylome during differentiation. *Genome Res,* **2010,** *20* (3), 320-31.

[39.] Guo, J. U.*, et al.*, Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell,* **2011,** *145* (3), 423-34.

[40.] Masala, L.*, et al.*, Methylation dynamics during folliculogenesis and early embryo development in sheep. *Reproduction,* **2017,** *153* (5), 605-619.

[41.] Kellinger, M. W.*, et al.*, 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nature structural & molecular biology,* **2012,** *19* (8), 831-833.

[42.] Messer, W.; Noyer-Weidner, M., Timing and targeting: the biological functions of Dam methylation in E. coli. *Cell,* **1988,** *54* (6), 735-737.

[43.] Zaleski, P.*, et al.*, The role of Dam methylation in phase variation of Haemophilus influenzae genes involved in defence against phage infection. *Microbiology,* **2005,** *151* (10), 3361-3369.

[44.] Sarnacki, S. H.*, et al.*, Dam methylation participates in the regulation of PmrA/PmrB and RcsC/RcsD/RcsB two component regulatory systems in Salmonella enterica serovar Enteritidis. *PLoS One,* **2013,** *8* (2), e56474.

[45.] Fu, Y.*, et al.*, N6-methyldeoxyadenosine marks active transcription start sites in Chlamydomonas. *Cell,* **2015,** *161* (4), 879-892.

[46.] Mondo, S. J.*, et al.*, Widespread adenine N6-methylation of active genes in fungi. *Nature genetics,* **2017,** *49* (6), 964-968.

[47.] Parashar, N. C.*, et al.*, N6-adenine DNA methylation demystified in eukaryotic genome: From biology to pathology. *Biochimie,* **2018,** *144*, 56-62.

[48.] Koziol, M. J.*, et al.*, Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nat Struct Mol Biol,* **2016,** *23* (1), 24-30.

[49.] Huang, W.*, et al.*, Determination of DNA adenine methylation in genomes of mammals and plants by liquid chromatography/mass spectrometry. *RSC Advances,* **2015,** *5* (79), 64046-64054.

[50.] Liu, J.*, et al.*, Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. *Nat Commun,* **2016,** *7*, 13052.

[51.] Liang, D.*, et al.*, The decreased N(6)-methyladenine DNA modification in cancer cells. *Biochem Biophys Res Commun,* **2016,** *480* (1), 120-125.

[52.] Law, J. A.; Jacobsen, S. E., Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics,* **2010,** *11* (3), 204-220.

[53.] He, Y.-F.*, et al.*, Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science,* **2011,** *333* (6047), 1303-1307.

[54.] Esteller, M., Epigenetics in cancer. *New England Journal of Medicine,* **2008,** *358* (11), 1148-1159.

[55.] Chistiakov, D. A.*, et al.*, Treatment of cardiovascular pathology with epigenetically active agents: Focus on natural and synthetic inhibitors of DNA methylation and histone deacetylation. *International Journal of Cardiology,* **2017,** *227*, 66-82.

[56.] Nestler, E. J.*, et al.*, Epigenetic basis of mental illness. *The Neuroscientist,* **2016,** *22* (5), 447-463.

[57.] Gassen, N. C.*, et al.*, Life stress, glucocorticoid signaling, and the aging epigenome: implications for aging-related diseases. *Neuroscience & Biobehavioral Reviews,* **2017,** *74*, 356-365.

[58.] Wahl, S.*, et al.*, Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature,* **2017,** *541* (7635), 81-86.

[59.] Ho, S.-M.*, et al.*, Environmental factors, epigenetics, and developmental origin of reproductive disorders. *Reproductive Toxicology,* **2017,** *68*, 85-104.

[60.] Lorenz, C.*, et al.*, tRNA Modifications: Impact on Structure and Thermal Adaptation. *Biomolecules,* **2017,** *7* (2).

[61.] Lyons, S. M.*, et al.*, The role of RNA modifications in the regulation of tRNA cleavage. *FEBS Lett,* **2018,** *592* (17), 2828-2844.

[62.] Grosjean, H.*, et al.*, Studies of the complex between transfer RNAs with complementary anticodons. I. Origins of enhanced affinity between complementary triplets. *J Mol Biol,* **1976,** *103* (3), 499-519.

[63.] Konevega, A. L.*, et al.*, Purine bases at position 37 of tRNA stabilize codon-anticodon interaction in the ribosomal A site by stacking and Mg2+-dependent interactions. *Rna,* **2004,** *10* (1), 90-101.

[64.] Cohn, W. E.; Volkin, E., Nucleoside-5'-Phosphates from Ribonucleic Acid. *Nature,* **1951,** *167*, 483-484.

[65.] Jühling, F., *et al.*, tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res,* **2009,** *37* (Database issue), D159-62.

[66.] Salditt-Georgieff, M., *et al.*, The addition of 5' cap structures occurs early in hnRNA synthesis and prematurely terminated molecules are capped. *Cell,* **1980,** *19* (1), 69-78.

[67.] Furuichi, Y., *et al.*, 5'-Terminal m-7G (5') ppp (5') Gmp in vivo: identification in reovirus genome RNA. *Proceedings of the National Academy of Sciences,* **1975,** *72* (2), 742-745.

[68.] Shimotohno, K.; Miura, K.-i., The process of formation of the 5'-terminal modified structure in messenger RNA of cytoplasmic polyhedrosis virus. *FEBS letters,* **1976,** *64* (1), 204-208.

[69.] Shuman, S., Capping enzyme in eukaryotic mRNA synthesis. *Progress in nucleic acid research and molecular biology,* **1995,** *50,* 101-129.

[70.] McIntyre, W., *et al.*, Positive-sense RNA viruses reveal the complexity and dynamics of the cellular and viral epitranscriptomes during infection. *Nucleic Acids Research,* **2018,** *46* (11), 5776-5791.

[71.] Courtney, D. G., *et al.*, Epitranscriptomic addition of m5C to HIV-1 transcripts regulates viral gene expression. *Cell host & microbe,* **2019,** *26* (2), 217-227. e6.

[72.] Tong, J., *et al.*, The Emerging Role of RNA Modifications in the Regulation of Antiviral Innate Immunity. *Frontiers in Microbiology,* **2022,** *13.*

[73.] Kaiser, S., *et al.*, The RNA methyltransferase Dnmt2 methylates DNA in the structural context of a tRNA. *RNA Biol,* **2017,** *14* (9), 1241-1251.

[74.] Goll, M. G., *et al.*, Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science,* **2006,** *311* (5759), 395-8.

[75.] Okano, M., *et al.*, DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell,* **1999,** *99* (3), 247-57.

[76.] Chédin, F., *et al.*, The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proceedings of the National Academy of Sciences,* **2002,** *99* (26), 16916-16921.

[77.] Marinus, M. G.; Morris, N. R., Isolation of deoxyribonucleic acid methylase mutants of Escherichia coli K-12. *J Bacteriol,* **1973,** *114* (3), 1143-50.

[78.] Geier, G. E.; Modrich, P., Recognition sequence of the dam methylase of Escherichia coli K12 and mode of cleavage of Dpn I endonuclease. *J Biol Chem,* **1979,** *254* (4), 1408-13.

[79.] May, M. S.; Hattman, S., Analysis of bacteriophage deoxyribonucleic acid sequences methylated by host- and R-factor-controlled enzymes. *J Bacteriol,* **1975,** *123* (2), 768-70.

[80.] Hattman, S., *et al.*, Comparative study of DNA methylation in three unicellular eucaryotes. *J Bacteriol,* **1978,** *135* (3), 1156-7.

[81.] Proffitt, J. H., *et al.*, 5-Methylcytosine is not detectable in Saccharomyces cerevisiae DNA. *Mol Cell Biol,* **1984,** *4* (5), 985-8.

[82.] Bass, B. L.; Weintraub, H., An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell,* **1988,** *55* (6), 1089-98.

[83.] Wagner, R. W., *et al.*, A double-stranded RNA unwinding activity introduces structural alterations by means of adenosine to inosine conversions in mammalian cells and Xenopus eggs. *Proc Natl Acad Sci U S A,* **1989,** *86* (8), 2647-51.

[84.] Xu, L., *et al.*, Three distinct 3-methylcytidine (m(3)C) methyltransferases modify tRNA and mRNA in mice and humans. *J Biol Chem,* **2017,** *292* (35), 14695-14703.

[85.] Ignatova, V. V*., et al.*, METTL6 is a tRNA m(3)C methyltransferase that regulates pluripotency and tumor cell growth. *Sci Adv,* **2020,** *6* (35), eaaz4551.

[86.] Wang, X*., et al.*, Structural basis of N(6)-adenosine methylation by the METTL3-METTL14 complex. *Nature,* **2016,** *534* (7608), 575-8.

[87.] Alexandrov, A*., et al.*, Two proteins that form a complex are required for 7-methylguanosine modification of yeast tRNA. *Rna,* **2002,** *8* (10), 1253-1266.

[88.] Anderson, J*., et al.*, The essential Gcd10p-Gcd14p nuclear complex is required for 1-methyladenosine modification and maturation of initiator methionyl-tRNA. *Genes Dev,* **1998,** *12* (23), 3650-62.

[89.] D'Silva, S*., et al.*, A domain of the actin binding protein Abp140 is the yeast methyltransferase responsible for 3-methylcytidine modification in the tRNA anti-codon loop. *Rna,* **2011,** *17* (6), 1100-10.

[90.] Noma, A*., et al.*, Actin-binding protein ABP140 is a methyltransferase for 3-methylcytidine at position 32 of tRNAs in Saccharomyces cerevisiae. *Rna,* **2011,** *17* (6), 1111-9.

[91.] Tomikawa, C., 7-Methylguanosine Modifications in Transfer RNA (tRNA). *International Journal of Molecular Sciences,* **2018,** *19* (12), 4080.

[92.] Leiros, I*., et al.*, Structural basis for enzymatic excision of N1-methyladenine and N3-methylcytosine from DNA. *The EMBO journal,* **2007,** *26* (8), 2206-2217.

[93.] Bélanger, F*., et al.*, Characterization of hMTr1, a human Cap1 2'-O-ribose methyltransferase. *J Biol Chem,* **2010,** *285* (43), 33037-33044.

[94.] Bouvet, M*., et al.*, RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proceedings of the National Academy of Sciences,* **2012,** *109* (24), 9372-9377.

[95.] Bouvet, M*., et al.*, In vitro reconstitution of SARS-coronavirus mRNA cap methylation. *PLoS pathogens,* **2010,** *6* (4), e1000863.

[96.] Chen, Y*., et al.*, Biochemical and structural insights into the mechanisms of SARS coronavirus RNA ribose 2'-O-methylation by nsp16/nsp10 protein complex. *PLoS pathogens,* **2011,** *7* (10), e1002294.

[97.] de Jager, T. L*., et al.*, Ultraviolet Light Induced Generation of Reactive Oxygen Species. *Adv Exp Med Biol,* **2017,** *996*, 15-23.

[98.] Reichle, V. F*., et al.*, NAIL-MS in E. coli determines the source and fate of methylation in tRNA. *ChemBioChem,* **2018**.

[99.] Yoluç, Y*., et al.*, The Stress-Dependent Dynamics of Saccharomyces cerevisiae tRNA and rRNA Modification Profiles. *Genes,* **2021,** *12* (9), 1344.

[100.] Jackson, A. L.; Loeb, L. A., The contribution of endogenous sources of DNA damage to the multiple mutations in cancer. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* **2001,** *477* (1-2), 7-21.

[101.] Friedberg, E. C*., et al.*, DNA repair and mutagenesis. American Society for Microbiology Press: **2005**.

[102.] Dizdaroglu, M., Oxidatively induced DNA damage: mechanisms, repair and disease. *Cancer letters,* **2012,** *327* (1-2), 26-47.

[103.] Jonkhout, N*., et al.*, The RNA modification landscape in human disease. *Rna,* **2017,** *23* (12), 1754-1769.

[104.] Moreira, P. I*., et al.*, Nucleic acid oxidation in Alzheimer disease. *Free Radical Biology and Medicine,* **2008,** *44* (8), 1493-1505.

[105.] Nunomura, A*., et al.*, Oxidative RNA damage and neurodegeneration. *Current Medicinal Chemistry,* **2007,** *14* (28), 2968-2975.

[106.] Goodman, L. S*., et al.*, Nitrogen mustard therapy. Use of methyl-bis(beta-chloroethyl)amine hydrochloride and tris(beta-chloroethyl)amine hydrochloride

for Hodgkin's disease, lymphosarcoma, leukemia and certain allied and miscellaneous disorders. *Jama,* **1984,** *251* (17), 2255-61.

[107.] Hansson, J.*, et al.*, Formation and removal of DNA cross-links induced by melphalan and nitrogen mustard in relation to drug-induced cytotoxicity in human melanoma cells. *Cancer Res,* **1987,** *47* (10), 2631-7.

[108.] Hsu, G. W.*, et al.*, Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase. *Nature,* **2004,** *431* (7005), 217-21.

[109.] Simms, C. L.*, et al.*, An active role for the ribosome in determining the fate of oxidized mRNA. *Cell Rep,* **2014,** *9* (4), 1256-64.

[110.] Ezaz-Nikpay, K.; Verdine, G. L., The effects of N7-methylguanine on duplex DNA structure. *Chem Biol,* **1994,** *1* (4), 235-40.

[111.] Zahn, K. E.*, et al.*, The miscoding potential of 5-hydroxycytosine arises due to template instability in the replicative polymerase active site. *Biochemistry,* **2011,** *50* (47), 10350-8.

[112.] Singer, B. G., D. , Molecular Biology of Mutagens and Carcinogens. Springer, New York, NY **1983**.

[113.] Tahiliani, M.*, et al.*, Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science,* **2009,** *324* (5929), 930-5.

[114.] Pfaffeneder, T.*, et al.*, Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat Chem Biol,* **2014,** *10* (7), 574-81.

[115.] Ito, S.*, et al.*, Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science,* **2011,** *333* (6047), 1300-3.

[116.] He, Y. F.*, et al.*, Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science,* **2011,** *333* (6047), 1303-7.

[117.] Zheng, G.*, et al.*, Nucleic acid oxidation in DNA damage repair and epigenetics. *Chem Rev,* **2014,** *114* (8), 4602-20.

[118.] Lindahl, T.*, et al.*, Regulation and expression of the adaptive response to alkylating agents. *Annual review of biochemistry,* **1988,** *57* (1), 133-157.

[119.] Mielecki, D.; Grzesiuk, E., Ada response – a strategy for repair of alkylated DNA in bacteria. *FEMS Microbiology Letters,* **2014,** *355* (1), 1-11.

[120.] Schaerer, O. D.*, et al.*, Specific binding of the DNA repair enzyme AlkA to a pyrrolidine-based inhibitor. *Journal of the American Chemical Society,* **1995,** *117* (24), 6623-6624.

[121.] Trewick, S. C.*, et al.*, Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature,* **2002,** *419* (6903), 174-8.

[122.] Falnes, P.*, et al.*, AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. *Nature,* **2002,** *419* (6903), 178-82.

[123.] Falnes, P., Repair of 3-methylthymine and 1-methylguanine lesions by bacterial and human AlkB proteins. *Nucleic Acids Res,* **2004,** *32* (21), 6260-7.

[124.] Delaney, J. C.; Essigmann, J. M., Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB Escherichia coli. *Proceedings of the National Academy of Sciences,* **2004,** *101* (39), 14051-14056.

[125.] Li, D.*, et al.*, Exocyclic Carbons Adjacent to the N6 of Adenine are Targets for Oxidation by the Escherichia coli Adaptive Response Protein AlkB. *Journal of the American Chemical Society,* **2012,** *134* (21), 8896-8901.

[126.] Reichle, V. F.*, et al.*, NAIL-MS reveals the repair of 2-methylthiocytidine by AlkB in E. coli. *Nat Commun,* **2019,** *10* (1), 5600.

[127.] Zhang, M.*, et al.*, Mammalian ALKBH1 serves as an N(6)-mA demethylase of unpairing DNA. *Cell Res,* **2020,** *30* (3), 197-210.

[128.] Liu, F*., et al.*, ALKBH1-Mediated tRNA Demethylation Regulates Translation. *Cell,* **2016,** *167* (7), 1897.

[129.] Westbye, M. P*., et al.*, Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA. *J Biol Chem,* **2008,** *283* (36), 25046-56.

[130.] Ma, C. J*., et al.*, AlkB Homologue 1 Demethylates N(3)-Methylcytidine in mRNA of Mammals. *ACS Chem Biol,* **2019,** *14* (7), 1418-1425.

[131.] Haag, S*., et al.*, NSUN3 and ABH1 modify the wobble position of mt-tRNAMet to expand codon recognition in mitochondrial translation. *Embo j,* **2016,** *35* (19), 2104-2119.

[132.] Duncan, T*., et al.*, Reversal of DNA alkylation damage by two human dioxygenases. *Proc Natl Acad Sci U S A,* **2002,** *99* (26), 16660-5.

[133.] Aas, P. A*., et al.*, Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature,* **2003,** *421* (6925), 859-63.

[134.] Zheng, G*., et al.*, ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Molecular cell,* **2013,** *49* (1), 18-29.

[135.] Zheng, G*., et al.*, Sprouts of RNA epigenetics: the discovery of mammalian RNA demethylases. *RNA Biol,* **2013,** *10* (6), 915-8.

[136.] Jia, G*., et al.*, N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol,* **2011,** *7* (12), 885-7.

[137.] Fu, Y*., et al.*, FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. *Nat Commun,* **2013,** *4*, 1798.

[138.] Wang, Z*., et al.*, The hDcp2 protein is a mammalian mRNA decapping enzyme. *Proc Natl Acad Sci U S A,* **2002,** *99* (20), 12663-8.

[139.] Mauer, J*., et al.*, Reversible methylation of m(6)A(m) in the 5' cap controls mRNA stability. *Nature,* **2017,** *541* (7637), 371-375.

[140.] Cheng, H*., et al.*, Human mRNA export machinery recruited to the 5′ end of mRNA. *Cell,* **2006,** *127* (7), 1389-1400.

[141.] Korvald, H*., et al.*, The Schizosaccharomyces pombe AlkB homolog Abh1 exhibits AP lyase activity but no demethylase activity. *DNA Repair,* **2012,** *11* (5), 453-462.

[142.] Korvald, H*., et al.*, Schizosaccharomyces pombe Ofd2 Is a Nuclear 2-Oxoglutarate and Iron Dependent Dioxygenase Interacting with Histones. *PLOS ONE,* **2011,** *6* (9), e25188.

[143.] Shivange, G*., et al.*, A role for Saccharomyces cerevisiae Tpa1 protein in direct alkylation repair. *The Journal of biological chemistry,* **2014,** *289* (52), 35939-35952.

[144.] Henri, J*., et al.*, Structural and Functional Insights into Saccharomyces cerevisiae Tpa1, a Putative Prolylhydroxylase Influencing Translation Termination and Transcription. *Journal of Biological Chemistry,* **2010,** *285* (40), 30767-30778.

[145.] Admiraal, S. J*., et al.*, Expansion of base excision repair compensates for a lack of DNA repair by oxidative dealkylation in budding yeast. *J Biol Chem,* **2019**.

[146.] Chang, M*., et al.*, A genome-wide screen for methyl methanesulfonate-sensitive mutants reveals genes required for S phase progression in the presence of DNA damage. **2002,** *99* (26), 16934-16939.

[147.] Berdal, K. G*., et al.*, Cloning and expression in Escherichia coli of a gene for an alkylbase DNA glycosylase from Saccharomyces cerevisiae; a homologue to the bacterial alkA gene. *Embo j,* **1990,** *9* (13), 4563-8.

[148.] Popoff, S. C.*, et al.*, Yeast structural gene (APN1) for the major apurinic endonuclease: homology to Escherichia coli endonuclease IV. *Proc Natl Acad Sci U S A,* **1990,** *87* (11), 4193-7.

[149.] Ramotar, D.*, et al.*, Cellular role of yeast Apn1 apurinic endonuclease/3'-diesterase: repair of oxidative and alkylation DNA damage and control of spontaneous mutation. *Molecular and Cellular Biology,* **1991,** *11* (9), 4537-4544.

[150.] Balk, J.; Pilon, M., Ancient and essential: the assembly of iron-sulfur clusters in plants. *Trends Plant Sci,* **2011,** *16* (4), 218-26.

[151.] Chen, Y.-Q.*, et al.*, MET18 Deficiency Increases the Sensitivity of Yeast to Oxidative Stress and Shortens Replicative Lifespan by Inhibiting Catalase Activity. *Biomed Res Int,* **2017,** *2017*, 7587395-7587395.

[152.] Wang, X.*, et al.*, The cytosolic Fe-S cluster assembly component MET18 is required for the full enzymatic activity of ROS1 in active DNA demethylation. *Sci Rep,* **2016,** *6*, 26443-26443.

[153.] Ito, S.*, et al.*, MMXD, a TFIIH-independent XPD-MMS19 protein complex involved in chromosome segregation. *Molecular cell,* **2010,** *39* (4), 632-40.

[154.] Askree, S. H.*, et al.*, A genome-wide screen for Saccharomyces cerevisiae deletion mutants that affect telomere length. *Proc Natl Acad Sci U S A,* **2004,** *101* (23), 8658-63.

[155.] Lauder, S.*, et al.*, Dual requirement for the yeast MMS19 gene in DNA repair and RNA polymerase II transcription. *Molecular and cellular biology,* **1996,** *16* (12), 6783-6793.

[156.] Chan, C. T.*, et al.*, A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. *PLoS Genet,* **2010,** *6* (12), e1001247.

[157.] Sakai, Y.*, et al.*, Biogenesis and growth phase-dependent alteration of 5-methoxycarbonylmethoxyuridine in tRNA anticodons. *Nucleic Acids Research,* **2015,** *44* (2), 509-523.

[158.] Heiss, M.*, et al.*, Observing the fate of tRNA and its modifications by nucleic acid isotope labeling mass spectrometry: NAIL-MS. *RNA Biology,* **2017,** *14* (9), 1260-1268.

[159.] Sanger, F.*, et al.*, DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* **1977,** *74* (12), 5463-7.

[160.] Ronaghi, M.*, et al.*, Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem,* **1996,** *242* (1), 84-9.

[161.] Crawford, J. E.*, et al.*, NMR studies of organic solvent denatured yeast phenylalanyl transfer RNA at 220 MHz. *Biochem Biophys Res Commun,* **1971,** *44* (1), 1-7.

[162.] Koehler, K. M.; Schmidt, P. G., NMR study of the modified base resonances of tRNAtyr-coli. *Biochem Biophys Res Commun,* **1973,** *50* (2), 370-6.

[163.] Kan, L. S.*, et al.*, NMR study on the methyl and methylene proton resonances of tRNA Phe yeast. *Biochem Biophys Res Commun,* **1974,** *59* (1), 22-9.

[164.] Kastrup, R. V.; Schmidt, P. G., 1H NMR of valine tRNA modified bases. Evidence for multiple conformations. *Nucleic Acids Res,* **1978,** *5* (1), 257-69.

[165.] Kumar, R. K.; Davis, D. R., Synthesis and studies on the effect of 2-thiouridine and 4-thiouridine on sugar conformation and RNA duplex stability. *Nucleic Acids Res,* **1997,** *25* (6), 1272-80.

[166.] Strebitzer, E.*, et al.*, 5-Oxyacetic Acid Modification Destabilizes Double Helical Stem Structures and Favors Anionic Watson-Crick like cmo5 U-G Base Pairs. *Chemistry,* **2018,** *24* (71), 18903-18906.

[167.] Stuart, J. W.*, et al.*, Naturally-occurring modification restricts the anticodon domain conformational space of tRNA(Phe). *J Mol Biol,* **2003,** *334* (5), 901-18.

[168.] Durant, P. C.*, et al.*, Structural effects of hypermodified nucleosides in the Escherichia coli and human tRNALys anticodon loop: the effect of nucleosides s2U, mcm5U, mcm5s2U, mnm5s2U, t6A, and ms2t6A. *Biochemistry,* **2005,** *44* (22), 8078-89.

[169.] Vendeix, F. A. P.*, et al.*, Human tRNA(Lys3)(UUU) is pre-structured by natural modifications for cognate and wobble codon binding through keto-enol tautomerism. *J Mol Biol,* **2012,** *416* (4), 467-85.

[170.] Huang, T. Y.*, et al.*, Top-down tandem mass spectrometry of tRNA via ion trap collision-induced dissociation. *Journal of the American Society for Mass Spectrometry,* **2010,** *21* (6), 890-8.

[171.] Taucher, M.; Breuker, K., Characterization of modified RNA by top-down mass spectrometry. *Angewandte Chemie (International ed. in English),* **2012,** *51* (45), 11289-92.

[172.] Jiang, T.*, et al.*, Oligonucleotide Sequence Mapping of Large Therapeutic mRNAs via Parallel Ribonuclease Digestions and LC-MS/MS. *Analytical chemistry,* **2019,** *91* (13), 8500-8506.

[173.] Jora, M.*, et al.*, Detection of ribonucleoside modifications by liquid chromatography coupled with mass spectrometry. *Biochimica et biophysica acta. Gene regulatory mechanisms,* **2019,** *1862* (3), 280-290.

[174.] Yu, N.*, et al.*, RNAModMapper: RNA Modification Mapping Software for Analysis of Liquid Chromatography Tandem Mass Spectrometry Data. *Analytical chemistry,* **2017,** *89* (20), 10744-10752.

[175.] Thakur, P.*, et al.*, Improved RNA modification mapping of cellular non-coding RNAs using C- and U-specific RNases. *The Analyst,* **2020,** *145* (3), 816-827.

[176.] Yu, N.*, et al.*, tRNA Modification Profiles and Codon-Decoding Strategies in Methanocaldococcus jannaschii. *J Bacteriol,* **2019,** *201* (9).

[177.] Wein, S.*, et al.*, A computational platform for high-throughput analysis of RNA sequences and modifications by mass spectrometry. *Nat Commun,* **2020,** *11* (1), 926.

[178.] Reichle, V. F.*, et al.*, Surpassing limits of static RNA modification analysis with dynamic NAIL-MS. *Methods,* **2018**.

[179.] Yoluç, Y.*, et al.*, Instrumental analysis of RNA modifications. *Crit Rev Biochem Mol Biol,* **2021,** *56* (2), 178-204.

[180.] Nordhoff, E.*, et al.*, Comparison of IR- and UV-matrix-assisted laser desorption/ionization mass spectrometry of oligodeoxynucleotides. *Nucleic Acids Res,* **1994,** *22* (13), 2460-5.

[181.] Nordhoff, E.*, et al.*, Mass spectrometry of nucleic acids. *Mass Spectrom Rev,* **1996,** *15* (2), 67-138.

[182.] Wyatt, G. R., Occurrence of 5-methylcytosine in nucleic acids. *Nature,* **1950,** *166* (4214), 237-8.

[183.] Gupta, R. C.*, et al.*, A double-labeling procedure for sequence analysis of picomole amounts of nonradioactive RNA fragments. *Nucleic Acids Res,* **1976,** *3* (11), 2895-914.

[184.] Stanley, J.; Vassilenko, S., A different approach to RNA sequencing. *Nature,* **1978,** *274* (5666), 87-9.

[185.] Gupta, R. C.; Randerath, K., Rapid print-readout technique for sequencing of RNA's containing modified nucleotides. *Nucleic Acids Res,* **1979,** *6* (11), 3443-58.

[186.] Grosjean, H.*, et al.*, Detection of enzymatic activity of transfer RNA modification enzymes using radiolabeled tRNA substrates. *Methods Enzymol,* **2007,** *425*, 55-101.

[187.] Keith, G., Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography. *Biochimie,* **1995,** *77* (1-2), 142-4.

[188.] Zhao, X.; Yu, Y. T., Detection and quantitation of RNA base modifications. *Rna,* **2004,** *10* (6), 996-1002.

[189.] Gehrke, C. W.*, et al.*, Chromatography of nucleosides. *Journal of chromatography,* **1980,** *188* (1), 129-47.

[190.] Gehrke, C. W.; Patel, A. B., Gas-liquid chromatography of nucleosides. Derivatization and chromatography. *Journal of chromatography,* **1976,** *123* (2), 335-45.

[191.] Gehrke, C. W.; Ruyle, C. D., Gas-liquid chromatographic analysis of nucleic acid components. *Journal of chromatography,* **1968,** *38* (4), 473-91.

[192.] Lakings, D. B.; Gehrke, C. W., Analysis of base composition of RNA and DNA hydrolysates by gas-liquid chromatography. *Journal of chromatography,* **1971,** *62* (3), 347-67.

[193.] Pomerantz, S. C.; McCloskey, J. A., [44] Analysis of RNA hydrolyzates by liquid chromatography-mass spectrometry. In *Methods in Enzymology*, Academic Press: **1990**; Vol. 193, pp 796-824.

[194.] Lobue, P. A.*, et al.*, Oligonucleotide analysis by hydrophilic interaction liquid chromatography-mass spectrometry in the absence of ion-pair reagents. *Journal of chromatography. A,* **2019,** *1595*, 39-48.

[195.] Sarin, L. P.*, et al.*, Nano LC-MS using capillary columns enables accurate quantification of modified ribonucleosides at low femtomol levels. *Rna,* **2018,** *24* (10), 1403-1417.

[196.] Tsuji, K.*, et al.*, Capillary electrophoresis-electrospray mass spectrometry for the analysis of recombinant bovine and porcine somatotropins. *Analytical chemistry,* **1992,** *64* (17), 1864-1870.

[197.] Geldart, S. E.; Brown, P. R., Analysis of nucleotides by capillary electrophoresis. *Journal of Chromatography A,* **1998,** *828* (1), 317-336.

[198.] Snyder, L. R., Kirkland, J.J. and Glajch, J.L. , Basics of Separation. In *Practical HPLC Method Development*, **1997**; pp 21-58.

[199.] Martin, A. J.; Synge, R. L., A new form of chromatogram employing two liquid phases: A theory of chromatography. 2. Application to the micro-determination of the higher monoamino-acids in proteins. *Biochem J,* **1941,** *35* (12), 1358-68.

[200.] Fenn, J. B.*, et al.*, Electrospray ionization for mass spectrometry of large biomolecules. *Science,* **1989,** *246* (4926), 64-71.

[201.] Waghmare, S. P.; Dickman, M. J., Characterization and quantification of RNA post-transcriptional modifications using stable isotope labeling of RNA in conjunction with mass spectrometry analysis. *Analytical chemistry,* **2011,** *83* (12), 4894-901.

[202.] Su, D.*, et al.*, Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry. *Nat Protoc,* **2014,** *9* (4), 828-41.

[203.] Cech, N. B.; Enke, C. G., Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom Rev,* **2001,** *20* (6), 362-87.

[204.] Cole, R. B.; Zhu, J., Chloride anion attachment in negative ion electrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry,* **1999,** *13* (7), 607-611.

[205.] Taylor, G. I., Disintegration of water drops in an electric field. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences,* **1964,** *280* (1382), 383-397.

[206.] Iribarne, J.; Thomson, B., On the evaporation of small ions from charged droplets. *The Journal of chemical physics,* **1976,** *64* (6), 2287-2294.

[207.] Thomson, B.; Iribarne, J., Field induced ion evaporation from liquid surfaces at atmospheric pressure. *The Journal of Chemical Physics,* **1979,** *71* (11), 4451-4463.

[208.] Dole, M.*, et al.*, Molecular beams of macroions. *The Journal of chemical physics,* **1968,** *49* (5), 2240-2249.

[209.] Constantopoulos, T. L.*, et al.*, Effects of salt concentration on analyte response using electrospray ionization mass spectrometry. *Journal of the American Society for Mass Spectrometry,* **1999,** *10* (7), 625-34.

[210.] Kebarle, P., A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J Mass Spectrom,* **2000,** *35* (7), 804-17.

[211.] Liigand, J.*, et al.*, pH Effects on Electrospray Ionization Efficiency. *Journal of the American Society for Mass Spectrometry,* **2017,** *28* (3), 461-469.

[212.] Wilm, M. S.; Mann, M., Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *International Journal of Mass Spectrometry and Ion Processes,* **1994,** *136* (2-3), 167-180.

[213.] De La Mora, J. F.; Loscertales, I. G., The current emitted by highly conducting Taylor cones. *Journal of Fluid Mechanics,* **1994,** *260*, 155-184.

[214.] Bruins, A. P.*, et al.*, Ion spray interface for combined liquid chromatography/atmospheric pressure ionization mass spectrometry. *Analytical chemistry,* **1987,** *59* (22), 2642-2646.

[215.] Mordehai, A. F., John, Agilent Jet Stream Thermal Gradient Focusing Technology. *Agilent Technologies,* **2009,** (5990-3494E).

[216.] Kebarle, P.; Tang, L., From ions in solution to ions in the gas phase - the mechanism of electrospray mass spectrometry. *Analytical chemistry,* **1993,** *65* (22), 972A-986A.

[217.] Kostiainen, R.; Bruins, A. P., Effect of Solvent on Dynamic Range and Sensitivity in Pneumatically-assisted Electrospray (Ion Spray) Mass Spectrometry. *Rapid Communications in Mass Spectrometry,* **1996,** *10* (11), 1393-1399.

[218.] Zook, D.; Bruins, A., On cluster ions, ion transmission, and linear dynamic range limitations in electrospray (ionspray) mass spectrometry. *International journal of mass spectrometry and ion processes,* **1997,** *162* (1-3), 129-147.

[219.] Bruins, A. P., ESI source design and dynamic range considerations. In *Electrospray ionization mass spectrometry: Fundamental, instrumentation and appliations*, Wiley: **1997**; pp 107-136.

[220.] McNaught, A. D. W., A. , BT IUPAC Compendium of Chemical Terminology. Blackwell Scientific Publications: Oxford, **1997**.

[221.] Hart-Smith, G. B., S. J. , Mass Spectrometry in Polymer Chemistry. Wiley-VCH Verlag & Co: Weinheim, Germany, **2012**.

[222.] Nakayama, H.*, et al.*, Method for Direct Mass-Spectrometry-Based Identification of Monomethylated RNA Nucleoside Positional Isomers and Its Application to the Analysis of Leishmania rRNA. *Analytical chemistry,* **2019,** *91* (24), 15634-15643.

[223.] Hua, Y.*, et al.*, Comparison of negative and positive ion electrospray tandem mass spectrometry for the liquid chromatography tandem mass spectrometry

analysis of oxidized deoxynucleosides. *Journal of the American Society for Mass Spectrometry,* **2001,** *12* (1), 80-87.

[224.] Dudley, E.*, et al.*, Study of the mass spectrometric fragmentation of pseudouridine: comparison of fragmentation data obtained by matrix-assisted laser desorption/ionisation post-source decay, electrospray ion trap multistage mass spectrometry, and by a method utilising electrospray quadrupole time-of-flight tandem mass spectrometry and in-source fragmentation. *Rapid Commun Mass Spectrom,* **2005,** *19* (21), 3075-85.

[225.] Shaffer, S. A.*, et al.*, Characterization of an improved electrodynamic ion funnel interface for electrospray ionization mass spectrometry. *Analytical chemistry,* **1999,** *71* (15), 2957-64.

[226.] Schneider, B. B.*, et al.*, Sampling Efficiency Improvement to an Electrospray Ionization Mass Spectrometer and Its Implications for Liquid Chromatography Based Inlet Systems in the Nanoliter to Milliliter per Minute Flow Range. *Journal of the American Society for Mass Spectrometry,* **2021,** *32* (6), 1441-1447.

[227.] Urban, P. L., Quantitative mass spectrometry: an overview. *Philos Trans A Math Phys Eng Sci,* **2016,** *374* (2079).

[228.] Kellner, S.*, et al.*, Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. *Nucleic Acids Res,* **2014,** *42* (18), e142.

[229.] Barker, S. A.*, et al.*, Intermediates in the photosynthetic cycle. *Biochim Biophys Acta,* **1956,** *21* (2), 376-7.

[230.] Meselson, M.; Stahl, F. W., THE REPLICATION OF DNA IN ESCHERICHIA COLI. *Proc Natl Acad Sci U S A,* **1958,** *44* (7), 671-82.

[231.] Ong, S. E.*, et al.*, Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics,* **2002,** *1* (5), 376-86.

[232.] Mann, M., Functional and quantitative proteomics using SILAC. *Nat Rev Mol Cell Biol,* **2006,** *7* (12), 952-8.

[233.] Schiesser, S.*, et al.*, Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. *Angewandte Chemie (International ed. in English),* **2012,** *51* (26), 6516-20.

[234.] Lu, K.*, et al.*, Isotope labeling strategies for NMR studies of RNA. *J Biomol NMR,* **2010,** *46* (1), 113-25.

[235.] Huber, S. M.*, et al.*, 2'-O-Methyl-5-hydroxymethylcytidine: A Second Oxidative Derivative of 5-Methylcytidine in RNA. *J Am Chem Soc,* **2017,** *139* (5), 1766-1769.

[236.] Dal Magro, C.*, et al.*, A Vastly Increased Chemical Variety of RNA Modifications Containing a Thioacetal Structure. *Angewandte Chemie (International ed. in English),* **2018,** *57* (26), 7893-7897.

[237.] Huber, S. M.*, et al.*, Formation and abundance of 5-hydroxymethylcytosine in RNA. *Chembiochem,* **2015,** *16* (5), 752-5.

[238.] Kellner, S.*, et al.*, Profiling of RNA modifications by multiplexed stable isotope labelling. *Chemical communications (Cambridge, England),* **2014,** *50* (26), 3516-8.

[239.] Reichle, V. F.*, et al.*, Surpassing limits of static RNA modification analysis with dynamic NAIL-MS. *Methods,* **2019,** *156*, 91-101.

[240.] Heiss, M.*, et al.*, Cell culture NAIL-MS allows insight into human tRNA and rRNA modification dynamics in vivo. *Nat Commun,* **2021,** *12* (1), 389.

[241.] Brandmayr, C.*, et al.*, Isotope-based analysis of modified tRNA nucleosides correlates modification density with translational efficiency. *Angewandte Chemie (International ed. in English),* **2012,** *51* (44), 11162-5.

[242.] Borland, K.*, et al.*, Production and Application of Stable Isotope-Labeled Internal Standards for RNA Modification Analysis. *Genes (Basel),* **2019,** *10* (1).

[243.] Heiss, M.*, et al.*, Quantification of Modified Nucleosides in the Context of NAIL-MS. In *RNA Modifications: Methods and Protocols*, McMahon, M., Ed. Springer US: New York, NY, **2021**; pp 279-306.

[244.] Yoluç, Y.*, et al.*, Systematic assessment of methylation damage in nucleic acids. *ChemBioChem,* **2022**.

[245.] Kim, D.*, et al.*, The Architecture of SARS-CoV-2 Transcriptome. *Cell,* **2020,** *181* (4), 914-921.e10.

[246.] Yang, S. L.*, et al.*, Comprehensive mapping of SARS-CoV-2 interactions in vivo reveals functional virus-host interactions. *Nature Communications,* **2021,** *12* (1), 5113.

[247.] Tao, K.*, et al.*, The biological and clinical significance of emerging SARS-CoV-2 variants. *Nat Rev Genet,* **2021,** *22* (12), 757-773.

[248.] Harvey, W. T.*, et al.*, SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology,* **2021,** *19* (7), 409-424.

[249.] Schäck, M. A.*, et al.*, Eukaryotic life without tQCUG: the role of Elongator-dependent tRNA modifications in Dictyostelium discoideum. *Nucleic Acids Res,* **2020,** *48* (14), 7899-7913.

[250.] Richter, F.*, et al.*, RNA marker modifications reveal the necessity for rigorous preparation protocols to avoid artifacts in epitranscriptomic analysis. *Nucleic Acids Research,* **2021,** *50* (8), 4201-4215.

[251.] Novikova, I. V.*, et al.*, Sizing up long non-coding RNAs: do lncRNAs have secondary and tertiary structure? *Bioarchitecture,* **2012,** *2* (6), 189-99.

[252.] Izadpanah, A.*, et al.*, Epitranscriptomics of SARS-CoV-2 Infection. *Front Cell Dev Biol,* **2022,** *10*, 849298.

[253.] Campos, J. H. C.*, et al.*, Direct RNA Sequencing Reveals SARS-CoV-2 m6A Sites and Possible Differential DRACH Motif Methylation among Variants. *Viruses,* **2021,** *13* (11).

[254.] Zhang, X.*, et al.*, Methyltransferase-like 3 Modulates Severe Acute Respiratory Syndrome Coronavirus-2 RNA N6-Methyladenosine Modification and Replication. *mBio,* **2021,** *12* (4), e0106721.

[255.] Liu, J. e.*, et al.*, The m6A methylome of SARS-CoV-2 in host cells. *Cell research,* **2021,** *31* (4), 404-414.

[256.] Li, N.*, et al.*, METTL3 regulates viral m6A RNA modification and host cell innate immune responses during SARS-CoV-2 infection. *Cell Reports,* **2021,** *35* (6).

[257.] Taiaroa, G.*, et al.*, Direct RNA sequencing and early evolution of SARS-CoV-2. *bioRxiv,* **2020**, 2020.03.05.976167.

[258.] Fleming, A. M.*, et al.*, Nanopore Dwell Time Analysis Permits Sequencing and Conformational Assignment of Pseudouridine in SARS-CoV-2. *ACS Central Science,* **2021,** *7* (10), 1707-1717.

[259.] Kaiser, S.*, et al.*, Strategies to Avoid Artifacts in Mass Spectrometry-Based Epitranscriptome Analyses. *Angewandte Chemie International Edition,* **2021,** *60* (44), 23885-23893.

[260.] Sloan, K. E.*, et al.*, Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biology,* **2017,** *14* (9), 1138-1152.

[261.] Cohn, W. E., 5-Ribosyl uracil, a carbon-carbon ribofuranosyl nucleoside in ribonucleic acids. *Biochim Biophys Acta,* **1959,** *32*, 569-71.

[262.] Dunn, D. B., Additional components in ribonucleic acid of rat-liver fractions. *Biochim Biophys Acta,* **1959,** *34*, 286-8.

[263.] Wacker, A.*, et al.*, Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Research,* **2020,** *48* (22), 12415-12435.

[264.] Züst, R.*, et al.*, Genetic interactions between an essential 3' cis-acting RNA pseudoknot, replicase gene products, and the extreme 3' end of the mouse coronavirus genome. *J Virol,* **2008,** *82* (3), 1214-28.

[265.] Goebel, S. J.*, et al.*, Characterization of the RNA components of a putative molecular switch in the 3' untranslated region of the murine coronavirus genome. *J Virol,* **2004,** *78* (2), 669-82.

[266.] Yang, J.*, et al.*, Mapping of Complete Set of Ribose and Base Modifications of Yeast rRNA by RP-HPLC and Mung Bean Nuclease Assay. *PLoS One,* **2016,** *11* (12), e0168873.

[267.] Hoffman, N. E.; Liao, J. C., Reversed phase high performance liquid chromatographic separations of nucleotides in the presence of solvophobic ions. *Analytical chemistry,* **1977,** *49* (14), 2231-4.

[268.] Pimenov, A. M.*, et al.*, Simultaneous separation of ribonucleotides, nucleosides and nitrogen bases by ion-pair reversed-phase high-performance liquid chromatography on columns with radial compression. *Journal of Chromatography A,* **1986,** *365*, 221-227.

[269.] Van der Boon, J.*, et al.*, Reversed-phase ion-paired HPLC of purine nucleotides from skeletal muscle, heart and brain of the goldfish, Carassius auratus L.—II. Influence of environmental anoxia on metabolite levels. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry,* **1992,** *101* (4), 583-586.

[270.] Zhao, J.; Fleet, G. H., Separation of twenty isomers of ribonucleotides and deoxyribonucleotides by reversed-phase ion-pairing high-performance liquid chromatography. *Journal of Chromatography A,* **1996,** *732* (2), 271-275.

[271.] Kromidas, S., The HPLC-MS Handbook for Practitioners. John Wiley & Sons, Incorporated: Weinheim, GERMANY, **2017**.

[272.] Neubauer, S.*, et al.*, Mass spectrometry based analysis of nucleotides, nucleosides, and nucleobases—application to feed supplements. *Analytical and Bioanalytical Chemistry,* **2012,** *404* (3), 799-808.

[273.] Preinerstorfer, B.*, et al.*, Metabolic profiling of intracellular metabolites in fermentation broths from β-lactam antibiotics production by liquid chromatography–tandem mass spectrometry methods. *Journal of chromatography A,* **2010,** *1217* (3), 312-328.

[274.] Kromidas, S., The HPLC Expert: Possibilities and Limitations of Modern High Performance Liquid Chromatography. Wiley: **2016**.

[275.] He, L.*, et al.*, Simultaneous Quantification of Nucleosides and Nucleotides from Biological Samples. *Journal of the American Society for Mass Spectrometry,* **2019,** *30* (6), 987-1000.

[276.] Kromidas, S., HPLC richtig optimiert: Ein Handbuch für Praktiker. Wiley: **2012**.

[277.] Snyder, L. R., Kirkland, J.J. and Glajch, J.L. , Ionic Samples: Reversed-Phase, Ion-Pair, and Ion-Exchange HPLC. In *Practical HPLC Method Development*, **1997**; pp 292-349.

[278.] Muthmann, N.*, et al.*, Quantification of mRNA cap-modifications by means of LC-QqQ-MS. *Methods,* **2022,** *203*, 196-206.

[279.] Wilhelm, O.*, et al.*, Electrospray evaporation and deposition. *Journal of Aerosol Science,* **2003,** *34* (7), 815-836.

[280.] Kageyama Kaneshima, A.*, et al.*, Influence of Solvent Composition and Surface Tension on the Signal Intensity of Amino Acids in Electrospray Ionization Mass Spectrometry. *Mass Spectrom (Tokyo),* **2019,** *8* (1), A0077.

[281.] Kromidas, S., HPLC optimal einsetzen: Konzepte und Strategien für Methodenentwicklung und -optimierung. Wiley: **2022**.

[282.] Wang, J.*, et al.*, Quantifying the RNA cap epitranscriptome reveals novel caps in cellular and viral RNA. *Nucleic Acids Res,* **2019,** *47* (20), e130.

[283.] Snyder, L. R., Kirkland, J.J. and Glajch, J.L. , Biochemical Samples: Proteins, Nucleic Acids, Carbohydrates, and Related Compounds. In *Practical HPLC Method Development*, **1997**; pp 479-536.

[284.] Mastovska, K. Z., John, Triggered MRM LC/MS/MS Method Development – Practical Considerations for MRM Optimization Using Agilent MassHunter Optimizer Software. *Agilent Technologies,* **2017,** (5991-7195EN).

[285.] Strzelecka, D.*, et al.*, Analysis of mononucleotides by tandem mass spectrometry: investigation of fragmentation pathways for phosphate- and ribose-modified nucleotide analogues. *Sci Rep,* **2017,** *7* (1), 8931.

[286.] Hudeček, O.*, et al.*, Dinucleoside polyphosphates act as 5′-RNA caps in bacteria. *Nature Communications,* **2020,** *11* (1), 1052.

[287.] Tang, L.; Kebarle, P., Dependence of ion intensity in electrospray mass spectrometry on the concentration of the analytes in the electrosprayed solution. *Analytical chemistry,* **1993,** *65* (24), 3654-3668.

[288.] 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,* **2020,** *10* (2), 190306.

[289.] Massé, E.; Gottesman, S., A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proceedings of the National Academy of Sciences,* **2002,** *99* (7), 4620-4625.

[290.] Hagelskamp, F.*, et al.*, Broadly applicable oligonucleotide mass spectrometry for the analysis of RNA writers and erasers in vitro. *Nucleic Acids Res,* **2020,** *48* (7), e41.

[291.] Hauenschild, R.*, et al.*, The reverse transcription signature of N-1-methyladenosine in RNA-Seq is sequence dependent. *Nucleic Acids Res,* **2015,** *43* (20), 9950-64.