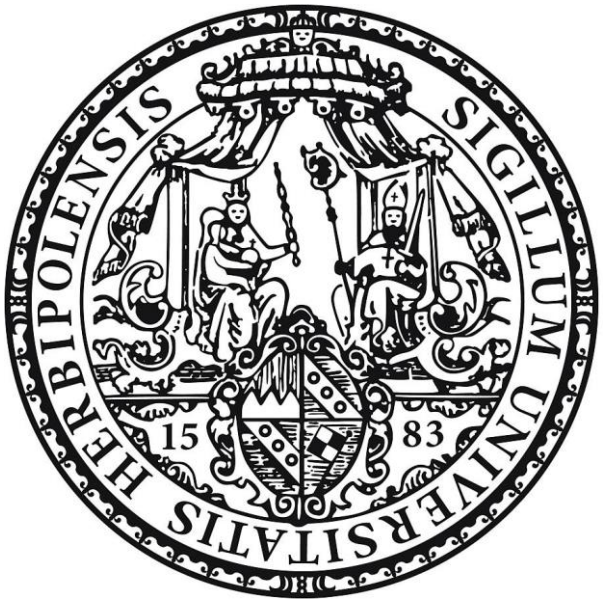
# In vitro selected ribozymes for RNA methylation and labeling




Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrades

der

Julius-Maximilians-Universität Würzburg

vorgelegt von

**Carolin P. M. Scheitl**

aus Würzburg

Würzburg, 2023

Eingereicht bei der Fakultät für Chemie und Pharmazie am:

_____

Gutachter der schriftlichen Arbeit:

1. Gutachter: _____

2. Gutachter: _____

Prüfer des öffentlichen Promotionskolloquiums:

1. Prüfer: _____

2. Prüfer: _____

3. Prüfer: _____

Datum des öffentlichen Promotionskolloquiums:

_____

Doktorurkunde ausgehändigt am:

_____

*“From so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved.”*

- Chares Darwin, On the Origin of Species -

# List of Publications

Parts of this thesis have been published (*authors contributed equally):

Scheitl, C.P.M.; Lange, S.; Höbartner, C. New Deoxyribozymes for the Native Ligation of RNA. *Molecules* **2020**, *25*, 3650.

Scheitl, C.P.M.; Ghaem Maghami, M.; Lenz, A.-K.; Höbartner, C. Site-specific RNA methylation by a methyltransferase ribozyme. *Nature* **2020**, *587*, 663-667.

Scheitl, C.P.M.*; Mieczkowski, M.*; Schindelin, H.; Höbartner, C. Structure and mechanism of the methyltransferase ribozyme MTR1. *Nat. Chem. Biol.* **2022**, *18*, 547-555.

Scheitl, C.P.M.; Okuda, T.; Adelmann, J.; Höbartner, C. Ribozyme-Catalyzed Late-Stage Functionalization and Fluorogenic Labeling of RNA. *Angew. Chem. Int. Ed.* **2023**, *62*, e202305463.

Chan, A.*; Naarman-de Vries, I.S.*; Scheitl, C.P.M.*; Höbartner, C.; Dieterich, C. mAFiA: Detecting $m^6A$ at single-molecular resolution via direct-RNA sequencing. *bioRxiv* **2023**, 2023.07.28.550944.

# Acknowledgements

First and foremost, I want to express my deepest gratitude to my thesis supervisor and mentor **Prof. Dr. Claudia Höbartner**. I am extremely grateful for your constructive feedback, your continuous support and guidance that I received throughout my time in your lab. It was a great experience to work on such an interesting and exciting topic. Your immense fascination and excitement for the chemistry of nucleic acids was contagious, and it turned me into a nucleic acid enthusiast as well.

I would also like to thank my thesis committee members **Prof. Dr. Hermann Schindelin** and **Prof. Dr. Utz Fischer** for agreeing to review my thesis. Furthermore, I want to thank Prof. Dr. Hermann Schindelin also for his invaluable assistance regarding data collection and analysis as well as X-ray crystallography in general.

I would also like to thank my cooperation partners **Prof. Dr. Christoph Dieterich**, **Isabel** and **Adrian** for the enjoyable and fruitful collaborative work, in which I gained many new insights into nanopore sequencing and data analysis.

A huge thank you also to all members of the mass department, especially **Sebastian**, **Michaela** and **Juliane**, for measuring my oligo samples, even if it was once more urgent and time critical.

Furthermore, I would like to thank my former apprentices, internship students, bachelor and master students, who have all been a great support and helped me advance my projects. Especially, I want to mention **Max**, who came in our lab twice. It was a lot of fun working with you and your twin brother.

I huge thank you goes to all technicians. Especially I want to mention **Celine** for providing me with so many RNAs for crystallization and **Manuela** for giving me support in the chemistry lab and sharing her knowledge about synthesis with me. Most importantly, I want to thank **Ann-Kathrin** for the great time we had together. You were always there with your positive energy whenever I was struggling and helped me so much with my experiments. I will for sure never forget the crazy weeks we had together throughout the nature-revision!

Special thanks also go to **Mateusz** for teaching me crystallography. In the beginning I never thought I would be able to handle such tiny crystals under the microscope.

I would also like to thank **Taku** for the great scientific discussions I had with you. I learned so much from you and you were always there with new ideas whenever I needed help. I also have great memories of our joint conference trip to Stockholm, even though it took us almost 24 h to come back home to Würzburg.

Thank you **Anam** for your positive energy you brought to our lab and all the fun we had together. After you graduated, we all really missed you a lot.

Furthermore, I want to say thank you to **Christian** and **Julia** for proofreading my thesis. Together with **Carina**, **Florian** und **Hermann** I also want to thank you all for your constant support in the lab, with the HPLC and especially with Origin. Thank you for organizing our group trips and Christmas parties. Working with you was great.

I also want to thank **Farbod** for introducing me into the world of *in vitro* selections and helping me throughout the characterization of MTR1.

Many thanks also to **Doris** for organizing everything in the background and always taking care of shipping our crystal-Dewar to Hamburg. Without you we would all be lost.

I would also like to thank all other former and current working group members for the wonderful time we spent together, which I will always remember. I genuinely enjoyed being with you - inside and outside the lab.

Importantly, I also want to thank my best friend **Julia** for continuous moral support. You are always there for me when I need someone to talk to or to join me for fun workout sessions.

A massive thank you also to my **parents** and **grandparents**, for their incredible support and the encouragement to always pursue my dreams. Without you, I would have never made it this far.

Finally, I want to thank **Philipp**. It's just crazy but also amazing how unexpectedly we both met. I'm so grateful for your constant support and love. Without you my life would not be as exciting and crazy as it is now. I love you.

# Table of Contents

**1 Theoretical Background ..... 1**

**1.1 The RNA Epitranscriptome ..... 1**

1.1.1 Post-transcriptional RNA modifications ..... 2

1.1.2 Detection of RNA modifications ..... 11

**1.2 The catalytic potential of nucleic acids ..... 23**

1.2.1 Natural ribozymes ..... 23

1.2.2 *In vitro* selected RNA and DNA catalysts ..... 23

1.2.3 Covalent RNA labeling catalyzed by nucleic acids ..... 35

1.2.4 Methyltransferase ribozymes: MTR1 and beyond ..... 38

**2 Research Objectives and Outline ..... 41**

**3 Site-specific RNA methylation by a methyltransferase ribozyme ..... 43**

**3.1 Introduction ..... 44**

**3.2 Results and Discussion ..... 45**

3.2.1 Search for methyltransferase ribozymes ..... 45

3.2.2 Identification of the methylation product ..... 47

3.2.3 RNA-catalyzed tRNA methylation ..... 49

3.2.4 Site-specific methylation of long RNAs ..... 52

**3.3 Discussion ..... 53**

**4 Structure and mechanism of the methyltransferase ribozyme MTR1 ..... 54**

**4.1 Introduction ..... 55**

**4.2 Results and Discussion ..... 56**

4.2.1 Design and activity of the crystallization construct ..... 56

4.2.2 Overall structure of the MTR1 ribozyme ..... 58

4.2.3 Mechanism of RNA-catalyzed RNA methylation ..... 60

4.2.4 Rate acceleration by ribose methylation ..... 62

4.2.5 Comparison of MTR1 guanine binding site with purine riboswitches ..... 64

**4.3 Discussion ..... 65**

**5 Ribozyme-catalyzed late-stage functionalization and fluorogenic labeling of RNA68**

**5.1 Introduction ... 69**
**5.2 Results and Discussion ... 70**
5.2.1 Identification of the reaction product ... 70
5.2.2 Generation of a new Hcy-chromophore directly on the RNA ... 72
5.2.3 Nebularine incorporation by *in vitro* transcription ... 73
5.2.4 tRNA transcription and labeling ... 75
**5.3 Discussion ... 75**

**6 *In vitro* selection of new ribozymes for RNA-methylation and -alkylation ... 76**
**6.1 DNAzyme based reselection of a methyltransferase ribozyme ... 78**
6.1.1 Introduction ... 78
6.1.2 General selection strategy ... 79
6.1.3 Selections CD - CG ... 81
6.1.4 CH- and CI-selection ... 88
6.1.5 Discussion ... 94
**6.2 Selection of new methyl- and alkyltransferase ribozymes ... 95**
6.2.1 C or G as modification site (CJ and CK selection) ... 96
6.2.2 Selecting an improved alkyltransferase ribozyme (CL selection) ... 109

**7 Conclusion and Outlook ... 115**

**8 Zusammenfassung und Ausblick ... 119**

**9 Appendix ... 125**
**9.1 Supporting information for chapter 3 ... 125**
9.1.1 Material and Methods ... 125
9.1.2 Supplementary tables ... 131
9.1.3 Extended data figures ... 135
**9.2 Supporting information for chapter 4 ... 144**
9.2.1 Material and Methods ... 144
9.2.2 Supplementary tables ... 148
9.2.3 Extended data figures ... 150
**9.3 Supporting information for chapter 5 ... 156**
9.3.1 Material and Methods ... 156

9.3.2 Supplementary tables ..... 164
9.3.3 Extended data figures ..... 165
**9.4 Supporting information for chapter 6.1 ..... 172**
9.4.1 Material and methods ..... 172
9.4.2 Supplementary tables ..... 176
9.4.3 Extended data figures ..... 178
**9.5 Supporting information for chapter 6.2 ..... 182**
9.5.1 Material and methods ..... 182
9.5.2 Supplementary tables ..... 185
9.5.3 Extended data figures ..... 188

**Individual contributions ..... 193**

**References ..... 196**

# Abbreviations

| | |
|---|---|
| A | adenosine |
| ab$^1$A | 1-(4-aminomethyl-benzyl)adenosine |
| Alk | alkyne |
| ALKBH | AlkB homolog |
| 2AP | 2-aminopurine |
| ASU | asymmetric unit |
| BAP | bacterial alkaline phosphatase |
| BC-biotin | biotinylated *O*$^2$-benzylcytidine |
| bio | biotin |
| bio-hyd | biotin-hydrazide |
| BG | *O*$^6$-benzylguanine |
| BG-biotin | biotinylated *O*$^6$-benzylguanine |
| BG-$NH_2$ | *O*$^6$-(4-aminomethyl-benzyl)guanine |
| bn$^1$A | 1-benzyladenosine |
| bn$^6$A | *N*$^6$-benzyladenosine |
| BuNA | butyl nucleic acid |
| C | cytidine |
| c$^7$A | 7-deazaadenine |
| c$^1$c$^3$A | 1,3-dideazaadenine |
| c$^3$dA | 3-deazadeoxyadenine |
| CDS | coding sequence |
| CeNA | cyclohexene nucleic acid |
| CPG | controlled pore glass |
| CV | column volume |
| DART-seq | deamination adjacent to RNA modification targets sequencing |
| dNTP | deoxynucleoside triphosphate |
| ddNTP | dideoxynucleoside triphosphate |
| DKC1 | Dyskerin |
| DMF | *N,N*-dimethylformamide |
| DMS | dimethyl sulfide |
| DMSO | dimethylsulfoxide |
| DNA | deoxyribonucleic acid |
| ds | double stranded |

| | |
|---|---|
| EDTA | ethylenediamine tetraacetate |
| EIC | extracted ion chromatogram |
| 2'F | 2' fluoro |
| FACS | fluorescence-activated cell sorting |
| FANA | 2' fluoro arabino nucleic acid |
| FBC | F30-Broccoli-*cis* |
| FBL | fibrillarin |
| FBT | F30-Broccoli-*trans* |
| FGE | formylglycine-generating enzyme |
| Flu-TSC | fluorescein thiosemicarbazide |
| FMN | flavin mononucleotide |
| FNA | flexible nucleic acid |
| G | guanosine |
| GNA | glycol nucleic acid |
| HAC-seq | Hydrazine-Aniline Cleavage sequencing |
| Hcy | hemicyanine |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| HNA | hexitol nucleic acid |
| HPLC | high-performance liquid chromatography |
| HR-ESI-MS | high resolution electrospray ionization mass spectrometry |
| I | inosine |
| IP | immunoprecipitation |
| IPTG | isopropyl β-d-1-thiogalactopyranoside |
| IVC | *in vitro* compartmentalization |
| LC-MS | liquid-chromatography mass-spectrometry |
| LNA | locked nucleic acid |
| lncRNA | long non-coding RNA |
| LY-CH | lucifer yellow carbohydrazide |
| 1M7 | 1-methyl-7-nitroisatoic anhydride |
| $m^1A$ | 1-methyladenosine |
| $m^6A$ | $N^6$-methyladenosine |
| $m^6A$-REF-seq | RNA endoribonuclease facilitated sequencing |
| $m^3C$ | 3-methylcytosine |
| $m^6G$ | $O^6$-methylguanine |
| $m^6H$ | $O^6$-methylhypoxanthin |
| MAC | $m^6A$-METTL complex |
| MACOM | $m^6A$-METTL associated complex |
| $m^6dG$ | $O^6$-methyl-2'-deoxyguanosine |

| | |
|---|---|
| MeRIP-seq | methylated RNA immunoprecipitation with next-generation sequencing |
| METTL | methyltransferase-like protein |
| miCLIP | $m^6A$ individual-nucleotide-resolution cross-linking and immunoprecipitation |
| miRNA | micro RNA |
| 2'MOE | 2' methyoxyethyl RNA |
| MPD | 2-Methyl-2,4-pentanediol |
| $m^6preQ_1$ | $O^6$-methyl pre-queuosine 1 |
| MR | molecular replacement |
| mRNA | messenger RNA |
| MS | mass spectrometry |
| $ms^6G$ | $S^6$-methylthioguanine |
| mt | mitochondrial |
| MTR1 | methyl transferase ribozyme 1 |
| $NAD^+$ | nicotinamide adenine dinucleotide |
| NAIM | nucleotide analogue interference mapping |
| ncRNA | non-coding RNA |
| ND | not determined |
| Ne | nebularine |
| NGS | next-generation sequencing |
| NHS | *N*-hydroxysuccinimide |
| *N,N*-DMED | *N,N*-dimethylethylenediamine |
| $N_m$ | 2'OMe nucleotide |
| NS | negative selection |
| nt | nucleotides |
| NTP | nucleoside triphosphate |
| 2'OMe | 2'-*O*-methoxy |
| ON | overnight |
| O6PG | $O^6$-propargylguanine |
| PAGE | polyacrylamide gel electrophoresis |
| PA-$m^6A$-seq | photo-crosslinking-assisted $m^6A$-sequencing |
| PDB | Protein Data Bank |
| PNA | peptide nucleic acid |
| PNK | T4 polynucleotide kinase |
| $preQ_1$ | pre-queuosine 1 |
| PS | phosphorothioate |
| Ψ | pseudouridine |
| Pu | purine |

| | |
|---|---|
| ref. | reference |
| Riboxi-seq | ribose oxidation sequencing |
| RIP | RNA immunoprecipitation |
| RNA | ribonucleic acid |
| rRNA | ribosomal RNA |
| RT | reverse transcription |
| Rz | ribozyme |
| SAM | *S*-adenosylmethionine |
| Sem | 2'-seleno-methyl |
| SMRZ-1 | SAM-ribozyme-1 |
| SD | standard deviation |
| SELEX | Systematic Evolution of Ligands by Exponential Enrichment |
| SHAPE | Selective 2′ Hydroxyl Acylation analyzed by Primer Extension |
| siRNA | small interfering RNA |
| sncRNA | small non-coding RNA |
| ss | single stranded |
| SVPD | snake venom phosphodiesterase |
| t | time |
| T | thymine |
| TET | ten-eleven translocation |
| THF | tetrahydrofuran |
| TM | transition mutated |
| TNA | threose nucleic acid |
| Tris | tris(hydroxymethyl)aminomethane |
| tRNA | transfer ribonucleic acid |
| TV | transversion mutated |
| U | uridine |
| UDG | uracil deglycosylase |
| UTR | untranslated region |
| wt | wildtype |
| WTAP | Wilms' tumor 1-associating protein |
| XNA | xeno nucleic acid |

# 1 Theoretical Background

## 1.1 The RNA Epitranscriptome

The complex molecular machinery found in living cells and organisms is strictly controlled by the flow of genetic information between DNA, RNA and proteins. This principle is conceptualized in the central dogma of molecular biology (Figure 1.1), which states that DNA is transcribed into RNA, which is then translated into proteins. However, RNA does not only exist in the form of messenger RNA (mRNA) to act as a simple template for protein synthesis. Instead, many classes of RNA can be distinguished that all play various vital roles in biological systems. The most widely known types are ribosomal RNA (rRNA) that catalyzes the formation of peptide bonds within the ribosome as well as transfer RNAs (tRNA) that translate the information encoded within the mRNA into the peptide sequence. However, the majority of RNAs, such as small non-coding RNAs (sncRNAs) and long non-coding RNAs (lncRNAs), micro RNAs (miRNAs) or small interfering RNAs (siRNAs), are involved in different regulatory processes regarding gene expression.[1]

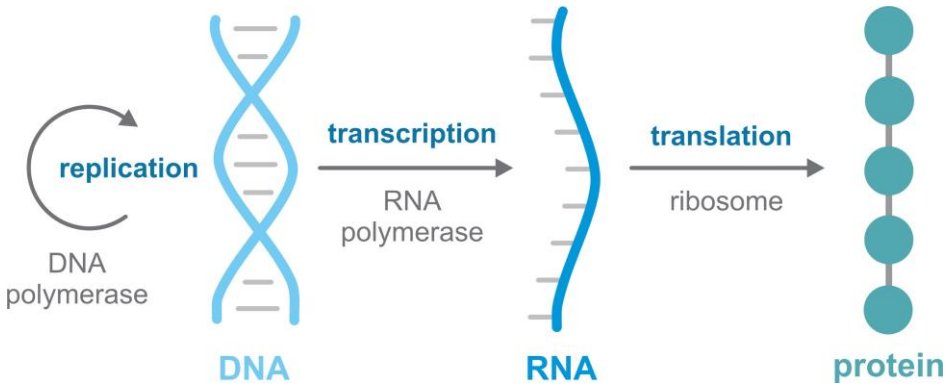



**Figure 1.1:** The flow of genetic information according to the central dogma of molecular biology. DNA is transcribed into RNA by RNA polymerases. The RNA is then translated into proteins at the ribosome. The storage form DNA can also be replicated by DNA polymerases.

Despite their different shapes and functions, RNA, just like DNA, is composed of only four genetic letters: adenosine (A), guanosine (G), cytidine (C), and uridine (U). However, all these nucleosides can also be modified post-transcriptionally in various ways, which is often also referred to as the ‘RNA epitranscriptome’. Those modifications are of special interest as they form an additional layer of information beyond the generic genetic alphabet. They have a major influence on RNA folding, localization, or its stability and interaction with other nucleic acids, proteins, or small molecules. Due to their diverse roles, post-transcriptional RNA modifications have also been linked to numerous human genetic diseases, which attracts huge interest and sparks the development of new epigenetic therapeutics.[2]

## 1.1.1 Post-transcriptional RNA modifications

### 1.1.1.1 The diversity of RNA modifications

Compared with DNA, a much broader spectrum of modifications can be found in RNA: More than 170 different modifications are known that are widely distributed in various types of RNA in all domains of life[3] and new modifications are constantly being discovered, reflecting the complexity and variety of functions that are fulfilled by RNA. Among other aspects, such chemical modifications influence the translation and stability of mRNAs, the structure of rRNAs, the regulatory properties of lncRNAs, sncRNAs and miRNAs and also play crucial roles for the structure and correct function of tRNAs.

The diversity of chemical alterations of RNA molecules is huge. Modifications are not just limited to the Watson-Crick face of the nucleobase (e.g., 1-methyladenosine, $m^1A$ or 3-methylcytosine, $m^3C$), where they can drastically alter base pairing properties and thus have a major influence on RNA structure and interaction with other cellular components. Instead, they are known at many different positions in all four canonical nucleosides (Figure 1.2).[4-5] While most post-transcriptional modifications are just small simple substituents like methyl groups that are added to the nucleobase (e.g., $N^6$-methyladenosine, $m^6A$) or the ribose 2'OH ($N_m$), also larger moieties can be installed. Examples include groups like isopentenyl (e.g., $N^6$-isopentenyladenosine, $i^6A$) or even bulkier substituents such as cyclo-pentanediol or guanidine (e.g., in the hypermodified queuosine, Q or archaeosine, $G^+$), which are often connected to the C7 of 7-deazapurine bases. Another important example of altered nucleosides is pseudouridine (Ψ), also often referred to as the 'fifth nucleoside', which is the most widespread modification in cellular RNA. Ψ is the C5-glycosidic isomer of uridine and thus shows an altered connectivity of the nucleobase. Consequently, Ψ provides an additional H-bond donor at position *N*1, while retaining the base pairing properties of uridine.[4, 6] Another example worth mentioning is modification of the phosphate backbone in the form of phosphorothioate (PS). Through the introduction of sulfur into the phosphate backbone, an additional stereocenter is created in the RNA. Even though this modification was already known for DNA, it was just recently discovered in RNA.[7]

Among all post-transcriptional RNA modifications, methylation is the most abundant one, accounting for nearly 100 different modified nucleosides.[3] The research field of RNA methylation was started around 100 years ago, when 5-methylcytosine ($m^5C$) was discovered as the first modified nucleotide in 1925.[8] Since then, owing to the great importance of methylations for RNA structure, function and metabolism in combination with new technological advances, research on methylated nucleosides is constantly rising. Especially the recent developments of antibody mediated enrichment strategies and high-throughput sequencing technologies,

enabled the identification and transcriptome wide mapping of various modifications (chapter 1.1.2.3). The following chapters will provide an overview of the distribution and functions (chapter 1.1.1.2) as well as the installation and removal (chapter 1.1.1.3) of post-transcriptional modifications with special focus on, but not limited to RNA methylation.

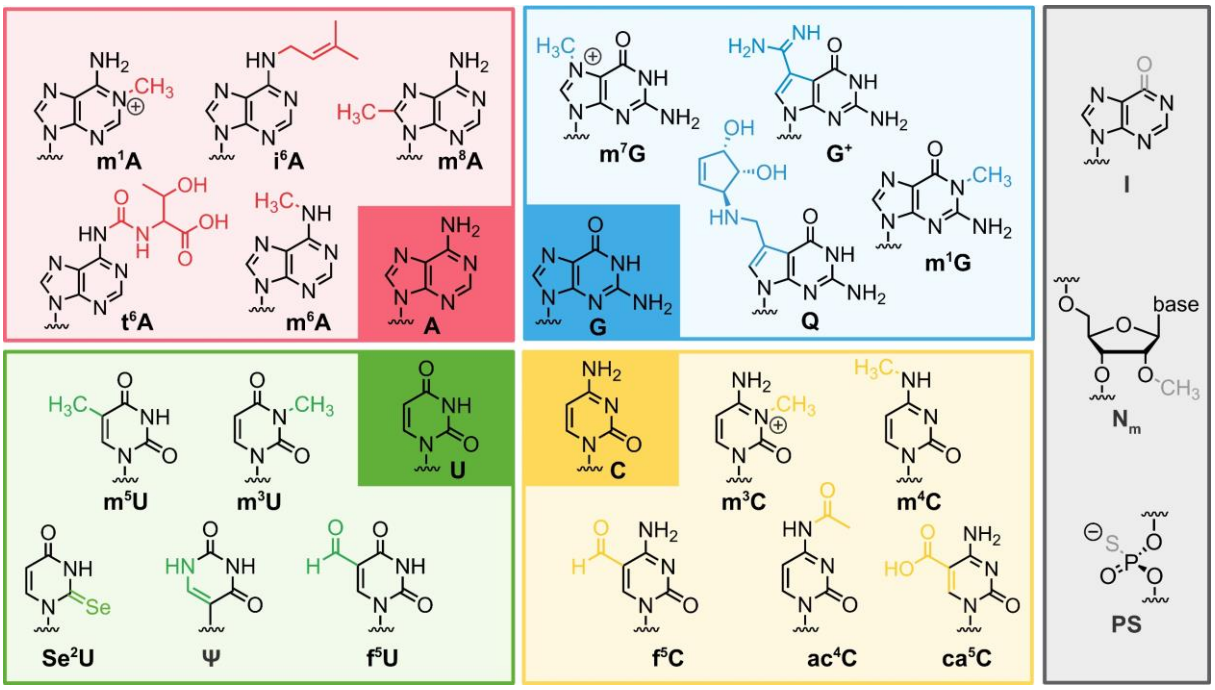



**Figure 1.2:** Selected examples of natural post-transcriptional RNA modifications at the nucleobase, ribose, or phosphate backbone. The respective modifications are highlighted.[4-5]

### 1.1.1.2 Functions of post-transcriptional RNA modifications

In general, RNA modifications can be found in all kinds of different RNAs (Figure 1.3), where they fine tune RNA structure and function thereby altering the interaction with other cellular components. While some modifications are present in many different RNA types, some only exist in one kind of RNA. Among all molecules, tRNAs are the most extensively modified nucleic acids with an average of 13 modifications per molecule.[9] Many of these modifications are located directly in or around the anticodon loop, where they expand the codon recognition and also enhance translation accuracy for example by preventing frameshift events. The strongest effects for mRNA decoding are obtained from (hyper-)modified nucleotides at the wobble position (position 34) and the nucleoside directly 3' of the anticodon residues (position 37).[10-11] Outside of the anticodon loop, modifications play crucial roles for the correct folding, stability, and rigidity of tRNAs. A notable example in this regard is the well-studied $m^1A9$ modification, which contains a methyl group at the Watson-Crick site and also carries a positive charge. It is found in almost all human tRNAs and for example impacts the folding of the human mitochondrial (mt) $tRNAs^{Lys}$: Through disruption of a Watson-Crick base pair between A9 and U64,

it shifts the equilibrium of the tRNA structure from a hairpin to the functional characteristic cloverleaf conformation.[12-13] Another common tRNA modification is pseudouridylation. Ψ occurs most frequently at position 55 in the TΨC-loop, the D-stem at position 13 as well as in the anti-codon stem (positions 38-39). Due to its strong base-stacking and -pairing properties, Ψ enhances the tertiary interactions within the tRNA and therefore improves its overall stability.[14] During translation, Ψ modifications in the anti-codon loop (positions 34-36) also enhances codon-anticodon recognition therefore modulating translation.[15]

Besides tRNA, rRNA is the second most heavily modified RNA with methylated 2' hydroxy groups ($N_m$) and Ψ as the two most abundant modifications.[16-17] These chemical alterations mostly cluster at functionally important regions. Those include the peptidyltransferase center where translation takes place as well as the ribosomal subunit interface, which is especially important for ribosome assembly and the ratcheting motion, which aids protein synthesis.[18-19] By stabilizing the rRNA at these crucial positions, RNA modifications enable efficient and correct mRNA translation.[20] Moreover, rRNA modification patterns are heterogenic and vary depending on environmental changes as well as during development of the organism or in diseases. Thus, they also play crucial roles in controlling translation as well as gene expression. In yeast, loss of multiple Ψ and $N_m$ sites has for example been reported to adversely affect folding and activity of the ribosome including higher rates of stop codon readthrough, slower amino acid incorporation and reduced rRNA levels.[21] Besides $N_m$ and Ψ, only seven more rRNA modifications have been found in human rRNAs that are located at critical positions between the ribosome subunits. Besides a single $m^7G$ residue, the 18S rRNA also contains two conserved $m^6{}_2A$ sites, two $ac^4C$ residues and a hypermodified Ψ. Less diverse modifications can be found within the large subunit including two $m^5C$ modifications, a single $m^1A$ site as well as one $m^5U$ residue.[20]

In mRNA, unlike in tRNA and rRNA, only a relatively small variety of post-transcriptional modifications has been identified. Among these, one of the most famous and well-studied modifications is 7-methylguanosine ($m^7G$) as it establishes the 5'-cap structure in mature eukaryotic mRNA, which is an essential structural motif for efficient mRNA translation. The cap-structure is generated co-transcriptionally and features $m^7G$, which is joined to the 5'-end nucleoside through a 5'-5'-triphosphate linkage. The ribose of the first transcribed nucleoside (most commonly A) can additionally be subjected to 2'-*O*-methylation resulting in a 5'-5' $m^7GpppN_m$ structure (Figure 1.3).[22] Furthermore, $A_m$ can also be methylated at the *N*6 position to generate $m^6A_m$, stabilizing the mRNA through resistance against the decapping enzyme DCP2.[23-24] The cap-structure also promotes the nuclear export of the mRNA as well as its polyadenylation, protects it against exonuclease degradation, defines the reading frame and regulates splicing, which all together has a profound impact on the cellular transcriptome and proteome.[25]

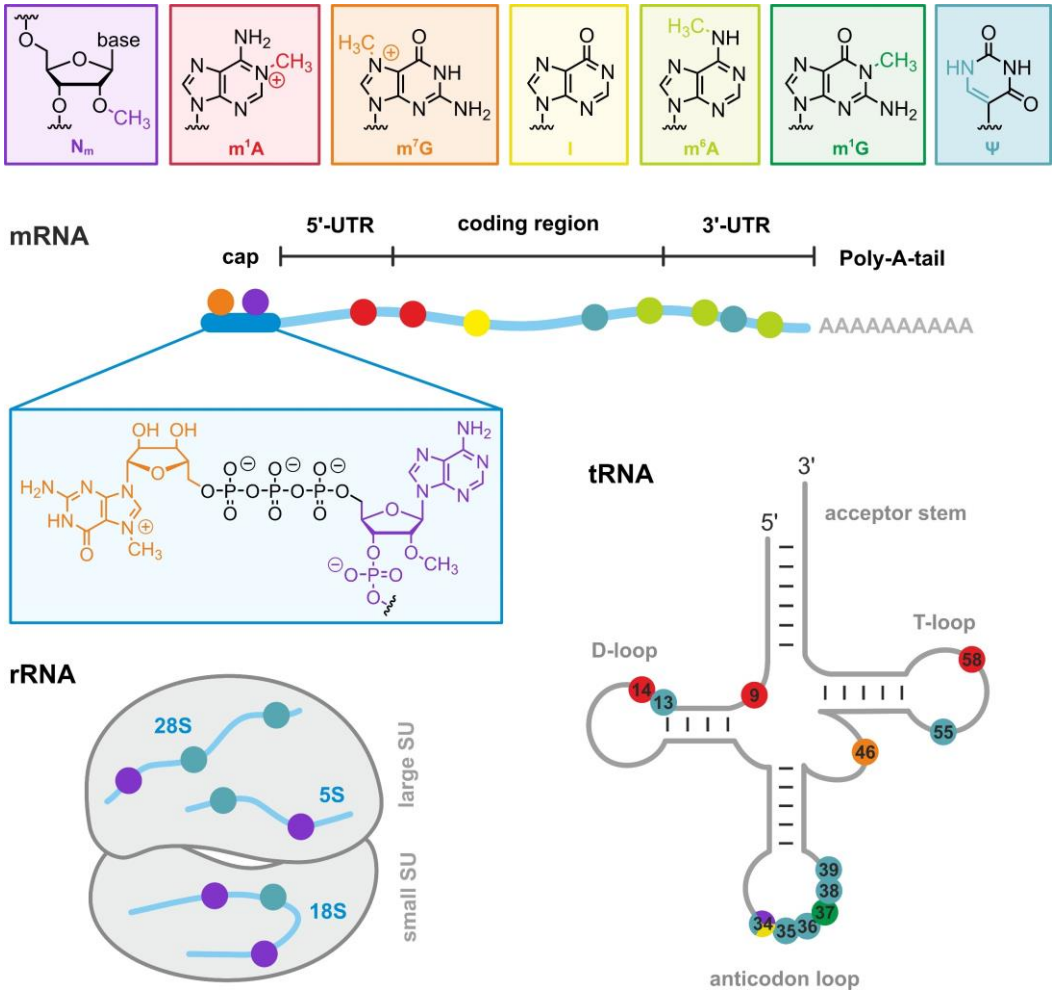



**Figure 1.3:** Schematic depiction of selected nucleoside modifications and their distribution in mRNA, rRNA and tRNA. Modification sites are shown in regions of particular enrichment and indicated with circles in different colors representing the respective modification. For tRNA, numbers show the respective modification position.

Methylation at *N*6 of A is however not only found in the mRNA 5'-cap: $N^6$-methyladenosine ($m^6A$) is the most abundant and therefore most heavily investigated modification in mRNA, which occurs at around three sites per molecule.[26-27] Initially discovered in 1974 in mRNA from rats[28], it has now been found in many eukaryotes, bacteria[29] and even viruses[30]. In mammals, $m^6A$ mostly occurs within the highly conserved DRACH-motif (D = A, G, U; R = A, G; H = A, C, U) [31] and is especially enriched in the 3' untranslated region (3' UTR) as well as around stop codons.[32] Investigations concerning the functions of $m^6A$ revealed its involvement in multiple processes like mRNA splicing[33], nuclear mRNA export[34], mRNA degradation[35], translation[36-37] or its impact on the secondary structure of mRNAs[38]. Analysis of different mRNAs showed that those highly modified with $m^6A$ often regulate developmental processes. In contrast, only few $m^6A$ sites were found in highly stable mRNAs encoding for 'house-keeping' genes like rRNAs.[39] In this regard, $m^6A$ installation and removal can be viewed as a kind of regulatory molecular switch for cell differentiation and development.[32, 36] Due to this function, $m^6A$ dysregulation has also been linked to a variety of diseases including different types of cancer, cardiovascular diseases, and obesity.[40-41]

In addition to 2'-*O*-methylation and $m^6A$, other methylated nucleosides can also be found in mRNA. Although well studied in tRNA and rRNA, $m^1A$ was only detected in human mRNA in 2016 by two independent groups using $m^1A$-immunoprecipitation ($m^1A$-IP) coupled to high throughput sequencing.[42-43] Both reports found that $m^1A$ is highly abundant in eukaryotic mRNAs as they identified nearly 1000 (from ~900 transcripts) or even >7000 $m^1A$ sites (from ~4500 transcripts), respectively. Additionally, both studies reported that $m^1A$ is especially enriched within the 5' UTR and in the coding sequence (CDS) around start codons. It was further found to be a highly dynamic, reversible modification that is especially installed under stress conditions and correlates positively with transcription initiation and protein expression, indicating an important regulatory role of $m^1A$. However, both reports failed to identify $m^1A$ at single nucleotide resolution and were not able to directly validate their results. Interestingly, a follow-up study of one of the groups used a conceptually similar approach based on $m^1A$-IP, but further validated their results using truncation and misincorporation patterns upon reverse transcription.[44] This now allowed for mapping $m^1A$ sites at single-nucleotide resolution. With these additional measures, only ~470 $m^1A$ sites could be detected in cytosolic mRNA. The modification sites were again especially enriched within the 5'UTR and often methylated at high levels. Additionally, the authors found 22 $m^1A$ sites in mt-mRNA, which were mostly located in the CDS, where they decreased translation efficiency. Remarkably, a later study used a highly similar method based on $m^1A$-IP coupled to reverse transcription but was only able to detect eight $m^1A$ sites in cytosolic mRNA and lncRNA, modified at low stoichiometries. In mt-mRNA, only a single $m^1A$ site could be identified. Additionally, the authors observe that the presence of $m^1A$ is associated with translation repression rather than enhancement.[45] As a consequence of the large discrepancy to the previous studies, the raw data of the earlier follow-up publication was reanalyzed:[46] It was found that a majority of $m^1A$ sites were detected incorrectly due to redundancies, mismapping and misannotation of reads or originate from sequencing errors. Only 53 out of the originally ~470 sites that were identified were confirmed as true $m^1A$ sites in cytoplasmic mRNA that were additionally only modified at very low to nearly undetectable levels. In summary, recent studies on $m^1A$ suggest its occurrence in human mRNA albeit at lower abundance and stoichiometry than reported in early studies. Additionally, $m^1A$ in mRNA is mostly present in loop structures resembling tRNA T-loops and might be negatively correlated with translation initiation.[45-46]

Another prevalent modification with regulatory function in human mRNAs is Ψ. Pseudouridylation mostly occurs inside the CDS and the 3' UTR of mRNAs and is found to be a highly dynamic moification.[47-48] Especially at stop codons, which all contain a uridine at the first position, conversion of U to Ψ is of particular significance. Pseudouridylation at these positions suppresses translation termination and instead generates sense codons coding for different amino acids.[49] Thus, pseudouridylation does not only provide a way to expand the genetic code:

Targeted U to Ψ isomerization also offers a promising approach for stop codon readthrough in the context of clinically relevant genetic diseases such as Duchenne muscular dystrophy[50] or cystic fibrosis[51] that are caused by single base nonsense mutations that lead to premature termination codons.[52]

Not only abundant in tRNAs, inosine (I) is another important mRNA modification. Generated by A-to-I editing, it is recognized as guanosine, therefore altering the coding sequence via introduction of a single point mutation. Through the same trait, it can also alter the secondary structure of mRNAs, and influence the interaction with different RNA binding proteins. Additionally, inosine is also recognized as G by the splicing machinery, thus being responsible for the generation of alternative mature mRNAs. In combination with the modulation of codon recognition, when it is present at the wobble position (I34) of tRNAs, inosine is a major translation modulator diversifying protein synthesis.[53]

#### 1.1.1.3 Installation and removal of methylation by writer and eraser enzymes

Post-transcriptional RNA modifications, and especially methylations are highly dynamic modifications, which are strictly regulated by methyltransferases ('writers') and demethylases ('erasers'). In order to understand the underlying mechanism of RNA epitranscriptome mediated gene regulation, identification and characterization of these RNA modifying proteins is key. Various enzymes from multiple classes are responsible for adding and removing methyl groups at different positions, but the general mechanism is similar for most of them: The methyl group is transferred via an acid-base catalysis from the universal donor substrate *S*-adenosylmethionine (SAM) onto a specific position within the target RNA nucleoside.[54]

##### 1.1.1.3.1 Writing and removing $m^6A$

For example, in case of the most abundant mRNA modification $m^6A$, the methyl group is installed highly specifically by a multi-subunit methyltransferase writer complex primarily in a co-transcriptional manner.[55] The most prominent and best-studied components of this complex are METTL3 (methyltransferase-like protein 3), METTL14 and Wilms' tumor 1-associating protein (WTAP) (Figure 1.4).[56-57] METTL3-METTL14 form a heterodimer, in which METTL3 is the catalytic component responsible for the methyl group transfer from SAM onto the *N*6 position of A. Deletion or inactivation of METTL3 results in nearly complete loss of all $m^6A$ sites in mRNA. METTL14, however, is the catalytically inactive partner of METTL3 with a disrupted SAM binding pocket. Instead, it acts as an allosteric activator that recognizes and binds to the target RNA.[56, 58-59] Similarly, WTAP is also a regulatory subunit for $m^6A$ synthesis

*in vivo*, essential for the formation of the functional methyltransferase complex. As the key METTL3 adaptor, it was found to stabilize the interaction between METTL3 and METTL14. Additionally, it is also responsible for recruiting the methyltransferase complex into nuclear speckles, where the pre-mRNA processing takes place.[55] Alongside WTAP, additional components of the writer complex form the so called m$^6$A-METTL associated complex (MACOM). These proteins aid the METTL3-METTL4 heterodimer (also known as m$^6$A-METTL complex, MAC) regarding complex assembly, stabilization, and methylation specificity,.[60-61]

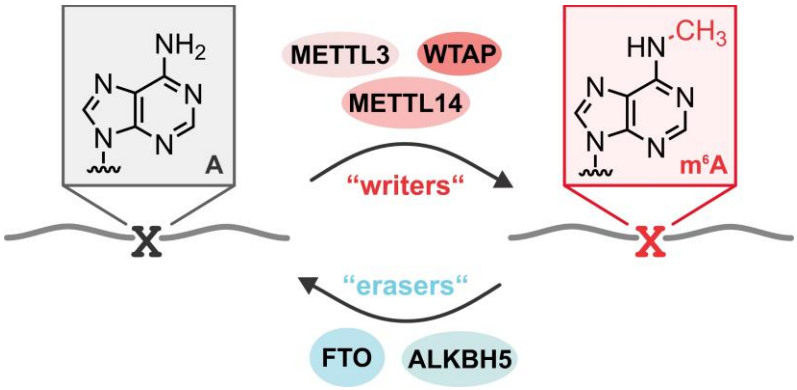



**Figure 1.4:** Schematic overview of the most important m$^6$A methyltransferases ('writers') and demethylase ('eraser') enzymes that regulate the m$^6$A content in mammalian mRNA.

Removal of m$^6$A in mRNA is mediated by two different enzymes. The first identified m$^6$A demethylase was the FTO (fat mass and obesity-associated) protein, which is also known as ALKBH9.[62] This non-heme Fe(II)- and α-ketoglutarate (α-KG) dependent dioxygenase enzyme was found to also demethylate other nucleosides like m$^3$T or m$^3$U, but with significantly lower efficiency. Hence, m$^6$A in nuclear mRNA was identified as its main biologically relevant target.[38, 62-63]

Additionally, m$^6$A$_m$ in the 5'cap structure was reported as another relevant substrate for FTO. Because FTO is localized in the nucleus, while m$^6$A-modified mRNA occurs within the cytosol some reports even questioned the significance of FTO for m$^6$A-demethylation.[64] This was further supported by the finding that m$^6$A$_m$ was demethylated even more efficiently than m$^6$A.[23, 65] However, more recent reports strongly corroborate the fact that both, m$^6$A and m$^6$A$_m$ are biologically relevant targets for FTO *in vivo*. This is further supported by the finding that a fraction of FTO also shuttles to the cytoplasm where it can then target m$^6$A in mature mRNA.[66] Additionally, overexpression of FTO in certain types of cancer, was for example also found to reduce m$^6$A levels in mRNA transcripts, further substantiating the role of FTO for removal of m$^6$A in mRNA transcripts.[67]

Mechanistically, demethylation by FTO follows an oxidative reaction pathway that proceeds through the formation of two intermediates (Figure 1.5): First, m$^6$A is oxidized to *N*6-hydroxymethyladenosine (hm$^6$A) that can be further oxidized to *N*6-formyladenosine (f$^6$A), which

is then hydrolyzed to adenosine. Alternatively, hm$^6$A can also directly decompose generating adenosine with release of formaldehyde. Due to their long half-lives under physiological conditions (~3 h), both intermediates, hm$^6$A and f$^6$A, can even be detected in isolated mRNA.[68-69]

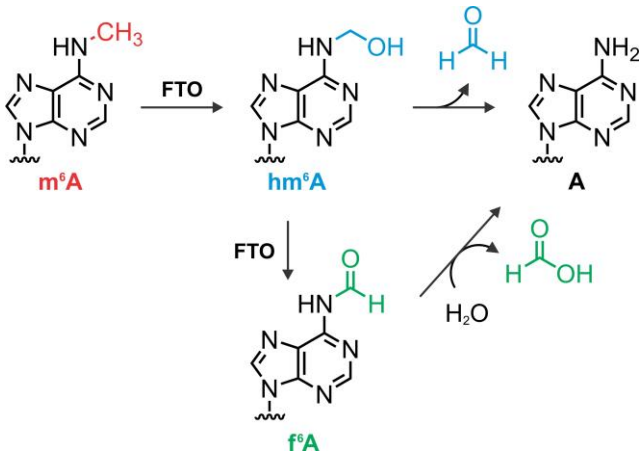



**Figure 1.5:** Demethylation of m$^6$A by FTO proceeding via an oxidative reaction mechanism with the formation of two different intermediates (hm$^6$A and f$^6$A).

Besides FTO, the second m$^6$A demethylase found in human cells is ALKBH5, which also belongs to the conserved family of Fe(II)/α-KG dependent dioxygenase enzymes. Consequently, FTO and ALKBH5 employ the same oxidative mechanism for removal of the *N*6-methyl group. However, unlike FTO, ALKBH5 catalyzes the direct demethylation of m$^6$A to yield adenosine without releasing the oxidized intermediates hm$^6$A and f$^6$A.[38] Moreover, it has no activity against m$^6$A$_m$ and is only localized within the nucleus where it demethylates m$^6$A in mRNA co-transcriptionally.[23]

#### 1.1.1.3.2 Epitranscriptomic role of other AlkB family dioxgenase enzymes

In addition to FTO and ALKBH5, seven more homologues of the AlkB family have been identified so far (ALKBH1-8). Those do not only catalyze RNA demethylation, but a wide range of different reactions, thereby distinctly influencing the DNA and RNA epitranscriptome.[70] For instance, ALKBH1 is a mitochondrial protein that was found to be responsible for demethylation of m$^3$C in ssDNA and mRNA, m$^1$A in tRNA or m$^6$A in DNA.[71-72] Additionally, it can also function as a demethylase of H2A histones.[73] The homologues ALKBH2 and ALKBH3 mostly act as nuclear DNA repair enzymes that protect cells through restoration of DNA alkylation.[74-75] ALKBH3 was additionally identified to demethylate m$^1$A and m$^3$C in RNA.[76] ALKBH8 on the other hand facilitates synthesis of further RNA modifications. It possesses an additional SAM-dependent methyltransferase domain and is involved in tRNA maturation through hypermodification of U at the wobble position of tRNAs (Figure 1.6). Through methylation of 5-carboxymethyluridine (cm$^5$U) it generates 5-methoxycarbonylmethyluridine (mcm$^5$U), which is further hydroxylated to yield 5-methoxycarbonylhydroxymethyluridine (mchm$^5$U).[77-79] In contrast,

ALKBH4 does not act on nucleic acids, but on proteins. It was found to demethylate a lysine residue in cytoplasmic actin, thereby modulating actin-myosin interactions.[80]

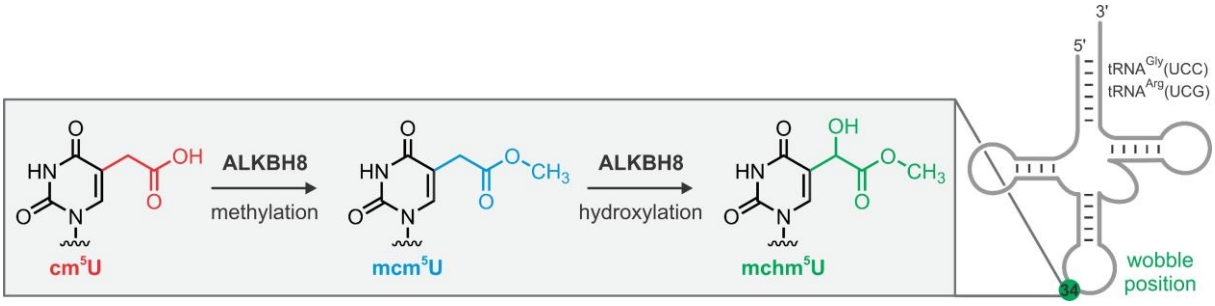



**Figure 1.6:** Sequential conversion of $cm^5U$ to $mchm^5U$ via $mcm^5U$ by ALKBH8 at the wobble position (U34) in $tRNA^{Gly}$ (UCC) or $tRNA^{Asp}$ (UCG).[79]

Despite intense research, the exact role of all AlkB homologues particularly regarding the RNA epitranscriptome are still not entirely uncovered and understood. Especially questions concerning the exact targets, the localization and the underlying reaction mechanisms of each enzyme are of special interest. Among others, this is of particular significance in the field of biomedical chemistry, as up- and down-regulation of AlkB homologues have been shown to be associated with the development of certain types of cancer. Hence, enzymes from this family are potential new targets for the development of novel anti-tumor agents.[81-82]

#### 1.1.1.3.3 Installation and removal of $m^1A$, $N_m$ and Ψ

Compared to the generation of $m^6A$, methylation of adenosine at the *N*1 position is mediated by a similar enzyme complex: In case of writing and removing $m^1A$ modifications, the best studied example is the human tRNA modification $m^1A58$. Its installation is also achieved by a two-component methyltransferase complex. Similar to the METTL3-METTL14 heterodimer, it consists of a catalytically active methyltransferase with SAM binding pocket (TRM61A) and an RNA binding protein responsible for tRNA recognition (TRMT6).[83] The TRMT6/TRMT61A complex was also found to catalyze $m^1A$ installation in human mRNAs. However, it requires a tRNA T-loop-like structure and consensus sequence for substrate recognition.[45] As mentioned above, demethylation of $m^1A$-modified RNAs is catalyzed by ALKBH3.[76] For removal of $m^1A$ in tRNA, ALKBH1 and FTO were also found to be important demethylases.[71-72]

In contrast to $m^6A$ and $m^1A$, $N_m$ and Ψ in rRNA are installed by small nucleolar ribonucleoproteins (snoRNPs). As the name already suggests, these consist of a snoRNA that directs the associated catalytically active protein units to the target modification site via antisense guide base pairing. In general, two major classes of snoRNPs can be distinguished: Box C/D and Box H/ACA, where the nomenclature refers to conserved sequence elements found within the snoRNAs. Associated with four different protein cofactors, Box C/D snoRNPs catalyze the

methylation of the ribose 2'OH[84] with fibrillarin (FBL) as the catalytically active methyltransferase enzyme.[85] BoxH/ACA snoRNPs on the other hand are responsible for pseudouridylation.[86] For the isomerization of U to Ψ, the snoRNA also needs to be associated with four different proteins with the enzyme Dyskerin (DKC1)/Cbf5 as the catalytic component.[87]

Even though the installation, removal, distribution, and functions of many RNA modifications have been investigated so far, the exact roles of many modified nucleosides are still indeterminate. Recent breakthroughs regarding the detection of RNA modifications, which are especially contributed to the development of high-throughput sequencing technologies may help to further expand the alphabet of RNA epigenetics. Those methods, that will be discussed below, may also help to gain deeper insights into the dynamics of the RNA epitranscriptome. This is also of particular importance in the context of genetic diseases considering the impact of post-transcriptional modifications for the regulation of cellular processes.

## 1.1.2 Detection of RNA modifications

Owing to the importance of post-transcriptional modifications for the structure, cellular functions and fate of RNA, a great interest was developed to study their abundance and distribution in more detail. As a result, numerous methods for the identification of chemically altered nucleotides in RNA have been established over time, and new methods are constantly being evolved and further refined. Those detection methods provide essential information for a more detailed understanding of RNA modifications and their roles in RNA metabolism.

### 1.1.2.1 Chromatographic methods

Due to their small size and low molecular weight, especially small modifications like methyl groups do not influence the physiochemical properties of cellular RNA enough to unequivocally discriminate modified from unmodified oligonucleotides. Thus, early attempts were based on digestion of RNA into single nucleotides or nucleosides that can then be analyzed more easily with different methods, as they then show highly distinct characteristics.

Historically, 2D thin layer chromatography (TLC) is a simple, cost-efficient method that can be used to separate modified from unmodified nucleosides based on their net charge, polarity and hydrophobicity.[88-89] Upon digestion into mononucleotides with nucleases such as RNase P1, RNase T2 or snake venom phosphodiesterase (SVPD), they can be separated from each other on cellulose plates using appropriate solvent mixtures. The resulting pattern can then either be visualized by UV shadowing or - in case of radioactively labeled RNAs - by autoradiog-

raphy.[89] The obtained mobilities of the single nucleotides are compared to reference maps to identify the specific modification status of single nucleosides. 2D-TLC can also be conducted in a preparative scale to allow further detailed analysis of specific methylated nucleosides by UV-spectroscopy or mass spectrometry. Using just three standard solvent mixtures, this simple method enabled the identification of more than 70 distinct RNA modifications.[89] Even though 2D-TLC in combination with UV shadowing is relatively simple to perform and cost-efficient with only little equipment required, it suffer from low sensitivity. Hence, it is usually only applied for detection of abundant modifications. Although isotope labeling greatly enhances the sensitivity of this method, it is often avoided due to the associated radioactivity.

Alternatively, the fully digested RNA can also be analyzed using high performance liquid chromatography (HPLC) as an advanced column chromatography based method.[90] Just like in 2D-TLC, separation of the single nucleotides is achieved based on their difference in hydrophobicity and polarity resulting in different retention times. The identity of the modification can then be determined through comparison with known (synthetic) standards. To increase the detection sensitivity, HPLC can also be coupled to mass spectrometry (LC-MS).[91] With this setup it is possible to determine the exact mass of each eluting species making it a powerful method to investigate RNA modifications on the nucleoside level. However, as the signal intensity in MS-based methods depends on the ionization potential of the compound rather than on its absolute amount, no real quantitative information can be obtained without the use of isotope-labeled reference material.[92-93] Additionally, since the RNA is fully digested, LC-MS and all other chromatographic methods discussed only provide a general transcriptome wide analysis of modified nucleosides and lack information about their sequence context.

#### 1.1.2.2 Locus-specific detection

However, site-specific detection of RNA modifications is key as their exact location is crucial for proper RNA folding, function and stability. Additionally, it is also desirable to obtain quantitative information about the abundance of a modification, as those with higher stoichiometry are more likely to have a profound influence on the fata and function of an RNA. One possibility to achieve locus-specific detection that is also TLC-based was developed by Liu *et al.* in 2013.[94] The method combines site-specific cleavage and radioactive labeling followed by ligation-assisted extraction and thin-layer chromatography (SCARLET) and can be used to accurately determine the exact position and modification level of $m^6A$ sites in RNA at single-nucleotide resolution. In this advanced multi-step method, the modified RNA of interest does not need to be purified first, instead SCARLET can be applied to total cellular RNA extract. However, the exact sequence context of the methylated nucleoside must be known.

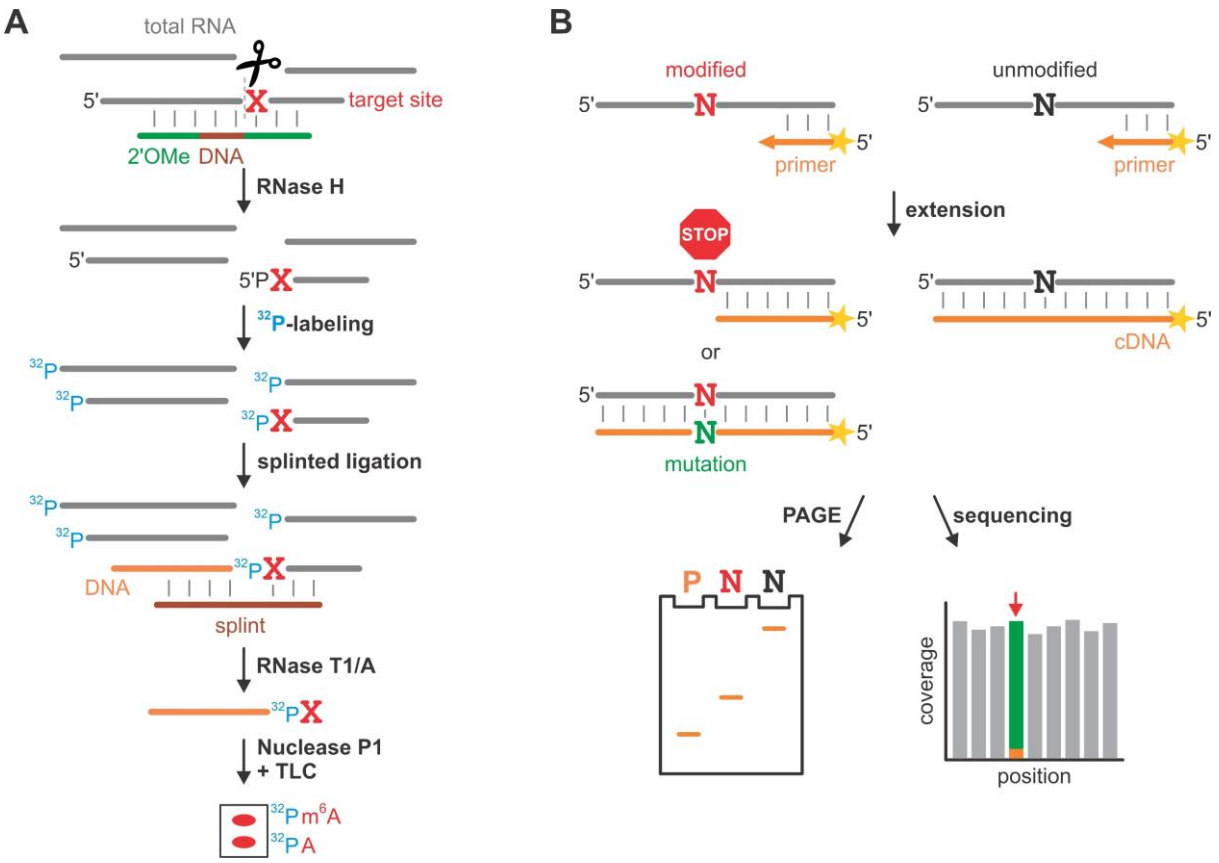



**Figure 1.7: A** Overview of SCARLET. Upon RNase H mediated cleavage at the target site (red), 5'-$^{32}$P-labeling is performed (blue). After splinted ligation to a DNA (orange), the RNA (gray) is fully digested. The isolated ligation product is then treated with nuclease P1 and the resulting nucleotides are analyzed via TLC. Adapted from ref.[94] **B** Reverse transcription of an RNA. At the modification site (red), primer extension is either blocked or misincorporation (green) occurs. Readout of the stop signal can be achieved via PAGE analysis. P = primer. Alternatively, mutations can be detected through sequencing.

In the first step, a 2'OMe-RNA/DNA chimera is annealed to the target RNA, which recruits RNase H resulting in site-specific cleavage 5' of the expected modification site. Next, the cleavage site is radiolabeled with $^{32}$P and then attached to a ssDNA using splinted ligation. Upon treatment with nucleases, all RNAs are fully digested leaving only the ligated DNA with the nucleoside of interest attached to the 3'-end intact. The m$^6$A modification status is then determined by nuclease-mediated digestion of the DNA followed by TLC (Figure 1.7A). The power of SCARLET is the sequence-dependent ligation step, enabling the use of RNA mixtures as the input material while retaining a high sensitivity. Starting from total cellular RNA extract, this method was successfully employed to study the abundancy of m$^6$A sites in human 18S rRNA, several mRNAs and lncRNAs[94] and even adapted for profiling other modifications like pseudouridine.[48] However, as a specific 2'OMe-RNA/DNA chimera oligonucleotide needs to be annealed, the exact sequence information of the modification site is required and modifications at the Watson-Crick site cannot be detected.

Instead, identification of methylations that disturb base pairing can be detected and localized using reverse transcription (RT). This method has been used extensively to study a wide variety of methylated nucleotides[95-99] including m$^1$A, m$^1$G, m$^3$C or m$^3$U. In this traditional, enzyme-

based approach, a specific 5'-end-labeled DNA primer is annealed to the RNA of interest and a reverse transcriptase enzyme is used to synthesize the complementary ssDNA (cDNA). In the absence of methylations, the extension proceeds until the 5'-end of the RNA is reached, generating a full-length product. However, when a modification is encountered, the extension is either blocked resulting in a shorter, abortive cDNA or misincorporation at the modification site will occur. Upon separation of the RT products on denaturing PAGE or analysis by sequencing, the exact position of the modification can then be determined (Figure 1.7B). Additionally, owing to the sequence-specific annealing of the RT primer, this method is very sensitive and only requires small amounts of input RNA. However, RT silent modifications such as $m^6A$ or $m^5C$ that do not affect the Watson-Crick face cannot be detected by RT primer extension.

#### 1.1.2.3 NGS-based methods

All detection methods discussed so far can only be used to investigate the modification status of a specific RNA. For a transcriptome-wide analysis, various high throughput sequencing approaches have been developed. Because of its high abundance and therefore great significance for correct RNA functions and gene regulation, early methods focused on detection of $N^6$-methylated adenosine. Due to the lack of chemical methods to differentiate A from $m^6A$, early $m^6A$ sequencing techniques, named $m^6A$-seq[100] and MeRIP-seq[32] (methylated RNA immunoprecipitation with next-generation sequencing) utilize modification-specific antibodies. In this approach, the total RNA is first fragmentated into pieces of around 100 bp followed by pre-enrichment of $m^6A$ containing RNAs by immunoprecipitation using modification-specific antibodies. The precipitated fragments are then submitted to next-generation sequencing (NGS) and the obtained reads are aligned to the transcriptome. RNA regions that show high read numbers are then assumed to contain at least one $m^6A$ modification site (Figure 1.8A). Using alternative antibodies, this sequencing method was adapted to detect alternative RNA modifications such as $m^1A$ ($m^1A$-seq and $m^1A$-ID-seq)[43, 101] or $m^5C$ ($m^5C$-RIP-seq)[102].

However, limited by the fragment size, both methods suffer from a low resolution (100-200 nt) and cannot detect more than one modification site in close proximity as they will all be covered by the same $m^6A$ peak. Additionally, the choice and quality of antibodies is crucial, as non-specific binding can generate false positive results. To tackle those issues, the immunoprecipitation-based methods were development further by using UV cross-linking of the antibody with a nearby nucleoside. In an approach termed PA-$m^6A$-seq (photo-crosslinking-assisted $m^6A$-sequencing), cells are provided with 4-thiouridine ($s^4U$), a photoactivatable ribonucleoside, prior to extraction of the RNA, resulting in incorporation of 0.1-1 % $s^4U$ at uridine sites during

transcription. After binding to m$^6$A-specific antibodies, the RNA is then treated with UV light resulting in cross-linking of s$^4$U to the antibody.[103] During library preparation, this linkage is read as a C, thus generating a U-to-C mutation, which can be used to determine the position of m$^6$A with a resolution of up to 23 nt. Further improvement was achieved with miCLIP (m$^6$A individual nucleotide resolution cross-linking and immunoprecipitation), that circumvents the need for introduction of modified nucleosides such as s$^4$U.[104] Instead, a natural nearby nucleoside in the RNA is crosslinked to the antibody upon irradiation with UV-light. During reverse transcription, mutations adjacent to the modifications site as well as truncations are introduced in the cDNA that enable mapping of m$^6$A sites (Figure 1.8B). Even though those strategies significantly improve the resolution, still no quantitative information can be extracted from the sequencing data. Additionally, comparably large amounts of input RNA and costly antibodies are required that often show strong off-target effects.[105] Therefore, alternative sequencing methods that do not rely on antibodies are desired.

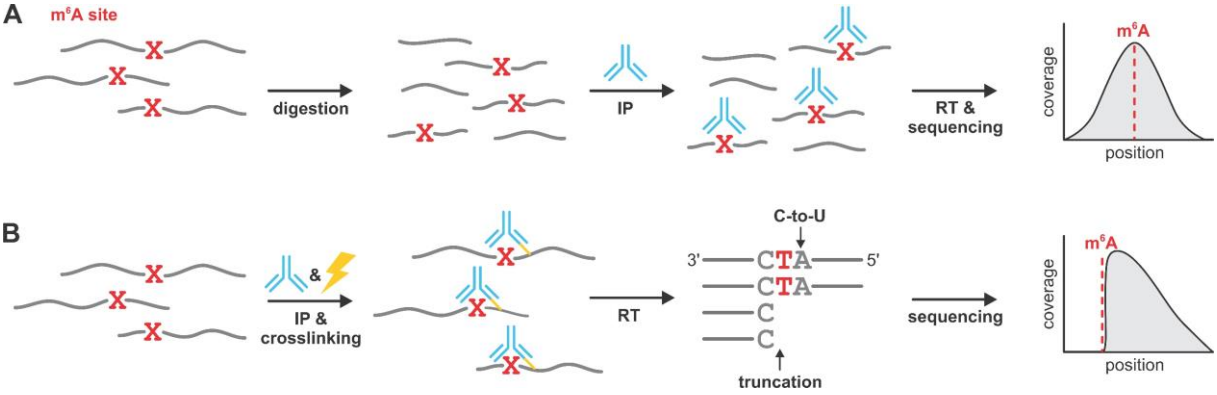



**Figure 1.8:** Schematic illustration of immunoprecipitation-based detection methods for m$^6$A. **A** In m$^6$A-seq and MeRIP-seq RNA is fragmented, followed by immunoprecipitation (IP), reverse transcription (RT) and sequencing. **B** For miCLIP, the antibody is crosslinked to the intact RNA introducing C-to-U mutations or truncations in the cDNA.

Just like antibodies, some modification-specific enzymes can also be employed for affinity pull-down or to induce base-specific editing, which can subsequently be read out by sequencing. Recent developments include DART-seq and m$^6$A-SEAL that can be used for detection of m$^6$A at high resolution in the natural DRACH consensus motif. DART-seq (deamination adjacent to RNA modification targets sequencing) uses APOBEC1, a cytidine deaminase that is fused to the m$^6$A-binding YTH domain of a native human m$^6$A reader protein.[106] Upon binding of YTH to a methylation site, APOBEC1 is directed to the adjacent C, where it introduces a C-to-U mutation, which is subsequently identified using NGS (Figure 1.9A). This approach was further expanded by integration of DART-seq into a single cell RNA sequencing platform, resulting in scDART-seq.[107] With this method, the m$^6$A modification profile in individual cells can be determined.

In contrast, m$^6$A-SEAL makes use of the natural eraser enzyme FTO, which catalyzes the oxidative demethylation of m$^6$A through the unstable intermediate hm$^6$A (see chapter

1.1.1.3.1).[108] Following fragmentation of the RNA and subsequent FTO treatment, hm$^6$A is further reacted and thereby stabilized with dithiothreitol (DTT). The *N*$^6$-dithiolthreitolmethyl-adenosine (dm$^6$A) thus generated contains a free thiol moiety, which can then be used for labeling with biotin followed by pull-down using streptavidin-coated magnetic beads. By sequencing of the enriched RNA fragments, accurate mapping of m$^6$A sites across the transcriptome is possible (Figure 1.9B).

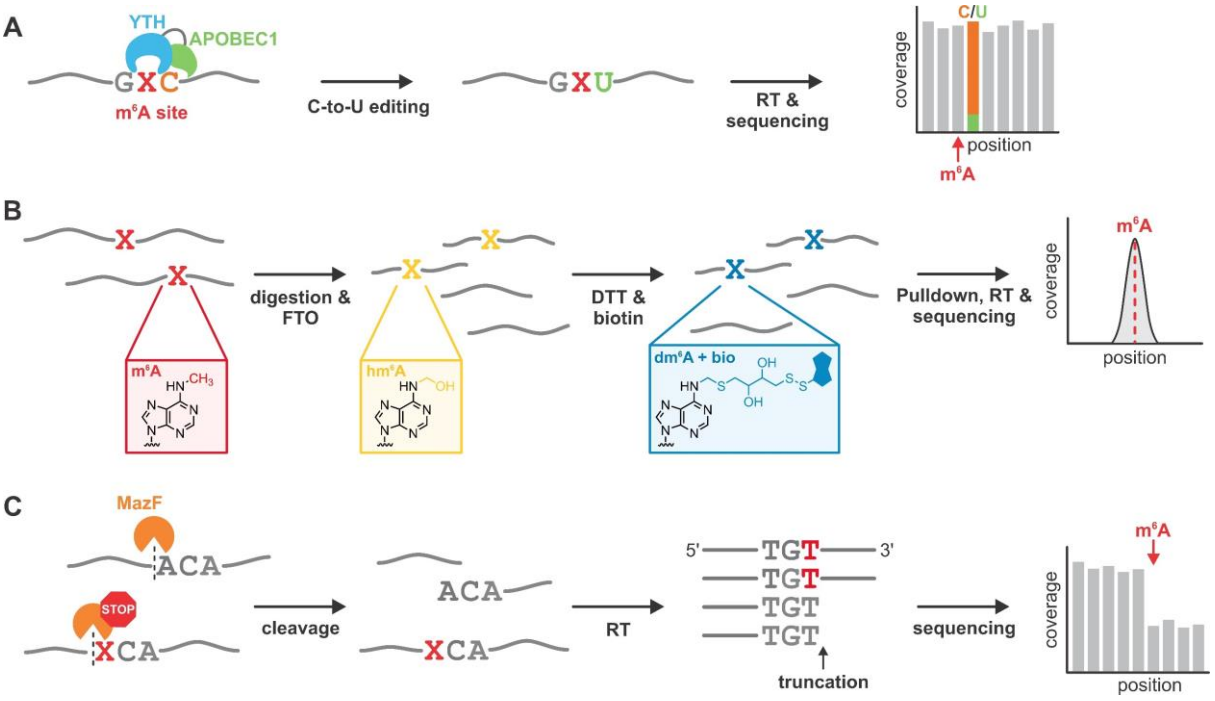



**Figure 1.9:** Schematic illustration of enzyme-based detection methods for m$^6$A. **A** DART-seq uses the m$^6$A-binding YTH domain fused to the cytidine deaminase APOBEC1 for site-specific C-to-U editing that is read out via NGS. **B** In m$^6$A-SEAL, FTO-mediated oxidation followed by treatment with DTT and conjugation with biotin allows pull-down of modified sequences. **C** MAZTER-seq employes MazF endonuclease for site-specific cleavage of unmodified RNA. Alignment of truncated fragments with 5'ACA allow for detection and quantification of m$^6$A.

Two more enzyme based approaches, m$^6$A-REF-seq (RNA endoribonuclease facilitated sequencing) and MAZTER-seq employ the *E. coli* RNA endoribonuclease MazF that specifically cleaves immediately upstream of unmodified ACA sequence motifs but is inhibited by (m$^6$A)CA.[109-110] After reverse transcription of MazF treated RNA and high-throughput sequencing, alignment of individual reads allow detection of m$^6$A sites (Figure 1.9C). Analyzing the number of reads that begin, end, and pass through an ACA site, provide additional quantitative information for each m$^6$A modification site. However, the recognition capability of MazF can be negatively influenced by RNA secondary structures and is also strongly dependent on the ACA sequence motif. Thus, when calculating the m$^6$A modification level, careful analysis is required to obtain accurate results. Consequently, in m$^6$A-REF-seq, all ACA sites, which are predicted to occur within a double stranded RNA region are removed from the sequencing data as this impairs the MazF cleaving efficiency.[110] In contrast, MAZTER-seq follows a slightly different approach: Here, the sequencing data is pre-filtered to for example remove any ACA sequences

that are too close to each other and can therefore not be distinguished accurately. Additionally, all ACA sites with a G at position +3 are discarded, as the MazF cleavage efficiency is diminished by this sequence context. As a result, only up to 25 % of $m^6A$ sites in yeast and 16 % within the mammalian transcriptome can be detected using this endonuclease-based approach.[109]

In addition to biochemical approaches, the distinct reactivity with specific chemicals can also be used to differentiate modified nucleosides from their unmodified counterparts. For example, the following sequencing approaches employ two different simple chemical reactions for selective fragmentation enrichment to profile ubiquitous 2' ribose methylation - one of the most abundant modifications in RNA.

RiboMeth-seq is based on the different stability of unmodified and 2'-*O*-methylated RNA nucleosides under alkaline conditions.[111-112] Typically, the 2'OH groups at the ribose are activated in basic environments and attack the neighboring phosphodiester linkage, which results in RNA fragmentation. 2'OMe nucleotides, however, are resistant to this nucleophilic cleavage mechanism. In RiboMeth-seq, total RNA is first fragmented under alkaline conditions, resulting in a pool in which RNAs with a 2'OMe modification close to the 3'- or 5'-end are strongly underrepresented. After library preparation, reverse transcription and high throughput sequencing, the obtained read ends are plotted against the sequence allowing localization of 2'OMe sites as those will generate a gap in the 5'-/3'-end coverage profile (Figure 1.10A). This sequencing method does not only yield single nucleotide resolution, but additionally even enables quantitative analysis of each methylation site. However, since RNA fragmentation under alkaline conditions is irregular and random, relatively high read coverage is required, limiting this approach to abundant RNA species.

To overcome these constrains, alternative sequencing methods were developed that rely on the different behavior of unmodified and 2'OMe modified ribose when treated with sodium periodate ($NaIO_4$). In RibOxi-seq[113] (ribose oxidation sequencing) and Nm-seq[114], fragmented RNA is submitted to iterative cycles of periodate oxidation (yielding 2',3' vicinal diols) followed by β-elimination and dephosphorylation to remove nucleosides from the 3'-end in a stepwise manner. As 2'OMe ribonucleosides are resistant towards oxidation with $NaIO_4$, this process leads to gradual enrichment of RNA fragments with 2' methylated ribose at the 3'-end. Furthermore, oxidized RNA is rendered incapable of linker ligation through library preparation, additionally eliminating RNA with unmodified 3'-ends from the library. After sequencing, alignment of the 3'-end reads to a reference genome allows for transcriptome-wide detection of 2'OMe modification with single-nucleotide precision (Figure 1.10B). Compared to RiboMeth-seq, these oxidation-based methods rely on detection of a positive signal rather than the absence of signals. In combination with the enrichment step, a high sensitivity can be achieved,

making RibOxi-seq and Nm-seq especially suitable for the detection of 2'OMe sites in rare RNAs.

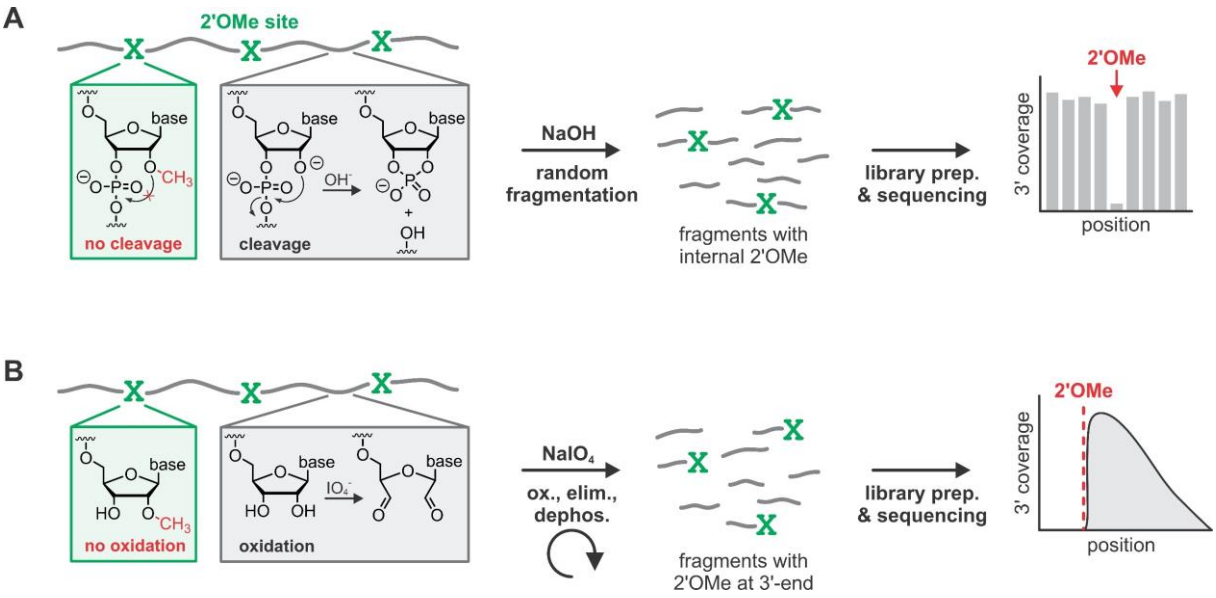



**Figure 1.10:** Schematic illustration of detection methods for 2'OMe modifications based on altered reactivity towards different chemicals. **A** In RiboMeth-Seq, alkaline treatment induces cleavage at unmodified positions resulting in fragments with internal 2'OMe modifications. Alignment of 3'-end read coverage results in gaps for 2'OMe modified nucleosides. **B** RibOxi-seq and Nm-seq use repeated cycles of $NaIO_4$-oxidation, elimination and dephosphorylation for stepwise removal of unmodified nucleotides resulting in fragments with 2'OMe at the 3'-end.

Similar approaches have also been applied for mapping other methylated nucleotides such as $m^7G$ or $m^3C$, that are prone to abasic site generation in an alkaline environment. AlkAniline-seq combines mild basic treatment with subsequent aniline cleavage of the RNA abasic sites generating fragments with a terminal 5' monophosphate.[115] Through selective adaptor ligation at these positions during library preparation, $m^7G$ and $m^3C$ residues can uniquely be detected as the beginning of sequencing reads will accumulate at positions carrying a modification.

Alternatively, $m^3C$-mapping is also possible combining hydrazine treatment with subsequent aniline cleavage. In HAC-seq (Hydrazine-Aniline Cleavage sequencing)[116], hydrazine is specifically added to $m^3C$ residues under high salt conditions followed by aniline treatment, which induces RNA cleavage at the modification sites. Detection of those cleavage positions through sequencing then allows mapping of $m^3C$ modifications across the entire transcriptome with high resolution.

#### 1.1.2.4 Nanopore sequencing

NGS-based detection methods have found wide application for the transcriptome wide mapping of various RNA modifications. However, they rely on very distinct and often complex reactions for detection of a specific modification, thus they cannot map multiple modifications at once. Those approaches are also only able to generate an indirect readout of the modification status of an RNA and can easily accumulate amplification bias during library preparation. Additionally, they require a laborious library preparation, and the sequencing length is limited to up to 300 bp.

As an alternative, a direct RNA-sequencing technology, namely nanopore sequencing, is heavily researched for detection of RNA methylation. Ever since the first nanopore sequencer MinION was released in 2014 by Oxford Nanopore Technologies (ONT)[117], this technology and its possible applications are constantly evolving. In short, nanopore sequencing is based on a nanoscale protein pore (nanopore), which is located in a membrane, surrounded by an electrolyte solution. By applying a constant voltage to the nanopore an ionic current is produced, generating a negatively charged 'cis' and a positively charged 'trans' side (Figure 1.11A). The oligonucleotide is then moving through the pore from cis to trans controlled by a motor protein, thereby generating changes in the ionic current. Those changes are characteristic for each nucleotide and can be decoded computationally ('base-calling') to obtain the full sequence information (Figure 1.11B).[118-119] Due to the helicase activity of the motor protein, also double stranded oligonucleotides can be sequenced. In contrast to NGS, no complicated multi-step sample preparation is required. After simple addition of a poly-A-tail and ligation of an adapter as the detection signals, the RNA is directly sent through the nanopore, which in turn decreases amplification bias.[120] Additionally, sequencing frames generated by nanopore sequencing are long enough to cover the entire length of a typical mature mRNA in human.[121]

Nanopore sequencing can also be used for the detection of modified nucleosides without the need for prior chemical or biochemical treatments of the oligonucleotide. This is either possible through analysis of altered ionic current intensities (Figure 1.11C) or systematic base-calling errors (Figure 1.11D) caused by the modification.[122] For instance, just recently, the nanopore technology enabled direct sequencing of full-length tRNAs including detection of multiple nucleoside modifications based on systematic base miscalling.[123] Other examples include successful mapping of $m^7G$ and ψ in *E. coli* 16S rRNA[124] or detection of $m^6A$ and $m^5C$ in native RNA sequences.[120, 125-126]

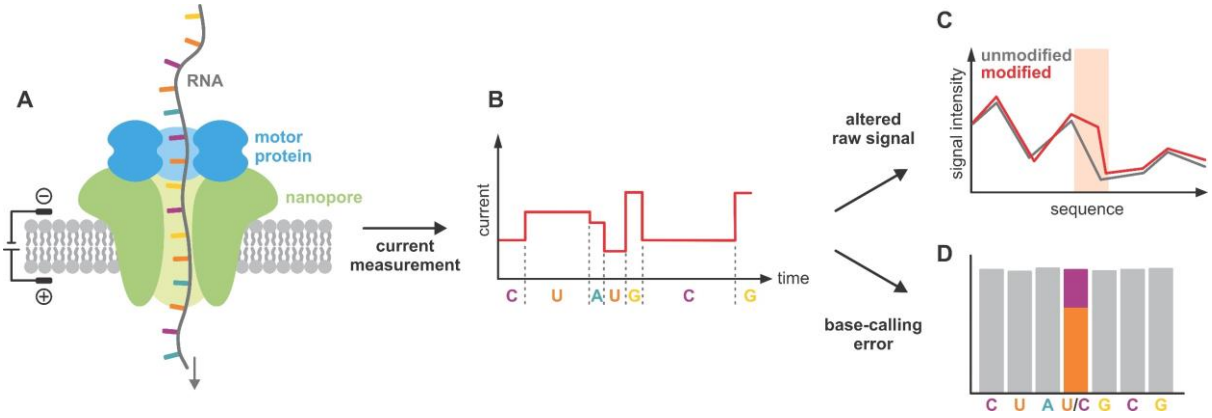



**Figure 1.11:** Schematic depiction of nanopore sequencing for detection of nucleoside modifications. **A** Overview of the nanopore construction. The nanopore (green) is associated with a motor protein (blue) and embedded within a membrane (grey). An ionic current is applied to the pore enabling the RNA to pass through the pore controlled by a motor protein. **B** The ionic current changes depending on which nucleoside passes through. RNA modifications can be detected through an altered intensity of the raw signal (**C**) or base-calling errors (**D**). Adapted from ref.[122, 127].

However, the field still faces some challenges like the comparably high error rate of up to 4-6 % or the relatively large amount of input material required.[118] Moreover, for detection of modifications using ionic current alterations, an initial training of machine learning algorithms is needed, for which biologically realistic datasets for both, modified and unmodified RNAs are required. Ideally, these reference oligonucleotides cover the modification of interest embedded within all possible sequence contexts, which is often laborious and difficult to obtain. Additionally, currently available algorithms can only be applied individually for the detection of one specific modification at a time and different algorithms are needed to analyze alternative modifications. When using base-calling errors for the detection of modified nucleosides, high precautions regarding the sequencing data quality have to be taken to obtain accurate results. In addition, high quality reference datasets are required for this approach as well. Only with these measures, random sequencing errors can accurately be distinguished from genuine modifications.[128]

Nevertheless, tools to directly extract information about RNA modifications from sequencing data are just being developed and are still constantly evolving. Considering that nanopore sequencing is a relatively new technique, the development of more powerful computational algorithms and improvement of the nanopore itself might soon improve data quality and analysis.

### 1.1.2.5 Deoxyribozymes

Besides sequencing methods, post-transcriptional RNA-methylation can also be read out using nucleic acids themselves, more precisely catalytically active DNA. Such DNA enzymes, also called deoxyribozymes, are short non-natural ssDNA strands that were evolved in the laboratory to catalyze a specific chemical reaction (see chapter 1.2.2.2). In general, RNA-

cleavage catalyzed by a DNA enzyme is mediated by the attack of a 2'-OH on the adjacent phosphodiester bond, resulting in the formation of a 2'-3'-cyclic phosphate and a free 5'-OH. For instance, this mechanism can easily be employed to read out the modification status of the attacking 2'-hydroxy group as a modification of this position (e.g., 2'-*O*-methylation) prevents RNA cleavage. With this approach, the RNA-cleaving DNAzyme 10-23 was for example utilized to identify $N_m$ modifications in ribosomal RNA.[129]

A more challenging task is the DNAzyme-based readout of methylation on the nucleobases. Recently, RNA-cleaving deoxyribozymes have been identified by *in vitro* selection that can sense the methylation status of a particular nucleobase and cleave the target RNA selectively depending on whether a methyl group is present or not. The first known set of deoxyribozymes with this function can distinguish between A and $m^6A$[130]. Two deoxyribozymes, VMA8 and VMA15, showed accelerated cleavage activity in the presence of $m^6A$ compared to A, but even higher selectivity could be achieved in the opposite direction: The DNAzyme VMC10 was strongly inhibited by the presence of the methyl group, while it efficiently cleaved unmodified RNA (Figure 1.12A). Thus, the activity of deoxyribozymes can strongly be modulated even by a modification as small as a methyl group. VMC10 can even be applied for analysis of the methylation status in more complex substrates like snoRNAs with different sequence contexts and the cleavage activity linearly correlates with the methylation level in the RNA. Furthermore, in combination with reverse transcription and real-time quantitative PCR (RT-qPCR) it was also possible to use VMC10 for the absolute quantification of $m^6A$ at specific sites in total RNA extract.[131]

Even more advanced is the task that can be fulfilled by another set of deoxyribozymes. Several highly orthogonal DNA enzymes were recently reported that can distinguish unmodified C from the three monomethylated cytidine isomers $m^3C$, $m^4C$ and $m^5C$.[132] Some deoxyribozymes like AK104 were severely inhibited by the presence of a methyl group and only cleaved unmodified RNAs (Figure 1.12B). The activity of others was strongly accelerated by the presence of one specific modified nucleobase. For example, AL112 only cleaved $m^3C$ containing RNA, but was strongly inhibited when unmodified, $m^4C$ or $m^5C$ containing substrates were used. However, how those deoxyribozymes exactly detect the methyl groups in RNA is still unknown. Thus, new RNA-cleaving DNAzymes that are sensitive for alternative post-transcriptional modifications cannot easily be engineered but must be selected from scratch.

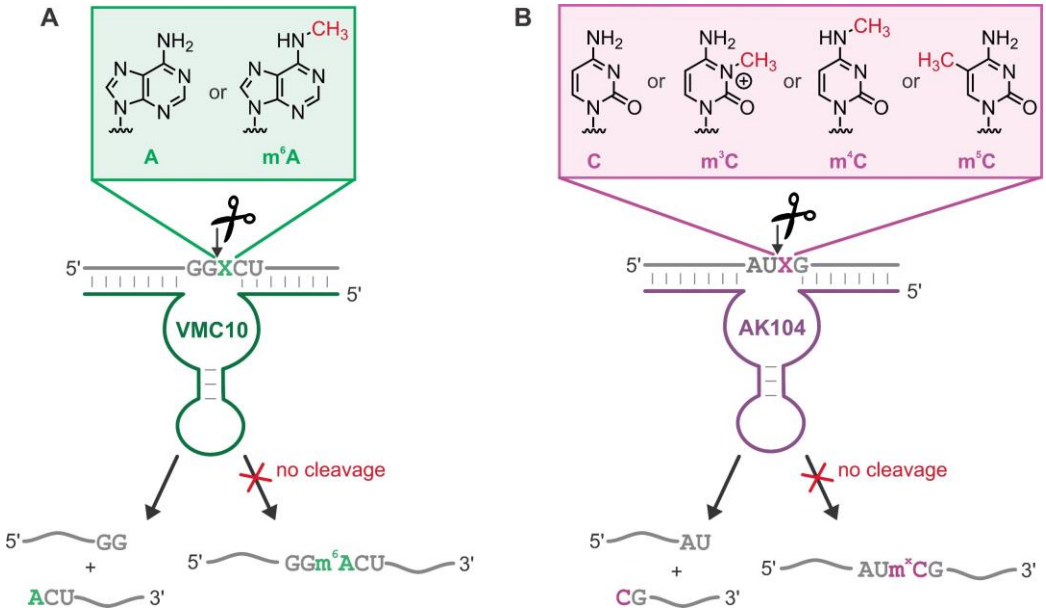



**Figure 1.12:** Examples for DNAzymes sensing RNA methylation. **A** VMC10 selectively cleaves unmethylated RNA and is inhibited by an $m^6A$ modification. **B** AK104 is sensitive against all monomethylated cytidine isomers (x = 3, 4, 5) and only cleaves unmethylated RNA.

Despite the recent increase of knowledge regarding installation, removal and functions of RNA modifications that was especially achieved due to the development of high throughput sequencing methods mentioned above, their biological functions are still not fully understood. Among others, more research is particularly needed to obtain more insights into the correlations and interaction between different modifications. Additionally, epitranscriptomic detection methods that can be applied to single cells (single cell epigenomics) are highly desirable to determine the contribution of a modification within each developmental state.[133-135] Moreover, to ability to site-specifically install and remove post-synthetic RNA methylation can also help to further investigate on the impact of those chemical alterations regarding RNA folding, stability and function. They may also contribute to the development of additional and more accurate detection methods for RNA modifications.

## 1.2 The catalytic potential of nucleic acids

For a long time, RNA was solely viewed as a messenger molecule that transfers the genetic information from DNA to the ribosome in form of mRNA and serves as adaptor molecule in protein synthesis as tRNA. However, he functions of RNA go far beyond. This became particularly clear as another important feature of RNA was discovered around 40 years ago: Similar to protein catalysts, some RNA molecules can act as enzymes, by folding into a defined 3D structure.

### 1.2.1 Natural ribozymes

The catalytic potential of RNA was first discovered in 1982, when Thomas Cech discovered that the class I intron of *Tetrahymena thermophila* is able to catalyze its own splicing reactions.[136] At about the same time, Sidney Altmann studied the ribonuclease P enzyme, which is involved in the maturation of tRNAs and could demonstrate that only the RNA part of the enzyme, but not any of the associated proteins is responsible for the cleavage of the pre-tRNA into the active tRNA.[137] In 1989, Cech and Altmann were even awarded the Nobel Prize in Chemistry 'for their discovery of catalytic properties of RNA'.[138] In 1992, investigation on an *E. coli* ribosome showed that the functional part that is responsible for the formation of the peptide bonds within the peptidyl transferase center is also fundamentally a ribozyme. The associated proteins only play structural roles.[139-140]

Since Cech's and Altmann's discovery, a total of nine different self-cleaving ribozyme motifs have been discovered in nature. Following the hammerhead[141-142] and hepatitis delta virus (HDV) ribozymes[143-145] that were initially found in RNA viruses and viroids, the hairpin[146-147], glucosoamine-6-phosphate synthase (glmS)[148-149] and Varkud satellite (VS) ribozymes[150] have been reported. Recently, with the help of bioinformatic sequence analysis, the twister[151], twister sister, hatchet, and pistol[152] motifs were also added to the list of natural small self-cleaving ribozymes.

### 1.2.2 *In vitro* selected RNA and DNA catalysts

Besides the ribosome, all known natural ribozymes catalyze phosphodiester cleavage or ligation reactions. But the catalytic repertoire of RNA and other nucleic acid polymers has been extended far beyond what can be found in nature using the powerful method of *in vitro* selection. Also often referred to as SELEX (Systematic Evolution of Ligands by Exponential Enrichment), the latter term is especially used for the selection of nucleic acids against ligand binding,

with the resulting molecules called aptamers. This method was first introduced independently by the groups of Jack Szostak[153] and Larry Gold[154] in 1990 and has now become a potent tool to obtain functional nucleic acids with desired features (chapter 1.2.2.3). Starting from a large number of random sequences, functional nucleic acids can be identified in the laboratory through an iterative evolutionary process. By applying an appropriate selection strategy, a wide range of nucleic acids with different functions can be evolved, ranging from non-covalent binding of small molecules[155-156], proteins[157] or even cells[158] by RNA or DNA aptamers to catalysis of various chemical reactions[159].

### 1.2.2.1 Artificial ribozymes

The very first non-natural ribozyme emerging from a test tube was discovered in 1990 by Robertson and Joyce.[160] Starting from several variants of the self-splicing *Tetrahymena* group I intron, they were able to evolve a catalytically active RNA that specifically cleaves a DNA substrate. Since then, many artificial ribozymes were selected that - similar to natural ribozymes - catalyze the cleave or ligation of phosphodiester bonds. However, a growing number of RNAs could be identified to also catalyze more complex reactions: Besides carbon-carbon bond formation through Diels-Alder cycloaddition[161-163] or aldol reactions[164], the scope of RNA-catalyzed reactions also comprises RNA capping[165], amino acid transfer reactions[166-168], peptide bond formation[169] or aminoacylation[170-173]. Some ribozymes can even act as RNA polymerases that can synthesize other complex functional ribozymes using natural NTPS.[174-177] Some other *in vitro* selected ribozymes are able to catalyze the covalent labeling of RNA using reactive epoxides[178-179], chloro- or iodoacetamides[180-182] or ATP derivatives[183-184] (chapter 1.2.3.2). Thus, ribozymes cannot only be used for RNA cleavage and ligation, but also for post-synthetic or post-transcriptional modification of RNA. In this context, another reaction type has just recently been added to the repertoire of ribozyme catalyzed reactions: Three different RNAs have been identified that catalyze the site-specific transfer of a single methyl group onto a target RNA thereby artificially generating post-transcriptional RNA modifications (chapter 1.2.4 and chapter 3).[95, 185-186]

### 1.2.2.2 Deoxyribozymes

Inspired by the discovery of enzymes composed of RNA, investigations into the catalytic potential of DNA were started as well. Even though no such DNA enzyme has been discovered in nature so far[187], Breaker and Joyce reported the first deoxyribozyme in 1994. Identified by the help of *in vitro* selection, the $Pb^{2+}$-dependent DNAzyme is able to specifically cleave an

RNA phosphodiester linkage within a target strand.[188] Following, a broad variety of deoxyribozymes with different catalytic properties have been selected, demonstrating that DNA can be as functional as RNA regarding catalysis, despite lacking the 2'-OH groups. Apart from RNA[188-189] and DNA[190-191] cleavage, the catalytic repertoire of DNAzymes is constantly growing. For instance DNA-catalyzed reactions comprise RNA ligation[192] - which was successfully repurposed for RNA labeling (see chapter 1.2.3.1) - depurination[193], formation of nucleopeptide linkages[194], nucleotide excision[130] and many more[195].

As DNAzymes are cost-effective to produce and the fact that the lack of 2'-OH groups renders them less susceptible for hydrolysis compared to RNA, their potential applications are the focus of intense investigations. In this regard, especially RNA-cleaving deoxyribozymes are of great interest as tools for fundamental biochemical research[196] or used to construct biosensors for various applications[197]. Some RNA-cleaving DNAzymes are even sensitive to certain small post-transcriptional nucleobase modifications and can be used as analytical tools (chapter 1.1.2.4). Additionally, their application as potential gene silencing tools is an emerging research field.[198] In this context, a major advantage of (deoxy-)ribozymes compared to protein enzymes is their programmability towards different targets, as their binding arms can be designed to anneal to nearly any mRNA of interest. This feature also greatly reduces potential off-target effects against non-homologous sequences.[197, 199] However, research on the possible therapeutical use of DNAzymes also faces some challenges, which are primarily attributed to the low catalytic activity under physiological conditions, where the cellular $Mg^{2+}$ concentration is the main limiting factor.[200]

### 1.2.2.3 How to select for catalytically active nucleic acids

#### 1.2.2.3.1 What is *in vitro* selection?

To obtain such catalytically active RNAs or DNAs, *in vitro* selection is the method of choice. The starting point of every selection experiment is a nucleic acid library comprising a constant part as well as a region with random (N) nucleotides. The constant elements either serve as primer binding sites for amplification or as hybridization elements for the substrate oligonucleotide. In addition, the random region is physically connected to the desired reaction position (Figure 1.13). With this design, the information about the catalytic potential of the randomized part is directly linked to the desired chemical reaction at the target site. Nowadays, the length of the random region is usually between 20 - 40 nt, directly determining the sequence space of the selection. In case of an $N_{40}$-pool, $4^{40} \approx 10^{24}$ different sequences are statistically possible. However, with a starting pool amount of 1 nmol, only ~$10^{14}$ oligonucleotides and therefore just a fraction of the actual sequence space can be covered. Despite this undersampling, many *in*

*vitro* selections result in active ribozymes or DNAzymes. This indicates that more than just one defined sequence has the desired catalytic activity and not all nucleotides within the catalytic core are essential. Rather only some key elements are important, whereas other nucleotides are non-critical and can be mutated while retaining the active conformation and therefore the full catalytic potential.[201-204]

In general, the appropriately designed starting pool is then sent through an iterative *in vitro* selection process in which the members of the library with desired activity are enriched until they outnumber the inactive fraction and dominate the pool (Figure 1.13). This is especially possible due to the cyclic nature of *in vitro* selection, which causes an exponential enrichment of the active fraction. The specific experimental strategy used for the enrichment step varies depending on the intended selection outcome and has to provide survival advantages for the active fraction, while rapidly eliminating the inactive proportion. Various strategies exist for the physical separation of the reacted from the many other unreacted species and will be discussed in more detail in chapter 1.2.2.3.2.

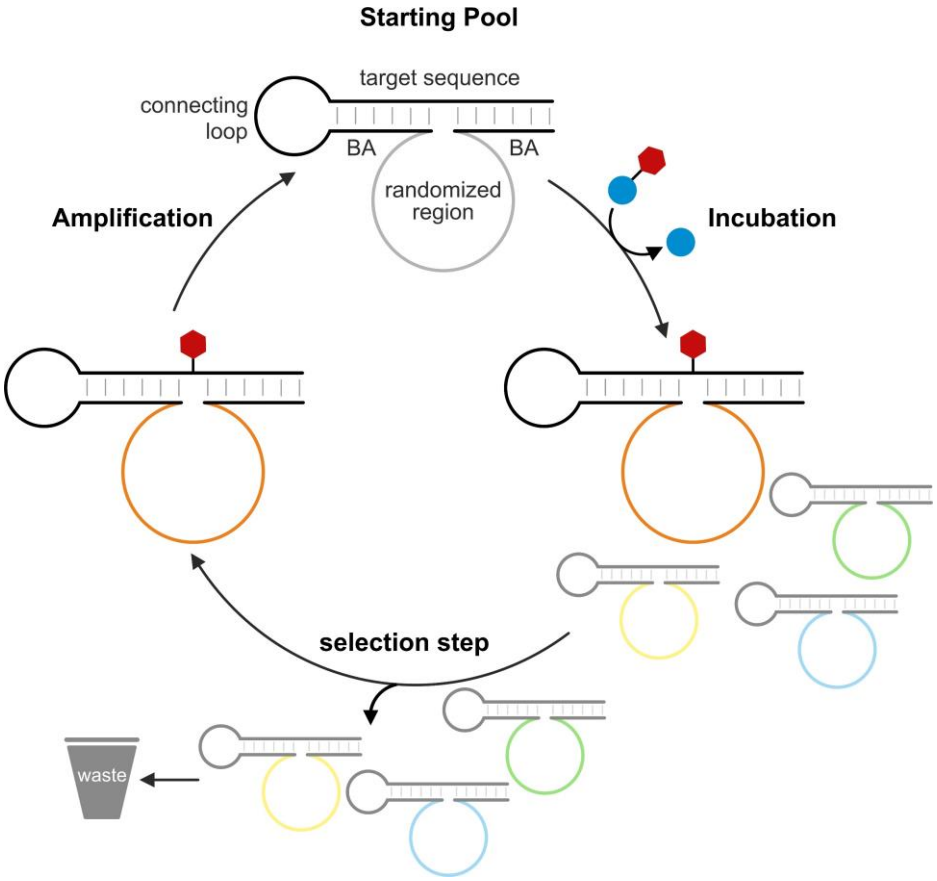



**Figure 1.13:** Schematic depiction of an exemplary *in vitro* selection cycle selecting for a self-modifying nucleic acid enzyme. The starting pool comprises a randomized region (gray) flanked by two constant binding arms (BA) hybridizing to the target sequence, which is connected to the catalytically active part via a connecting loop. During the incubation step, only a small fraction will modify itself (orange catalytic core) and will be amplified to enter the next selection round. The unreacted and therefore unmodified sequences are removed and discarded during the selection step.

After separation functional from inactive sequences, the selection cycle is completed by amplifying the active fraction and re-synthesizing the now enriched pool, which is then sent through the next round. For DNA, this is commonly achieved through PCR amplification, while in case of RNA, additional reverse transcription and *in vitro* transcription steps are needed. In theory, it seems possible that a selection may be complete after just a single round. However, some inactive sequences always survive the separation step even though they do not show the desired active. Thus, the selection process has to be repeated for multiple rounds to remove those false-positive sequences until only a few variants with high catalytic activity are left from the initial pool. The number of selection rounds required to obtain a sufficiently high enrichment level is often between 10-20.[205] However, it can be highly variable between different selection experiments as it also strongly depends on the specific selection design. To further eliminate sequences that do not catalyze the desired reaction but are carried over the next rounds, occasional 'negative selections' can be performed. In those additional rounds, an alternative selection pressure is applied to actively eliminate species that are inactive or catalyze an undesired site reaction. To obtain variants with the most ideal catalytic characteristics, the stringency on individual selection parameters, such as incubation time, cofactor concentration, buffer composition or temperature is gradually increased until no further enrichment is detectable. The pool is then sequenced, and the activity of individual sequences is tested and further characterized.[201-204]

The ribozymes or DNAzymes obtained by *in vitro* selection may not be the most optimal catalysts. To further improve their activity, a reselection can be performed, in which a starting pool is used that is based on an already active RNA or DNA sequence. Such library can either be prepared with mutagenic PCR, in which random mutations are introduced in the initial sequence using nonstandard deoxynucleotide triphosphates[206] or altering buffer conditions and concentrations, which increases the polymerase misincorporation rate.[207] Alternatively, a partially randomized pool can be used that is obtained by rational design, which is for example based on secondary structure information of active sequences.

#### 1.2.2.3.2 How to separate active from inactive sequences

A key point of every *in vitro* selection experiment is the separation of active from the large number of inactive sequences. While various different methods exist, a general distinction can be made between 'gel-based' and 'bead-based' methods (Figure 1.14). The first approach relies on the altered electrophoretic mobility of active and inactive species, when separated using PAGE purification. This method is particularly common for the selection of DNAzymes with ligase activity. Catalytically active ligases will attach an oligonucleotide to themselves,

which results in slower electrophoretic mobility and enables the separation from the inactive library members.[208-210] The same principle can be applied to obtain RNA- or DNA-cleaving (deoxy-)ribozymes, but in the opposite direction (Figure 1.14, top). Examples of DNAzymes found by this methods include the famous versatile DNA catalysts 8-17 and 10-23 that both possess ribonuclease activity[189] as well as deoxyribozymes that are sensitive to RNA modifications[130, 132, 211]. Alternatively, oligonucleotides can also by circularized to obtain an electrophoretic mobility alteration as it was for example used to search for ribozymes with aminoacyl-AMP synthetase activity.[212]

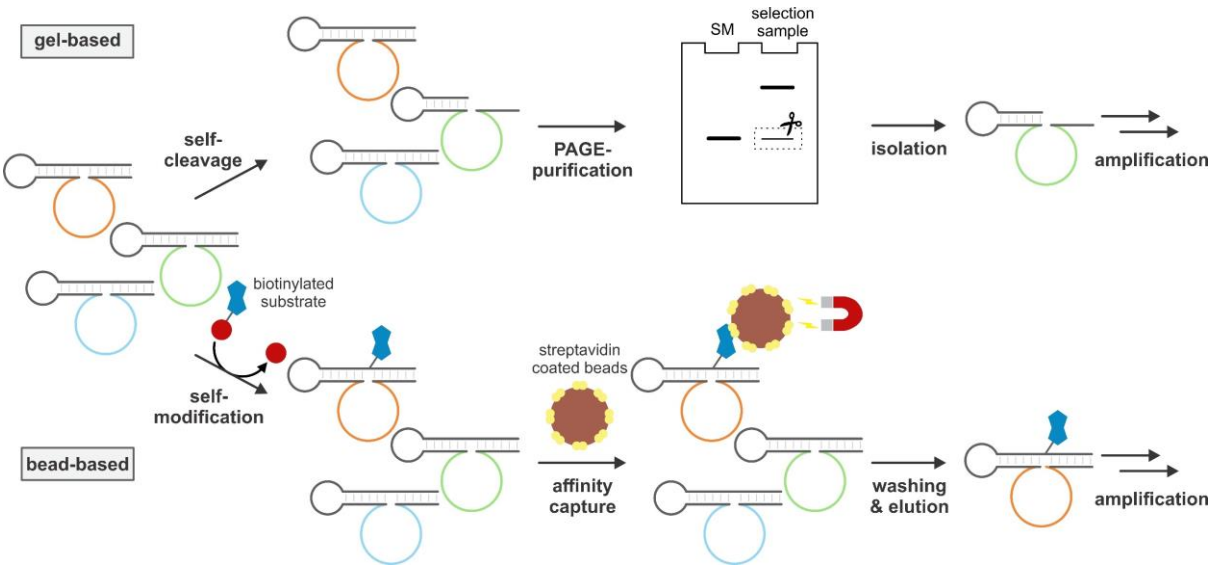



**Figure 1.14:** Schematic depiction of a gel-based (top) and bead-based (bottom) selection steps. In the gel-based strategy, self-cleaving ribozymes (green) are separated from the inactive species by PAGE according to a size marker (SM). In the bead-based selection, active sequences (orange) biotinylate themselves followed by capture with streptavidin coated magnetic beads.

In contrast, 'bead-based' selection strategies rely on the installation of a specific tag to capture reactive pool members. This method is commonly used to select for self-modifying ribozymes that possess the activity to covalently attach a modification, including an affinity tag, to themselves. Subsequently, tagged sequences will be captured on a solid support based on strong non-covalent interactions, while inactive pool members are washed away. Examples include the first alkyltransferase ribozyme that employed a biotinylated iodoacetamide as small molecule substrate, which enables capture of the active library members with agarose immobilized streptavidin.[181] This interaction is especially suited for efficient enrichment of the low abundant active sequences due to the extremely high affinity of biotin for streptavidin ($K_d$ = $10^{-15}$ M).[213] Similar biotin/streptavidin-based enrichment strategies were used to identify adenylyltransferase ribozymes[184] (Figure 1.14, bottom) or RNAs that used 2,3-disubstituted epoxides for covalent self-labeling.[178] Another popular affinity tag is the use of thiol groups, that can react with pyridyl disulfide functionalized beads to form disulfide bonds. This method was for example

applied to select for kinase ribozymes that could thiophosphorylate themselves using ATPγS as substrate.[214-215]

Alternatively, RNA or DNA itself can be used as a tag that facilitates hybridization of another short oligonucleotide. Application of this strategy for instance enabled the selection of an RNA ligase ribozyme, which was the first ribozyme that originated from a random sequence pool: Only active pool members catalyzed the reaction between their 5' triphosphate with a free 3' OH of another RNA sequence. The thus self-tagged active library members were then purified using beads coupled to complementary oligonucleotides. Additionally, the ligated sequence also acted as a primer binding site for further amplification to ensure enrichment of only active RNA sequences, which results in an extremely powerful dual-selection procedure (Figure 1.15).[216]

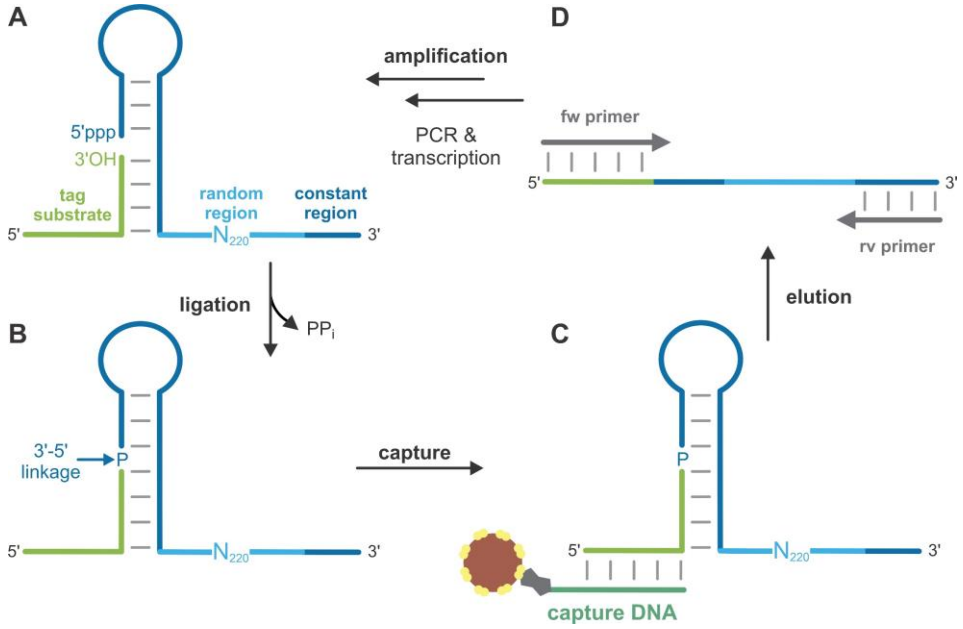



**Figure 1.15:** Schematic overview of the *in vitro* selection of an RNA ligase ribozyme using a tag substrate that enables isolation of active sequences. **A** Active library members catalyze the ligation of a substrate RNA (light green) to the RNA pool (blue) generating a 3'-5' phosphodiester bond (**B**). **C** A biotinylated capture oligo (dark green), immobilized on streptavidin coated agarose beads (brown) is used to separate active from inactive species. **D** The tag substrate is also used as primer binding site in the subsequent PCR amplification additionally ensuring only active sequences are enriched. Adapted from ref.[216].

Even though countless new RNA and DNA catalysts were selected using the traditional methods described so far, they sometimes only act in *cis* (that is they do only perform the catalyzed reaction on themselves) or lack multiple turnover activity. To improve on these characteristics, the method of *in vitro* compartmentalization (IVC) can be used. This emulsion-based technology, which was initially applied for protein enzyme evolution, is based on trapping single ribozyme variants, the corresponding substrate strand, and the reaction product within the same small compartment. Those are usually generated in the form of droplets in a water-in-oil emulsion. In this way, the genotype (the sequence information) and phenotype (the catalytic activity) of each RNA molecule are directly linked, which allows to select for true catalysts that can act

in *trans* with multiple turnover reactivity.[217] Such a selection starts by encapsulating single copies of a ribozyme-encoding DNA template. These are then amplified and transcribed *in situ*, generating multiple ribozyme copies within each droplet. During the incubation step, active library members will catalyze the desired reaction that introduces a tag to the DNA template. In this way, after breaking the emulsion, the ssDNA encoding for active ribozymes can be captured and sent through the next selection round. An additional advantage of this method is the redundancy of the reverse transcription step, which reduces amplification bias and can help to enrich for ribozymes with more complex secondary structures. This is because reverse transcriptases show preferences for less structured RNAs, therefore eliminating more stable and highly structured species.[218-219]

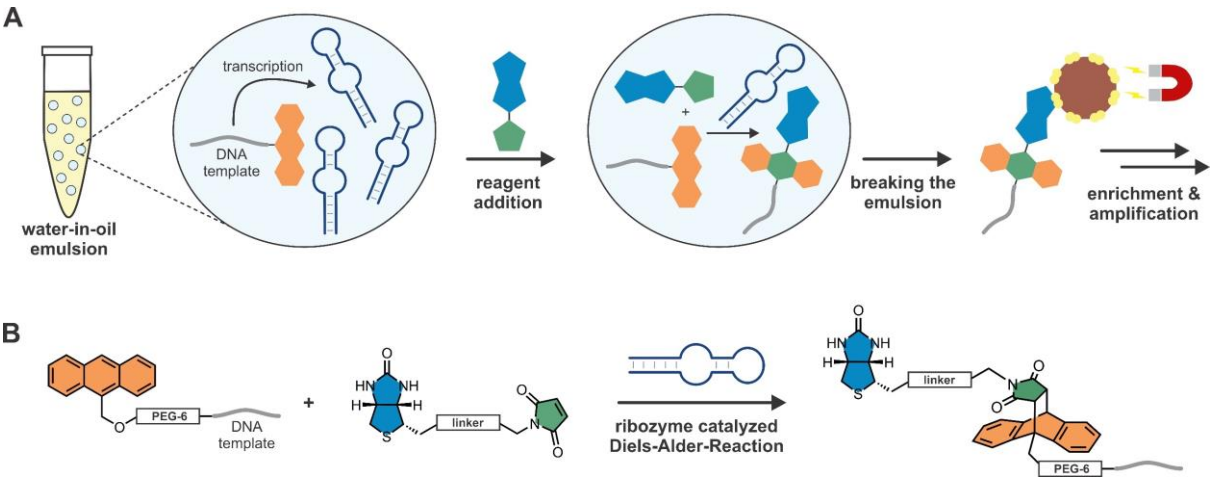



**Figure 1.16:** Selection of Diels-Alderase ribozymes as an example for IVC-based selection strategies. **A** Schematic depiction of the selection procedure. A single DNA template (grey), coupled to anthracene (orange) is compartmentalized in an aqueous drop of a water-in-oil emulsion. The DNA is transcribed within the drop, followed by addition of the reagent biotin-maleimide (blue and green). Active ribozymes (dark blue) catalyze the cycloaddition, thereby biotinylating the DNA template. The emulsion is broken, active sequences are enriched via affinity purification and then amplified. **B** Cycloaddition reaction between anthracen and maleimide catalyzed by active Diels-Alderase ribozymes. Adapted from ref.[218].

Among others, an IVC-based *in vitro* selection that was coupled to traditional SELEX has, for instance, helped to improve the turnover number of a Diels-Alderase ribozyme that works in a true bimolecular fashion. Here, active library members catalyzed the cycloaddition reaction of the small molecule substrate biotin-maleimide with anthracene, which was coupled to the DNA template. Therefore, templates encoding for active ribozymes were biotinylated and could then be isolated using streptavidin-coated magnetic beads (Figure 1.16).[218] Furthermore, IVC also helped to drastically improve the activity of an RNA polymerase ribozyme: The new catalyst was able to extend a primer approximately three-fold faster than the wild type ribozyme. In the selection design, active sequences elongated a DNA template that was then hybridized to a biotinylated probe for subsequent capturing.[220]

Alternatively, instead of affinity purification, fluorescence-activated cell sorting (FACS), can be used to discriminate active from inactive species. In this method, individual fluorescently

labeled cells or particles are sorted one by one starting from a heterogenous suspension by the help of a flow cytometer. Using a laser beam, the fluorescence signal of each species is determined as they pass through the device. They are then encapsulated in single droplets, which are charged by an electrode according to the individual detected signal. By the help of an electrostatic deflection system, the droplets are then collected in appropriate tubes (Figure 1.17).[221] This advanced method was for example employed to select for RNA ligase ribozymes: The DNA template for the RNA pool as well as one half of the substrate RNA were immobilized on beads. Active sequences catalyzed the ligation of fluorescently labeled RNA to the other substrate strand. In this way, the reaction product was immobilized on the same bead as the gene encoding the active ribozyme. After breaking the emulsion, beads carrying a fluorescent label were isolated using FACS and the respective RNA was amplified (Figure 1.17).[222]

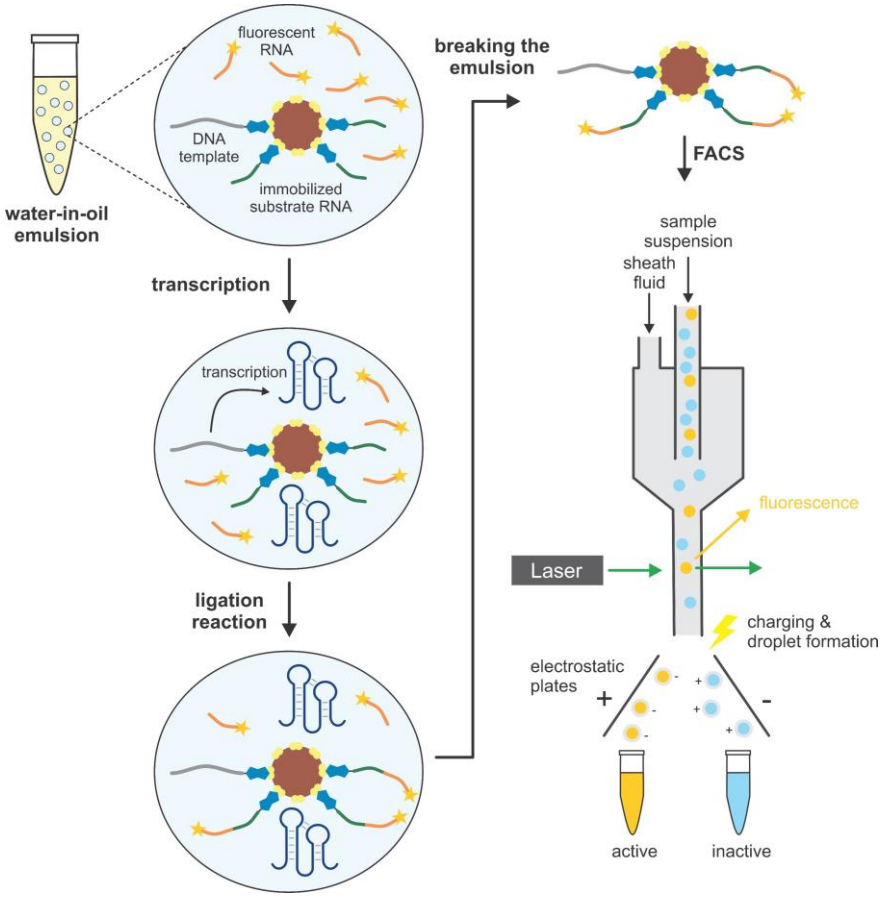



**Figure 1.17:** Schematic depiction of a FACS-based selection to obtain *trans*-active RNA ligase ribozymes. A single DNA template (gray) is immobilized on a magnetic bead and compartmentalized in an aqueous drop of a water-in-oil emulsion. Following the transcription of the DNA within the droplet, active ribozymes (dark blue) catalyze the ligation of a fluorescently labeled RNA (orange) to another RNA substrate (green), which is also attached to the magnetic bead. After the emulsion is broken, beads are sorted using FACS: With a steady flow of sheath fluid, the sample is flushed into a thin tube and thereby singularized. Using a laser, fluorescently tagged beads can be identified, which are then encapsulated in single droplets and charged accordingly to enable sorting using an electrostatic deflection system. Adapted from ref. [222].

#### 1.2.2.4 Implications for prebiotic life on earth

The selection methods presented above led to the discovery of countless new artificial ribozymes with multiple functions. These findings do not only contribute to a better fundamental mechanistic understanding of nucleic acid enzymes, but also open up new perspectives for prebiotic chemistry and the origin-of-life research. In this context, they especially provide support for the so called 'RNA-world-hypothesis', which states that RNA was the key macromolecule in prebiotic self-replicating systems by acting as both, carrier of the genetic information and catalyst.[223-224] Thus, early life on earth would have solely been based on RNA without the need for DNA or proteins. Only in the course of evolution RNA was gradually replaced by DNA as a more robust storage form for genetic information and by proteins as the more flexible and versatile catalysts. Although this hypothesis may never be fully proven, evidence can be collected through the discovery of ribozymes catalyzing reactions like RNA polymerization or C-N and C-C bond formations that would have been essential for early metabolic reactions and RNA-based early forms of life. In this context, also the exploration of secondary structures as well as reaction mechanisms of *in vitro* selected ribozymes is of great interest. This may help to develop a better understanding of how ancient ribozymes could have been able to catalyze complex chemical reactions for which only protein enzymes are known in modern day biology.

#### 1.2.2.5 Expanding beyond RNA and DNA as catalysts: XNAzymes

Recently, evolution and rational design strategies have expanded beyond (deoxy-)ribozymes consisting of just the canonical nucleosides and led to the development of catalytically active polymers with novel backbone architectures. Such molecules, called xeno nucleic acids (XNAs), differ from their natural counterparts by the identity of the backbone or sugar moiety (Figure 1.18). Albeit to varying degrees, they have the ability to store genetic information, form duplex structures and can also be amplified enzymatically, just like natural nucleic acids.[225-228] Especially in the fields of oligonucleotide therapeutics, synthetic biology and nanotechnology, XNA polymers are of particular interest, as they exhibit special characteristics including improved stability *in vivo*[229-230], altered helical structures[231], and increased thermal resistance[232-233]. For example, the ribose in locked nucleic acids (LNA) contains an additional methylene bridge connecting the 2' oxygen with the 4' carbon. This modification 'locks' the modified ribose in a constrained 3'-endo sugar pucker conformation, causing the oligonucleotide to adapt a stable A-form helix. This LNA structure leads to enhanced nuclease resistance compared to the non-modified natural counterparts and offers the advantage of improved base pairing specificity as well as stability.[234] Similar structural and biophysical features are provided by 2'-methoxyethyl (2'MOE) RNA, which has an additional methoxy-ethyl residue attached to the 2'-

oxygen[235], 2'-fluoro arabino nucleic acid (FANA)[236] or 2'-methoxy (2'OMe) RNA[237], a modification also found in nature.

Other XNAs possess an acyclic backbone architecture, which vastly differs from the natural nucleoside structure. One of the most famous examples in this regard are probably peptide nucleic acids (PNA). Here, an amino acid residue replaces the ribose sugar and the phosphate group, thus they can easily be synthesized using standard peptide chemistry.[238] Like many other XNAs, PNA shows high stability against enzymatic degradation, improved selectivity and increased hybridization energy in binding RNA and ssDNA. However, due to its uncharged backbone, PNA is much more hydrophobic compared to standard DNA or RNA, which has to be considered when it comes to its cellular uptake.[239]

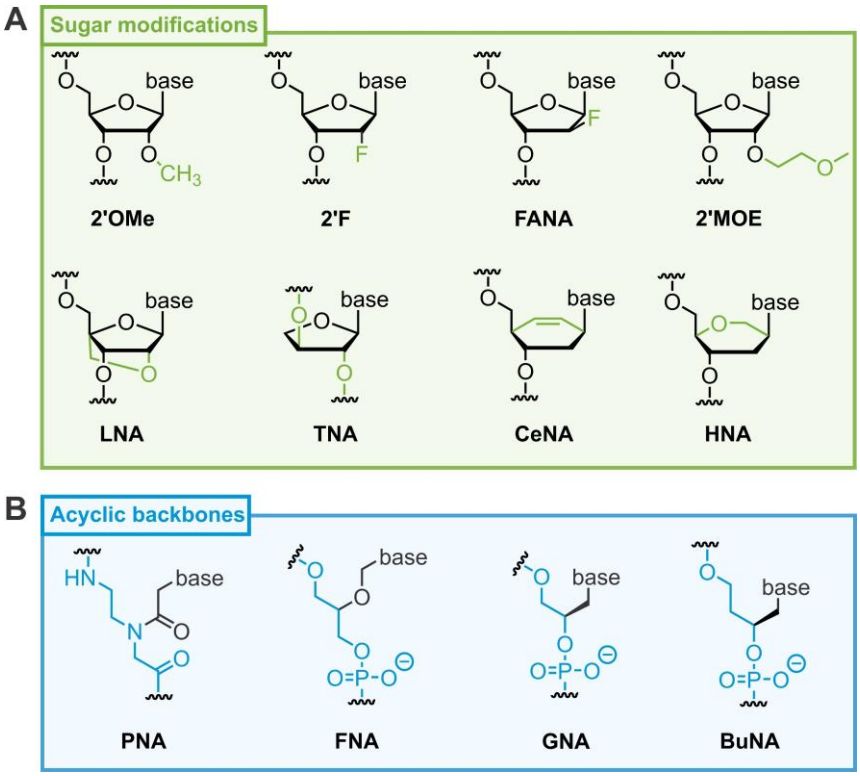



**Figure 1.18:** Overview of common XNA chemical structures**. A** Xeno nucleic acids with modified sugar moieties. Shown are the structures of 2'-methoxy RNA (2'OMe), 2'-fluoro RNA (2'F), 2'-fluoro arabino nucleic acid (FANA), 2'-methoxyethyl RNA (2'MOE), locked nucleic acid (LNA), threose nucleic acid (TNA), cyclohexene nucleic acid (CeNA) and hexitol nucleic acid (HNA). **B** Nucleic acids with acyclic backbone architectures. Shown are the structures of peptide nucleic acid (PNA), flexible nucleic acid (FNA), (*R*)-glycol nucleic acid (GNA) and (*S*)-butyl nucleic acid (BuNA).

The improved binding properties, enhanced stability and altered helical conformations make XNAs attractive building blocks not only for their therapeutical application, but also for generating a novel type of catalytic nucleotides: XNAzymes. In 2015, Taylor *et al.* first demonstrated that synthetic genetic polymers can fold into defined 3D structures that renders them catalytically active.[240] Through *in vitro* selection, they identified XNAzymes with four different backbone chemistries, namely arabino nucleic acids (ANA), 2′-fluoroarabino nucleic acids (FANA), hexitol nucleic acids (HNA) and cyclohexene nucleic acids (CeNA) that possessed ligase and

endonuclease activity. Further examples of *in vitro* selected XNA catalysts that are entirely composed of non-natural building blocks include a general RNA-cleaving FANAzyme[226] that outcompetes the well-known DNAzyme 10-23 when targeting a chimeric DNA substrate[241] and TNA enzymes (threozymes) with RNA-ligase[242] or endoribonuclease activity.[227] Very recently, an endonuclease FANAzyme that was further engineered to specifically cleave highly structured mRNAs *in vivo*[228] as well as an RNA-cleaving 2'OMe RNA enzyme (2'OMezyme)[243] were added to the list of XNA catalysts. However, carrying out *in vitro* selections with XNAs is not straightforward as often special engineered polymerases and reverse transcriptases are required to copy genetic information back and forth between DNA and XNA.[225-228] Furthermore, many of the respective triphosphates are not readily commercially available, but need to be synthesized in the lab.

Alternatively, the activity of known nucleic acid enzymes can also be improved by precise incorporation of distinct XNA building blocks. Several groups used this approach for optimization of the classic RNA-cleaving 10-23 DNAzyme due to its potential therapeutic application as gene silencing tool. For example, in early attempts deoxynucleosides in the binding arms were replaced by LNA to improve hybridization with the target strand resulting in an increased RNA cleavage rate.[244] However, the types of XNA and their incorporation sites have to be selected carefully, as increased binding affinity can also result in product inhibition. Just recently, two more heavily modified 10-23 variants with enhanced catalytic activity and biostability have been reported: In 10-23X the binding arms and two positions in the catalytic core were replaced with FANA residues, while TNA was used at the 3' and 5' termini for protection against nucleases.[245] In case of Dz46, iterative cycles of chemical evolution guided by structural information were employed to yield a 10-23 variant modified with 2'OMe, MOE, LNA and phosphorothioate (PS) that exhibits a high turnover number and target specificity.[246]

Research on XNAzymes is an emerging field that has just started to develop. More detailed structural and functional investigations to comprehend how certain XNA modifications exactly influence the activity of an XNA catalyst are still in high demand. Also, the lack of efficient polymerases for the incorporation of XNA building blocks to perform more *in vitro* selections is currently still an issue. However, considering the high *in vivo* stability and performance of XNA catalysts, they hold high promise for future biochemical and biomedical applications.

### 1.2.3 Covalent RNA labeling catalyzed by nucleic acids

With the help of *in vitro* selection also more complex and advanced chemical reactions such as RNA labeling based on binding followed by transfer of a specific cofactor have been discovered for both RNA and DNA. This is an especially useful catalytic activity, as the ability to directly site-specifically label or tag an RNA of interest can provide invaluable information regarding its localization, interactions, stability, and conformational dynamics. Especially being able to track a particular RNA may ultimately help to develop a more profound understanding of the functions of both, coding and non-coding RNAs.

#### 1.2.3.1 Labeling using DNAzymes

When selecting for RNA-ligating DNAzymes some were found to possess 2'-5' ligase activity[247] at an internal nucleoside instead of creating a natural 3'-5' phosphodiester linkage. Thus, a branched RNA is formed rather than a linear oligonucleotide possessing a native connection. One RNA-ligating deoxyribozymes generating a 2'-5' linkage is the well-studied 10DM24.[247] It was first employed as a useful tool for RNA labeling by ligating a short, labeled RNA to a specific 2'OH at an internal adenosine thereby generating a labeled oligonucleotide product.[248]

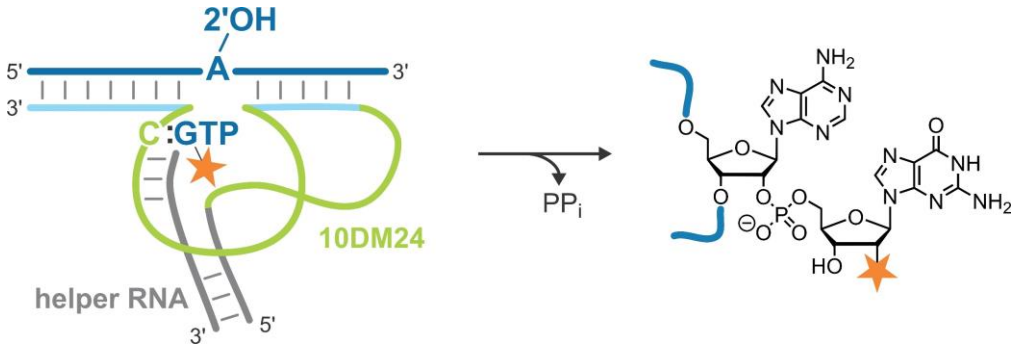



**Figure 1.19:** 10DM24 catalyzed labeling of an RNA of interest at the 2'OH position of an internal adenosine upon release of $PP_i$. The DNAzyme (green) recognizes its target RNA (dark blue) through hybridization with the binding arms (light blue). The substrate GTP carrying a label (fluorophore, affinity tag, functional group) at the 2'OH hybridizes with an internal cytosine. A helper RNA that pairs to a third binding arm (gray) is required for labeling, which results in the formation of a three-helix-junction. Adapted from ref.[249].

Further engineering resulted in its ability to ligate a simple mononuclotide triphosphate (e.g., GTP) instead of an entire oligonucleotide to the target site, resulting in a 2'-5' mononucleotidyl branch at the labeling position (Figure 1.19).[250] This reaction is mediated by a helper oligonucleotide that is required for formation of the active secondary structure and is strongly accelerated in the presence of $Tb^{3+}$. Recognition of the GTP substrate is achieved through formation of a Watson-Crick base pair with a cytidine in the catalytic core. Mutation of this C even enabled the ligation of other NTPs, although labeling reactions using purine triphosphates were more efficient than with pyrimidines. This labeling approach was further expanded in 2014, when it

was demonstrated, that a wide range of GTP analogs with various functional groups at the 2'-position such as azides, biotin or fluorescent dyes, are accepted by 10DM24.[249] However, due to the requirement of toxic $Tb^{3+}$ and the need of an additional helper oligo, this nucleic acid catalyst can only be employed for *in vitro* labeling. As another option, ribozymes with a comparable RNA labeling strategy are promising alternatives to DNAzymes when it comes to the installation of covalent RNA modifications.

#### 1.2.3.2 Labeling using ribozymes

The work on RNA-labeling ribozymes was started in 1995, when the Szostak lab reported the first self-alkylating ribozyme named BL8-6, which used a biotinylated α-iodoacetamide substrate as small molecule cofactor (Figure 1.20A).[181] Following this work, α-chloroacetamide[180] or disubstituted epoxides[178] were also successfully employed as agents for self-modifying ribozymes. Further investigations of the modification products revealed that all these ribozymes specifically biotinylate the *N*7 position of an internal guanine.[178, 180-181]

The RNA-catalyzed RNA modification approach was then further expanded from biotinylation to fluorescent labeling in 2014 by Sharma *et al*.[182] Using fluorescein-iodoacetamide, two ribozymes were identified that are able to alkylate themselves at an undetermined internal position. However, the practical application of these catalytic RNAs as labeling tools is limited owing to their strict dependence on fluorescein-iodoacetamide as cofactor and the fact that they are restricted to self-modification (e.g., only work in *cis*).

To overcome these constraints, a targeted *in vitro* selection strategy was employed to obtain more versatile RNA-labeling ribozymes.[184] Using $N^6$-biotin-ATP as substrate in combination with a structured RNA pool to guide the catalytic activity to a desired bulged modification site adenosine, several adenylyltransferase ribozymes could be identified. All ribozymes catalyzed the formation of a 2'-5'-phosphate bond at the intended labeling site (Figure 1.20B), showed a broad RNA substrate scope and readily accepted a variety of different $N^6$-modified ATP analogs. The most active ribozyme FH14 was even successfully employed for fluorescent labeling of long, complex, and heavily structured RNAs such as 5S, 16S and 23S ribosomal RNA in total cellular RNA extract. The generated natural 2'-5' phosphate linkage, however, is strongly susceptible to cleavage by debranching enzymes, which limits their application in cellular contexts.

To enhance biorthogonality of this labeling approach und also generate a more stable linkage, a biotinylated antiviral nucleoside was used as small molecule substrate in a follow-up *in vitro* selection.[183] Using the same selection setup but replacing $N^6$-biotin-ATP with an analog of the

antiviral drug tenofovir, the next generation of RNA-labeling ribozymes was obtained: The most active ribozyme, named FJ1, also targets the 2'OH of the desired modification site adenosine but generates a more stable phosphonate ester bond (Figure 1.20C). Just like FH14, the catalytically active RNA can be designed to target a broad RNA substrate scope and can utilize a wide variety of fluorescent tenofovir analogs. Additionally, FJ1 and other ribozymes from this family are completely orthogonal to FH14 enabling dual-color RNA-labeling. *In vivo* application of the FH and FJ ribozymes is however limited due to the high $Mg^{2+}$ concentrations of around 10 mM required for efficient labeling (compared to cellular concentration of 0.25 - 1 mM[251]). Additionally, the delivery and availability of the tenofovir substrate in cells is another limiting factor. To obtain RNA-labeling ribozymes with improved performance that can potentially be used efficiently in a cellular context, targeted reselection strategies or *de novo in vitro* evolution with high stringency towards *in vivo* conditions are potential options. This is of particular interest as ribozymes, in contrast to DNAzymes, are vector encodable and therefore especially desirable. By cloning them into an appropriate expression vector, they can directly be transcribed within the cell circumventing the need for transfection of the oligonucleotide as it is the case for deoxyribozymes.[252-253]

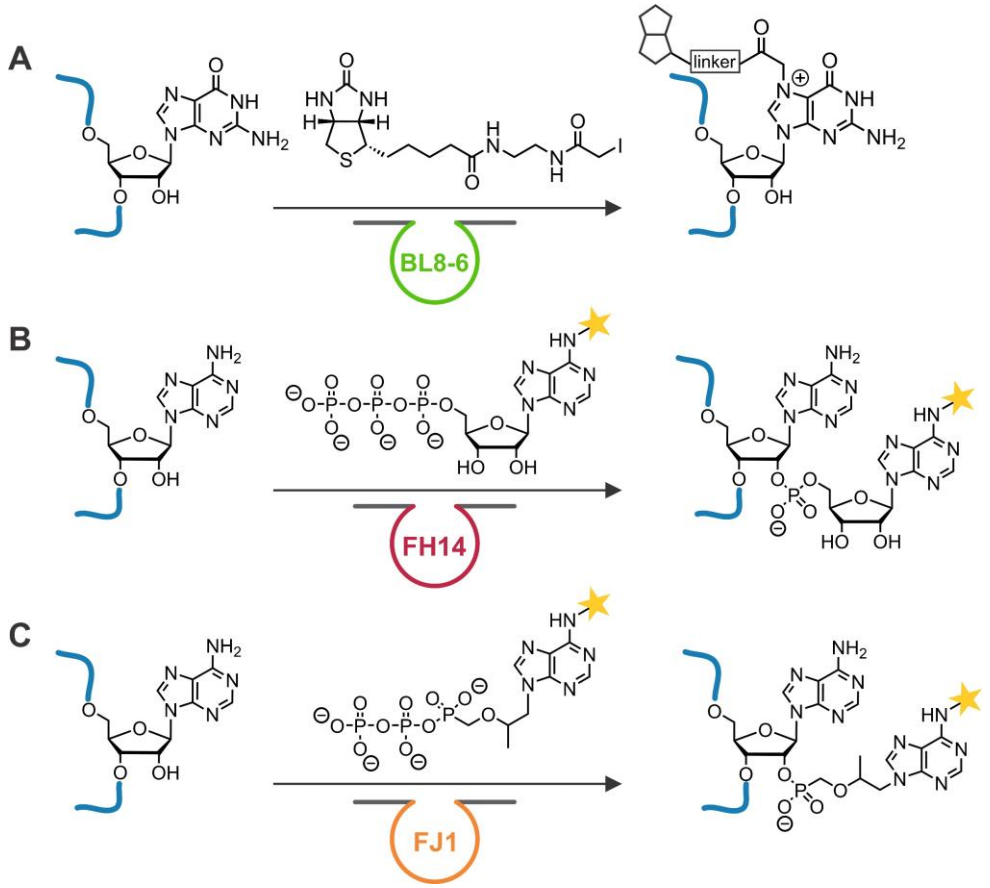



**Figure 1.20:** Overview of RNA-modifying ribozymes. **A** BL8-6 uses biotin-iodoacetamide to biotinylate *N*7 of G. **B** FH14 employs a fluorescent $N^6$-modified ATP-substrate to selectively label the 2'OH of an internal adenosine generating a phosphodiester linkage. **C** FJ1 utilizes a fluorescently labeled tenofovir analog to generate a phosphonate ester linkage in the RNA product.

### 1.2.4 Methyltransferase ribozymes: MTR1 and beyond

After demonstrating that ribozymes can catalyze RNA labeling upon binding of a small molecule cofactor, it was debated whether ribozymes can also conduct more advanced chemical reactions such as RNA methylation. Considering the variety and importance of methylations in all kinds of coding and non-coding RNAs, the discovery of a ribozyme with methyltransferase activity would be of vast significance, especially in the context of the RNA-world-hypothesis and research concerning the origin of life.[223-224] The question regarding the existence of RNA with methyltransferase activity, however, remained unanswered for a long time. It was not until 2020, when the first ribozyme catalyzing the transfer of a single methyl group, named MTR1, was finally described (chapter 3).[95] Since then, two more RNA methyltransferases were discovered that use different cofactors as methyl group donors.

#### 1.2.4.1 $preQ_1$ riboswitch

One year after the discovery of MTR1, Flemmich *et al.* reported the second catalytically active RNA with methyltransferase activity. In contrast to MTR1, however, it was not identified through *in vitro* selection.[185] Rather, the authors adapted a natural riboswitch - a non-coding regulatory RNA motif in bacteria that is used to sense different metabolites: Upon ligand binding, conformational changes occur in the RNA secondary structure, resulting in an altered expression of the associated gene.[254] The authors found that the non-natural ligand $O^6$-methyl pre-queuosine 1 ($m^6preQ_1$) can be bound by a native $preQ_1$ class I riboswitch. Upon binding, the methyl group of the small molecule cofactor is transferred to the *N*3 position of C15 to generate the natural RNA modification $m^3C$ (Figure 1.21A). Optimizing the reaction conditions resulted in a maximum self-methylation yield of 50 % after 24 h at pH 6.0. Since the leaving group is also tightly bound by the riboswitch, this is the maximum yield that can possibly be achieved with this setup. Moreover, the highest reactivity was observed at slightly acidic pH. This is consistent with protonation of $m^6preQ_1$ at the 7-aminomethyl group, which improves cofactor binding within the active ribozyme center (Figure 1.21B). The *N*1 position of $m^6preQ_1$ is also protonated under these conditions, further stabilizing the $preQ_1$ leaving group.

When other near cognate cofactors were tested for activity, only minor or no alkylation could be observed and representatives of other well studied $preQ_1$ riboswitch classes lacked methyltransferase activity as well. However, engineering the riboswitch for *trans*-activity (e.g., methylating an external RNA substrate) was successful. By dividing the scaffold into two separate oligonucleotides, 38 % methylated RNA could be obtained after 48 h, which is significantly slower compared to the MTR1 ribozyme. However, this is not entirely unexpected considering that MTR1 was evolved *de novo* in the laboratory as a methyltransferase ribozyme. In contrast,

the $preQ_1$ riboswitch evolved in nature for strong cofactor binding rather than for the transfer of a methyl group.

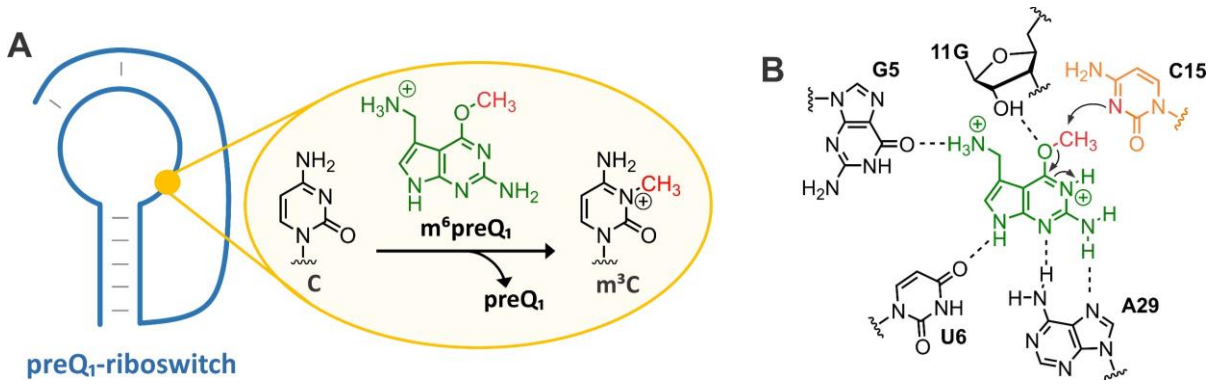



**Figure 1.21:** Methyltransferase reaction catalyzed by the $preQ_1$ riboswitch. **A** The methyl group of $m^6preQ_1$ is transferred to the *N*3 of an internal cytidine generating $m^3C$. **B** Proposed reaction mechanism. The *N*1-protonated $m^6preQ_1$ is held in place by several hydrogen bonds with neighboring nucleotides. The mechanism involves a nucleophilic attack of *N*3 of C15 (orange) on the methyl carbon (red), generating $preQ_1$ as the leaving group. Adapted from ref.[185].

### 1.2.4.2 SMRZ-1

So far, both methyltransferase ribozymes described herein (MTR1 and $preQ_1$) employed guanine-based small molecule substrates for their reaction. However, in present day biology protein enzymes use a different ribonucleotide cofactor as the universal methyl group donor: SAM.[255] While multiple SAM-binding aptamers and natural riboswitches have been discovered[256-257], none of these RNAs were found to exhibit self-methylation activity. However, very recently, Murchie *et al.* identified the first SAM-dependent catalytic RNA through *in vitro* selection.[186] The ribozyme named SAM-ribozyme-1 (SMRZ-1), was selected from a randomized RNA library using $m^7G$ antibody pull-down. As expected, analysis of the ribozyme revealed that it site-specifically methylates *N*7 of a guanine within a target RNA and is furthermore dependent on $Mg^{2+}$ and $Cu^{2+}$. Overall, it folds into a hairpin structure, but can be divided into two separate RNAs, allowing it to work in *trans* (Figure 1.22A). SMRZ-1 shows good reactivity over a wide temperature and pH range with an optimum at 60 °C and pH 6.0.

An X-ray crystal structure of the ribozyme in complex with SAM captures the pre-catalytic state and shows that the $Cu^{2+}$ ion is located in the active site of SMRZ-1. It coordinates to SAM and a guanosine close to the modification site but is not positioned near the target G or the activated methyl group of SAM (Figure 1.22B). Bioinformatic searches helped to identify numerous examples of the SMRZ-1 motif across a range of organisms. Those natural ribozyme candidate sequences showed methylation activity as high as the original ribozyme when tested *in vitro*. This indicates that other natural ribozymes with methyltransferase activity may still exist in nature or might have lost their catalytic activity over the course of evolution.

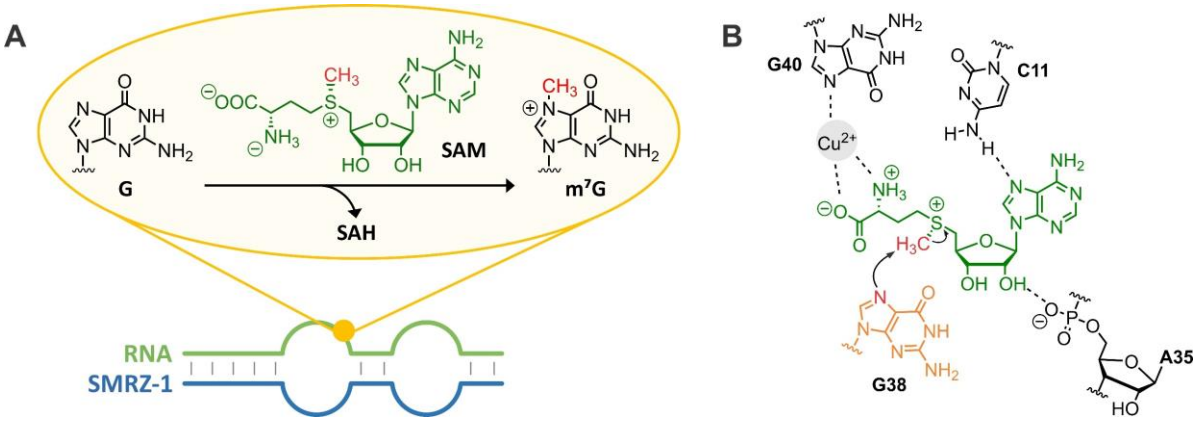



**Figure 1.22:** Methyltransferase reaction catalyzed by the SMRZ-1 ribozyme. **A** The methyl group of SAM is transferred onto *N*7 of an internal G generating $m^7G$ (blue: SMRZ-1 ribozyme, green: target RNA). **B** Proposed reaction mechanism. SAM is held in place through coordination with $Cu^{2+}$ and hydrogen bonds to *N*4 of C11 and the phosphate of A35. The mechanism involves a nucleophilic attack of *N*7 of G38 (orange) on the methyl carbon (red) generating SAH as the leaving group. Adapted from ref.[186].

Overall, the existence of ribozymes with methyltransferase activity does not only have a profound impact concerning the origin of life research, but such ribozymes are also important tools for fundamental research on RNA methylation. The ability to site-specifically generate a methylated nucleotide within a defined sequence can help to elucidate its impact on RNA structure, function, and interaction of the RNA with other cellular components. Considering that research on RNA-methylating ribozymes has only started to develop (the studies on $preQ_1$ and SMRZ-1 mediated methylation were just recently published after launching this project), this field of research holds high promise to develop a deeper understanding of RNA methylations.

# 2 Research Objectives and Outline

Post-transcriptional RNA modifications, in particular methylations, are important regulators that fine tune the structure, function, and stability of all types of RNA. In present day biology, methyl groups are transferred onto the RNA by protein methyltransferases that utilize the universal methyl group donor SAM as small molecule cofactor.[54] However, methylated nucleotides are believed to be ancient relicts that were already present in a hypothetically prebiotic RNA-world. Among others, this idea is for instance supported by the fact that SAM itself is a nucleotide-derived cofactor. Additionally, six different classes of SAM-binding riboswitches can still be found in nature today.[258-260]

In an ancient RNA world, the methyl transfer reaction would have been catalyzed by RNA itself. Such RNA enzymes (ribozymes) can still be found in nature today, where they mainly catalyze the cleavage or formation of phosphodiester bonds.[261-262] However, *in vitro* selection led to the discovery of artificial ribozymes evolved from random nucleic acid libraries that catalyze a wide range of reactions including covalent RNA-labeling utilizing a specific small molecule cofactor.[178, 180-184] In nature, only one cofactor utilizing ribozyme has been discovered so far: The glmS riboswitch-ribozyme requires binding of glucosamine-6-phosphate for cleavage of RNA phosphodiester backbones.[149] Moving one step further in the context of RNA labeling and evolving a ribozyme that mimics natural methyltransferase protein enzymes would have a profound impact on the RNA-world-hypothesis and origin of life research. Moreover, such a ribozyme would also have important prospective applications for fundamental biochemical research: As a novel tool for targeted manipulation of RNA methylation, it can be used to gain deeper insights into the dynamics and function of the RNA methylome. At the launch of this project no such ribozyme capable of single carbon metabolism has been reported.

This thesis therefore focuses on the discovery, characterization, structural investigation, and utilization of a ribozyme for targeted RNA methylating. Chapter 3 describes the development of the first methyltransferase ribozyme (MTR1) followed by the determination of its exact modification site. It also discusses its application for methylation of long, complex RNAs such as tRNAs and its prospective uses for *in vivo* RNA methylation. To elucidate the structure and mechanism of MTR1, X-ray crystallography was performed (chapter 4). The obtained structure further guided the development of an MTR1 variant with severely accelerated reaction rate. Subsequent utilization of MTR1 as an alkyltransferase ribozyme in combination with nebularine as alternative target nucleoside, it was also successfully repurposed as a tool for site-specific aldehyde generation in RNA (chapter 5). The obtained formylimidazole nucleotide analogue was used for bioconjugation to various nucleophiles and also yielded a novel

hemicyanine fluorophore when conjugated to an indole moiety. Finally, chapter 6.1 focuses on reselection of MTR1 following a DNAzyme based strategy that aims to improve its methyltransferase activity. Additionally, two more *in vitro* selections were conducted to evolve alternative RNA-modifying ribozymes targeting nucleosides different from the MTR1 modification site or to obtain an RNA-labeling ribozyme, which works under cellular conditions (chapter 6.2).

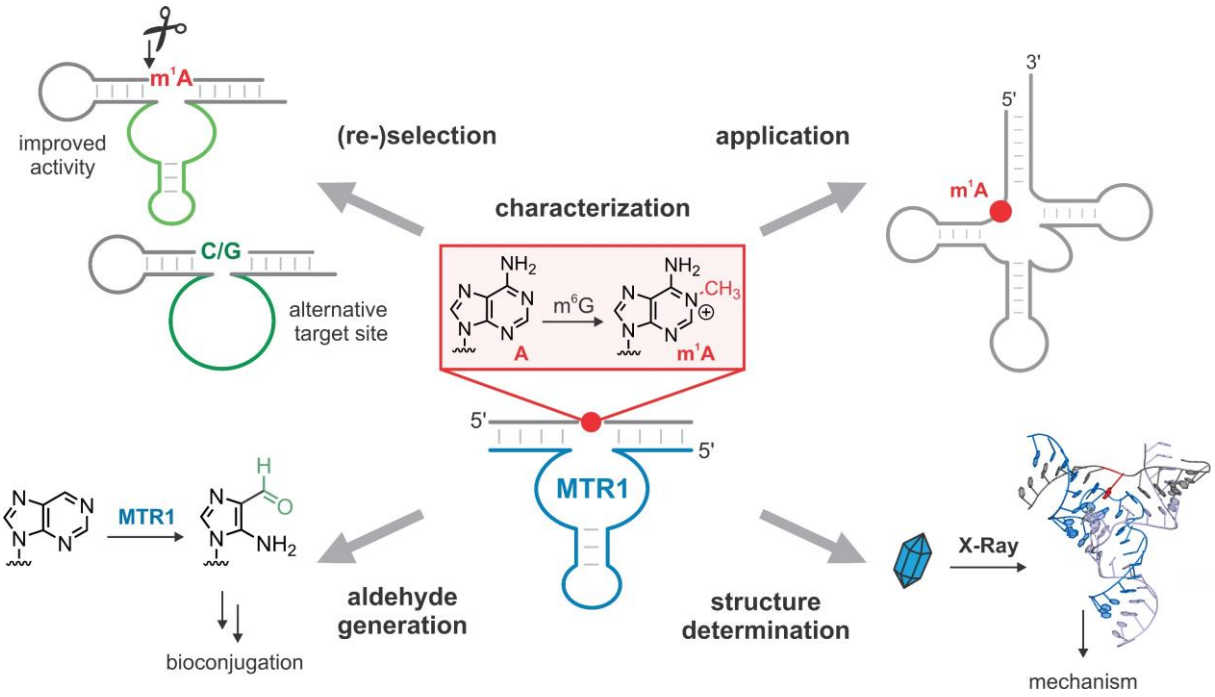
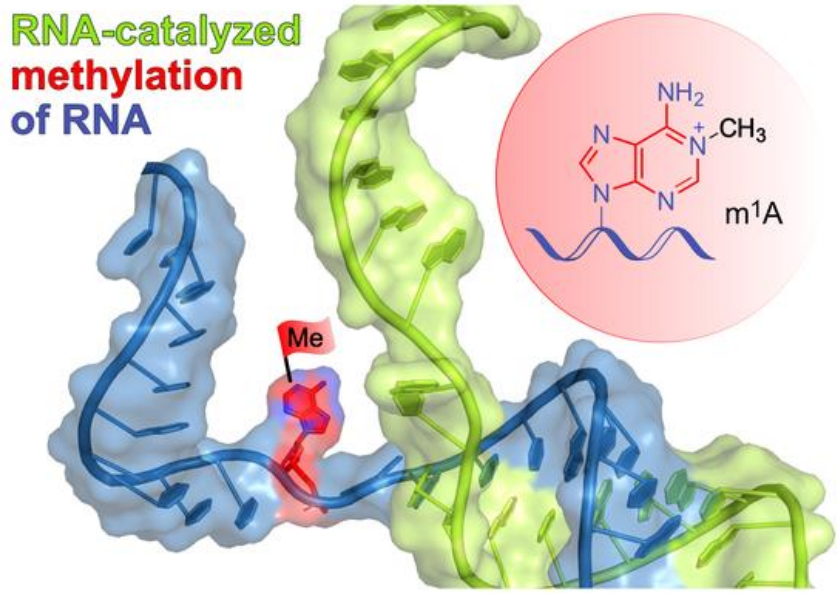



**Figure 2.1:** Overview of the different projects covered in this thesis, which are all centered around the methyltransferase ribozyme MTR1. Following its characterization and application (chapter 3), its 3D crystal structure was determined (chapter 4). MTR1 was also repurposed for site-specific aldehyde generation in RNA (chapter 5). Finally, selections to obtain methyltransferase ribozyme variants with improved activity or alternative target sites were performed (chapter 6).

# 3 Site-specific RNA methylation by a methyltransferase ribozyme




This chapter and the corresponding supporting information (chapter 9.1) were published in:

C.P.M Scheitl, M. Ghaem Maghami, A.-K. Lenz, C. Höbartner, *Nature* **2020,** *587* (7835), 663-667.

(https://www.nature.com/articles/s41586-020-2854-z)



For the sake of unity of this thesis, several editorial changes have been made, which, however, do not affect the contents of the thesis.

**Abstract**

Nearly all classes of coding and non-coding RNA undergo post-transcriptional modification, including RNA methylation. Methylated nucleotides are among the evolutionarily most-conserved features of tRNA and rRNA.[263-264] Many contemporary methyltransferases use the universal cofactor *S*-adenosylmethionine (SAM) as a methyl group donor. SAM and other nucleotide-derived cofactors are considered to be evolutionary leftovers from an RNA world, in which ribozymes may have catalyzed essential metabolic reactions beyond self-replication.[265] Chemically diverse ribozymes seem to have been lost in nature, but may be reconstructed in the laboratory by *in vitro* selection. Here we report a methyltransferase ribozyme that catalyzes the site-specific installation of 1-methyladenosine in a substrate RNA, using $O^6$-methylguanine as a small-molecule cofactor. The ribozyme shows a broad RNA-sequence scope, as exemplified by site-specific adenosine methylation in various RNAs. This finding provides fundamental insights into the catalytic abilities of RNA, serves a synthetic tool to install 1-methyladenosine in RNA and may pave the way to *in vitro* evolution of other methyltransferase and demethylase ribozymes.

## 3.1 Introduction

More than 70 different methylated nucleotides play important functional roles in present-day RNA.[266-267] Mostly known for shaping the structures and tuning the functions of non-coding rRNA, tRNA, and snRNA, some modifications also influence gene expression programmes by regulating the fate and function of mRNA.[268-270] The majority of methylated nucleotides currently known in RNA are installed by post-synthetic (i.e., post- or co-transcriptional) methylation by protein enzymes that use *S*-adenosylmethionine (SAM) as the universal methyl group donor. Methyl transferases are considered ancient enzymes, and methylated nucleotides are also discussed as molecular fossils of the early Earth produced by prebiotic methylating agents.[271-272] In an era preceding modern life based on DNA and proteins, RNA was thought to function both as primary genetic material and as catalyst.[273] Ribozymes have been discovered in nature, where they catalyze RNA cleavage and ligation reactions, mostly in the context of RNA splicing and retrotransposition.[274-276] *In vitro* selected ribozymes have been evolved as RNA ligases and replicases that are able to reproduce themselves or their ancestors, and are able to produce functional RNAs, including ribozymes and aptamers.[277-279] Self-alkylating ribozymes have been described using reactive iodo- or chloroacetyl derivatives,[181-182, 280] or electrophilic epoxides,[178] but the design of earlier *in vitro* selection strategies prevented the emergence of catalysts capable of transferring a one-carbon unit. Thus, ribozymes that catalyze RNA methylation have so far remained elusive. This lack of methyltransferase ribozymes

seems surprising, since numerous natural aptamers are known to specifically bind nucleotide-derived metabolites associated with methyl group transfer or one-carbon metabolism, including SAM, methylene tetrahydrofolate (THF), and adenosylcobalamin (vitamin B12)).[281-282] These aptamers are found as components of riboswitches that regulate the expression of associated genes, often involved in the biosynthesis of the respective metabolite or its transport across membranes.[283] Interestingly, six different classes of SAM-binding riboswitches accommodate the ligand with its reactive methyl group in various different conformations.[258-259] However, these RNAs apparently avoid self-methylation.

Therefore, it remained an open question if RNA can catalyze site-specific methylation reactions to produce defined methylated RNA products. Previously, *in vitro* selection efforts have identified SAM-binding aptamers, but methyl transfer reactions were not observed, likely because the aptamer established a binding site for the adenine moiety of the cofactor but did not specifically interact with its 5' substituent.[256]

We speculated that alternative methyl group donors other than SAM or methylene-THF could be substrates for RNA-catalyzed RNA methylation, and took inspiration from an enzyme class that is responsible for repair of alkylated DNA (that is, catalyzes the demethylation of DNA). The $O^6$-methylguanine ($m^6G$) DNA methyltransferase releases unmodified DNA, accompanied with irreversible methylation of the protein.[284] By analogy, we hypothesized that RNA-catalyzed methyl transfer would result in methylated RNA upon release of guanine. Using *in vitro* selection, we identified a ribozyme that utilizes $m^6G$ as a small-molecule methyl group donor and catalyzes site-specific methylation of adenosine at position *N*1, resulting in position-specific installation of 1-methyladenosine ($m^1A$) in the target RNA (Figure 3.1A).

## 3.2 Results and Discussion

### 3.2.1 Search for methyltransferase ribozymes

*In vitro* selection is a powerful method to enrich functional RNAs by repeated cycles of selection and amplification from a random RNA library. We used a structured RNA pool containing 40 random nucleotides that was designed according to our previously used strategy to direct RNA-catalyzed labeling of a specific adenosine in a target RNA.[184, 285] RNA methylation would most likely occur at an O or N nucleobase heteroatom, on the 2'-OH group or on the phosphate backbone. In either case, attachment of a single methyl group would hardly enable physical separation of the active sequences based on size or charge.

We therefore searched for alkylating ribozymes that catalyzed the transfer of a biotin group attached via a benzyl linker to the target RNA, and speculated that resulting ribozymes could later be engineered to enable RNA methylation. After incubation with biotinylated $O^6$-benzylguanine (BG-biotin), the biotinylated products were separated via streptavidin (or neutravidin) affinity purification on magnetic beads, and amplified the enriched candidates by reverse transcription and PCR. Then, we used *in vitro* transcription with T7 RNA polymerase to generate the enriched library that was used in the next round of selection; the *in vitro* selection scheme is shown in Figure 9.1. After 11 rounds of *in vitro* selection, we identified two candidate alkyltransferase ribozyme (which we named CA13 and CA21) that contained a predicted internal hairpin structure with a partially complementary stem, and showed high sequence similarity in the flanking regions (Figure 3.1B). Both ribozymes were able to catalyze alkylation of the target RNA in a bimolecular setup (referred to as *trans* activity), in which the ribozyme and the target RNA interacted via Watson-Crick base pairing. Moreover, the biotin moiety was not essential: fast and efficient alkylation of the target RNA was achieved with $O^6$-(4-aminomethylbenzyl)guanine (BG-$NH_2$) as well as with $O^6$-benzylguanine (BG) (Figure 3.1B). Inspired by the natural or engineered promiscuity of protein methyltransferases that tolerate SAM cofactors with extended alkyl groups,[286-287] we examined the opposite direction for the *in vitro* selected ribozyme and asked if the transferred alkyl group could be a simple methyl group: that is, whether $m^6G$ could serve as cofactor for the new ribozymes. The target RNA (and the Watson-Crick binding arms of the ribozyme) were shortened to simplify the analysis of the reaction product. We stabilized the predicted stem in the ribozyme core and introduced an extra-stable UUCG tetraloop. The engineered ribozyme (which we named MTR1) (Figure 3.1C) used a 13-nt or a 17-nt RNA as target and $m^6G$ as cofactor to generate methylated RNA products in 80-90 % yield after 23 h incubation at 37 °C, pH 7.5. The reaction rate was dependent on $m^6G$ concentration with an apparent $K_m$ of ca 100 µM. The presence of the stem-loop in the core of MTR1 was confirmed by compensatory mutations of individual base pairs, which retained catalytic activity. The stem was shortened with only slightly reduced activity while deletion of the stem resulted in inactive ribozymes (Figure 9.2). RNA structure probing experiments (using dimethylsulfate (DMS) and selective 2'-hydroxyl acylation analysed by primer extension (SHAPE)) also confirmed the overall architecture of the ribozyme (Figure 9.3).

The RNA-catalyzed reaction was strictly dependent on $m^6G$ as demonstrated by control experiments in which $m^6G$ was replaced by dimethyl sulfoxide (DMSO) or guanine (Figure 3.1D). We observed residual activity with $O^6$-methyl-2'-deoxyguanosine ($m^6dG$), while $S^6$-methylthioguanine ($ms^6G$), $O^6$-methylhypoxanthin ($m^6H$), $N^6$-methyladenine ($m^6A$), and SAM could not serve as methyl group donors under the conditions tested. The methylated product (P) was easily separable from the unmodified RNA substrate (S) by denaturing polyacrylamide gel electrophoresis (PAGE) (Figure 3.1A) and anion exchange high-performance liquid

chromatography (HPLC) (Figure 9.2A). Addition of a single methyl group to the target RNA was confirmed by high resolution electrospray ionisation mass spectrometry (HR-ESI-MS) (Figure 3.2B).

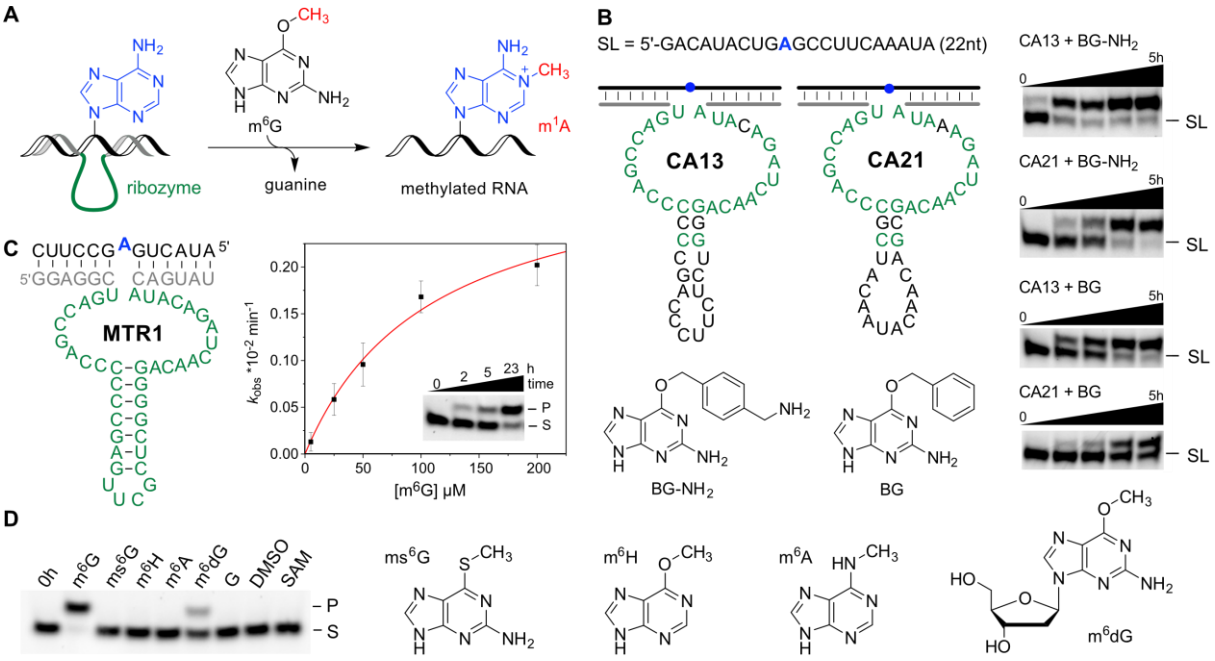



**Figure 3.1:** Methyltransferase ribozyme-catalyzed synthesis of $m^1A$ in RNA using $m^6G$ as methyl group donor. **A** Reaction scheme with intermolecular hybridization of ribozyme to target RNA. **B** Sequences and predicted secondary structure of CA13 and CA21 ribozymes identified by *in vitro* selection, and their *trans*-activity for modification of a 22-nt RNA (designated SL) with BG-$NH_2$ or BG, analyzed by 20 % denaturing PAGE (100 µM guanine derivative, 40 mM $MgCl_2$, pH 7.5, 37 °C, and timepoints 0, 0.5, 1, 2 and 5 h). Representative images of three independent experiments with similar results. **C** Methyltransferase ribozyme MTR1 with stabilized stem-loop shows efficient methyl group transfer. The insert shows a gel image of a 3′-fluorescein-labeled 13-mer RNA substrate reacted with MTR1 and $m^6G$ (10 µM). The observed rate constant $k_{obs}$ was determined with a 3′-fluorescein-labeled 17-mer RNA at five concentrations of $m^6G$, ranging from 5 to 200 µM. The red line represents a curve fit to $k_{obs} = k_{max}[m^6G]/(K_{m,app} + [m^6G])$, mean ± s.e.m. (black). P, product; S, substrate. **D** Structures of $m^6G$ analogues tested. Gel image shows that product formation occurs with $m^6G$, and to a minor extent with $m^6dG$ (24 h reaction time, 25 °C, with 100 µM $m^6G$ or analogue). Representative image from two independent experiments. G, guanine.

## 3.2.2 Identification of the methylation product

Our next goal was to identify the chemical constitution of the methylated RNA product. We obtained the first indication that the reaction happened at the bulged adenosine using mutated target RNAs, as RNA substrates with adenosine changed to guanosine, inosine or cytidine were not modified (Figure 9.4). Adenosine has several possible nucleophilic positions, and several isomeric methylated adenosines are known to be native RNA modifications - including $m^6A$, $m^1A$ and 2′-*O*-methyladenosine. Other possible methylation sites are *N*7, *N*3 and the non-bridging oxygen atoms of the phosphodiester backbone. We performed atomic mutagenesis with various modified adenosines in the target RNA, which revealed the substrate requirements (Figure 3.2D). We used BG-$NH_2$ in these reactions, because the larger electrophoretic shift upon transfer of a 4-aminomethylbenzyl group simplified the analysis. RNA

oligonucleotides with 2′-deoxyadenosine or 2′-*O*-methyladenosine - as well as 3′-methylphosphate and 3′-methylphosphonate linkages - were tolerated, and showed that the reaction occurred on the nucleobase. This conclusion was further corroborated by alkaline hydrolysis and RNase T1 digestion of the isolated product, which revealed the presence of the cleavage product at the bulged adenosine and up-shifted digestion products beyond this position (Figure 3.2C); by contrast, alkylation at the 2′-OH would have caused a missing band in the hydrolysis pattern. Instead, we observed an extra hydrolysis band close to the adenosine position, which could not be explained by counting the number of nucleotides. To solve this puzzle, we collected additional hints from analysis of the ribozyme-catalyzed alkylation of RNAs that contained different nucleobase analogues (Figure 3.2D, Figure 9.4).

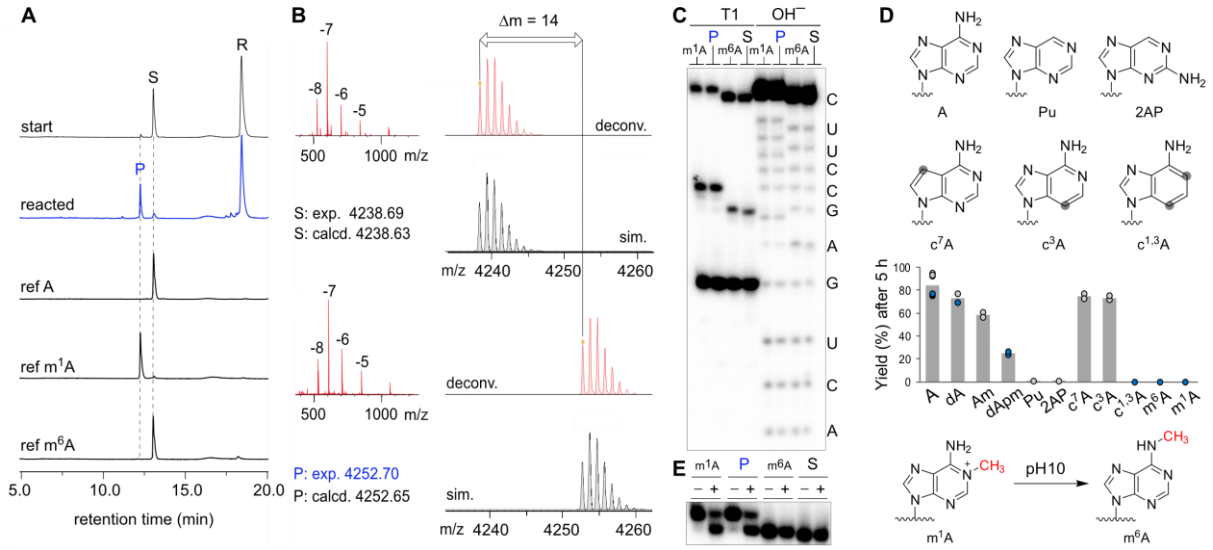



**Figure 3.2:** Characterization of reaction products. **A** Anion exchange HPLC analysis of MTR1-catalyzed reaction of 13-mer RNA substrate (5′-AUACUGAGCCUUC-3′) with m$^6$G at 25 °C for 23 h; 10 µM RNA substrate, 12 µM MTR1, 100 µM m$^6$G and 40 mM $MgCl_2$, pH 7.5. HPLC traces of unmodified reference oligonucleotide substrate (ref. A), and the corresponding m$^6$A- and m$^1$A-modified synthetic RNAs (ref. m$^6$A and ref. m$^1$A, respectively) are shown for comparison. m$^1$A-modified RNA elutes earlier, whereas unmodified and m$^6$A-modified RNAs cannot be distinguished. P, reaction product; Rz, ribozyme; S, substrate RNA. **B** HR-ESI-MS of substrate (S)* (top) and product (P)* (bottom) (* denotes 3′-aminohexyl RNA). The measured *m/z* spectra, the deconvoluted mass spectrum (red) and the simulated isotope pattern (grey) are shown. Exp., experimentally determined. **C** RNase T1 digestion and alkaline hydrolysis of reaction product (P) in comparison to RNA substrate (S), and m$^6$A and m$^1$A references, demonstrate that P contains m$^1$A. **D** Atomic mutagenesis of RNA substrate. Individual data points ($n$ = 6 for ref. A, $n$ = 2 for all others) and average (grey bar) shown. White, 17-nt RNAs; blue, 13-nt RNAs. Gel images and a detailed description are provided in Figure 9.4. **E** Incubation under Dimroth rearrangement conditions (pH 10, 65 °C and 1 h) produced m$^6$A from m$^1$A, as seen in the '+' lanes of m$^1$A reference and MTR1 reaction product (P).

Our observation that 2-aminopurine and purine could not be efficiently alkylated suggested that the $N^6$-amino group is essential - in contrast to $N$7 and $N$3, both of which could be removed individually without compromising the alkylation efficiency. By contrast, the RNA containing $N$1,$N$3-dideaza-adenosine was not alkylated. These results narrowed down the possible reaction sites to the $N^6$ or $N$1 of adenosine. This conclusion was supported by the observation that synthetic RNAs that contained m$^6$A or m$^1$A could not be further alkylated by the ribozyme (Figure 3.2D). We attributed the retarded electrophoretic mobility of m$^1$A-modified RNA compared to m$^6$A-modified RNA to the positive charge on the m$^1$A nucleobase. Indeed, m$^6$A-modified

and unmodified 13-nt RNA could not be separated by PAGE or anion exchange HPLC, in contrast to the reaction product (which was observed as a separated band or peak in both assays), which suggested that the MTR1 reaction product contained $m^1A$. The presence of $m^1A$ also explained the extra band in the alkaline hydrolysis lane: $m^1A$ is susceptible to Dimroth rearrangement under alkaline conditions, resulting in partial formation of $m^6A$. Under alkaline hydrolysis conditions, RNAs containing $m^6A$ or $m^1A$ produced distinctly different bands. This was confirmed by comparison of the alkaline hydrolysis patterns of authentic reference RNAs containing $m^6A$ or $m^1A$ (Figure 3.2C). Furthermore, Dimroth rearrangement was induced by incubation of the MTR1 reaction product at pH 10 and 65 °C for 1 h, resulting in greater than 60 % conversion to $m^6A$ without concomitant hydrolysis of the RNA backbone (Figure 3.2E). In combination, these results firmly establish $m^1A$ as the sole product of MTR1-catalyzed RNA methylation using $m^6G$ as methyl group donor.

### 3.2.3 RNA-catalyzed tRNA methylation

The methylated nucleoside $m^1A$ is a native tRNA modification that is found in all domains of life at positions 9, 14, 22 and 57 or 58, and is installed by 2 distinct families of methyltransferases that use SAM as a cofactor (the SPOUT family and Rossman-fold MTases).[288] We asked whether MTR1 could install $m^1A$ on *in vitro* transcribed tRNA. The prerequisite for such an application is a general RNA sequence scope of the ribozyme. Therefore, we first examined the ability of MTR1 to catalyze alkylation of transition and transversion mutants of the parent target RNA, and checked whether the flanking guanosines could also be mutated (Figure 9.4). All of the RNA substrates we tested were alkylated - although with varying efficiency (between 10 and 90 %). These results suggested G<u>A</u>G and A<u>A</u>G (underlining denotes the reactive nucleotide) as the preferred methylation sites. Next, we chose three natural tRNA sequences that contain $m^1A$ flanked by purines and synthesized 13-nt tRNA fragments that enclose the $m^1A$ site. Using MTR1 derivatives with binding arms complementary to these tRNA fragments, we obtained the corresponding methylated RNAs upon incubation with $m^6G$ (Figure 9.5). These results encouraged us to test MTR1 on full-length tRNAs that were prepared by *in vitro* transcription. We chose to install $m^1A$ at position 9 of *Rattus norvegicus* tRNA-Lys, $m^1A$ at position 22 of *Bacillus subtilis* tRNA-Ser and $m^1A$ at position 58 of *Thermus thermophilus* tRNA-Asp (Figure 9.5). The synthetic tRNAs were annealed with the corresponding ribozymes, and the incubation with $m^6G$ was carried out for 22 h. All three synthetic tRNAs were successfully methylated by the corresponding MTR1 ribozymes (Figure 3.3B), as revealed by the strong abort bands in primer extension experiments (in which $m^1A$ blocks the reverse transcriptase).

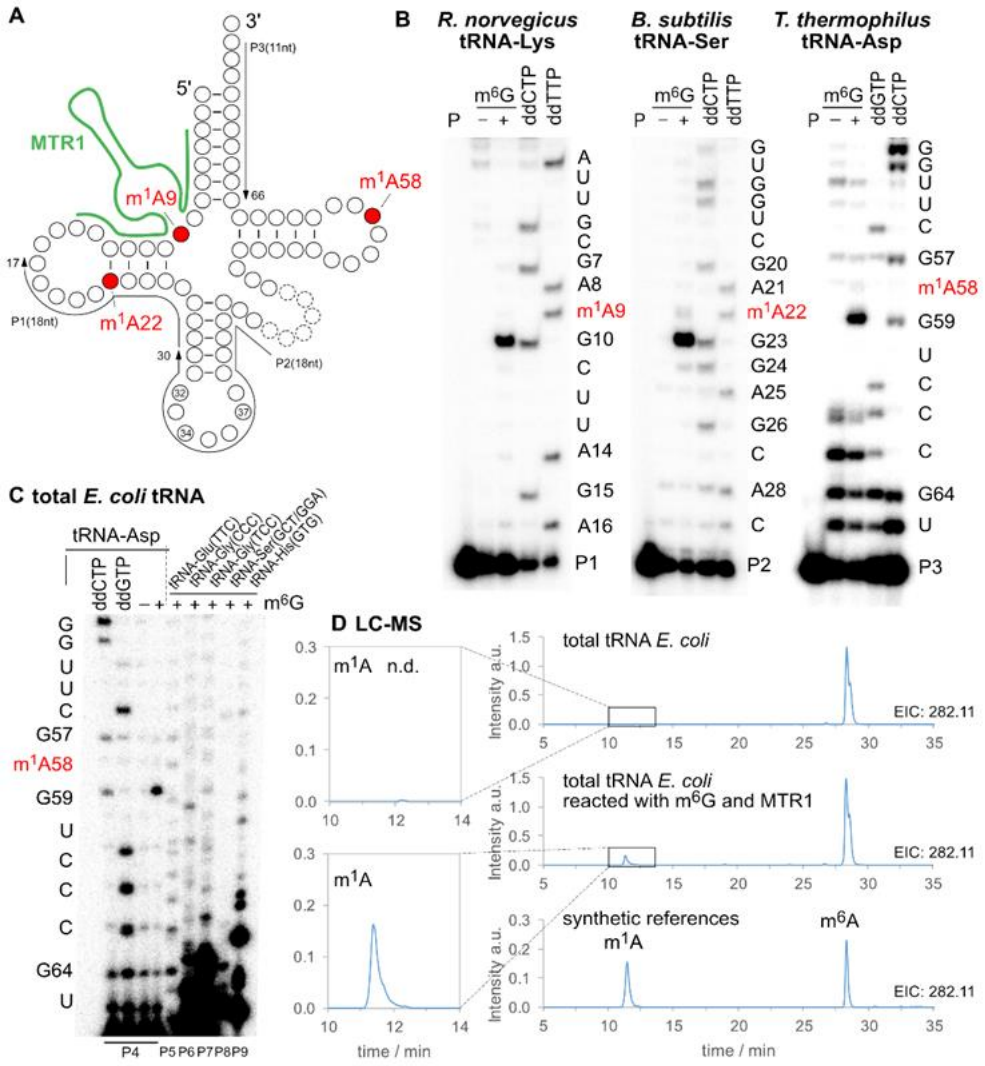



**Figure 3.3:** MTR1-catalyzed methylation of tRNA. **A** Native $m^1A$ sites at positions 9, 22 and 58 are shown in a generic tRNA scaffold. **B** *In vitro* transcribed tRNAs were incubated with the corresponding complementary ribozymes in the presence (+) or absence (-) of $m^6G$. The installation of $m^1A$ was probed by primer extension experiments. Primer binding sites are indicated on the tRNA scheme. Sequencing reactions were run in parallel to assign the position of the abort bands. **C** Total *E. coli* tRNA was incubated with MTR1 specific for A58 of tRNA-Asp, and specific methylation of tRNA-Asp was probed by primer extension with six tRNA-specific primers (P4-P9). Sequences of tRNAs are given in Figure 9.6. **D** LC-MS analysis of MTR1-catalyzed methylation of total *E. coli* tRNA. Extracted ion chromatograms (EIC) (detecting $MH^+$ (*m/z* 282.11 ± 0.05), corresponding to methylated adenosines) are shown for digested tRNAs (before and after treatment with MTR1), and for synthetic $m^1A$ and $m^6A$ nucleosides. AU, arbitrary units; ND, not detected.

Successful methylation of *in vitro* transcribed tRNA stimulated the test of MTR1 for specific methylation of one target tRNA in total *E. coli* tRNA, since $m^1A$ has not been found as natural modification in *E.coli* tRNAs.[289-290] After treatment with MTR1 and $m^6G$, primer extension assays with six different tRNA-specific primers confirmed methylation of the target tRNA-Asp at A58, whereas other tRNAs with highly similar TψC-stem-loop sequences (tRNA-Glu, two tRNA-Gly, tRNA-Ser and tRNA-His) were not methylated (Figure 3.3C, Figure 9.6). Additionally, $m^1A$ was unequivocally detected by liquid chromatography-mass spectrometry (LC-MS) in the total tRNA nucleosides of MTR1-treated total *E. coli* tRNA, but not in native *E. coli* tRNA (Figure 3.3D).

To investigate the potential for future application of MTR1 *in vivo*, we designed plasmids for expression of MTR1 in *E. coli*. One construct contained the *cis*-active MTR1 ribozyme in the stabilizing F30-scaffold together with the fluorogenic aptamer Broccoli,[291] which was used to confirm ribozyme expression by staining with DFHBI (Figure 3.4A). Successful methylation was shown by primer extension and LC-MS after incubation of isolated total *E. coli* RNA with $m^6G$ (Figure 9.7). These results confirm the correct folding of the ribozyme in the Broccoli-F30 construct. Direct *in situ* methylation was limited by the availability of $m^6G$ and the required $Mg^{2+}$ level in *E. coli*. A second plasmid contained a *trans*-reactive F30-Broccoli-MTR1 construct that was targeted against *E. coli* tRNA-Asp, and the corresponding transcript was tested on total *E. coli* tRNA (Figure 3.4B). Primer extension assays with the *E. coli* tRNA-specific primers confirmed that the specificity of the MTR1 ribozyme was maintained when incorporated into the F30-scaffold (Figure 9.8). These results establish the MTR1 ribozyme as a promising tool for the installation of $m^1A$ at a specific target RNA, and may therefore aid in the validation of predicted and controversially discussed $m^1A$ sites in eukaryotic mRNA[292-293], as well as enable the study of $m^1A$ biology (readers and erasers) in RNAs for which corresponding methyltransferase enzymes have not yet been identified.[294-295] Moreover, we notice that these ribozymes could serve as highly promising tools for site-specific labeling of RNA, using fluorescently labeled benzylguanine derivatives as cofactors for RNA-catalyzed RNA alkylation.

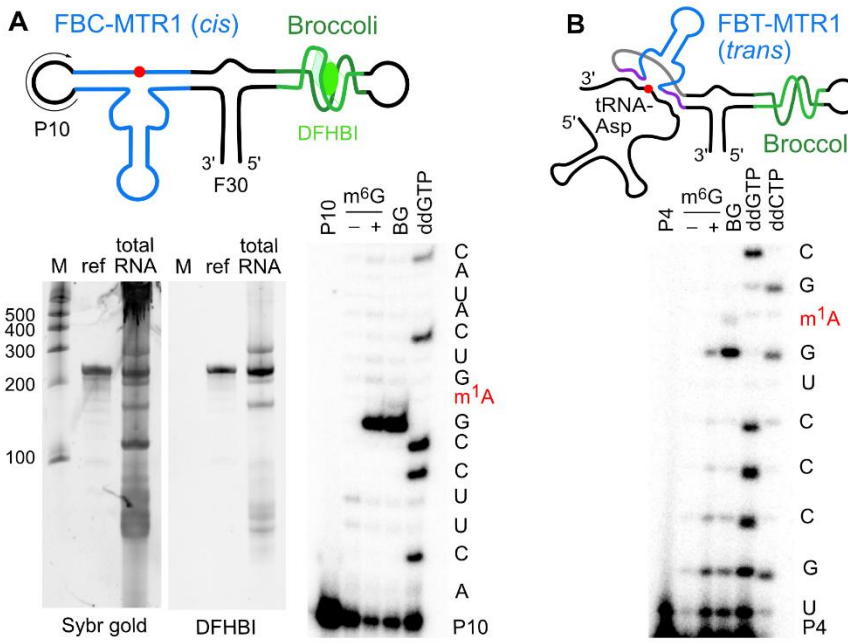



**Figure 3.4:** Plasmid-encoded *cis*- and *trans*-active MTR1. **A** F30-Broccoli-cis (FBC)-MTR1 (modification site indicated as a red dot) is transcribed in *E. coli*. Total RNA isolated 1 h after isopropyl β-d-1-thiogalactopyranoside (IPTG) induction, analyzed on PAGE next to an *in vitro* transcribed reference (ref.) and a size marker (M), and stained by DFHBI (10 % denaturing PAGE, 20 µM DFHBI) and SYBR gold. Both the *in vitro* and *in vivo* transcripts contain an active MTR1 ribozyme, as revealed by the primer extension stops after incubation with $m^6G$ or BG (here shown with *in vitro* transcript and primer P10; data with *in vivo* transcript are shown in Figure 9.7). **B** F30-Broccoli-trans (FBT)-MTR1 with binding arms specific for hybridization to *E. coli* tRNA-Asp. The activity was reduced with $m^6G$ but retained with BG. Formation of $bn^1A$ was confirmed by LC-MS and specifically detected only in tRNA-Asp (Figure 9.8).

### 3.2.4 Site-specific methylation of long RNAs

Possible future applications of MTR1 as a tool for fundamental research on methylated nucleosides will require the highly specific installation of multiple $m^1A$ sites, especially in long and complex RNAs. Encouraged by the successful methylation of several tRNAs we challenged MTR1 to target even more complex RNA molecules. A total of eight MTR1 variants with binding arms targeting six different adenosines in the 658 nt long tetR and two different adenosines in the 399 nt long blaR transcript were designed. After treatment of the RNAs with the corresponding MTR1 variants in the presence of $m^6G$ at 25 °C for 4.5 h, the RNAs were purified from the ribozymes on denaturing PAGE. Successful and highly site-specific generation of $m^1A$ was confirmed by RT primer extension assays on the isolated transcripts, revealing a strong abortive band corresponding to every targeted position (Figure 3.5, Figure 9.9). Following, nanopore sequencing further unveiled the presence of $m^1A$ at the desired sites and confirmed that the RNAs were still intact. This demonstrates that our methylation protocol is compatible even with complex RNAs longer than 650 nt and that MTR1 variants targeting different sites within the same transcript can act simultaneously without any obstructive crosstalk.

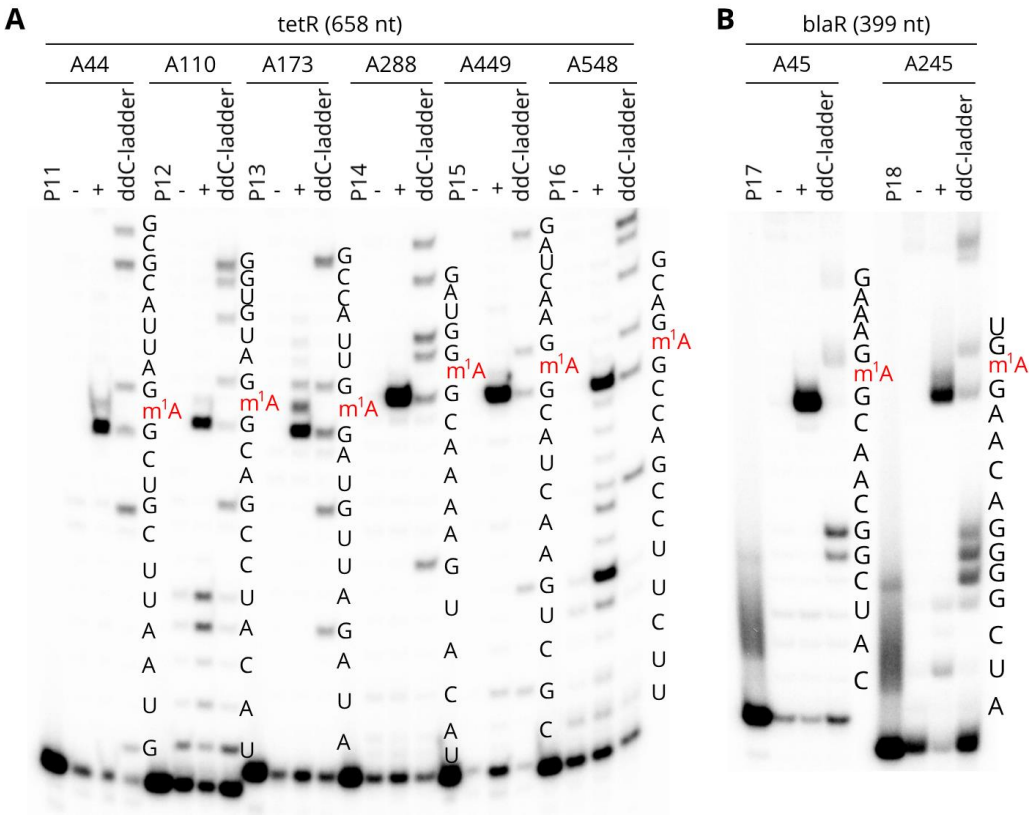
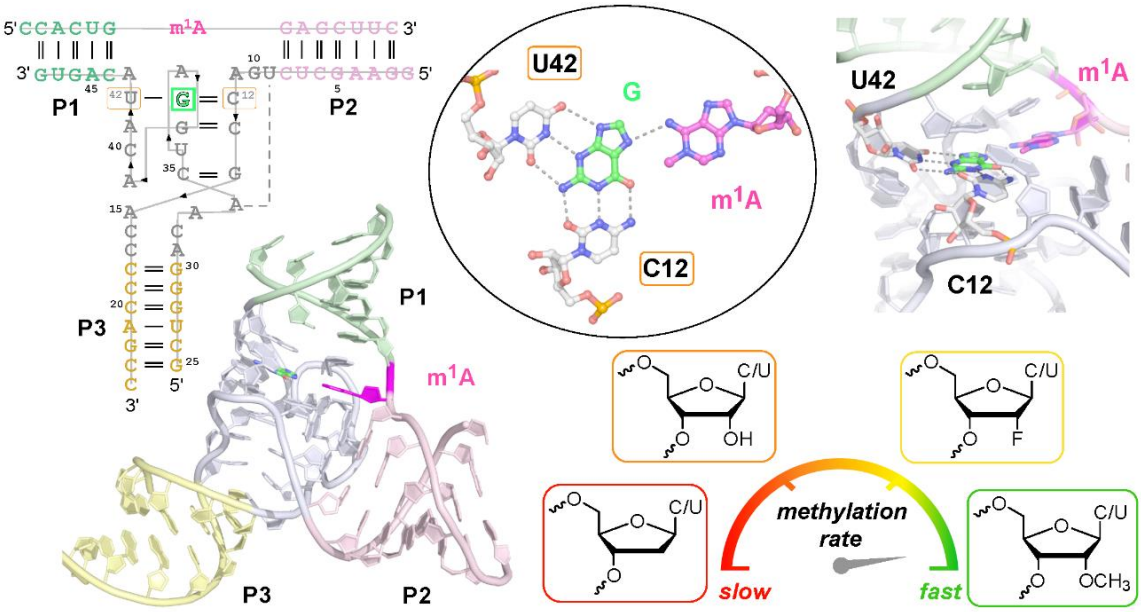



**Figure 3.5:** MTR1-catalyzed methylation of long RNAs. *In vitro* transcribed 658 nt long tetR (**A**) and 399 nt long blaR (**B**) were incubated with six or two different ribozymes respectively in the presence (+) or absence (-) of $m^6G$. Successful generation of $m^1A$ was probed by RT primer extension. Sequencing reactions using ddC (indicating guanines) were run in parallel to assign the position of the termination product bands. Full gel images are provided in Figure 9.9.

## 3.3 Discussion

In summary, we report a ribozyme with methyltransferase activity for the site-specific methylation of adenosine. The methyl group donor for the MTR1 methyltransferase ribozyme is a simple methylated nucleobase. Conceptually, the ribozyme mimics RNA-guided RNA methylation by RNA-protein complexes, such as CD-box RNPs involved in 2′-*O*-methylation of ribosomal RNA.[296] Here, the ribozyme combines both functions - guide and enzyme - in a single molecule of RNA. The cofactor-binding site in the catalytic core of the *in vitro* selected ribozyme may imitate the binding site of guanine or $m^6G$ in purine riboswitches.[297-298] Thus, it is conceivable that methyltransferase ribozymes could be evolved from riboswitch RNAs that are known to bind modern methyltransferase cofactors, including SAM and THF derivatives. Given the activity of MTR1 with $m^6dG$, it seems likely that an analogous ribozyme activity can be developed to catalyze the removal of a methyl group from RNA (or DNA), thus mimicking repair enzymes of alkylation damage-response pathways. Such hypothetical RNA repair ribozymes could have been beneficial catalysts in an RNA world, aiding the evolution of RNA replicases by releasing mutagenic methylation blocks that originated from environmental damage and interfered with faithful Watson-Crick base pairing. Our work has also demonstrated that MTR1 enables the site-specific synthesis of $m^1A$ in defined RNA targets. Thus, the reported findings have implications for scrutinizing the evolution of catalytic RNA as well as studying fundamental aspects of RNA methylation in contemporary biology.

# 4 Structure and mechanism of the methyltransferase ribozyme MTR1




This chapter and the corresponding supporting information (chapter 9.2) were published in:

C.P.M. Scheitl, M. Mieczkowski, H. Schindelin, C. Höbartner, *Nat. Chem. Biol.* **2022**, *18* (5), 547-555.

https://www.nature.com/articles/s41589-022-00976-x



For the sake of unity of this thesis, several editorial changes have been made, which, however, do not affect the contents of the thesis.

**Abstract**

RNA-catalyzed RNA methylation was recently shown to be part of the catalytic repertoire of ribozymes. The methyltransferase ribozyme MTR1 catalyzes the site-specific synthesis of 1-methyladenosine ($m^1A$) in RNA, using $O^6$-methylguanine ($m^6G$) as a methyl group donor. Here, we report the crystal structure of MTR1 at a resolution of 2.8 Å, which reveals a guanine-binding site reminiscent of natural guanine riboswitches. The structure represents the post-catalytic state of a split ribozyme in complex with the $m^1A$-containing RNA product and the demethylated cofactor guanine. The structural data suggest the mechanistic involvement of a protonated cytidine in the methyl transfer reaction. A synergistic effect of two 2′-*O*-methylated ribose residues in the active site results in accelerated methyl group transfer. Supported by these results, it seems plausible that modified nucleotides may have enhanced early RNA catalysis and that metabolite-binding riboswitches may resemble inactivated ribozymes that have lost their catalytic activity during evolution.

## 4.1 Introduction

Natural ribozymes are known to catalyze a narrow range of chemical reactions, while a large diversity of natural riboswitches are known to bind nucleotide metabolites, including *S*-adenosylmethionine (SAM), cobalamin, nicotinamide adenine dinucleotide ($NAD^+$), flavine mononucleotide (FMN), and purines, such as guanine and pre-queuosine ($preQ_1$).[159, 300] Several of these metabolites are essential coenzymes for methyltransferase enzymes in contemporary biology. These coenzymes are thought to have evolved in the RNA world, and primordial RNA catalysts may have utilized coenzymes to achieve greater chemical diversity.[301-305] However, only one natural cofactor-utilizing ribozyme has yet been found, that is the glmS riboswitch-ribozyme, which uses glucosamine-6-phosphate to assist the site-specific cleavage of an RNA phosphodiester bond.[149] In contrast, ribozymes generated in the laboratory by *in vitro* selection from random nucleic acid libraries have been shown to catalyze a wider variety of reactions,[159, 187] and some synthetic ribozymes use natural cofactors.[306-309]

While natural RNA-cleaving ribozymes are well characterized, only a few structures are known of artificial ribozymes that catalyze other reactions,[310-313] limiting our general structural and mechanistic view of the scope of RNA catalysis. Recently, we identified the synthetic methyl transferase ribozyme MTR1 that catalyzes the site-specific methylation of RNA to generate 1-methyladenosine ($m^1A$) using $O^6$-methylguanine ($m^6G$) as methyl group donor.[95] Subsequently, a natural $preQ_1$ riboswitch was shown to enable a similar methyl transfer reaction, using the non-natural $O^6$-methyl-$preQ_1$ ligand as cofactor.[185] Micura and co-workers

demonstrated the formation of 3-methylcytidine ($m^3C$) in the $preQ_1$ riboswitch RNA, with the methyl group being transferred from the ligand to a precisely positioned cytidine in the ligand binding site.[185] In analogy, one may anticipate that the active site of MTR1 could mimic the ligand binding site of guanine riboswitches,[314] which was earlier shown to accommodate $m^6G$ with only slight structural perturbations,[298] but a transfer of the methyl group to the RNA had not been observed.

To address the question of how the laboratory-evolved ribozyme MTR1 binds and activates $m^6G$ for synthesis of $m^1A$ in the target RNA, we solved the crystal structure of the MTR1 ribozyme and examined the chemical mechanism of the methyl transfer reaction. We report the co-crystal structures of MTR1 bound to the methylated RNA product and to the remaining guanine at a resolution of 2.8 Å. The sample for crystallization was prepared with unmethylated RNA and $m^6G$, however, the methyl group was stoichiometrically transferred to the RNA during crystallization. Therefore, the structure represents a post-catalytic state. The catalytic core harbours $m^1A$ and guanine in the centre of a 3-helix junction structure, and the H-bonding interactions of the core nucleotides with the bound guanine resemble those found in natural guanine riboswitch aptamers. Additional post-catalytic structures were obtained from experiments with $m^6G$ derivatives that mimic the substrate used for in vitro selection. We found the same guanine binding site architecture with the alkyl groups from $O^6$-benzylguanine (BG) and $O^6$-(4-aminomethyl-benzyl)guanine (BG-$NH_2$) being transferred to adenine. These structural insights suggested the chemical mechanism of MTR1-catalyzed RNA methylation, which was supported by structure-guided mutations of the ribozyme core and analysis of the methylation rates. We found evidence for general acid catalysis and discovered a methylated ribozyme variant (MTR1$m_2$) with strongly enhanced catalytic activity. Our findings uncovered a surprising similarity in the architecture of natural guanine riboswitches[314] and the *in vitro* selected ribozyme, and provide experimental support for the hypothesis that the presence of modified nucleotides may have provided a catalytic advantage for early RNA catalysis.[315] Moreover, these results nourish the thoughts that more diverse ribozymes may have existed in an RNA world.

## 4.2 Results and Discussion

### 4.2.1 Design and activity of the crystallization construct

The MTR1 ribozyme contains two binding arms (P1 and P2) that hybridize to the RNA substrate (R1) with the methylation site being located at the single adenosine in the junction between P1 and P2 (Figure 4.1A). An internal stem-loop in MTR1, earlier confirmed by structure probing and covariation,[95] is connected to the binding arms via the 9-nt and 13-nt long junctions

J2 and J3 that form the catalytic core. The stem P3 is variable, and the loop can be removed, resulting in a split ribozyme that assembles into an active ribozyme from two 24-nt long RNAs (R2 and R3).

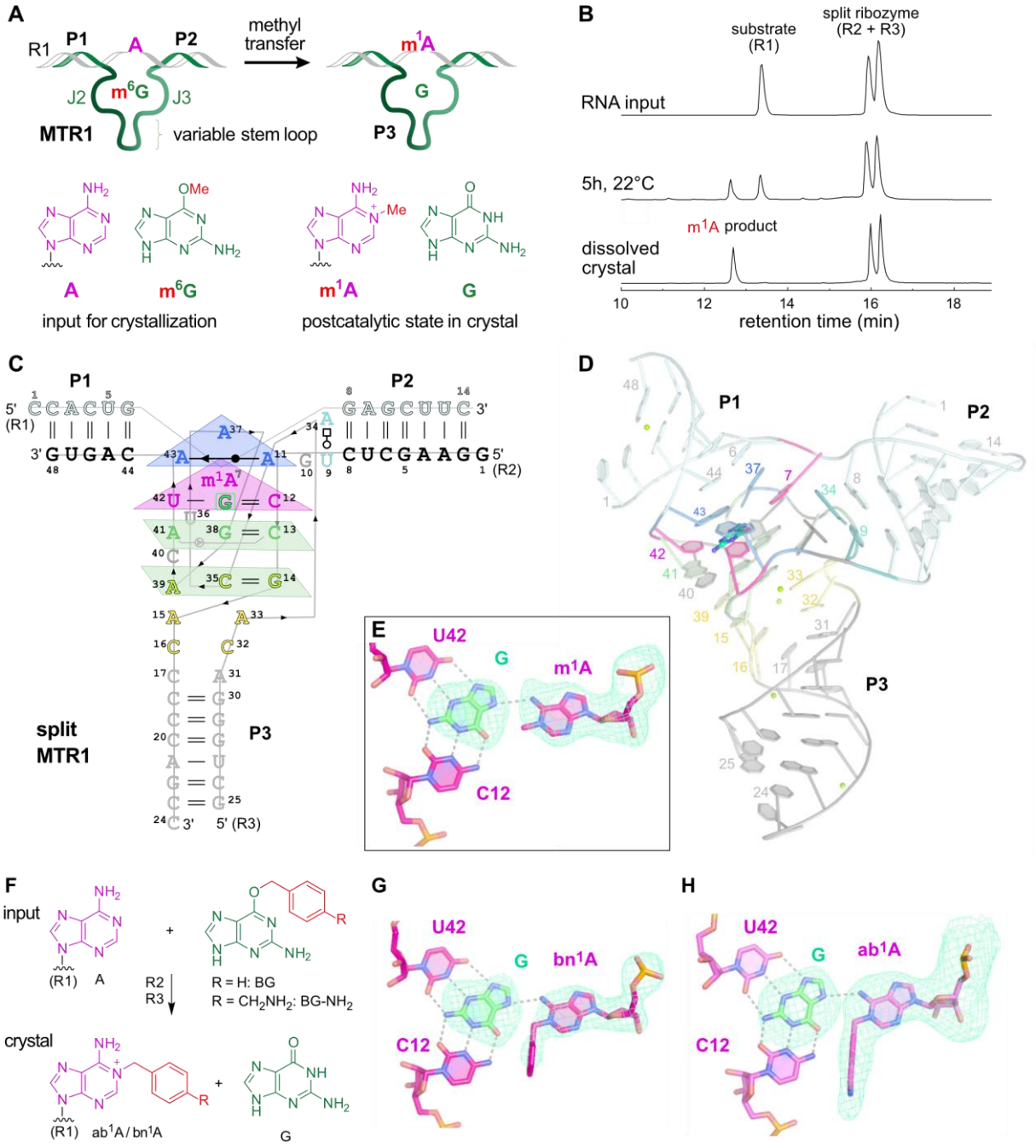



**Figure 4.1:** MTR1 ribozyme-catalyzed methyl transfer reaction and overall structure. **A** Schematic depiction of unmodified RNA (R1) hybridized to MTR1, which binds $m^6G$ and catalyzes methyl transfer to *N*1 of adenosine. Guanine and $m^1A$ are present in the post-catalytic state captured in the crystal. **B** Anion exchange HPLC analysis of a dissolved crystal confirms methyl transfer and shows the presence of three RNA strands. **C** Sequence and scheme of split MTR1 bound to $m^1A$-RNA and guanine (boxed). **D** Ribbon diagram of the three-dimensional (3D) structure, color-coded and numbered as in **C**. **E** Zoom-in of the guanine-binding site. **F** MTR1 also acts as an alkyltransferase ribozyme with cofactors $bn^6G$ and $ab^6G$. **G,H** MTR1 active sites show guanine and alkylated adenosines. The green mesh represents $F_o$–$F_c$ omit maps for guanine and $m^1A$ (**E**), $bn^1A$ (**G**) and $ab^1A$ (**H**) contoured at 3σ.

We screened crystallization conditions for several bi- and trimolecular constructs with variation in binding arm length, composition and overhangs, and found highly reproducible crystallization for the split MTR1 ribozyme complex (R2 and R3) bound to a 14-nt long substrate RNA R1. Diffraction quality crystals were formed when the complex was supplemented with $O^6$-methylguanine ($m^6G$), $O^6$-benyzlguanine (BG) or $O^6$-(4-aminomethyl)benzylguanine (BG-$NH_2$), but not with guanine (G). Anion exchange HPLC analysis of a dissolved crystal grown with $m^6G$ confirmed the presence of all three RNA strands and revealed that product formation of $m^1A$ RNA had occurred (Figure 4.1B). Initial phases were obtained by iterative molecular replacement with A-from double helix fragments,[316] followed by model building and refinement. Additional crystals were grown with heavy atom derivatives, including a 2'-selenomethyluridine modified RNA substrate,[317] and co-crystallization with thallium acetate. Anomalous diffraction data were used for structure solution by MR-SAD and for checking the final structures for consistency. The structure of the $m^1A$ containing MTR1 complex was refined at a resolution of 2.8 Å and revealed the bound guanine in the active site. Likewise, the structures of MTR1 with bound guanine and the RNA products containing $bn^1A$ and $ab^1A$ were refined at resolutions of 2.9 and 3.3 Å, respectively (Table 9.6, Figure 9.10). The electron density maps clearly indicated the presence of guanine and the alkyl groups always attached to *N*1 of adenine (Figure 4.1E, Figure 9.10). The structures with $m^1A$ and $bn^1A$ contained one copy in the asymmetric unit (ASU), while the $ab^1A$-modified complex had two copies in the ASU, and their arrangement was found to be similar to the $m^1A$ containing structure of a crystal grown in the absence of $Mg^{2+}$ (Figure 9.11).

### 4.2.2 Overall structure of the MTR1 ribozyme

The overall structures of the MTR1 ribozyme reflects a 3-helix junction with the three A-form helices P1, P2 and P3 radiating from the central catalytic core, which positions the adenosine substrate (A7 of R1) and the $m^6G$ cofactor. In the crystal structure, the reaction products $m^1A$ and guanine are held in close proximity and form extensive tertiary contacts with the ribozyme. A schematic of the tertiary fold of MTR1 is depicted in Figure 4.1C, and its 3D structure is shown in a ribbon representation in Figure 4.1D. The complementary 5' overhanging nucleotides in P1 and P2 form a semi-continuous duplex in the crystal lattice (Figure 9.11). The catalytic domain contains four layers of stacked base pairs and base triples that are connected to P1 and P3 via continuous $\pi$ stacking interactions, while P2 is oriented almost perpendicularly to the core domain, and is connected via stacking with the reverse Hoogsteen U9:A34 base pair, followed by G10. The top layer of the core domain is formed by a *cis* Watson-Crick sugar edge base pair A11:A43 and A37 (Figure 4.2A). The second layer contains the products of the RNA catalyzed reaction, i.e., $m^1A7$ of the 14-mer RNA strand R1 and the demethylated

guanine ligand, which is immobilized by extensive H-bonding with C12 and U42 (Figure 4.2B). In the third layer, A41 forms an A-minor interaction with the Watson-Crick base pair C13:G38 (Figure 4.2C), and the fourth layer contains the Watson-Crick base pair G14:C35, which is contacted in the minor groove by the Hoogsteen edge and phosphate of A39 (Figure 4.2D). In this way, A39 mediates a continuous staple of C42, A41, C40, A15 and C16, which transitions into P3 via the C17:A31 base interaction (Figure 4.1D). A parallel staple is formed on top of A31, involving C32 and A33. In this arrangement, A33 stacks with the G14:C35 base pair and is placed in the same layer as A15, with their Hoogsteen edges oriented toward each other, but too distant for H-bonding. Similarly, C16 and C32 are oriented toward each other, but are too distant for direct interactions. Instead, this loose junction between the core domain and P3 is stabilized by coordinating metal ions. The structures suggest a magnesium ion interacting with C32 and A15, and additional metal ion binding sites were confirmed by anomalous signals from bound $Tl^+$ ions (Figure 9.10). However, magnesium is not essential for structure formation and activity of the ribozyme, since the reaction products $m^1A$ and guanine were also found in the crystal grown in a buffer that contained only monovalent ions $Na^+$, $K^+$ and $Li^+$ (Figure 9.11).

The overall fold seen in the crystal structure is consistent with results from in-line probing experiments in solution, which monitor changes in backbone flexibility upon RNA folding and ligand binding events.[318] When 5'-$^{32}$P-labeled MTR1 was hybridized to R1 and incubated with varying concentrations of $m^6G$ at pH 8.0, increasing backbone cleavage was observed at U36, which is not involved in any tertiary interactions in the MTR1 structure (Figure 4.2E). However, the cleavage bands at A37 and G38 strongly decreased, consistent with their stacked positions in the catalytic core. Comparable cleavage patterns were obtained with MTR1 hybridized to $m^1A$-RNA and incubation with increasing concentrations of guanine, resulting in an apparent $K_d$ of 2 µM (Figure 4.2F, Figure 9.12). Surprisingly, in the absence of $m^1A$ (*i.e.* when MTR1 was hybridized to unmodified RNA substrate), the affinity for guanine was weaker by almost two orders of magnitude (Figure 4.2G,H), as observed by the in-line probing pattern of A37 which stacks directly on $m^1A$ in the crystal structure (Figure 4.2A).

The functional significance of individual base pairs observed in the crystal structure was supported by activity analysis of several ribozyme mutants (Figure 9.13). Individual base changes in the bubble connecting the core domain and P3 (such as C17U (M1) or A15G (M2)) as well as deletion of the loose C17:A31 base pair (M3) were tolerated, but resulted in 5-10-fold reduced methyl transfer rates at pH 7.5. Attempts to close the stem by base pairing (M4), as well as transition mutations of the AC stacks (M5) impaired the ribozyme activity. The compensatory mutation of the base pair G14:C35 (M6) was tolerated, while changing the base pair C13:G38 (M7) was detrimental.

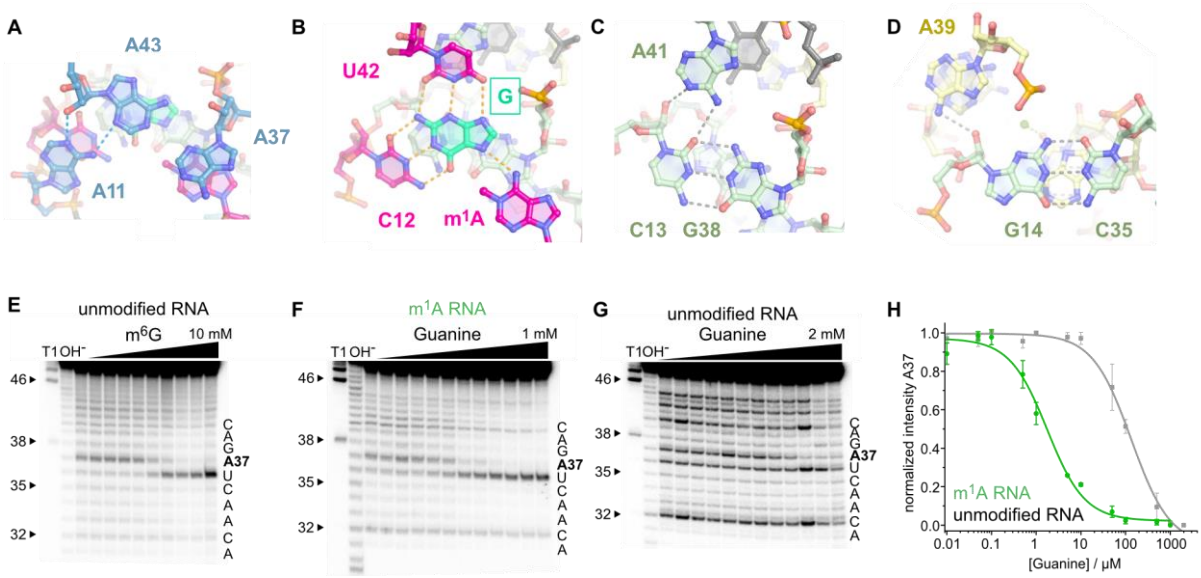



**Figure 4.2:** The catalytic core domain of MTR1. **A**-**D** Stick representations of the four layers of the core domain from top to bottom. The adenosine triple A11, A37 and A43 (**A**), active site with $m^1A$ and guanine interacting by base pairing with C12 and U42 (**B**), C13:G38 base pair with A-minor interaction of A41 (**C**) and G14:C35 base pair interacting with A39 (**D**) are shown. **E**-**G** Representative excerpts of in-line probing gels of MTR1 (pH 8.0, 20 °C, 36 h). Gels of MTR1 hybridized to unmodified R1 in the absence and presence of $m^6G$ (0.5 µM-10 mM) (**E**), hybridized to methylated $m^1A$-RNA (**F**) or unmodified R1 in the absence and presence of guanine (**G**) are shown (0.01-1,000 µM in **F** and 2,000 µM in **G**); T1, RNase T1 digestion; $OH^-$, alkaline hydrolysis. **H** Normalized in-line probing band intensities for A37 in the $m^1A$-containing RNA complex (**F**; green) compared to unmodified RNA (**G**; gray). The lines represent a fit to a one-site-binding model. Error bars denote ±s.d. of the mean for $n$ = 3 (gray) or 4 (green) independent replicates. See Figure 9.12 for full gel images and additional analysis for U36 and G38.

### 4.2.3 Mechanism of RNA-catalyzed RNA methylation

The structure represents the post-catalytic state with the methyl group transferred to $N^1$ of A7, which is sandwiched between A37 and G38 of MTR1. The key observation that the bound guanine in the centre forms extensive hydrogen bonds with C12 and U42 of MTR1 and $m^1A7$ of the methylated RNA product suggests a plausible mechanism for the methyl transfer reaction. The Watson-Crick interaction of guanine with C12 in the product state provides a favourable arrangement of the reactive groups, with the nucleophilic $N^1$ of A7 in line with the electrophilic $CH_3$ of $m^6G$ in the pre-catalytic state. This scenario likely requires a protonated C12 as H-bonding partner for $m^6G$ to maintain a Watson-Crick like orientation, thus suggesting a protonated cytidine as general acid catalyst in the methyl transfer reaction (Figure 4.3A). Consistent with this hypothesis, the MTR1-catalyzed methyl transfer was accelerated by lowering the pH of the reaction medium (Figure 4.3B). The pH dependence of the reaction was analyzed with the full-length MTR1 as well as with the split ribozyme construct used for crystallization, and in both cases, the reaction rate was ca. 8-fold enhanced at pH 6.0 compared to pH 7.5 (Figure 4.3D,E). The pH rate profile showed an optimal pH range near pH 5.5 (Figure 4.3C). The analogous alkyl transfer reaction with $bn^6G$ and $ab^6G$ were also accelerated at pH 6.0 (Figure 9.14). These results likely reflect a modest increase in $pK_a$ for C12 in MTR1 compared

to free cytidine, similar to the apparent p$K_a$ for protonation of a m$^6$G:C base pair in a DNA duplex.[319]

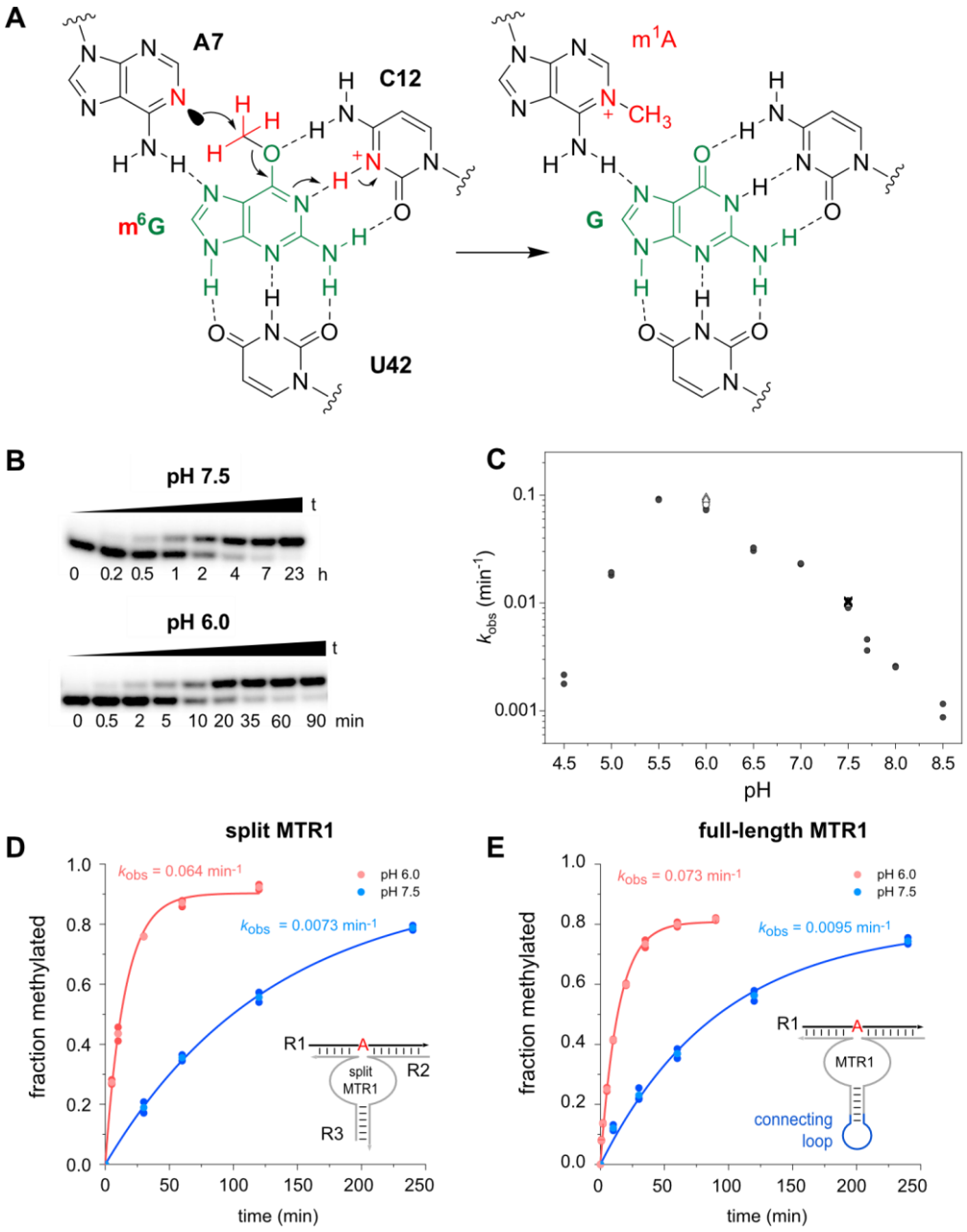



**Figure 4.3:** The methyltransferase ribozyme uses general acid catalysis. **A** A protonated cytidine is suggested in the active site for RNA-catalyzed methyl group transfer. The mechanism involves the nucleophilic attack of the nitrogen lone pair (*N*1 of A7) on the methyl carbon (at *O*6 of m$^6$G). The guanine leaving group accepts a proton from cytidine C12$^+$, resulting in a positive charge on m$^1$A and a standard G:C hydrogen-bonding pattern in the product state. **B** Representative PAGE images of the kinetic assays from three independent experiments with split MTR1 at pH 6.0 and pH 7.5 under single turnover conditions; t, time. **C** pH rate profile for split MTR1 ($n$ = 2 or 3 individual data points for each buffer (different symbols for the same pH indicate different buffers used; for details, see Material and Methods section). **D**,**E** Comparison of methyl transfer rates for split (**D**) and full-length (**E**) MTR1 reaction with m$^6$G at pH 6.0 (red) and pH 7.5 (blue). The following are the conditions used in **B**-**E**: 5′-$^{32}$P-labeled R1, 10 µM MTR1, 100 µM m$^6$G and 40 mM $MgCl_2$ at 25 °C.

### 4.2.4 Rate acceleration by ribose methylation

Consistent with the functional importance of the base triple, both the C12U mutant and the U42C mutant showed strongly reduced reaction rates (Figure 4.4A). At pH 6.0, the U42C mutant still generated 65 % methylated product after 7 h, but less than 15 % product was formed by the C12U mutant after 24 h. Replacing C12 by $m^5C$ lowered the rate by less than a factor of two, while a variant containing $m^4C$ was also significantly impaired (30-fold slower rate than wt), likely due to steric crowding in the active site (Figure 4.4B,C). Replacing C12 and U42 individually or simultaneously with the corresponding 2'-deoxynucleotides was tolerated, although the reaction rates were slightly reduced, suggesting that nucleotides with a 3'-endo conformation are preferred at positions 12 and 42. This was further confirmed by installation of the analogous 2'-fluoro-2'-deoxy nucleotides, which restored the activity, when inserted individually, but enabled 2-3-fold faster methyl transfer when incorporated simultaneously. An even more surprising rate acceleration was observed with 2'OMe nucleotides at C12 and U42. Individual substitutions were slightly favourable (less than 2-fold higher $k_{obs}$), but replacing both nucleotides at the same time and allowing them to act in concert enhanced the $k_{obs}$ by a factor of 15. In other words, the $C_m12$-$U_m42$-MTR1 variant (hereafter called MTR1$m_2$; Figure 4.4B) produced 90 % $m^1A$ RNA within 5 min under single turnover conditions at pH 6.0. The $K_{1/2}$ (apparent $K_m$) for $m^6G$ is on the order of 130 µM (obtained from $k_{obs}$ at variable $m^6G$ concentrations; Figure 4.4D). Importantly, the synergistic effect of the two ribose methylations was specific for C12 and U42, both of which are directly involved in ligand/cofactor binding and activation. The 2'OMe substitution was also tolerated at A41, but its combination with C12 was only additive at best. Moreover, we found that MTR1$m_2$ retained significant activity at a strongly reduced $Mg^{2+}$ concentration down to 0.5 mM $Mg^{2+}$, where MTR1wt barely produced any methylated product (Figure 4.4E). These observations strongly confirm the specific roles of C12 and U42 for the activity of the MTR1 ribozyme in recruiting, aligning and activating the cofactor $m^6G$ (Figure 4.4F).

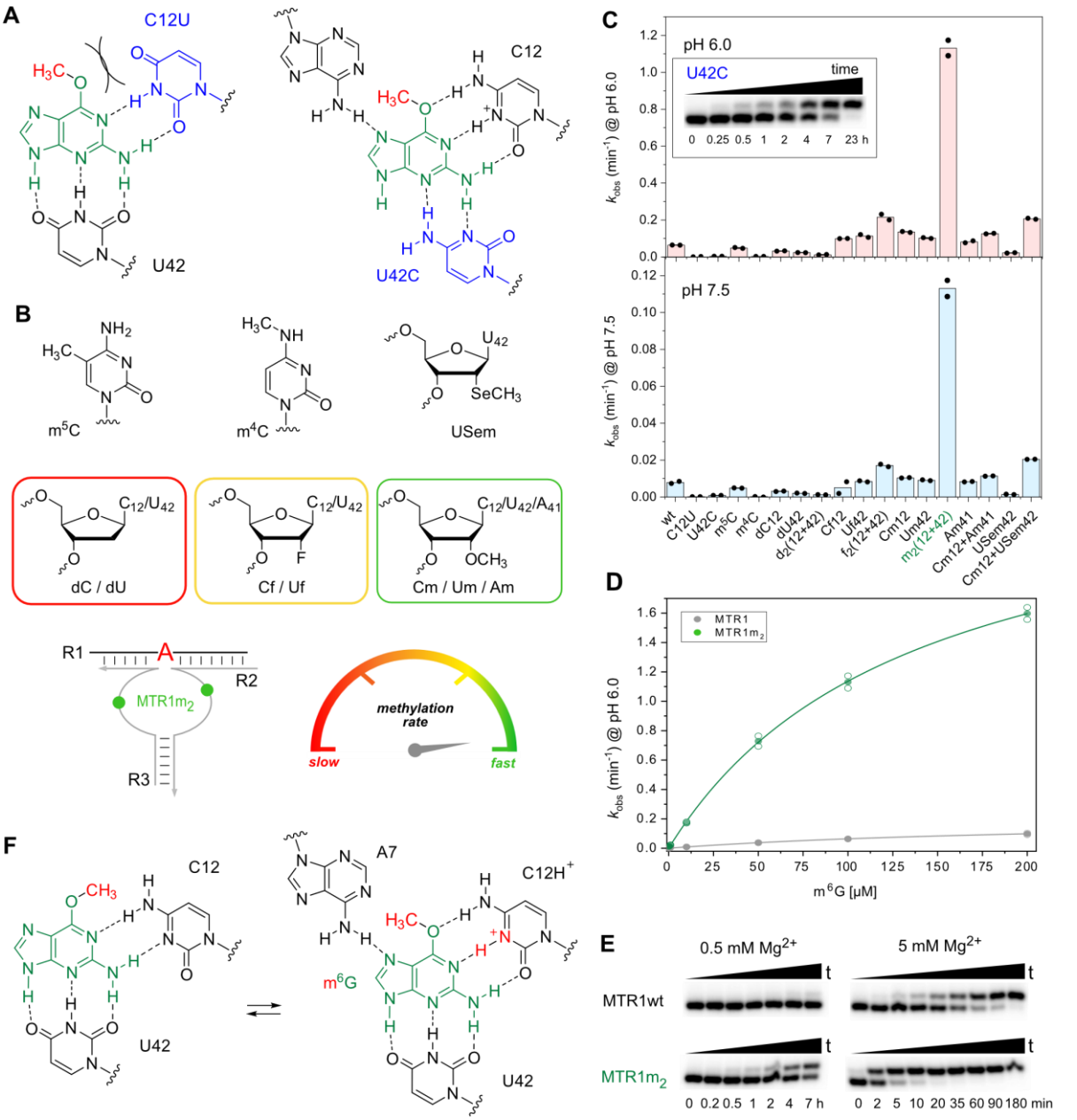



**Figure 4.4:** Acceleration of methyl transfer by synergistic effects of $C_m12$ and $U_m42$. **A** Schematic of C12U and U42C active-site mutants with strongly reduced activity. **B** Structures of modifications used at positions 12 and 42, including schematic representation of MTR1 with simultaneous installation of 2′-OMe nucleotides at C12 and U42 (green symbols in MTR1m$_2$) and color code (red to green) for observed rate accelerations; 'd' indicates 2′-deoxynucleotides; 'f' indicates 2′-fluoro-2′-deoxynucleotides; 'USem' indicates 2′-SeMe-U. **C** Comparison of $k_{obs}$ values for modified split MTR1 ribozymes at pH 6.0 (top) and at pH 7.5 (bottom); individual data points are shown for two independent replicates. Inset shows a representative gel image (of two replicates) for slow but complete reaction of the U42C mutant. **D** MTR1m$_2$ kinetics (green) at various m$^6$G concentrations at pH 6.0 in comparison to unmodified MTR1 (gray). The lines represent the curve fits $k_{obs} = k_{max}\,[m^6G]/(K_{1/2} + [m^6G])$ (where [m$^6$G] is the m$^6$G concentration) to the mean (filled symbol) of two individual replicates shown as empty data points. **E** Representative PAGE images (from two independent experiments) of kinetic assays with MTR1 wild type and MTR1m$_2$ at 0.5 mM $Mg^{2+}$ (left) and 5 mM $Mg^{2+}$ (right) with 5′-$^{32}$P-labeled R1, 10 μM split MTR1 wild type or split MTR1m$_2$, 100 μM m$^6$G, pH 6.0, 25 °C. **F** Possible conformational equilibrium in the active site involving *syn*/*anti* orientation of m$^6$G and C12 in an unprotonated or a protonated state.

### 4.2.5 Comparison of MTR1 guanine binding site with purine riboswitches

The bound guanine cofactor is located in the same position in all three structures, immobilized by hydrogen bonds to the Watson-Crick edges of C12 and U42 and the Hoogsteen side of the $N^1$-alkylated adenine. This active-site architecture has striking similarities to guanine-binding sites found in natural riboswitches (Figure 4.5). In the purine riboswitch family, guanine recognition is mediated by two universally conserved pyrimidine residues[314, 320] such as U51 and C74 in the Bacillus subtilis *xpt-pbuX* riboswitch (Protein Data Bank (PDB): 1Y27). Mutation of these two residues changed the specificity of the guanine riboswitch. While C74U allowed binding of adenine instead of guanine, changing U51 to cytidine shifted the base pairing pattern and created room for binding of *N*9-substituted guanine, similar to 2′-deoxyguanosine (2′dG) riboswitches[321]. In the 2′dG riboswitch from *Mesoplasma florum*, the ligand forms a Watson-Crick base pair with C80, and C58 contacts *N*3 and *N*2 at the sugar edge of the purine nucleoside (PDB: 3SKI; Figure 4.5B). In MTR1, the C12U mutant strongly reduced the methyltransferase activity, while the U42C mutation was better tolerated, suggesting a similarly shifted hydrogen-bonding arrangement, as seen in the 2′dG riboswitch.

Interestingly, the guanine riboswitch was previously shown to bind $m^6G$, but with strongly decreased binding affinity relative to guanine[298], while $bn^6G$ was not bound, and no alkyl transfer was reported from any of the guanine derivatives. The crystal structure with $m^6G$ revealed that C74 shifted downward to form new hydrogen bonds with *N*1 and *N*2 of $m^6G$ and make room for the methyl group in the otherwise tight binding pocket, in which the Hoogsteen edge of the ligand is contacted by the 2′-OH of U22 (Figure 4.5D). In principle, a similarly shifted hydrogen-bonding pattern for C12 with $m^6G$ is conceivable in the MTR1 precatalytic state (Figure 4.4F). However, the relative arrangement of $m^1A$ and guanine in the post-catalytic state suggest that the nucleophilic attack is likely to happen to the methyl group in the anti-orientation (Figure 4.4F, right). Moreover, the protonated $C12^+$:$m^6G$ base pair is more consistent with the observed pH activity profile.

Interestingly, riboswitches also make use of protonated nucleobases, for example, for binding of cyclic-diGMP, which involves a stably protonated adenine (Figure 9.15A,B).[322] Protonated nucleobases with highly shifted p$K_a$ values were also found as key structural elements in synthetic RNA aptamers, including the GTP-binding RNA aptamer.[323] Alternatively, the p$K_a$ of the functional groups in the ligand can be changed following binding to RNA, as, for example, described for aminoglycosides, including neomycin B.[324]

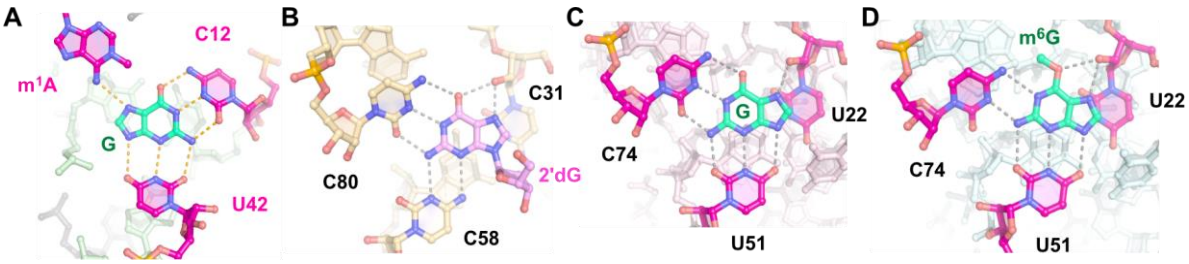



**Figure 4.5:** Comparison of MTR1 guanine-binding site to purine riboswitches. **A** MTR1 active site with $m^1A$ and guanine (this work, PDB: 7Q7X). **B** *M. florum* riboswitch with 2'dG (PDB: 3SKI). **C** The *B. subtilis xpt-pbuX* riboswitch with guanine (PDB: 1Y27). **D** *B. subtilis xpt-pbuX* with $m^6G$ (PDB: 3FO6).

A conceptually similar scenario is likely operating in the $preQ_1$ riboswitch, which was recently shown to mediate a methyl group transfer from an $O^6$-methylated cofactor analog to *N*3 of C15, which is involved in a Watson-Crick base pair with the non-methylated cofactor $preQ_1$ (Figure 9.15C). With $m^6preQ_1$, the methyl transfer reaction was accelerated by shifting the pH of the reaction medium from pH 7.5 to pH 6.0, consistent with protonation of the ligand and stabilization of the leaving group. The methyl transfer rate of the $preQ_1$ riboswitch is, however, much slower than observed for MTR1. $PreQ_1$ yields ~50 % methylated RNA after 48 h (at pH 6.0 with 100 µM $m^6preQ_1$). For comparison, MTR1 provides 90 % methylated RNA in 50 min under comparable conditions (pH 6.0, 100 µM $m^6G$). This difference is not too surprising, given that the $preQ_1$ riboswitch evolved in nature for tight binding to the unmethylated ligand $preQ_1$, which has also been shown to be a strong competitive inhibitor for the $preQ_1$-mediated RNA methylation.[185] By contrast, MTR1 was evolved *in vitro* by enrichment based on catalytic activity and not based on tight binding or high affinity. This is also reflected in the $K_d$ for guanine, which was found by in-line probing to be in the range of 2 µM for the methylated ribozyme-product complex and >100 µM for the unmethylated ribozyme-substrate complex, in contrast to a $K_d$ of 5 nM for guanine binding to the natural riboswitch.[325-326] The weak affinity for guanine also supports the potential of MTR1 to perform multiple-turnover catalysis of RNA methylation, especially in the trimolecular format, with the RNA substrate transiently hybridized to the split ribozyme.

## 4.3 Discussion

The crystal structure of the MTR1 ribozyme indicates that the RNA forms an intricate cofactor-binding site that enabled highly efficient methyl transfer from $m^6G$ to the RNA before or during crystallization. The structure revealed the post-catalytic state of the RNA catalyst, with the $m^1A$-containing methylated RNA product and unmodified guanine bound in the catalytic core. Analogous results were obtained with two additional $O^6$-alkylated guanine co-substrates that transferred a benzyl group to *N*1 of A7 in R1. The benzyl/(4-aminomethyl)benzyl groups

appeared in a perpendicular orientation to the adenine nucleobase and pointed outward from the core domain, which is consistent with their likely orientation during in vitro selection, when biotin was conjugated to the 4-aminomethyl group.

We found that the methyl group transfer is accelerated at pH 6.0 compared to at pH 7.5, which hints at the involvement of a protonated nucleobase in the catalytic mechanism. Indeed, the organization of the active site is highly consistent with a protonated cytidine in the precatalytic state that facilitates binding through hydrogen bonding of $m^6G$ and at the same time serves as a donor for protonation of the leaving group, resulting directly in the stable G•C base pair found in the product state. The pH rate profile suggests that the p$K_a$ of C12 is moderately enhanced compared to the p$K_a$ of cytidine in single-stranded RNA. Shifted p$K_a$ values toward neutrality are commonly observed in ribozyme catalysis of RNA cleavage[327], and the p$K_a$ of nucleobases are considered key determinants of chemical evolution.[328]

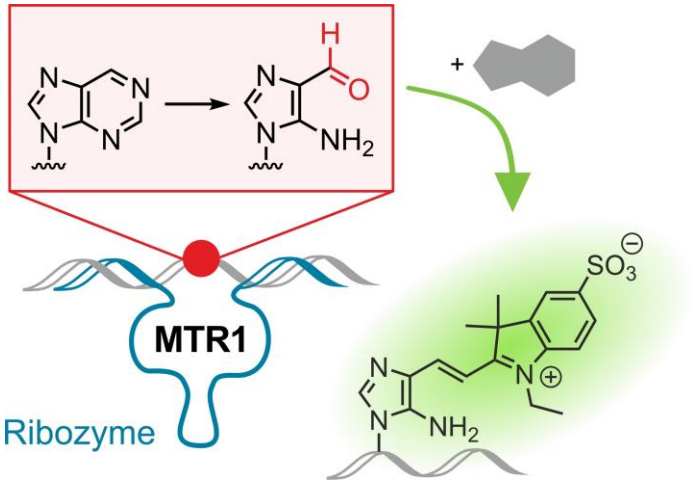
Moreover, the methyl transfer rate of MTR1 was strongly enhanced following the simultaneous presence of two 2′-*O*-methylated nucleosides C12 and U42 in the catalytic core. In combination with the lower pH, the methylated ribozyme achieved at least a 120-fold faster methylation than the rates observed under *in vitro* selection conditions. Analogous replacement with 2′-fluoro-2′-deoxynucleosides was also beneficial, but the accelerating effect was much less pronounced. Ribose methylation is known to affect the preferred ribose pucker and the flexibility and dynamics of conformational states[329], and, in addition, 2′-ribose modifications have a small effect on the p$K_a$ of the nucleobase.[330] A combination of these forces likely causes the unexpected synergistic effect of $C_m12$ and $U_m42$. To our knowledge, a comparable accelerating effect by two natural methylated nucleotides acting in concert to enhance cofactor binding and catalytic activity has not been previously observed for any other ribozyme. In previous work, 2′-modified nucleotides were used to explore the importance and the involvement of the 2′-OH groups in the catalytic mechanisms of ribozymes.[331-332] While placement of 2′-OMe or locked nucleic acid modifications in the ribozyme binding arms usually increased activity of RNA-cleaving ribozymes and deoxyribozymes by enhancing the hybridization strengths, introduction into the catalytic core was often detrimental or in the best case tolerated. Recently a synergistic effect of two 2′-fluoroarabino nucleosides introduced into an engineered version of the RNA-cleaving 10-23 deoxyribozyme was shown to be beneficial for multiple turnover under cellular conditions.[333]

Our data indicate that MTR1 does not require divalent metal ions for catalysis. The post-catalytic state of MTR1 also occurred in a crystal that was grown in a buffer without added $Mg^{2+}$ ions. On first glance, this was surprising, because previous kinetic experiments revealed a pronounced drop in the methyl transfer rate below 5 mM $Mg^{2+}$ (ref.[95]). Likely, the higher concentrations and the longer reaction time during crystallization rescued the slower rate in the

absence of $Mg^{2+}$. In principle, $Mg^{2+}$ ions could have assisted in cofactor binding, as was recently shown in the structure of a xanthine riboswitch.[334] However, none of the six structures analyzed in this work revealed any $Mg^{2+}$ ions in contact with the purine cofactor. Our data instead support the view that $Mg^{2+}$ assists folding and/or affects RNA dynamics in the ribozyme core. The observation that the methylated ribozyme MTR1$m_2$ retained high activity with less than 1 mM $Mg^{2+}$ is consistent with the interpretation that the 2′-OMe nucleotides impart some degree of preorganization to the catalytic core.

In summary, we reported the structure and the chemical mechanism of the methyltransferase ribozyme MTR1. The crystal structure captured the post-catalytic state with the methylated RNA product and the released guanine bound in close proximity. We found that the guanine-binding site shares common structural features with natural purine riboswitches. The ligand RNA interactions were confirmed by mutagenesis. The synergistic effect of the two methylated pyrimidine nucleosides directly in contact with the cofactor suggests that small changes of a few key nucleotides can have profound beneficial effects for catalysis. Based on these results, we anticipate that a wide range of activities for cofactor-utilizing ribozymes can be identified in the laboratory. Our results also support the hypothesis that methylated nucleotides in RNA may be evolutionary leftovers from a time when RNA catalysts were engaged in a wider scope of chemical reactivity.

# 5 Ribozyme-catalyzed late-stage functionalization and fluorogenic labeling of RNA




This chapter and the corresponding supporting information (chapter 9.3) were published in:

C.P.M. Scheitl, T. Okuda, J. Adelmann, C. Höbartner, *Angew. Chem. Int. Ed.* **2023**, *62*, e202305463.

(https://onlinelibrary.wiley.com/doi/10.1002/anie.202305463)



For the sake of unity of this thesis, several editorial changes have been made, which, however, do not affect the contents of the thesis.

**Abstract**

Site-specific introduction of bioorthogonal handles into RNAs is in high demand for decorating RNAs with fluorophores, affinity labels or other modifications. Aldehydes represent attractive functional groups for post-synthetic bioconjugation reactions. Here, we report a ribozyme-based method for the synthesis of aldehyde-functionalized RNA by directly converting a purine nucleobase. Using the methyltransferase ribozyme MTR1 as an alkyltransferase, the reaction is initiated by *N*1-benzylation of purine, followed by nucleophilic ring opening and spontaneous hydrolysis under mild conditions to yield 5-aminoimidazole-4-carbaldehyde site-specifically and in high yields. The modified nucleotide is accessible to aldehyde-reactive probes, as demonstrated by the conjugation of biotin or fluorescent dyes. Moreover, upon fluorogenic condensation with a 2,3,3-trimethylindole donor, a novel hemicyanine chromophore was generated directly on the RNA oligonucleotide. This work expands the MTR1 ribozyme's area of application from a methyltransferase to a tool for site-specific late-stage functionalization of RNA.

## 5.1 Introduction

Site-specific modifications of biomolecules for labeling and crosslinking have become increasingly important for studying structures, functions and interactions of nucleic acids[336-339] and proteins[340-342]. Aldehydes are attractive modifications for bioconjugation, mostly by reaction with nitrogen nucleophiles through condensation or reductive amination.[343-344] Masked aldehydes can be installed synthetically, e.g., by solid-phase synthesis[345-348] or enzymatic primer extension,[349-350] and released by deprotection of acetals, or generated by oxidative cleavage of 1,2-diols[351-355] of furan residues.[356-357] Alternatively site-specific enzymatic transformations of amino acids or modified nucleotides can give rise to aldehyde functional groups for post-synthetic modification and bioconjugation with aldehyde reactive probes. For example, in DNA, aldehydes are introduced through the local generation of abasic sites that are in equilibrium with the open aldehyde form (Figure 5.1A), e.g., by uracil deglycosylase (UDG) or other lyases.[358-359] Natural modified nucleotides containing aldehyde functional groups such as 5-formylcytidine, are generated enzymatically by Fe(II)/α-KG dependent dioxygenases, e.g., TET enzymes[360] or ALKBH1[361] (Figure 5.1B). In proteins, internal site-specific generation of an aldehyde side chain has been described by formylglycine generating enzyme (FGE), which oxidizes a cysteine in a specific sequence context (Figure 5.1C).[362] The aldehyde-tag[363] is used as short genetically encoded peptide tag that is fused to a protein of interest, and upon treatment with FGE the resulting formylglycine is further chemically modified. Here, we developed a conceptually similar modification strategy for RNA that employs a synthetic ribozyme

as mimic of an aldehyde generating enzyme. A purine nucleobase (**1**) in RNA is converted to a 4-formylimidazolyl nucleoside analog (**3**) in a site-specific manner, enabled by an RNA-alkylating ribozyme that generates 1-benzylpurine (**2**) as an intermediate (Figure 5.1D).

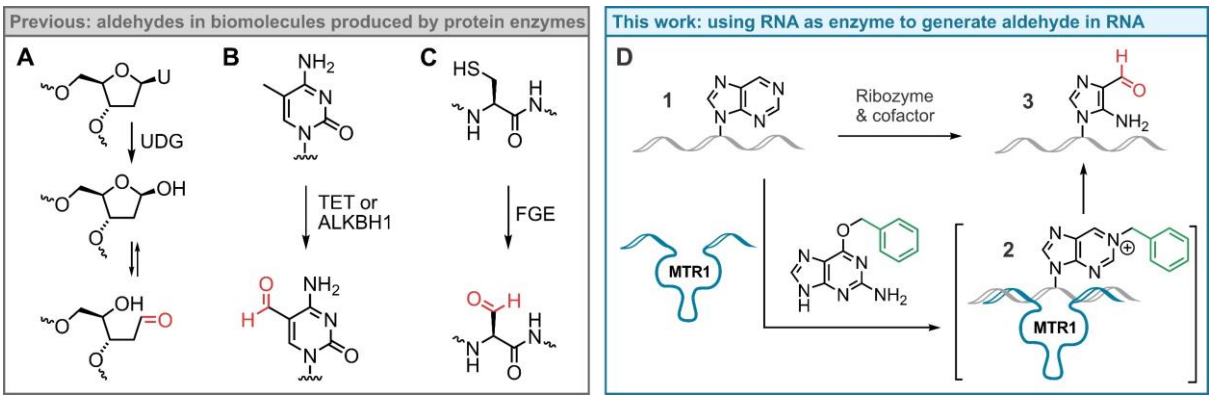



**Figure 5.1:** Enzymatic generation of aldehydes in nucleic acids and proteins. **A** Formation of abasic site in DNA by glycosylases (e.g., UDG). **B** Oxidation of $m^5C$ to $f^5C$ in DNA/RNA by dioxygenase enzymes of the AlkB family (e.g., TET in DNA or ALKBH1 in tRNA). **C** Oxidation of cysteine to formylglycine in a peptide tag consensus motif (CxPxR) by formylglycine-generating enzyme (FGE). **D** RNA-catalyzed synthesis of 5-amino-4-formylimidazole nucleobase analog **3** in RNA via *N*1 benzylation of a specific purine nucleobase.

## 5.2 Results and Discussion

### 5.2.1 Identification of the reaction product

The aldehyde generation in RNA is enabled by our recently reported methyltransferase ribozyme MTR1,[95, 299] which catalyzes the methylation of adenosine (A) at the *N*1 position using $O^6$-methylguanine ($m^6G$) as the methyl group donor. Crystal structures revealed the mechanism of the MTR1 ribozyme[299, 364], which originated from an *in vitro* selection using biotinylated $O^6$-benzylguanine as alkyl donor. Thus, MTR1 is also an active alkyltransferase ribozyme with $O^6$-benzylguanine (BG) as the cofactor. Regarding the substrate scope of MTR1, previous atomic mutagenesis studies revealed that besides adenosine (A), 2'-deoxyadenosine (dA), 2'-*O*-methyladenosine ($A_m$), 3-deazaadenosine ($c^3A$) and 7-deazaadenosine ($c^7A$) were tolerated as target nucleosides, while 2-aminopurine (2AP) and purine ribosides (i.e., Nebularine, Ne) did not show any shifted product bands on denaturing PAGE. However, more detailed analysis of the reaction products by HPLC and ESI-MS revealed that the Ne-containing RNA was not unchanged. We studied the formation of the new product in detail and turned it into a useful RNA labeling approach that involves the formation of the aminoimidazole carbaldehyde **3**.

MTR1-catalyzed benzylation of unmodified RNA (i.e. containing adenosine at the target position) showed a single product corresponding to 1-benzyladenosine, which was detected by denaturing PAGE and anion exchange HPLC, while RNAs that contained 2-aminopurine or

2,6-diaminopurine at the target site were unreactive (Figure 9.17). In contrast, when adenosine was replaced by nebularine, anion exchange HPLC of the reaction product showed an additional later eluting peak (Figure 5.2A), with increasing intensity after longer reaction time. The new product was lighter by nine mass units than the untreated Ne-RNA, suggesting that the purine ring was opened by nucleophilic attack of water, followed by hydrolysis of imine and formamide intermediates to generate the 5-amino-4-formyl-imidazol-1-yl riboside **3** in the otherwise intact RNA. The suggested mechanism is depicted in Figure 9.18 and supported by the detection of several partially hydrolyzed reaction intermediates by mass spectrometry. Nucleophilic ring opening of *N*1-alkylated purine nucleotides occurs in the context of the Dimroth rearrangement, which converts e.g., *N*1-alkylated to $N^6$-alkylated adenosines.[365-366] On the nucleoside level, similar ring-opening of *N*1-alkylated inosine and other purine nucleosides has been used e.g., during atomic mutagenesis and synthesis of isotope-labeled nucleotides,[367-368] and other modified nucleotides.[369-370] However, the reaction mediated by MTR1 is the first observation of site-specifically generating 5-aminoimidazole-4-carbaldehyde in an intact RNA oligonucleotide.

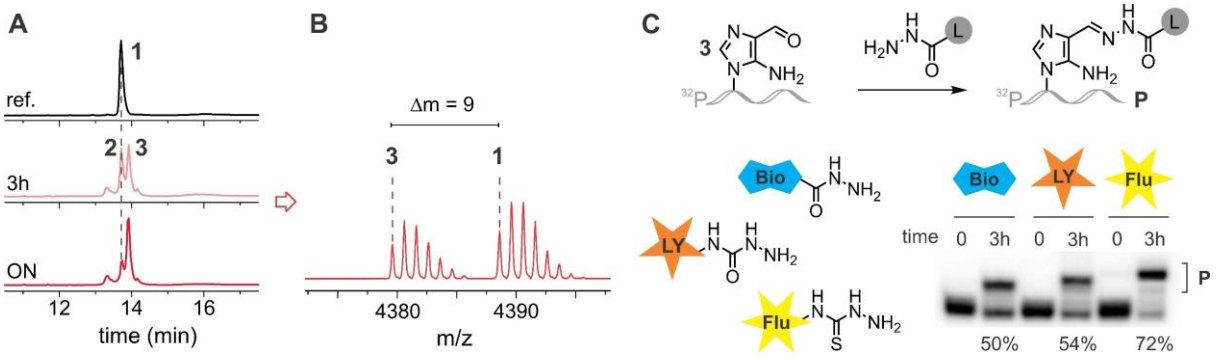



**Figure 5.2: A** Anion exchange HPLC analysis of a 14-nt long Nebularine-containing RNA R1 (reference; ref) after incubation with MTR1 and BG for 4 h or overnight at 25 °C, in MES buffer pH 6.0. **B** Deconvoluted HR-ESI-MS spectrum of RNA isolated after 4 h shows formation of a product that has lost 9 mass units (mol wt. of R1 calc. 4388.63, found mass of the new product **3**: 4379.61 g/mol). **C** Bioconjugation to aldehyde-reactive probes carrying various labels (L): Biotin hydrazide (Bio), Lucifer Yellow carbohydrazide (LY) or fluorescein thiosemicarbazide (Flu). *N,N*-dimethylethylenediamine at pH 7.5, 37 °C, 3 h, 20 % denaturing PAGE using 5'-$^{32}$P-labeled RNA.

This newly generated internal formyl group in RNA is accessible for labeling with aldehyde reactive probes. We used Biotin-hydrazide (Bio-hyd), Lucifer Yellow carbohydrazide (LY-CH) or Fluorescein-thiosemicarbazide (Flu-TSC) and optimized the reaction conditions by varying pH and buffer composition. *N,N*-dimethylethylenediamine (*N,N*-DMED)[371] buffer at pH 7.5 increased the yield to 72 % (Figure 5.1c), while other buffers or lower pH gave only moderate conversion (Figure 9.19). These results establish the MTR1-mediated aldehyde unmasking of Ne-RNA as a potent tool for internal RNA labeling at physiological pH.

### 5.2.2 Generation of a new Hcy-chromophore directly on the RNA

Building upon our previous work on fluorogenic labeling of 5-formylpyrimidines,[372] we hypothesized that a new hemicyanine chromophore could possibly be generated in RNA containing the aldehyde nucleobase **3**. We examined the reaction with *N*-ethyl-2,3,3-trimethylindolenium-5-sulfonate **4** under various conditions and found efficient formation of a new product band on PAGE after 3 h at 45° C (Figure 5.3 A,B). High-resolution ESI-MS analysis of the isolated full-length RNA product confirmed the expected aldol condensation product in the RNA oligonucleotide with a monoisotopic mass of 4628.7 Da (Figure 5.3C). For further analysis, the isolated RNA was enzymatically digested and dephosphorylated to mononucleosides and analyzed by LC-MS. The new nucleoside was readily detected in the UV trace at 260 nm and in the extracted ion chromatogram (Figure 5.3D). Thus, ESI-MS and LC-MS suggested that only one product was formed. Surprisingly, anion exchange HPLC of the full-length RNA isolated from the reaction at pH 8.5 showed two product peaks **5a** and **5b**. Only **5a** showed absorbance at 486 nm, suggesting that **5b** is an isomer likely formed by spirocyclization. Longer incubation in aqueous buffered solution increased the intensity of **5b** (Figure 5.3E). In contrast to spiropyran photoswitches[373-374], which can undergo ring opening to the merocyanine form even when incorporated into DNA[375], in our preliminary screening we did not find conditions for efficient photochemical or thermal restoration of **5a**.

For characterizing the open hemicyanine **5a** by optical spectroscopy, we used freshly PAGE-purified RNA that contained less than 20 % cyclized isomer. The UV-VIS absorption spectrum of freshly purified modified RNA revealed the hemicyanine with an absorption maximum at 512 nm and a shoulder at 486 nm (Figure 5.3F). Fluorescence emission upon excitation at 486 nm showed an emission maximum at 531 nm in both single- and double-stranded RNAs. The emission intensity was sensitive to the nucleoside directly opposite the hemicyanine in the duplex. With adenosine, the duplex showed the same emission intensity as the single strand, while C, G, or U in the duplex opposite to **5a** showed up to 50 % enhanced emission (Figure 5.3G). Given the promise for fluorogenic labeling of RNA, the reaction conditions were optimized to allow ribozyme-catalyzed alkylation, aldehyde generation and labeling to occur in one pot without the isolation of intermediates (Figure 9.20).

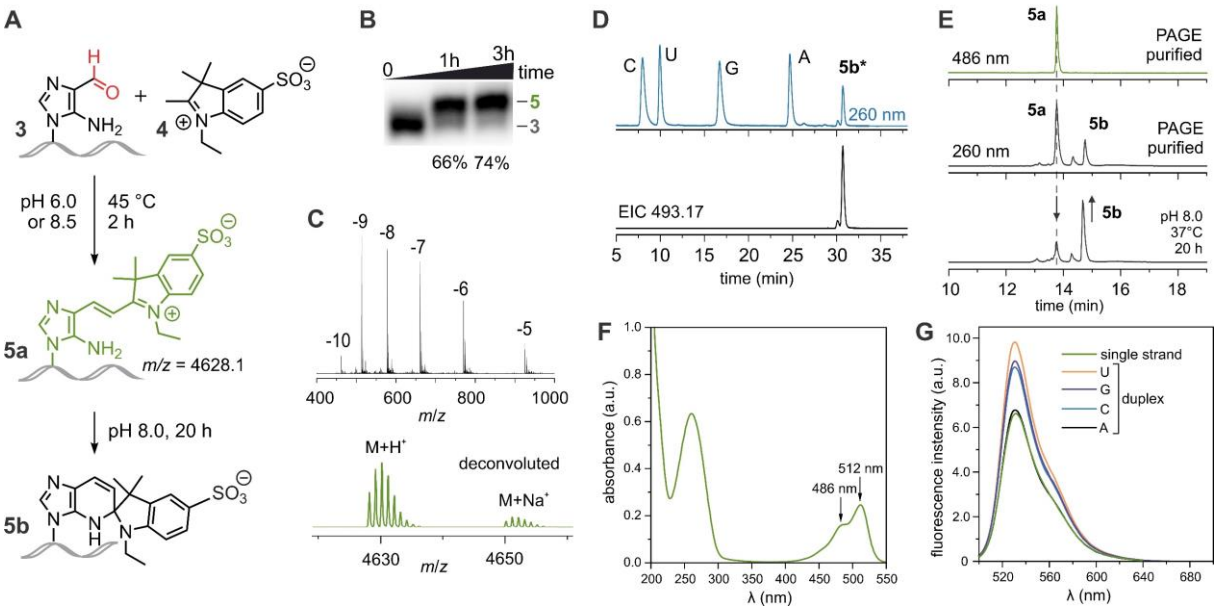



**Figure 5.3: A** Fluorophore synthesis on RNA by condensation of **3** and **4**. The hemicyanine **5a** forms efficiently and can be isolated, but is prone to spirocyclization to **5b**. **B** Gel image of RNA R1 containing **3** reacted with 75 mM **4** at pH 6.0 and 45 °C. **C** Negative mode HR-ESI-MS (top) and deconvoluted spectrum (bottom) of the isolated product after reaction at pH 8.5 ($\Delta m$ = 22, $Na^+$ salt). **D** LC-MS analysis of digested isolated RNA from reaction with **4** at pH 8.5. UV traces at 260 nm and extracted ion chromatogram (EIC) corresponding to the modified nucleoside marked by asterisk (**5***) (m/z ($M+H^+$) = 493.17 ± 0.01, for $C_{22}H_{28}N_4O_7S$). **E** Anion exchange HPLC of freshly isolated RNA with detection at 260 nm and 486 nm, and re-analysis after 20 h of incubation at pH 8.0 and 37°C (bottom). **F** UV/Vis absorption spectrum of freshly isolated RNA (5 µM) showing absorption of **5a** in the visible spectral range. **G** Fluorescence emission spectra of freshly isolated RNA as single strand, or hybridized to a complementary strand, containing one of the four standard nucleotides placed opposite to the chromophore ($\lambda_{ex}$ = 486 nm).

## 5.2.3 Nebularine incorporation by *in vitro* transcription

For general application of the RNA-catalyzed labeling strategy, facile accessibility of Ne-containing RNA is required. Although the nebularine phosphoramidite is commercially available, solid-phase synthesis is limited to rather short RNAs. These can be integrated into larger RNAs by enzymatic ligation.[376-378] Alternatively, *in vitro* transcription provides direct access to long RNAs, but modified nucleotides cannot easily be introduced site-specifically. However, even if nebularine was randomly incorporated during *in vitro* transcription by T7 RNA polymerase, the MTR1 ribozyme can be used to precisely select one position for derivatization, and this is achieved by appropriate design of the binding arms.

It has been reported that 2'-deoxynebularine acts as an analogue of adenosine in DNA and predominantly forms base pairs with thymidine.[379] Nebularine has also been used during nucleotide analogue interference mapping (NAIM) of RNA, which requires its incorporation during *in vitro* transcription.[380] Thus, when ATP was replaced by NeTP (Figure 5.4A) during *in vitro* transcription of the 19-nt RNA Tr1, T7 RNA polymerase was able to generate a full-length product, which contained a single Ne nucleoside, as shown by HR-ESI-MS analysis and by LC-MS after digestion (Figure 9.23). Transcription yields were comparable to unmodified

RNAs, and only dropped for consecutive nebularines, or when Ne was close to the transcription start site. MTR1-catalyzed conversion of Ne in the transcript and labeling of the generated aldehyde were similar to the synthetic RNAs made by solid-phase synthesis (Figure 5.4B,C). Importantly, we showed that MTR1 can site-specifically modify only the desired position in a transcript RNA that contained more than one nebularine, even if these are at neighbouring positions. This was confirmed by RNase T1 digestion and alkaline hydrolysis as shown in Figure 5.4D. A nebularine-guanosine mismatch in the ribozyme binding arm reduced the labeling efficiency, but good reaction yield could be restored by placing uridine across Ne in the binding arm (Figure 9.23).

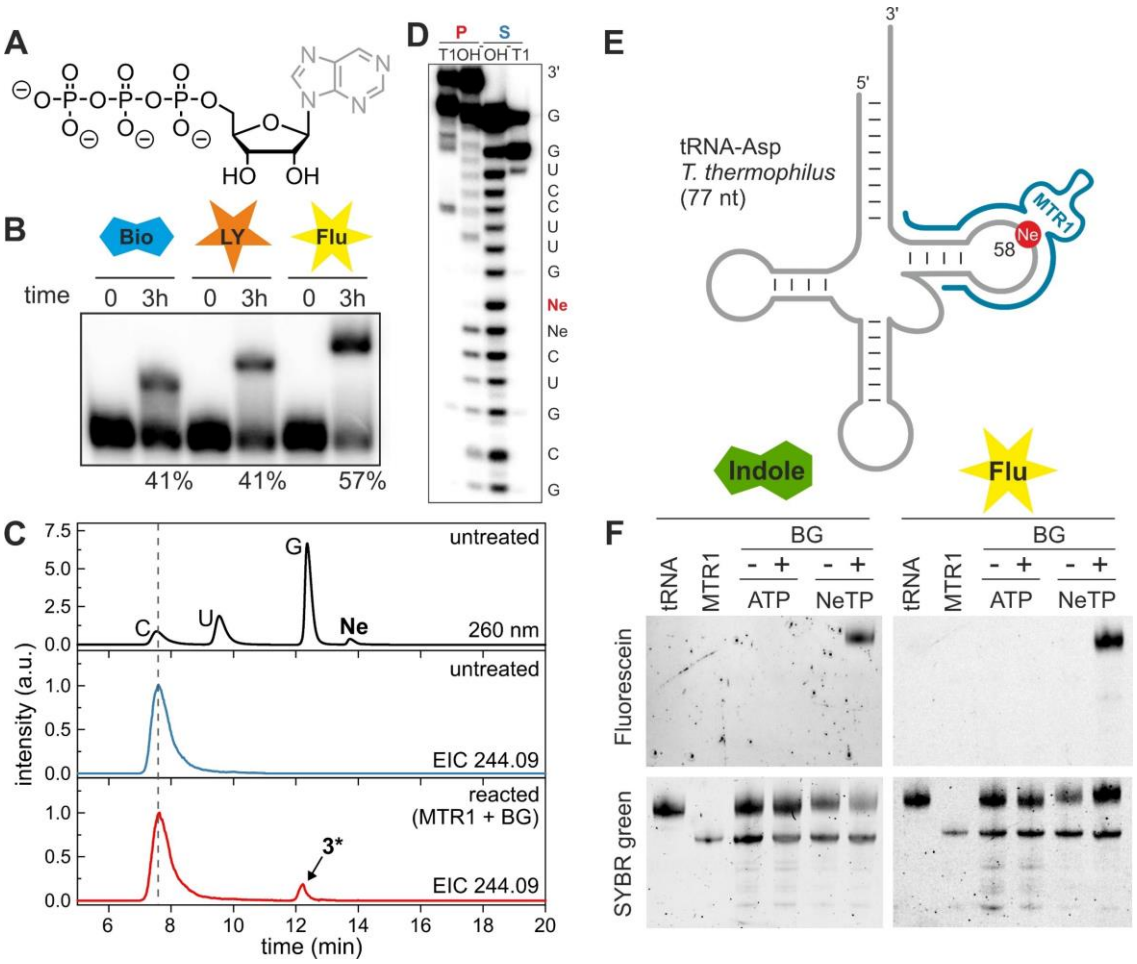



**Figure 5.4:** MTR1-catalyzed labeling of *in vitro* transcribed nebularine-containing RNA. **A** Structure of NeTP. **B** Aldehyde generation and labeling in transcribed RNA Ne-Tr1 (19 nt) conjugated product; P. **C** LC-MS analysis of digested unreacted and reacted (MTR1 + BG treated) Ne-Tr1. UV trace at 260 nm and EIC (m/z 244.09 ± 0.01; note that **3** is a constitutional isomer of cytidine with the same sum formula $C_9H_{13}N_3O_5$). **D** RNase T1 digestion and alkaline hydrolysis of untreated (substrate, S) or MTR1-modified and consecutively Flu-TSC labeled (product, P) transcript Ne-Tr2 (19 nt) with two consecutive Ne, of which only the desired one was modified site-specifically. **E** Schematic representation of the labeling site at position 58 in *T. thermophilus*. tRNA-Asp. **F** *T. thermophilus* tRNA-Asp was transcribed using ATP or NeTP, treated with MTR1 in the presence (+) or absence (-) of BG followed by conjugation with **4** or Flu-TSC. The gels were imaged in the fluorescein-channel (top) to detect the labeled RNAs, and then stained with SYBR green (bottom) to visualize tRNA and MTR1.

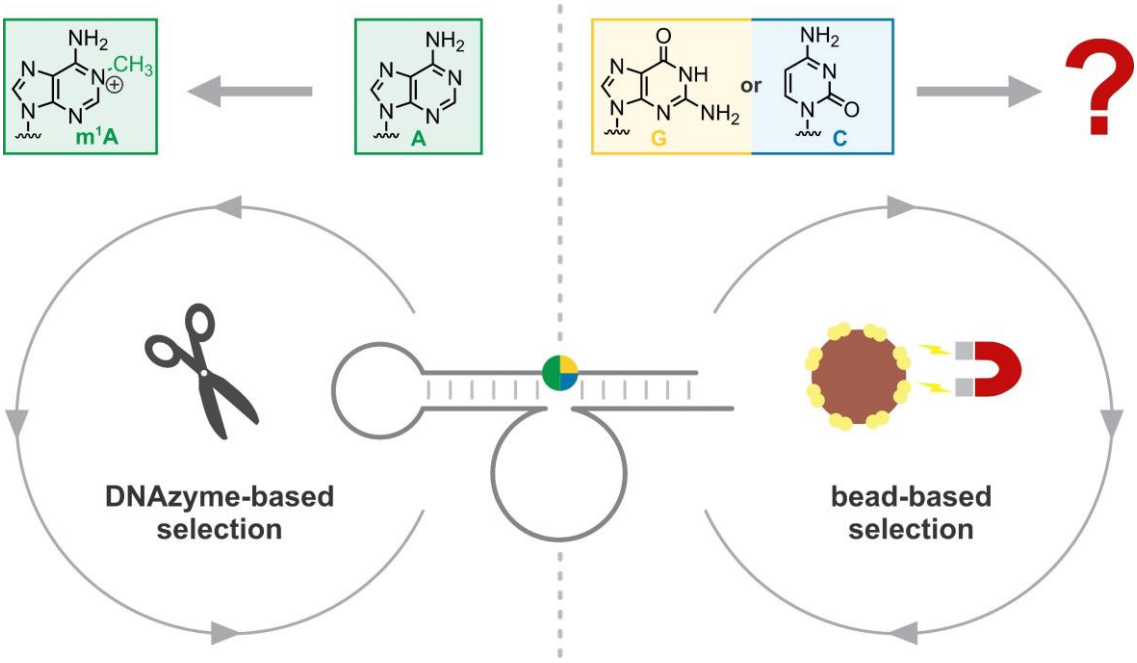
### 5.2.4 tRNA transcription and labeling

Next, we moved to longer RNAs and chose *in vitro* transcribed tRNAs as substrates, which have previously been shown to be accessible for MTR1-catalyzed installation of $m^1A$.[95] The *Thermus thermophilus* tRNA-Asp was transcribed in the presence of NeTP, and modified by an MTR1 ribozyme targeting position 58 (Figure 5.4E). Successful generation of **3** in tRNA was revealed by the presence of a strongly fluorescent band upon conjugation to Flu-TSC, while no fluorescence was detected in the negative control reactions (Figure 5.4F). Similarly, generation of fluorophore **5** allowed for selective visualization of the tRNA (Figure 5.4F).

## 5.3 Discussion

In summary, we report the site-specific generation of a formylimidazole nucleotide analogue in RNA initiated by an alkyltransferase ribozyme. Using $O^6$-benzylguanine as the small molecule cofactor, the MTR1 ribozyme alkylates nebularine to generate 5-amino-4-formylimidazole-ribonucleoside at a specific internal position in the target RNA. The aldehyde can easily be further conjugated with various nucleophiles. We demonstrated the versatility of this approach by fluorescent labeling of a full-length tRNA transcript. In addition, through condensation of the 5-aminoimidazole-4-carbaldehyde **3** with a 2,3,3-trimethylindole moiety, we generated a novel environmentally sensitive hemicyanine fluorophore directly on the RNA. This reaction opens the possibility to synthesize alternative chromophores with altered fluorescent properties that could serve as probes for elucidating nucleic acid structures and dynamics. Furthermore, the MTR1-ribozyme catalyzed aldehyde generation holds potential for nucleic acid-protein cross-linking[381-383] via nucleophilic amino acid side chains to covalently trap proteins that are associated with their DNA or RNA targets. Overall, this work establishes nebularine as aldehyde precursor in RNA and demonstrates its post-synthetic functionalization by an alkyl transferase ribozyme, thus expanding the ribozyme toolbox for RNA labeling.[184, 384]

# 6 *In vitro* selection of new ribozymes for RNA-methylation and -alkylation




This chapter and the corresponding supporting information (chapters 9.4 and 9.5) contain unpublished work.

**Abstract**

Besides proteins, RNA can also fold into defined structures and act as a catalyst for chemical reactions. In nature, these so-called ribozymes are known to catalyze the cleavage or formation of phosphodiester bonds as well as the peptide bond formation within the center of the ribosome.[274-276] However, *in vitro* selection techniques have led to the development of a wide variety of synthetic ribozymes, which are able to catalyze a much broader range of reactions, including the covalent labeling of RNA.[181-184, 385] Recently, the evolution of the first ribozyme with methyltransferase activity, named MTR1, added a whole new type of reaction to the repertoire of RNA-catalyzed reactions.[95] This finding clearly demonstrates the immense catalytic potential of ribozymes and also stimulates the search for more RNA-methylating ribozymes.

Building on this work, we report here on two different selection strategies used to obtain new methyltransferase ribozymes with improved activity or new target sites. The first strategy, which was used for a reselection of MTR1, followed a novel approach for the direct selection of active methyltransferase ribozymes. The enrichment step used the $m^6A$-sensing RNA-cleaving deoxyribozyme VMC10 to distinguish active from inactive variants.[130] Although no ribozyme with superior activity to MTR1 was obtained, the selection outcome helped to identify a structural feature crucial for the high activity of MTR1. Additionally, it provided important insights into the pitfalls of DNAzyme-based ribozyme selections. The second strategy used starting pools with cytosine or guanosine as target sites in combination with an established selection protocol to obtain ribozymes that generate methylated nucleosides other than $m^1A$. Even though no methyltransferase was obtained following this approach, both selections yielded active ribozymes, which are able to benzylate external target RNAs in high yields. Preliminary characterization, however, revealed once more adenosine as the modification site of the obtained ribozymes. Nevertheless, the selection succeeded in identifying new alkyltransferases that modify an A in a new RNA context sequence, which is only a poor substrate for MTR1. Additionally, the findings provide valuable insights into the design of *in vitro* evolution experiments and may guide the design of future selections.

## 6.1 DNAzyme based reselection of a methyltransferase ribozyme

### 6.1.1 Introduction

Methylation is an essential RNA modification that is crucial for tuning the folding, function, localization and stability of both coding and non-coding RNAs. Thereby, methylated nucleotides do not only play a structural role, but are also important regulators regarding translation and gene-expression.[386-387] In contemporary biology, most of the almost 100 different methylated nucleosides are installed post-synthetically by protein enzymes, which use SAM as the methyl group donor.[3, 54] Methylated RNA residues as well as the nucleotide derived cofactor SAM itself are viewed as ancient leftovers from a time in which life on earth was not based on DNA and proteins, but rather RNA was the primary biopolymer.[223, 272] Thus, in such a prebiotic RNA-world, also catalytically active RNAs capable of single carbon transfer reactions may have existed. The existence of such methyltransferase ribozymes would have greatly enhanced the structural and functional variety of RNA molecules. Although this theory has been discussed for a long time, only recently, the first methyltransferase ribozyme, called MTR1 was obtained by our group using the method of *in vitro* selection (chapter 3).[95] Following, two more active methyltransferase ribozymes have been reported that were obtained by either repurposing a natural riboswitch scaffold[185] or by *in vitro* evolution[186].

MTR1 uses the small molecule substrate $O^6$-methylguanine ($m^6G$) to generate an $m^1A$-modification within an external target RNA in a highly site-specific manner. The ribozyme was initially selected as an alkyltransferase with biotinylated $O^6$-benzylguanine (BG-biotin) as cofactor, which enabled the capturing of active sequences using streptavidin or neutravidin functionalized magnetic beads (chapter 3.2.1). Only afterwards it was repurposed as a methyltransferase ribozyme by changing the transferred residue from a biotin moiety to just a single methyl group (e.g., using $m^6G$ as substrate). Although MTR1 efficiently catalyzes RNA methylation reactions with this cofactor, the indirect selection method by which it was evolved is still reflected in its substrate preferences: MTR1 shows much higher activity towards $O^6$-benzylated guanine cofactors (e.g., BG-$NH_2$ and BG) compared to $m^6G$. To obtain a ribozyme that is, however, primarily optimized for RNA methylation rather than for benzylation, a direct method to select for methyltransferase activity would be desirable. The challenge, however, is the separation of methylated (active) from unmethylated (inactive) sequences, especially in the context of a ~100 nt long RNA library. As no affinity tag such as biotin is added during the ribozyme catalyzed methylation reaction, it is not possible to capture active species on functionalized solid supports, as it was done in the initial selection approach. Moreover, the additional mass of +15 (and possibly extra charge) introduced by a single methyl group is not sufficient to generate a detectable shift on denaturing PAGE. If the methylation occurs at the

Watson-Crick face, abortive RT primer extension products can be generated and used for detection of active pool members. However, polymerases often read through such modifications and only introduce mutations.[388] Thus, by only isolating the abortive RT products in the selection step, a significant amount of active species will be lost as false negatives in each round. Alternatively, immunoprecipitation with modification-specific antibodies may be considered.[182, 186] Although this is a powerful enrichment tool, antibodies are always biased: They can only be used to select for a specific, predefined modification that is recognized by the antibody. Instead, however, an unbiased selection, from which any modification can emerge, is desirable. Additionally, unspecific off-target binding as well as the high costs of antibodies are also major limitations of this approach.[389-390]

### 6.1.2 General selection strategy

Thus, we devised an alternative selection strategy based on the modification sensitive RNA-cleaving DNAzyme VMC10, which prefers unmethylated over $m^6A$-containing sequences.[130] The goal was to obtain a more optimized version for the methyltransferase reaction by directly using $m^6G$ as a cofactor for an MTR1 reselection. Using VMC10 in the enrichment step, inactive and therefore unmodified library members will be cleaved and hence show a vastly different electrophoretic mobility when resolved on PAGE. This allows easy separation of methylated from unmethylated species. However, in terms of cleavage preferences, VMC10 is not the most ideal DNA catalyst for this selection design: Although it is completely inhibited by RNA methylation, unmodified RNA is not quantitatively cleaved by VMC10, resulting in some carry-over of inactive species in each round. Rather, it would be preferable to use a DNAzyme that discriminates in the opposite direction (e.g., cleaves only modified RNA). Existing DNA enzymes with this activity are, however, not completely inhibited by the lack of methylation. They only cut modified RNA faster than unmodified RNA.[130] Therefore, VMC10 was chosen for this selection due to its superior discriminatory ability.

As VMC10 was initially selected to cleave RNA containing $m^6A$ embedded in a DRACH motif, it was first tested if the DNAzyme also accepts the sequence context of the desired selection substrate. Additionally, the product of the MTR1 catalyzed reaction is not $m^6A$, for which VMC10 was selected, but rather $m^1A$. Thus, the first step was to test whether VMC10 can also recognize and distinguish *N*1-methylated A from unmodified A. Therefore, the 13-mer $m^1A$- or A-containing RNA R1 was treated with VMC10 in a single turnover time course experiment (Figure 6.1B). Satisfyingly, even after overnight incubation, nearly complete inhibition of the cleavage reaction by $m^1A$ could be observed, while the unmodified control RNA was cleaved to 56 %. Increasing the length of the target RNA and therefore the number of base pairs

between VMC10 and the unmethylated RNA R2 (17-mer; Figure 6.1A) resulted in an even higher proportion of cleaved RNA (76 %; Figure 6.1B). Thus, VMC10 can be used to distinguish A from m$^1$A even if they are embedded in a non-DRACH motif.

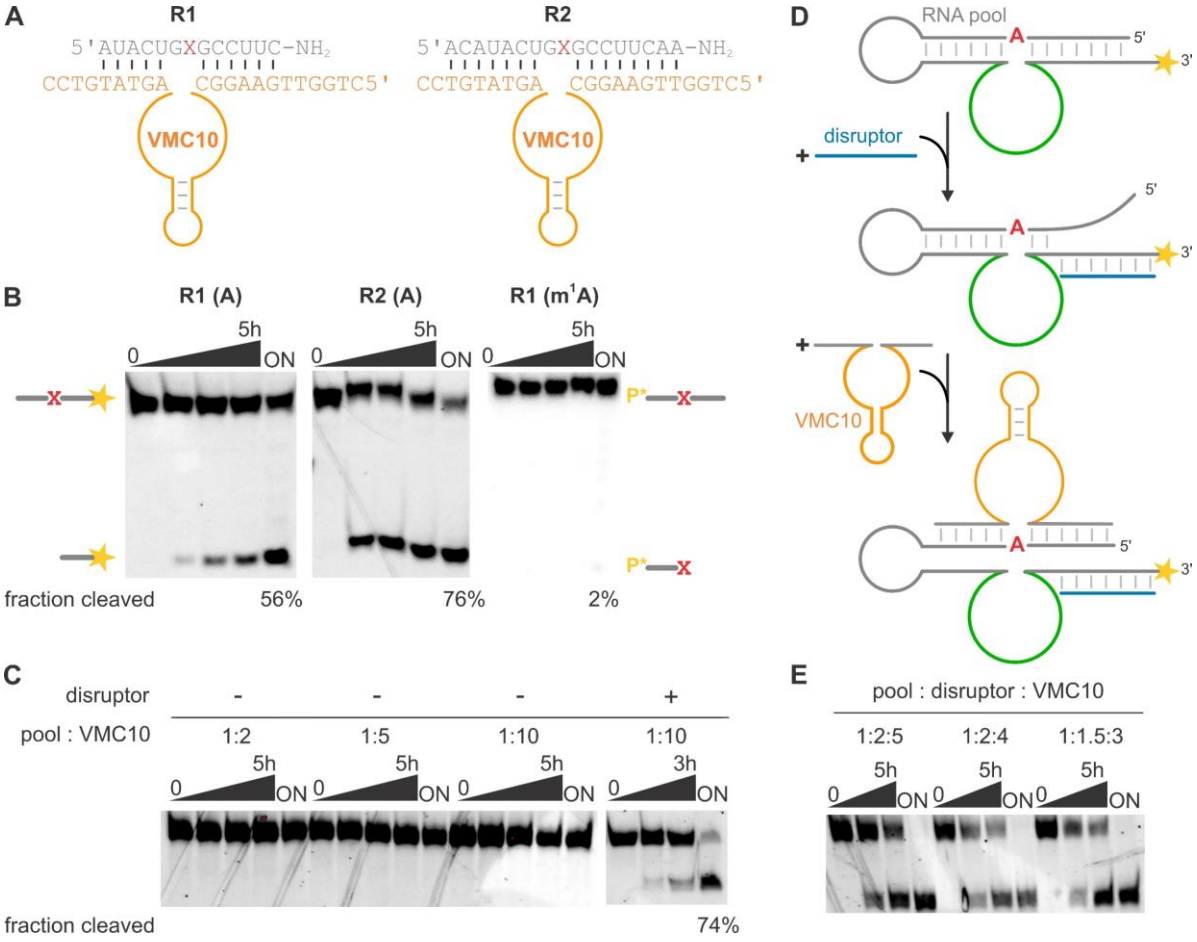



**Figure 6.1:** Cleavage of unmodified or m$^1$A containing RNA using VMC10. **A** Sequences of RNAs R1 and R2 annealed to VMC10 (X = A or m$^1$A). **B** Cleavage of single stranded unmodified RNA (R1 and R2; 3'-fluorescein-labeled) and m$^1$A-containing RNA (R1; 5'-$^{32}$P-labeled) using a 10-fold excess of VMC10 over RNA. Timepoints: 0, 1h, 3h, 5h, ON. **C** Cleavage of *cis*-active ribozyme pool with different ratios of pool to VMC10. In the reaction in which a disruptor oligonucleotide was present (+), the ratio of pool:disruptor:VMC10 was 1:2.5:10. Timepoints: 0, 1h, 3h, 5h, ON (without disruptor) or 0, 1h, 3h, ON (with disruptor). **D** Schematic depiction of annealing sites of disruptor DNA (blue) and VMC10 (orange) within the ribozyme pool. **E** VMC10 cleavage assay of *cis*-RNA using different ratios of disruptor DNA and VMC10 over the RNA pool. Timepoints: 0, 2h, 5h, ON.

However, when applied to a full-length *cis*-reactive ribozyme pool, no cleavage could be observed, even when a tenfold excess of VMC10 over the targeted RNA was used (Figure 6.1B). This is most likely because the target RNA is heavily base-paired to the ribozymes binding arms and therefore not accessible for the DNAzyme. To circumvent this issue and facilitate efficient annealing of VMC10, a disruptor oligo was added to the annealing mixture. This short DNA binds to the 3' binding arm of the RNA pool, which weakens the ribozyme-target RNA complex and enables binding of VMC10 (Figure 6.1D). With this setup, 74 % cleaved RNA was obtained. However, 25 pmol of the disruptor DNA and 100 pmol of VMC10 were used to cleave 10 pmol of *cis*-reactive RNA. When scaled up to approximately 250-500 pmol of RNA pool, which is the minimum amount of library required to initiate the selection, this ratio is

unrealistic and thus was further optimized (Figure 6.1E). Using a ratio of pool:disruptor:VMC10 = 1:1.5:3, quantitative cleavage of the RNA was achieved after overnight incubation in a small-scale test reaction. Thus, this ratio was then used to initiate the methyltransferase reselection.

### 6.1.3 Selections CD - CG

A total of four different reselections were performed, all following the same VMC10-based enrichment strategy, but using different RNA library designs. Each of the starting pools was pre-structured and based on the previously selected CA13 and CA21 ribozymes, interspersed with a stretch of random RNA sequences (Figure 6.2): While for the CD selection the initial stem-loop structure was only partially replaced by an $N_{16}$ random region, it was entirely substituted by an $N_{10}$ or $N_{20}$ insert in selections CE and CF, respectively. In CG, the starting pool was designed to contain a 4 bp long stem-loop structure, while an $N_8$ sequence stretch was inserted directly within the 3' flanking region of the catalytic core. With these eight random nucleotides, the CG starting pool has the same length as the libraries CD and CF, which later simplifies the isolation of the cleaved product on PAGE during the selection step. The insertion site was chosen based on previous SHAPE- and DMS-probing results (Figure 9.3), which revealed that the nucleotides in this region are highly protected and may thus contribute to the formation of the ribozymes active site. Insertion of a random region at this position may therefore facilitate the generation of an alternative, improved substrate binding site, thereby accelerating the methyltransferase reaction.

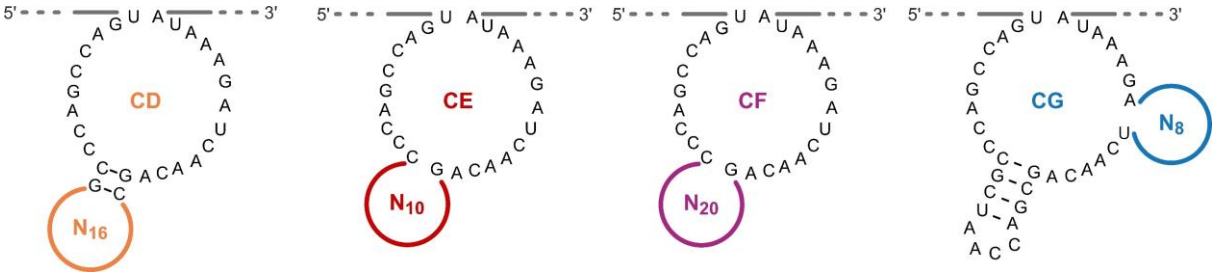



**Figure 6.2:** Starting pool design for selections CD-CG, which are based on ribozymes CA13 and CA21.

Thus, with random regions of $N_8$ - $N_{20}$, the DNA templates used for T7 *in vitro* transcription of the library had a theoretical complexity of $4^8 = 6.5 \times 10^4$ (CG) up to $4^{20} = 1.1 \times 10^{12}$ (CF) unique RNA species. However, by submitting 300 pmol dsDNA template to each transcription reaction, a maximum of $1.8 \times 10^{14}$ unique RNA sequences can be expected, which is already larger than the entire sequence space of the $N_{20}$ CF-pool. In the first selection round, 400 pmol of each pool were subjected to the self-methylation reaction. Consequently, in case of the $N_{20}$ random region, every variant should statistically be present more than 200 times within the reaction mixture.

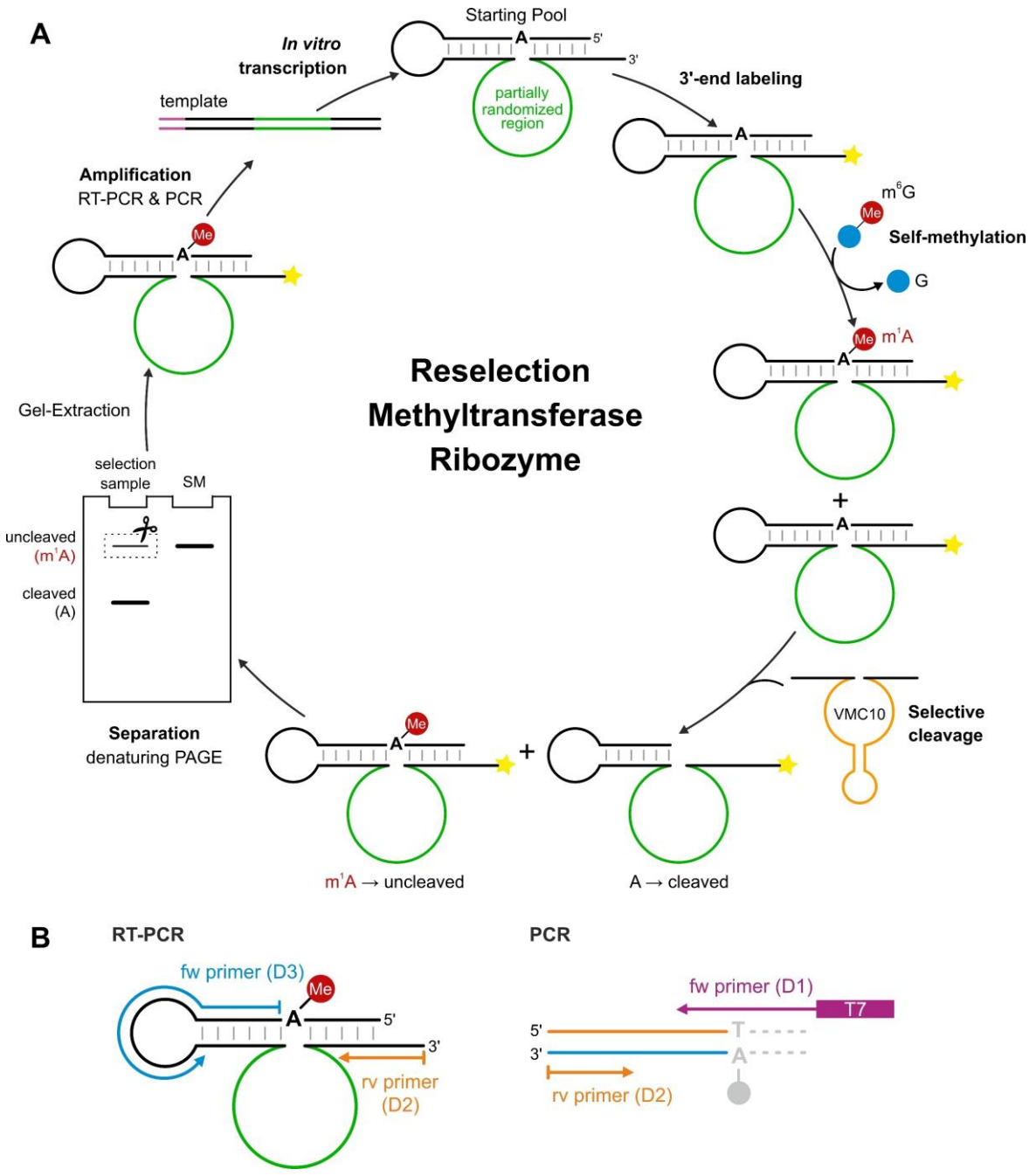



**Figure 6.3:** Schematic overview of the DNAzyme-based selection setup for reselection of a methyltransferase ribozyme. **A** The selection cycle starts with incubation of the library containing a partially randomized region (green) with the cofactor $m^6G$. Active variants catalyze the self-methylation reaction (e.g., attachment of a methyl group, red) under release of guanine (blue). Unmodified variants are then cleaved by VMC10 (orange). After separation on PAGE, uncleaved RNA is isolated, amplified using RT-PCR, PCR and *in vitro* transcription and submitted to the next round of selection. **B** Primer design for amplification of the enriched pool by RT-PCR followed by PCR. The reverse primer (orange) is identical for both reactions and extents until it encounters the methylated A. The forward primer of the RT-PCR (blue) anneals directly 3' of the desired methylation site. After RT-PCR, the missing 5' part of the substrate sequence is added by the help of another forward primer (purple) in a second PCR, which also introduces a T7-promotor for subsequent *in vitro* transcription.

To facilitate easy visualization of even small amounts of the RNA libraries when they are resolved on denaturing PAGE, the starting pools were first fluorescently labeled at the 3'-end using periodate glycol oxidation. After ON incubation at 37 °C in the presence of 100 µM $m^6G$ and 40 mM $MgCl_2$, the RNA pools were precipitated from the reaction mixture and subjected

to the VMC10-catalyzed reaction. Inactive variants were cleaved followed by separation from the active library members on denaturing PAGE. The intact RNA fraction was recovered, followed by reverse transcription and amplification by RT-PCR. Subsequently, the DNA thus generated was further amplified in a second PCR, in which also a T7-promotor was added. After the following *in vitro* transcription, the enriched pool was submitted to the next round of selection. A schematic overview of the selection cycle is provided in Figure 6.3.

#### 6.1.3.1 Selection progress

The CD-CG selections were carried out for a total of eight rounds. The conditions for the self-methylation reaction were kept constant throughout the entire selection process. Solely the amount of input RNA was reduced to 350 pmol from the second round and further decreased to just 300 pmol starting from round four. An overview of the enrichment observed in each round is provided in Figure 6.4A and Table 9.16.

Already in the first round, a band for the full-length RNA (5-10 %), representing the active fraction, was detectable for each of the four selections (Figure 9.25A). Therefore, active methyltransferase ribozymes were most likely already present at high abundance within each of the partially randomized starting pools. Over the following rounds, the enrichment level kept increasing steadily until 40-50 % of the library members were resistant to VMC10 mediated cleavage in round four. Subsequently, to remove non-truly active variants from the libraries a negative selection round was performed: The pools were directly subjected to the VMC10 reaction, without performing a self-methylation beforehand. The cleaved fraction, which made up only ~50 % of the pool, was isolated and further amplified (Figure 9.25B). RNA variants that remained intact in this setup presumably withstand the cleavage reaction for reasons other than self-methylation activity. For example, they may form a stable secondary structure that prevents efficient annealing of VMC10. As expected, the negative selection led to a sudden drop in the amount of uncleaved RNA. Until round eight, however, the enrichment kept increasing again until it reached ~45 % uncleaved RNA for selections CD and CG and even up to >55 % for selections CE and CF.

After seven rounds, an RT primer extension assay was performed to test the pools (7CD - 7CG) for activity in *cis*. However, this experiment only revealed substantial methyltransferase activity for the enriched library of selection CD. All other pools showed hardly any detectable abortive extension products (Figure 6.4B). Additionally, the methyltransferase activity of the CD pools from rounds five (5CD) and eight (8CD) was comparable. Hence, the selection was stopped after the eighth round and only the enriched library from selection CD was subjected to cloning and sequencing.

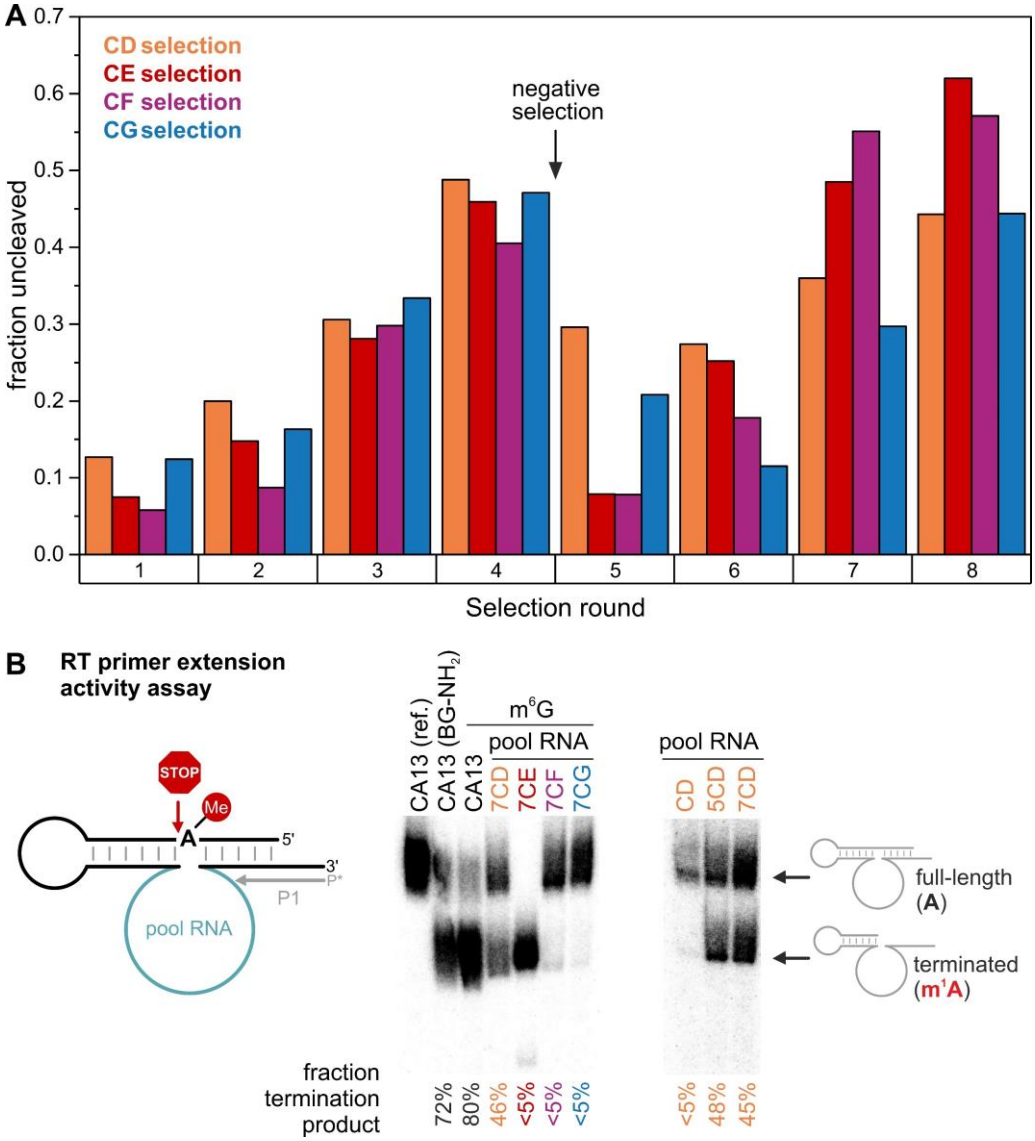



**Figure 6.4:** Progress of selections CD-CG. **A** Overview of the enrichment. Shown is the proportion of uncleaved RNA for each selection round, which was subsequently isolated and further amplified. A negative selection to remove non-truly active species from the libraries was performed between round four and five (arrow). The source data of the enrichment is provided in Table 9.16. **B** Activity assays of all enriched pools from round 7 (7CD - 7CG) as well as CD pools from round 0 (CD), 5 (5CD) and 7 (7CD) using RT primer extension. Following ON incubation in the presence of $m^6G$ under selection conditions, 5'-$^{32}$P-labeled primer P1 was annealed to the 3' binding arm and extended by the reverse transcriptase. Active methyltransferase ribozymes contain an $m^1A$ that causes an extension block resulting in a shorter terminated cDNA band. The active ribozyme CA13 that was either untreated (ref.), reacted with $BG\text{-}NH_2$ or $m^6G$ was included as positive reference. Note that the pool of CE is shorter than the other libraries.

### 6.1.3.2 Sanger sequencing results of CD selection

The plasmids of ten randomly selected colonies were submitted for Sanger sequencing. A total of nine clones were sequenced successfully. All sequences obtained were unique. The results are summarized in Table 6.1.

**Table 6.1:** Sequencing results obtained from the CD selection. Shown are the ribozymes binding arms (green), the constant flanking region of the catalytic core (blue) and the fully randomized $N_{16}$ region (orange). Base paired nucleosides forming the ribozymes stem-loop structure (according to secondary structure prediction, Figure 9.26) are underlined. Mutations that occurred in constant regions are marked in light blue.

| ribozyme | 5'-sequence-3' |
|---|---|
| CD pool | UUGAAGGC UGACCGACCCG NNNNNNNNNNNNNNNN CGACAACUAGAAAUA CAGUAUGUCC |
| CA13 | UUGAAGGC UGACCGACCCC CCGACCCUUCUCUG GGACAACUAGACAUA CAGUAUGUCC |
| CA21 | UUGAAGGC UGACCGACCCG CUACAAUACAACAG CGACAACUAGAAAUA CAGUAUGUCC |
| MTR1 | UUGAAGGC UGACCGACCCC CCGAGUUCGCUCGG GGACAACUAGACAUA CAGUAUGUCC |
| CD2 | UUGAAGAC UGACCGACCCG CCACCAAUAUAAAAAACG CGACAACUAGAAAUA CAGUAUGUCC |
| CD5 | UUGAAGAC UGACCGACCCG GCACUACAGCUCGUUC CGACAACUAGACAUA CAGUAUGUCC |
| CD8 | UUGAAGAC UGACCGACCCG CCCGCGAGACAAAGCG CGACAACUAGACAUA CAGUAUGUCC |
| CD10 | UUGAANAC UGACCGACCCG CUCCUCUUAAAACCCG CGACAACUANAAAUA CNGUANGUCC |
| CD28 | UUAAAGGC UGACCAACCGA AGGCCCAGCGCCGCC CGACAACUAGAAAUA CAGUAUGUCC |
| CD29 | UUGAAGGC UGACCGACCCG GGGGCUCAGACACGUC CGACAACUAAAAAUA CAGUAUGUCC |
| CD30 | UUGAAGAC UGACCGACCCG UACAAAAAACCCAAAUGA CGACAACUAGAAAUA CAGUAUGUCC |
| CD32 | UUGAAGGC UGACCGACCCG UUACGUCCGAGCGCAG CGACAACUAGAAAUA CAGUAUGUCC |
| CD40 | UUGAAGAC UGACCGACCCG GACCAAACUCACAGAC CGACAACUAGAAAUA CAGUAUGUCC |

Remarkably, six clones shared the same G-to-A mutation in their 5' binding arm that changes a G-C Watson-Crick base pair to an A-C mismatch. Two of these sequences (CD5 and CD8) exhibited an additional A-to-C mutation in the 3' flanking region of the catalytic core, which can also be found in CA13 and MTR1. Interestingly, the length of the random sequence insert shrank from initially 16 nt to 15 nt (in CD28) or expanded up to 18 nt (in CD2 and CD30). Despite the lack of sequence complementarity within this region, a common feature shared by six out of the nine variants was the extension of the stem-loop structure from initial 2 bp to at least 3 bp. This clearly indicates the importance of a stable stem for the methyltransferase activity. Strikingly, in CD5 and CD30, the stem was elongated to even 5 bp, therefore closely resembling the architecture of the original CA13 ribozyme (Figure 9.26). An even more stable stem-loop with a length of 7 bp was found in CD32, further hinting towards the importance of this structural motive.

#### 6.1.3.3 Activity assay

All ribozyme variants were first tested for methyltransferase activity in *cis*. Following overnight incubation at 37 °C and pH 7.5 in the presence of 100 µM $m^6G$, an RT primer extension was performed to read out the individual activities. Using this assay, successful self-methylation

was detected for the majority of the variants tested, albeit to varying degrees. The fraction of methylated RNA for each of the ribozymes was quantified by comparing the intensity of the full-length and terminated product bands (Figure 6.5A). However, it should be noted that this method may underestimate the true activity of the ribozymes, since the reverse transcriptase is not completely blocked by $m^1A$ modifications but may also read through some of the methylated nucleosides. Nevertheless, by including the active methyltransferases CA13 and CA21 as positive references, a reliable estimate of the *cis*-activity of individual ribozymes can still be obtained.

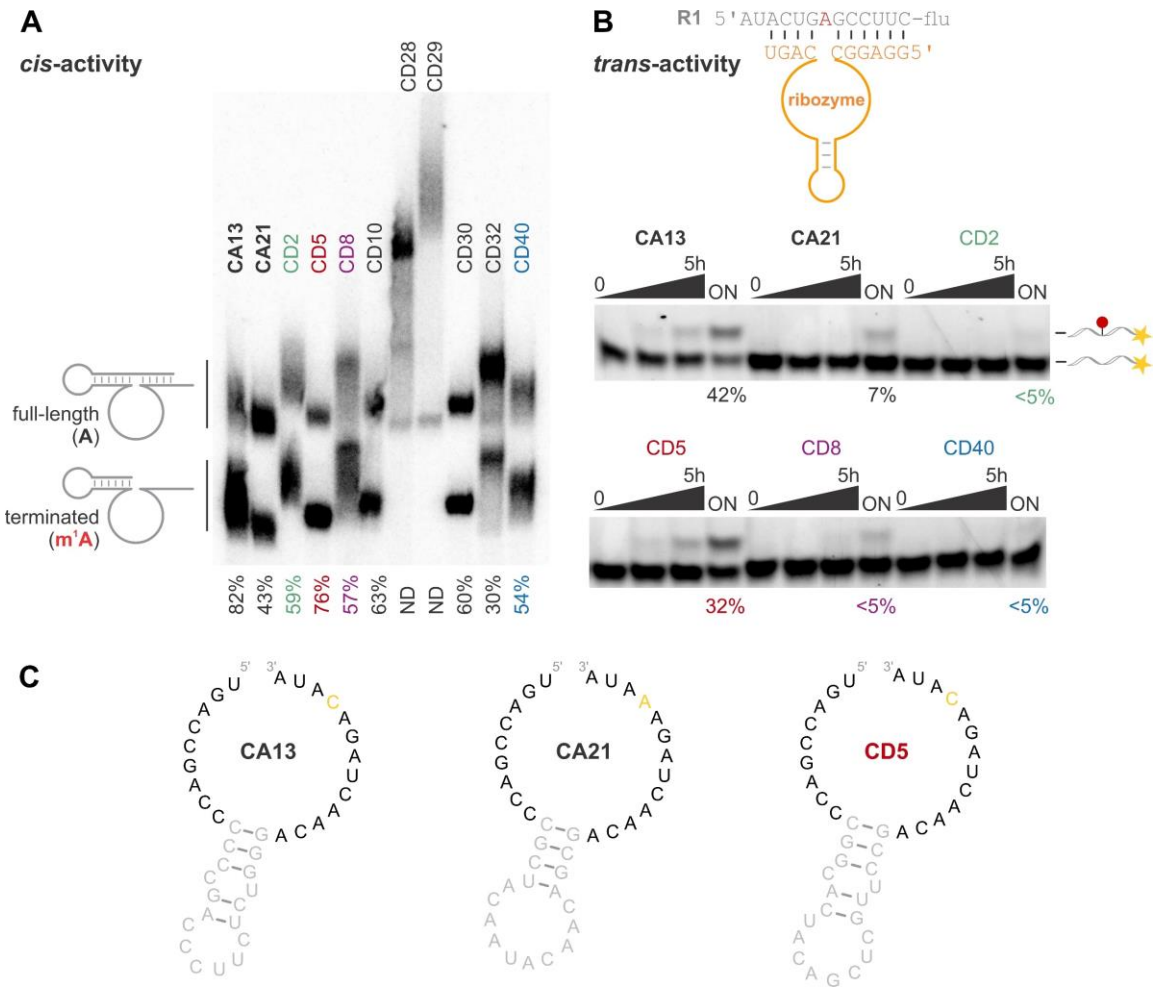



**Figure 6.5:** Methyltransferase activity assays of individual CD ribozymes together with CA13 and CA21 as positive references. **A** *Cis*-activity assay by RT primer extension using 5'-$^{32}$P-labeled primer P1 for readout. Ribozymes were incubated ON at 37 °C in the presence of 100 µM $m^6G$ beforehand. The proportion of the termination product band was quantified to obtain an estimate of the activity. **B** *Trans*-activity assay for methylation of 3'-fluorescein labeled RNA R1. The ribozymes were transcribed from ssDNA templates. Timepoints: 0, 2h, 5h, ON. **C** Sequences and secondary structure predication of ribozymes CA13, CA21 and CD5. The mutated nucleoside in the 3' flanking region is shown in yellow, the stem-loop structure in gray and identical sequences in the catalytic core are depicted black. The secondary structures were predicted using the RNAfold web server.[391]

Among all CD ribozymes obtained, CD5 was the most active methyltransferase with an activity comparable to that of CA13 (76 % and 82 %, respectively). For five more variants (CD2, CD8, CD10, CD30 and CD40) the primer extension resulted in approximately 50-60 % truncated

product. Analysis of CD32, however only yielded 30 % abortive DNA. Finally, primer extension on self-methylated CD28 and CD29 resulted in the least intense termination product band. Additionally, the extension products for the latter ribozymes were much longer than expected and the gel image was very smeary, exacerbating a proper analysis. The reason for such behavior could be an unusually stable secondary structure interfering with primer extension.

Next, four of the most active variants were tested for reaction in *trans* (e.g., methylation of an external substrate RNA in a bimolecular fashion) and the activity was compared to CA13 and CA21. The substrate RNA and the connecting loop were removed and the 5'- and 3'-binding arms were shortened to modify an external 13-mer substrate RNA (R1). With 32 % methylated RNA after overnight incubation, only CD5 showed true methyltransferase activity, whereas CD2, CD8 and CD40 yielded <5 % RNA product (Figure 6.5B). Under identical reaction conditions, however, CA13 was able to generate 42 % $m^1A$-containing RNA. Consequently, no ribozyme with higher methyltransferase activity than the initial CA13 ribozyme could be obtained.

Nevertheless, some conclusions for the structure-activity relationship of CA13-derived methyltransferase ribozymes can be drawn from the results presented above: First, the sequencing results clearly hint towards the importance of a stable stem-loop structure for efficient catalysis. This is further supported by the finding that - according to secondary structure predictions - the most active variant (CD5) has the longest and therefore most stable stem among all CD ribozymes tested for *trans*-activity. Overall, the stem-loop of CD5 has high structural analogy to the similarly active predecessor ribozyme CA13 (Figure 6.5C, Figure 9.26). Moreover, further stabilization of the stem later resulted in generation of MTR1 (chapter 3.2.1), the most active methyltransferase ribozyme from this group to date. The exact sequence and length of the loop, however, seem to be insignificant for the activity, as all *cis*-active ribozymes are highly diverse within this region (Figure 9.26).

## 6.1.4 CH- and CI-selection

### 6.1.4.1 Starting pool design and selection strategy

In the search for a highly active methyltransferase ribozyme, an improved selection strategy was devised taking into account the results obtained from the first set of reselections (CD - CG). Regarding the starting pools, the previous designs may not have been ideal. By simply introducing random nucleotides within the stem-loop structure in selections CD-CF, the RNA sequence that actually forms the catalytic center remained unchanged. Since the exact sequence of the stem-loop proofed to be irrelevant for catalysis, no improvement in methyltransferase activity could be achieved with this library design. In the case of selection CG, insertion of the $N_8$ region close to the 3'-end of the catalytic core may have disrupted an essential structural element, indispensable for its methyltransferase activity. This is further supported by the crystal structure of MTR1: the $N_8$ region was inserted between A37 and U36, directly in the catalytic core of the ribozyme. A37 is part of the adenosine triple located on top of the active site. Thus, correct folding of this layer is probably compromised in the CG starting pool, which likely negatively effects cofactor binding and catalysis of the ribozyme.

Therefore, another strategy regarding the starting pool design was pursued for the next reselections: The stem-loop structure was kept constant with three G-C base pairs in the stem and a 6 nt long loop, while the remaining catalytic core sequence was randomized (Figure 6.6). For the CH selection, a doped library was used, which was generated based on the CA13 ribozyme. It contained the wild-type CA13 sequence with a frequency of 76 %, while the other three nucleotides were present at a frequency of 8 % each. With a total of 22 partially randomized nucleotides, 80 % of the pool should theoretically contain between three and seven mutations relative to the parent ribozyme. Sequences exhibiting five mutated nucleosides are expected to be the most abundant ones within the starting library (20 %).[392] A detailed calculation of the extend of pool randomization is provided in Figure 9.27 and Table 9.15. For selection CI, the 22 nucleotides were completely randomized yielding a theoretical complexity of $4^{22} = 1.8 \times 10^{13}$ unique variants. By submitting 400 pmol of RNA to the first selection round, every sequence should statistically be present more than ten times in this selection.

Another issue of the previous selections was the carry-over, and therefore enrichment, of non-truly active library members. The sudden drop in the amount of cleaved fraction that was observed after the negative selection round suggests that a large portion of the pool could withstand the VMC10 mediated cleavage, even though it lacked self-methylation activity. Most likely, some sequences folded into a complex, stable secondary structure that prevented annealing of the DNAzyme or inhibited its cleavage reaction. To avoid this issue in the new

reselections, negative selections were performed after each of the first three rounds and when a peak in enrichment was encountered.

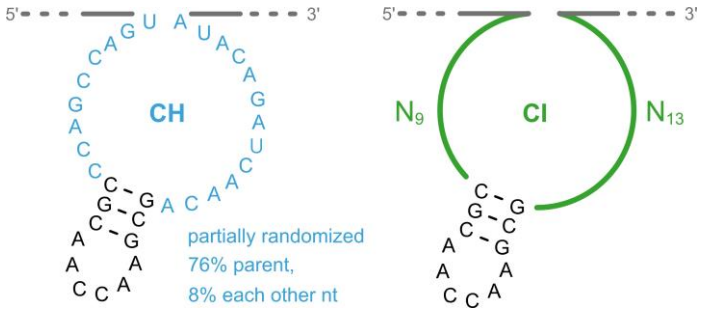



**Figure 6.6:** Starting pool designs for selections CH and CI based on the outcome of the previous reselection CD. While the pool for the CH selection contains 22 partially randomized nucleosides, those nucleosides were fully randomized for selection CI.

### 6.1.4.2 Selection progress

In the first selection round, 400 pmol of each starting library containing 100 pmol 3'-fluorescein-labeled pool were submitted to the self-methylation reaction. The incubation was performed under the same reaction conditions that were used for selections CH-CI (100 µM $m^6G$, 40 mM $MgCl_2$, pH 7.5, 37 °C, overnight incubation). An overview of the selection progress is provided in Figure 6.7A and Table 9.17.

In the case of selection CH, the fraction of uncleaved RNA remained below 10 % during the first four rounds, which were constantly alternated with negative selections. Starting from round five, the enrichment level began to rise to about 15 %, further increasing to 30 % in the sixth round. In contrast to the previous reselections, more stringent reaction conditions were applied in the following rounds to select for the most efficient catalysts: The incubation time was reduced from overnight to 6 h in rounds seven and eight and further decreased to 4 h in the following two rounds. When an enrichment of >50 % was reached after the seventh round, an additional round of negative selection was performed, which reduced the uncleaved fraction to 26 % in the subsequent round. This indicated that, in spite of the many initial negative selections, a large proportion of the pool still consisted of inactive sequences, which were resistant to VMC10 cleavage. The selection was then further continued for a total of ten rounds, during which negative selections were again performed regularly. For the CI selection, the first six rounds were performed analogously to selection CH. However, the first enrichment was not detected until the sixth round, when the amount of uncleaved RNA spiked from 2 % to 43 %. Despite altering negative and positive selections in the following rounds and decreasing the reaction time in round eight from overnight to 6 h, the enrichment level remained constant at around 30-40 %. Thus, the selection was stopped after a total of eight rounds.

Throughout the selection process, regular RT primer extension assays were performed to determine the methyltransferase activity of the enriched pools (Figure 6.7B). The incubation conditions for the methyltransferase reaction were identical to those used in the respective selection rounds. Promisingly, in the case of selection CI, >40 % of the RNA pool was uncleaved after rounds six and eight. However, RT primer extension assays showed no detectable methyltransferase activity for these pools. Hence, no further research was conducted on this selection. For selection CH, a low activity of 4 % was detectable for the pool of round six as indicated by a weak abortive RT product band. However, even though the selection pressure was subsequently enhanced, the actual methyltransferase activity determined by primer extension showed no further increase. The fraction of uncleaved RNA in the last round was among the highest detected during the entire selection process. The corresponding RT assay, however, only resulted in 3 % abortive product. Therefore, the round ten pool is expected to contain a large proportion of inactive RNAs. Consequently, only the enriched libraries of round six and eight (hereafter named 6CH and 8CH, respectively) were cloned and submitted to Sanger sequencing.

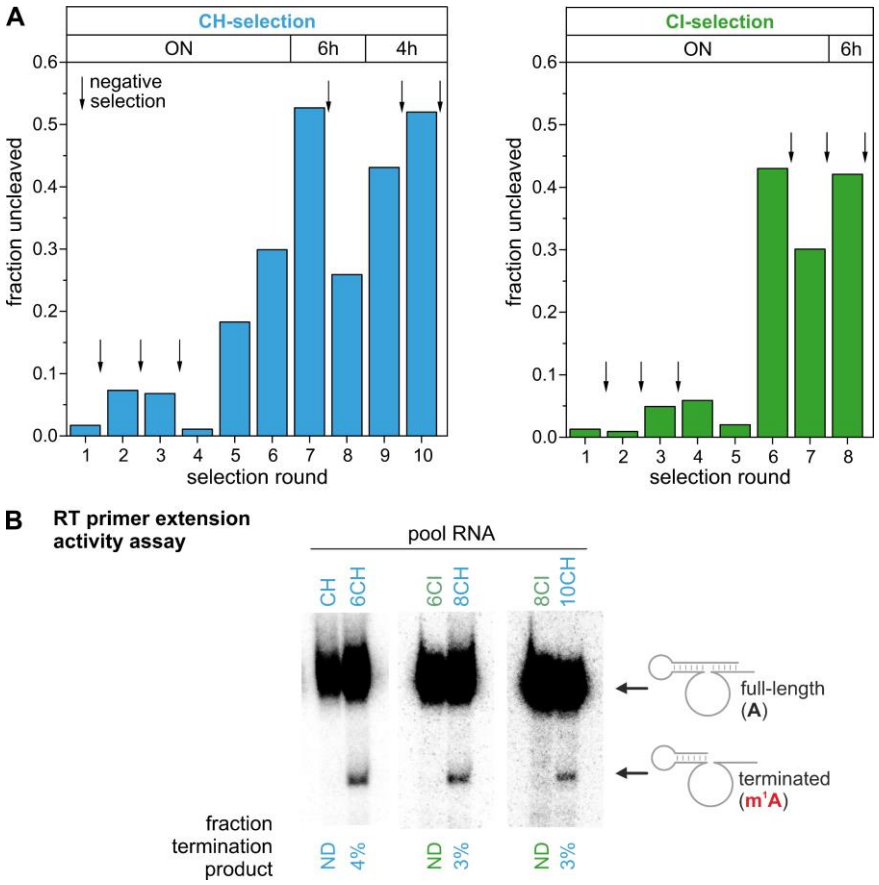



**Figure 6.7:** Progress of selections CH-CI. **A** Overview of the enrichment and selection parameters. Depicted is the proportion of uncleaved RNA for each selection round, which was isolated and further amplified. Arrows indicate negative selection rounds to remove non-truly active species from the libraries. The source data for the enrichment levels is provided in Table 9.17. **B** Activity assay of the CH pools of round 0 (CH), 6 (6CH), 8 (8CH) and 10 (10CH) and for the CI pools of round 6 (6CI) and 10 (10CI) using RT primer extension. The assay was performed as depicted and described in Figure 6.4 using the reaction conditions of the respective selection round.

### 6.1.4.3 Sanger sequencing results of CH selection

For both selection pools, each of the eleven randomly selected clones submitted were sequenced successfully. However, the identity of some nucleotides within four of the sequences could not be detected unambiguously. This can for example be due to contamination with other plasmids or due to poor DNA quality. The sequencing results are summarized in Table 6.2.

**Table 6.2:** Sequencing results obtained from round 6 (6CHX) and round 8 (8CHX) of the CH selection together with a possible consensus sequence. Shown are the ribozymes binding arms (green), the partially randomized region (blue) and the constant stem-loop (orange). Base-paired nucleotides in the pool, in CA13 and in MTR1 are underlined. Mutations that occurred in the partially randomized region are indicated in green (A), blue (C), gray (G) and red (U). Mutations of constant regions and nucleotides undetectable by sanger sequencing are marked yellow.

| ribozyme | 5'-sequence-3' |
|---|---|
| CH pool | UUGAAGGC UGACCGACC CGCAACCAAGCG ACAACUAGACAUA CAGUAUGUCC |
| CA13 | UUGAAGGC UGACCGACCCCCCGACCCUUCUCUGGGACAACUAGACAUA CAGUAUGUCC |
| MTR1 | UUGAAGGC UGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUA CAGUAUGUCC |
| 6CH3 | UUGAUGGC UGAUUGACC CGCAACCAAGCG AAAACUCGACAGA CAGUAUGUCC |
| 6CH4 | UUGAAGGC UCAGUGAUC CGCAACCAAGCG ACAACUGGCCAUG CAGUAUGUCC |
| 6CH6 | UUGUAGGC UGACGGACC CGCAACCAAGCG ACAACUAGUCAUA CAGUAUGUCC |
| 6CH8 | UUGUAGGC UUACCGACC CGCAACCAAGCG ACACAUGGACAGA CAGUAUGUCC |
| 6CH11 | UUGAAGGC CGAUCGUUC CGCAACCAAGCG ACAACUAGAACAGACAGUAUGUCC |
| 6CH15 | UUGAAGGC UCAGCGUCA CGCAACCAAGCG ACUACUAGACCGA CAGUAUGUCC |
| 6CH16 | UUGAAGGC GGACCGCCC CGCAACCAAGCG ACAACUAGUUGCA CAGUAUGUCC |
| 6CH18 | UUGAAGAC UGACCGAUC CGCAGCCAAGCG ACAACUAGACAUA CAGUAUGUCC |
| 6CH21 | UUGAAGGC GSGCCGMCC CGCAACCAAGCG ACAWCUCGACMWA CAGUAUGUCC |
| 6CH31 | UUGAUGGC AGACUGGGC CGCAACCAAGCG AAAACUAAAGAUA CAGUAUGUCC |
| 6CH35 | UUGAAGGC GGGCCGAUC CGCAACCAAGCG AUGACUAGUCUCA CAGUAUGUCC |
| 8CH1 | UUGAAGGC UGAGUGGCU CGCAACCAAGCG ACAACCGGACAUA CAGUAUGUCC |
| 8CH2 | UUGAAGGC UUAGUGUGC CGCAACCAAGCG ACAGCUGGCCGCA CAGUAUGUCC |
| 8CH3 | UUGAAGGC CGGNAUGCC CGCAACCAANCG ACACCNGCNCAGA CNGUAUGUCC |
| 8CH4 | UUGAAGGC UCACCGMYC CGCAACCAAGCG ACUAUUCGACCUG CAGUAUGUCC |
| 8CH17 | UUGAAGGC UCACUGACC CGCAACCAAGGG ACAUCCGCACAUG CAGUAUGUCC |
| 8CH20 | UUGAAGGC UCACCCGUU CGCAACCAAGCG ACAGCUACACAGA CAGUAUGUCC |
| 8CH25 | UUGACGGC UGCCCGACG CGCAACCAAGCG CCAACUAGCCCUG CAGUAUGUCC |
| 8CH30 | UUGAAGGC UCAGCCAAA CGCAACCAAGCG ACAACUUGACCUG CAGUAUGUCC |
| 8CH32 | UUGAAGGC UUAACGUUC CGCAACCAAGCG ACGAGUAGACAUG CAGUAUGUCC |
| 8CH35 | UUGAAGGC UCAUCGCCA CGCAACCAAGCG ACACCAAGACGUG CAGUAUGUCC |
| 8CH36 | UUGAAGGC UUAGCKRUC CGCAACCAAGCG ACAACUGGGUMUC CAGUAUGUCC |
| consensus | U.A.CG..C ACAACU.GAC... |

Analysis of the sequencing data reveals 22 unique sequences. Similar to the CD ribozyme selection, six different variants exhibited a mutation in the 5' binding arm, while the 3' binding arm remained unchanged for all sequences, which is probably because it also serves as primer binding site in the amplification step. Moreover, the stem-loop, which was kept constant in the starting pool did also not encounter any major mutations that would disrupt this structural motif. Regarding the 22 nt long partially randomized sequence, the number of mutations that appeared relative to CA13 varied drastically from only one point mutation (in 6CH18) up to ten mutations (in 8CH2). None of the obtained sequences was completely identical to the parent ribozyme. Additionally, some nucleosides showed higher mutation rates than others. For instance, the adenosine immediately 3' of the stem-loop was mutated in only a single variant (8CH25), suggesting that this residue plays an essential role in catalysis. In contrast, the cytosine at position four within the catalytic core was changed to any of the other nucleosides in a total of ten different variants. Due to the large variety of sequences obtained from this selection, it is challenging to derive a conclusive consensus sequence. However, some positions show low mutation rates and therefore seem to be conserved, allowing the deduction of at least a partial consensus motif (Table 6.2; nucleotides with mutation rates <35 % were included in the consensus sequence).

#### 6.1.4.4 Activity assay

As no ribozyme families could be determined from the sequencing results, each individual RNA sequence was subsequently tested for methyltransferase activity in *cis*. After overnight incubation with $m^6G$ at 37 °C, the methylation status was read out using RT primer extension with primer P1, which anneals to the 3' binding arm of the ribozyme. Unfortunately, only three out of all 22 sequences tested showed detectable activity: 6CH18 was the most active ribozyme with 85 % abortive extension product followed by 8CH36 with 16 % and 6CH2 with <5 % (Figure 6.8A). For all other RNA sequences, the primer extension assays only revealed the full-length DNA as well as some overextended products.

Next, the two most active CH ribozymes were tested for *trans*-activity and compared to the active methyltransferases CA13 and MTR1. Therefore, 6CH18 and 8CH36 were both transcribed from a ssDNA template encoding the catalytic core flanked by the binding arms that were designed to methylate a 3'-fluorescein-labeled 17-mer substrate RNA (R1). In the case of 6CH18, the G-to-A mutation in the 5' binding arm was restored to achieve full complementarity between the ribozyme and the target sequence. For 8CH36, the most probable sequence according to the Sanger chromatogram trace file was used, as it was not fully resolved during sequencing (Table 9.13). While no methylated product was detected for the 8CH36 ribozyme,

at least 8 % was obtained after overnight incubation with 6CH18 (Figure 6.8B). In comparison, MTR1 yielded 79 % methylated RNA product under identical reaction conditions (100 µM $m^6G$, 40 mM $MgCl_2$, pH 7.5, 37 °C). While no meaningful secondary structure could be predicted for 8CH36, analysis of the 6CH18 fold reveals high structural similarity to CA13 (Figure 6.8B). However, this was to be expected as both ribozymes only differ by a single point mutation which proves to be crucial for the activity. In summary, the nucleotides in the catalytic core of CA13 outside the stem-loop structure appear to be highly conserved, as almost all mutations within this region rendered the ribozymes completely inactive. This selection outcome indicates that the active site sequence of CA13 may already fold into a structure that is very well suited for the transfer of a methyl group from $m^6G$ to the *N*1 position of adenosine.

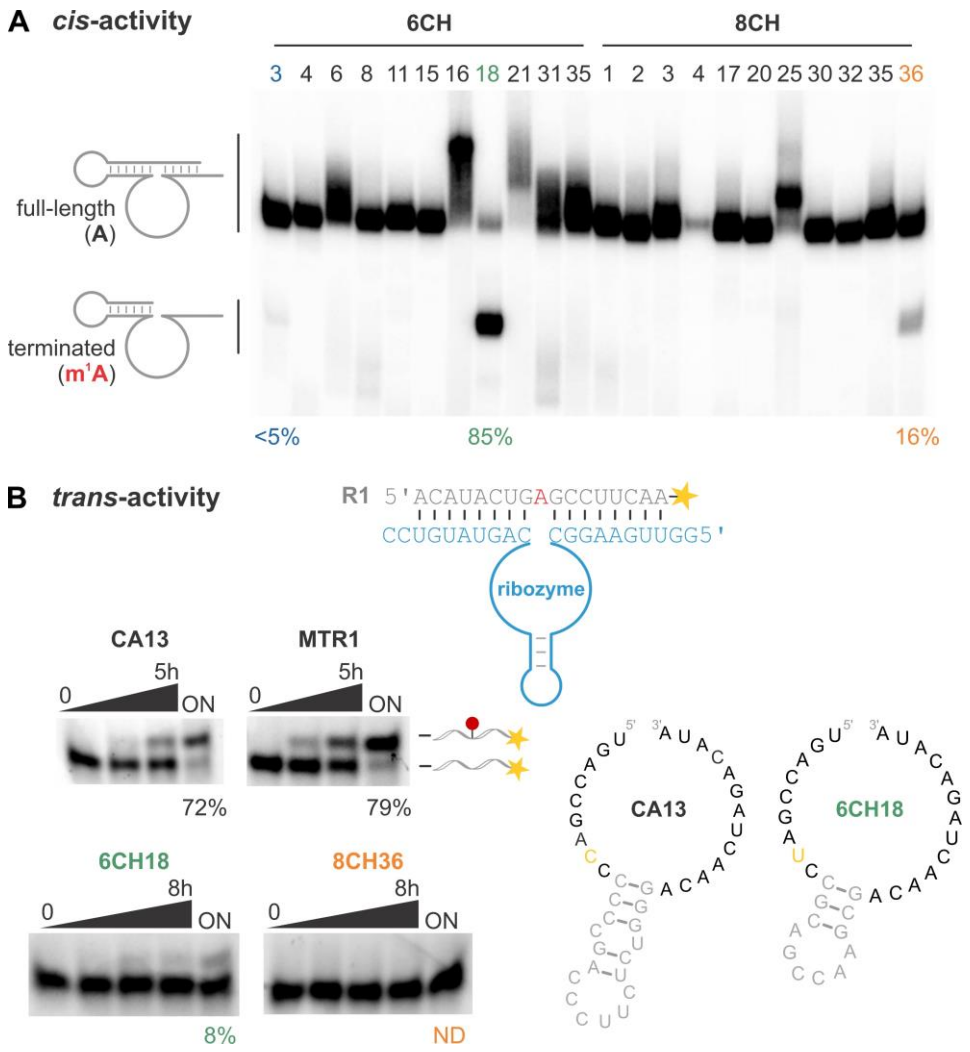



**Figure 6.8:** Methyltransferase activity assays of individual CH ribozyme variants. **A** *Cis*-activity assay by RT primer extension using 5'-$^{32}$P-labeled primer P1 for readout. The reaction was performed after ON-incubation of the ribozymes at 37 °C in the presence of 100 µM $m^6G$. The amount of termination product band was quantified and used as an indication for the methyltransferase activity. **B** *Trans*-activity assay for methylation of 3'-fluorescein labeled RNA R2. The ribozymes were transcribed from ssDNA templates. Timepoints: 0, 2h, 5h, ON for CA13 and MTR1; 0, 2h, 5h, 8h, ON for 6CH18 and 8CH36. Sequences and secondary structure predications of ribozymes CA13 and 6CH18 reveal high similarities. The mutated nucleoside is shown in yellow and the stem-loop structure in gray. Identical sequences in the catalytic core are depicted black. The secondary structures were predicted using the RNAfold web server.[391]

### 6.1.5 Discussion

In summary, both VMC10-based reselections failed to yield a more active methyltransferase ribozyme than the predecessor CA13. However, the outcome of the CD selection helped to determine the stability of the stem-loop structure as a key factor for high catalytic activity, which later guided the rational design of MTR1. Further attempts to improve the methyltransferase activity by partial or complete randomization of the 22 nucleotides flanking the stem-loop structure only yielded a single *trans*-active ribozyme (6CH18), albeit with very low activity. These findings showcase that CA13 may already be a very efficient catalyst for this specific reaction.

A detailed sequence analysis of 6CH18 and 8CH36 and comparison to the MTR1 crystal structure (chapter 4.2.2, Figure 4.1) provides an explanation for the low catalytic activity with respect to the parent ribozyme: In 6CH18, nucleoside C16 is mutated to U. This residue is part of the $Mg^{2+}$ binding site, located in between the stem loop structure and the catalytic center. It was previously shown that a point mutation of another residue within this bubble (A15G) severely impairs the activity of MTR1 (Figure 9.13). Therefore, it is conceivable that 6CH18 also exhibits a greatly reduced methyl transfer rate. In the case of 8CH36, a total of ten mutations can be identified throughout the entire catalytic core. Some of them (e.g., G10U and C40U) are located at positions that are neither heavily structured nor involved in any crucial hydrogen bonding and may therefore not be essential for the activity. Others, however, are found at more critical positions: Most importantly, mutation of the crucial cytosine residue at position 12 (C12G), which is in direct contact with the substrate and acts as a general acid catalyst in the methyl transfer reaction, is expected to be detrimental for the activity. In addition, A37 and A39, which are both changed to a G, are located in the top and bottom layer of the catalytic core, respectively. Both nucleosides therefore play an important role for the correct folding of the ribozymes active site. Finally, as discussed for 6CH18, two more mutations within the $Mg^{2+}$ ion binding pocket (A15G and C16U) are also expected to negatively impact the catalytic activity. Considering all these critical mutations, the lack of *trans*-activity of 8CH36 is certainly to be expected. However, it is remarkable that the ribozyme shows at least some detectable *cis*-activity, which may be due to the unimolecular nature of the construct that facilitates folding of the ribozyme.

Regarding the VMC10-based enrichment step, altering positive and negative selections certainly helped to prevent the accumulation of non-truly active variants as judged by the low enrichment level observed in the first four selection rounds of CH and CI. However, since the cleavage of unmodified sequences by VMC10 is not quantitative, some inactive species will always be carried over to the next round, delaying the enrichment of active methyltransferases. Therefore, this approach may be more suitable for a possible selection of demethylating ribozymes, as the cleavage reaction is entirely blocked by the presence of $m^6A$ or $m^1A$.

## 6.2 Selection of new methyl- and alkyltransferase ribozymes

The original CA *in vitro* selection, which employed BG-biotin as the small molecule cofactor, used a selection pool with a bulged adenosine as the target modification site. Characterization of the obtained ribozymes revealed that all of them indeed transferred the benzyl-biotin moiety to this specific A.[95] Earlier *in vitro* selections, which used the same selection strategy and pool design succeeded in obtaining highly specific adenylyl and tenofovir transferase ribozymes modifying the 2'OH of the same adenosine.[183-184] However, targeting nucleosides other than A is highly desirable, as this will provide access to a wider variety of modifications. If the selected ribozymes can again be readily converted into methyltransferases this would even enable the ribozyme-catalyzed generation of natural methylated nucleosides other than $m^1A$, e.g., modifications such as $m^7G$, $m^3C$ or $m^5C$. In addition, such ribozymes would generally expand the target RNA scope by facilitating the labeling of sequences that do not contain an adenosine suitable for modification by existing ribozymes. Another limitation of the above mentioned adenosine modifying ribozymes is their high $Mg^{2+}$ requirement (10 mM for FH14 and FJ1[183-184], 5 mM for MTR1[95]), which limits their application under cellular conditions (intracellular $Mg^{2+}$ concentration of 0.25 - 1.0 mM[251]). With alternative selection conditions, novel ribozymes may be evolved that work under reduced concentrations of $Mg^{2+}$.

In order to obtain such catalytic RNAs that target nucleosides other than A or work under cellular conditions, three different selections were initiated using BG-biotin as the alkyl donor substrate (Figure 6.9). In general, every starting pool contained a randomized $N_{40}$ region that was expected to evolve into the catalytic core. The overall structure as well as the exact sequence and primer binding sites of the libraries were the same as used in the previous FH, FJ and CA selections, except for the targeted bulged nucleoside: For selections CJ and CK, a cytosine or guanosine embedded within a G<u>C</u>G or G<u>G</u>G sequence context (underlining indicates the intended modified nucleoside) was used as the desired modification site, respectively (Figure 6.9). The progress and outcome of these selections are discussed in chapter 6.2.1. Selection CL was reinitiated from the second round of the CA selection, but in the presence of lower $MgCl_2$ and BG-biotin concentrations. The aim of this strategy was to obtain ribozymes that catalyze the same reaction as CA13/MTR1, but function under cellular conditions. The corresponding pool consequently contained a bulged adenosine within a G<u>A</u>G context as the desired modification site. The progress and outcome of selection CL is covered in section 6.2.2. Exemplary gel images of all selections are provided in Figure 9.30.

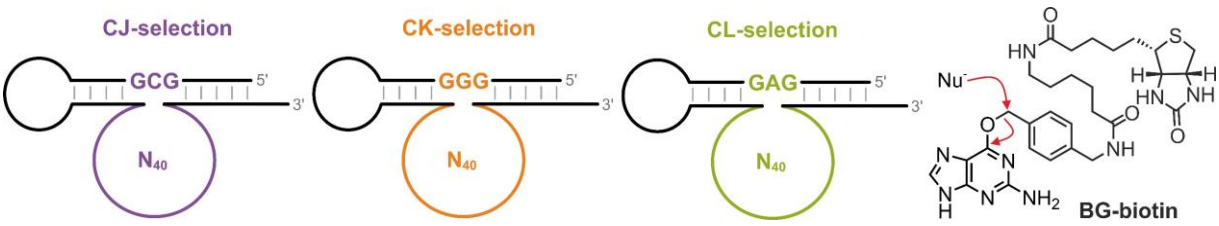



**Figure 6.9:** Schematic depiction of starting pools for selections CJ - CL together with the structure of the selection substrate BG-biotin used in the selections. $Nu^-$ denotes a nucleophilic group within the RNA of interest.

## 6.2.1 C or G as modification site (CJ and CK selection)

### 6.2.1.1 Selection progress

The starting pools to obtain cytosine (CJ) or guanosine (CK) alkylating ribozymes contained a random region of 40 nt. Thus, a theoretical complexity of $4^{40} = 1.2 \times 10^{24}$ unique molecules for the corresponding DNA templates are expected. With 500 pmol of this template used for transcription, a maximum of $3.0 \times 10^{14}$ different RNA variants can be obtained. In the first round, 3.15 nmol of this RNA library was used for each selection, thus the average copy number of every RNA is six.

The selection pools were incubated overnight at 37 °C at pH 7.5 in the presence of 100 µM BG-biotin and 40 mM $MgCl_2$, followed by pull down of the reactive sequences using magnetic beads functionalized with streptavidin. The captured RNAs were then eluted, followed by reverse transcription, PCR amplification and *in vitro* transcription to yield the next generation library (Figure 9.30). The bead-type was switched regularly between streptavidin and neutravidin to prevent enrichment of RNA aptamers that bind to one of the proteins. Additionally, a negative selection was performed after two rounds, in which the pool was incubated under selection conditions, but without the addition of BG-biotin. After the capture step, only the RNA that was free in solution was further amplified. This ensured that bead-binding during selection is actually due to pool biotinylation rather than non-specific interactions with streptavidin or neutravidin.

A total of 14 and 15 rounds were performed for the CJ and CK selections, respectively. Figure 6.10A as well as Table 9.21 provide an overview of the progress and reaction conditions of each round for both selections. The first ten (CJ) or eleven (CK) rounds were performed with overnight incubation in the presence of 40 mM $MgCl_2$ and 100 µM BG-biotin. The enrichment level in every round was monitored using fluorescence measurements: Before setting up the self-biotinylation reaction, ~150 pmol of each pool was 3'-end labeled with lucifer yellow carbohydrazide (LY-CH) using periodate glycol oxidation. By measuring the fluorescence intensity

of the eluted sample and comparing it to the initial selection sample, the amount of active RNA sequences within each round can be estimated.

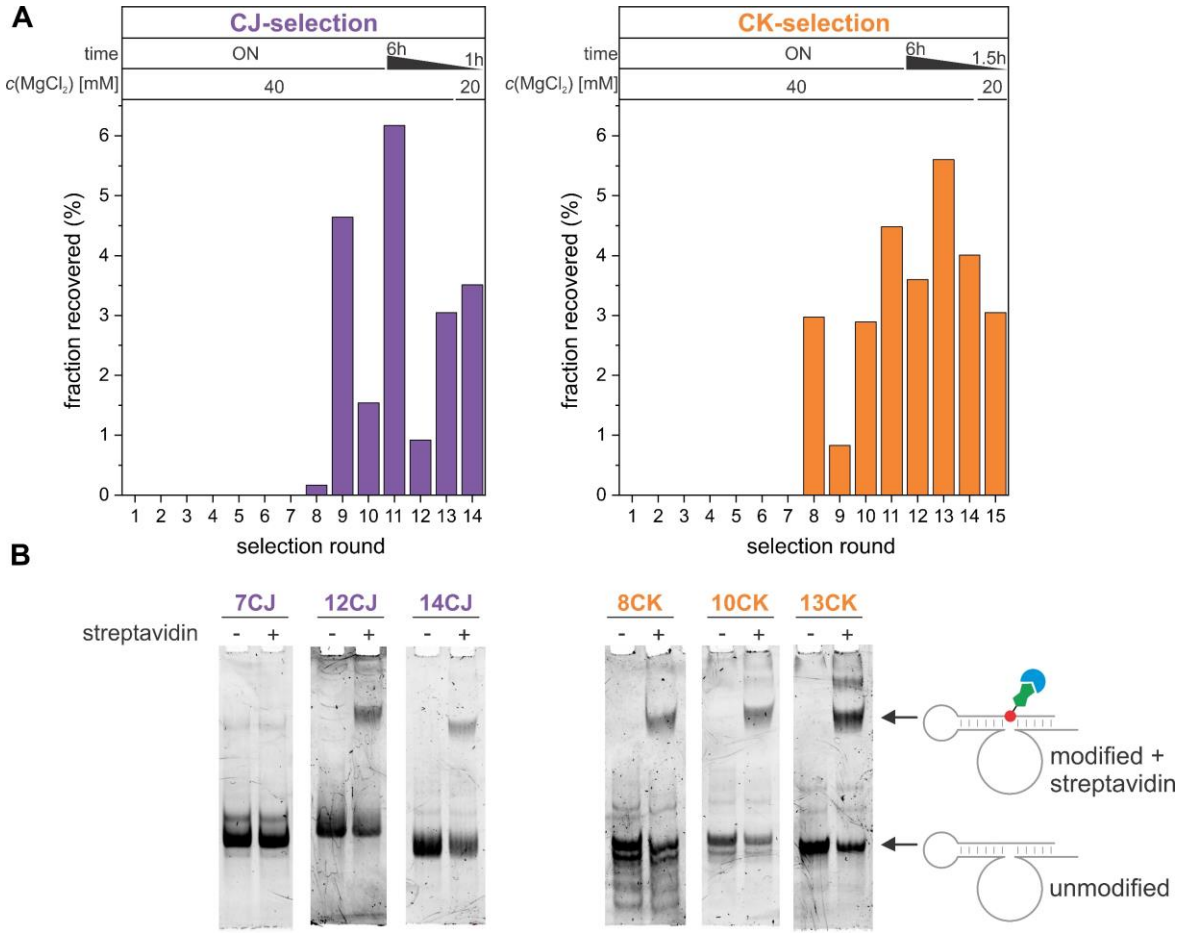



**Figure 6.10:** Progress of selections CJ and CK. **A** Overview of the enrichment and selection parameters. The fraction recovered in each round was determined using fluorescence measurements of the initial and eluted 3'-LY-labeled library. A detailed summary of the reaction conditions and enrichment level in each round is provided in Table 9.21. **B** Streptavidin gel shift assays of the round 7 (7CJ), 12 (12 CJ) and 14 (14CJ) of selection CJ and round 8 (8CK), 10 (10CK) and 13 (13CK) of selection CK to determine the activity of the enriched RNA pools. The RNAs were reacted with BG-biotin using the incubation conditions of the respective round and then coupled to streptavidin (blue). The desired modification site is depicted in red; biotin is shown in green. The gels were stained with SYBR gold.

For the first seven rounds, no enrichment was observed in either selection. In the eighth round, first signs of enrichment became apparent as 0.2 % and 3.0 % active species were observed for CJ and CK, respectively. In case of selection CJ, the fraction of eluted RNA then increased to 4.6 % in round nine, but then suddenly dropped again to 1.5 % in the following round, despite invariant reaction conditions. Similarly, only 0.1 % of the RNA was eluted from the beads in the ninth round of the CK selection, increasing again to 2.9 % and 4.5 % in subsequent rounds. Due to this fluctuation, it was decided to increase the stringency on both selections to only obtain ribozyme variants with highest labeling efficiency and fastest kinetics: The incubation time was gradually reduced from overnight to just 60 min (CJ) or 90 min (CK) in the final round. Additionally, the $Mg^{2+}$ concentration was halved from 40 mM to 20 mM in the last round of both selections. However, the increased pressure hardly effected the activity within selection CK:

The fraction of eluted RNA remained largely constant at approximately 4 % - 5 %. For CJ, although the reaction time was reduced to 6 h, the activity initially increased to 6.1 %. However, in the following round, when the pool was incubated for only 3 h, it dropped to less than 1 %. Subsequently, the fraction of eluted species then remained at a constant level of approximately 3 %.

During the *in vitro* selection process, streptavidin gel shift assays were performed regularly (Figure 6.10B), which provided a second method of monitoring the enrichment levels in addition to the fluorescence measurements. Although precise quantification is hampered by the fact that the gels were only stained with SYBR gold and the RNA itself was not labeled, the assays still provide a good estimate of the pool activity and the enrichment levels. In general, the assays revealed a higher catalytic activity of the enriched CK pools compared to the CJ pools. In the last round of CJ (14CJ), only about 20 % of the RNA was shifted and therefore active, whereas in round 13CK, about half of the pool was biotinylated. As no further enrichment could be observed, the selections were stopped after 14 (CJ) or 15 rounds (CK) and the PCR products derived from these pools were cloned. A total of twelve and nine plasmids from randomly selected clones were sent for Sanger sequencing for selections CJ and CK, respectively.

### 6.2.1.2 Sanger sequencing results of CJ and CK selections

In the case of selection CJ, eleven plasmids were sequenced successfully revealing two distinct sequence families that vastly differ from the CA13 ribozyme, which was previously evolved using the same cofactor. Each family consists of several identical or highly related sequences, which are summarized in Table 6.3.

**Table 6.3:** Sequencing results obtained from the CJ selection. Shown are the ribozymes binding arms (green) and the randomized region, which evolved into the catalytic core (black). Base paired nucleotides forming the ribozymes stem-loop structure in CA13 and MTR1 are underlined. Point mutations are marked in blue. Secondary structure predictions of CJ07 and CJ18 are provided in Figure 6.11.

| ribozyme | 5'-sequence-3' |
|---|---|
| pool | UUGAAGGC NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN CAGUAUGUCC |
| CA13 | UUGAAGGC UGACCGACCCCCCGACCCUUCUCUGGGACAACUAGACAUA CAGUAUGUCC |
| MTR1 | UUGAAGGC UGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUA CAGUAUGUCC |
| CJ10 | UUGAAGGC SCACUGGACGAAUACAUCCCGCUGCUGCACCGUAGUUGUG CAGUAUGUCC |
| CJ13 | UUGAAGGC GCACUGGACGAAUACAUCCCGCUGCUGCGCCGUAGUUGUG CAGUAUGUCC |
| CJ18 | UUGAAGGC CCACUGGACGAAUACAUCCCGCUGCUGCACCGUAGUUGUG CAGUAUGUCC |
| CJ27 | UUGAAGGC UCACUGGACGAAUACAUCCCGCUGCUGCACCGUAGUUGUG CAGUAUGUCC |
| CJ30 | UUGAAGGC GCACUGGACGAAAACAUCCCGCUGCUGCGCCGUAGUUGUG CAGUAUGUCC |
| CJ32 | UUGAAGGC GCMCUGGACGAAUACAUCCCGCUGCUGCRCCGUAGUUGUG CAGUAUGUCC |
| CJ01 | UUGAAGGC GCCCUGGACGAAUACAUCCUACCUGCCUGUUAUAAUCCGG CAGUAUGUCC |
| CJ07 | UUGAAGGC GCCCUGGACGAAUACAUCCUACCUGCCUGUUAUAAUCCGG CAGUAUGUCC |
| CJ08 | UUGAAAGC GCCCUGGACGAAUACAUCCUACCUGCCUGUUAUAAUCCGG CAGUAUGUCC |
| CJ09 | UUGAAGGC GCCCUGGACGAAUACAUCCUACCUGCCUGUUAUAAUCCGG CAGUAUGUCC |
| CJ25 | UUGAAGGC GCCCUGGACGAAUACAUCCUACCUGCCUGUUAUAAUCCGG CAGUAUGUCC |

First, the six extremely similar RNA sequences CJ10, CJ13, CJ18, CJ27, CJ30 and CJ32 can be identified. Their sequences are identical except for a mutation of the first nucleotide within the catalytic core and an A-to-G transversion mutation at position 29, which occurs in three variants. In the case of CJ30, an U13A mutation can be identified in the catalytic core. The sequences of the second ribozyme family obtained, comprising ribozymes CJ01, CJ07, CJ08, CJ09 and CJ25, are completely identical. Only CJ08 contains a G-to-A mutation in the 5' binding arm. Additionally, both ribozyme families share the same 19 nucleotides within the 5'-end of the catalytic core, except for an A at position 3 that is a C in family two.

In contrast, sequencing of the CK selection only yielded one group of ribozymes. Except for the variants CK05 and CK35, which comprise a G-to-A mutation in the catalytic core, all sequences are identical. Only in CK31 a uridine in the 5' binding arm was changed to a cytosine.

Furthermore, also the CK ribozymes show no sequence similarity to CA13 or MTR1. However, comparison of the CK and CJ sequencing results reveals that CJ18 is highly similar to the CK ribozymes and differs from CK03 only by two point mutations (Table 6.4).

**Table 6.4:** Sequencing results obtained from the CK selection. Shown are the ribozymes binding arms (green) and the randomized region, which evolved into the catalytic core (black). Base paired nucleotides forming the ribozymes stem-loop structure in CA13 and MTR1 are underlined. Point mutations are marked in blue. The secondary structure prediction of CK03 is provided in Figure 9.31B.

| ribozyme | 5'-sequence-3' |
|---|---|
| pool | UUGAAGGC NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN CAGUAUGUCC |
| CA13 | UUGAAGGC UGACCGACCCCCCGACCCUUCUCUGGGACAACUAGACAUA CAGUAUGUCC |
| MTR1 | UUGAAGGC UGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUA CAGUAUGUCC |
| CJ18 | UUGAAGGC CCACUGGACGAAUACAUCCCGCUGCUGCACCGUAGUUGUG CAGUAUGUCC |
| CK03 | UUGAAGGC CCACUGGACGAAGACAUCCCGCUGCUGCGCCGUAGUUGUG CAGUAUGUCC |
| CK05 | UUGAAGGC CCACUGGACGAAAACAUCCCGCUGCUGCGCCGUAGUUGUG CAGUAUGUCC |
| CK08 | UUGAAGGC CCACUGGACGAAGACAUCCCGCUGCUGCGCCGUAGUUGUG CAGUAUGUCC |
| CK11 | UUGAAGGC CCACUGGACGAAGACAUCCCGCUGCUGCGCCGUAGUUGUG CAGUAUGUCC |
| CK18 | UUGAAGGC CCACUGGACGAAGACAUCCCGCUGCUGCGCCGUAGUUGUG CAGUAUGUCC |
| CK20 | UUGAAGGC CCACUGGACGAAGACAUCCCGCUGCUGCGCCGUAGUUGUG CAGUAUGUCC |
| CK21 | UUGAAGGC CCACUGGACGAAGACAUCCCGCUGCUGCGCCGUAGUUGUG CAGUAUGUCC |
| CK31 | UCGAAGGC CCACUGGACGAAGACAUCCCGCUGCUGCGCCGUAGUUGUG CAGUAUGUCC |
| CK35 | UUGAAGGC CCACUGGACGAAAACAUCCCGCUGCUGCGCCGUAGUUGUG CAGUAUGUCC |

### 6.2.1.3 Activity assay of CJ ribozymes

For the CJ selection, one ribozyme from each family (CJ07 and CJ18) was first tested for *cis*-activity by RT primer extension (Figure 6.11A). The ribozymes were incubated overnight under selection conditions using either BG-$NH_2$ or $m^6G$ as the small molecule substrates (Figure 6.11C). This allows to test whether $O^6$-modified guanine cofactors other than BG-biotin are accepted by the ribozymes. Using primer P1 for readout, which binds in the connecting loop and extends over the substrate RNA, this experiment simultaneously enables the detection of the modification site within the target RNA with single nucleotide resolution. With BG-$NH_2$ as cofactor, an intense termination band can be detected for both ribozymes, which indicates that biotin is not essential for the reaction. Notably, in combination with the respective ladders generated using dideoxynucleoside triphosphates (ddNTPs), the primer extension reveals that both ribozymes do not modify the bulged cytosine, but rather target adenosine A7, which was originally designed to hybridize to the ribozymes 3' binding arm. However, to determine the exact chemical composition of the modified adenosine, further research is required. This may

for instance include the investigation of atomic substrate RNA mutants or LC-MS analysis of the digested RNA product.

Using $m^6G$ as the cofactor, no termination product band can be detected for neither of the ribozymes, which may be due to two reasons: First, $m^6G$ is not a substrate of CJ07 and CJ18, because the ribozymes are no active methyltransferases. Alternatively, the ribozymes may be able to use $m^6G$ as the substrate, but they install the methyl group at an RT silent position. This would allow the reverse transcriptase to read through the modification and renders it undetectable by an RT primer extension experiment. Alternative analytical methods such as ESI-MS measurements of a short RNA treated with one of the ribozymes in the presence of $m^6G$ would be one possibility to further investigate on this issue.

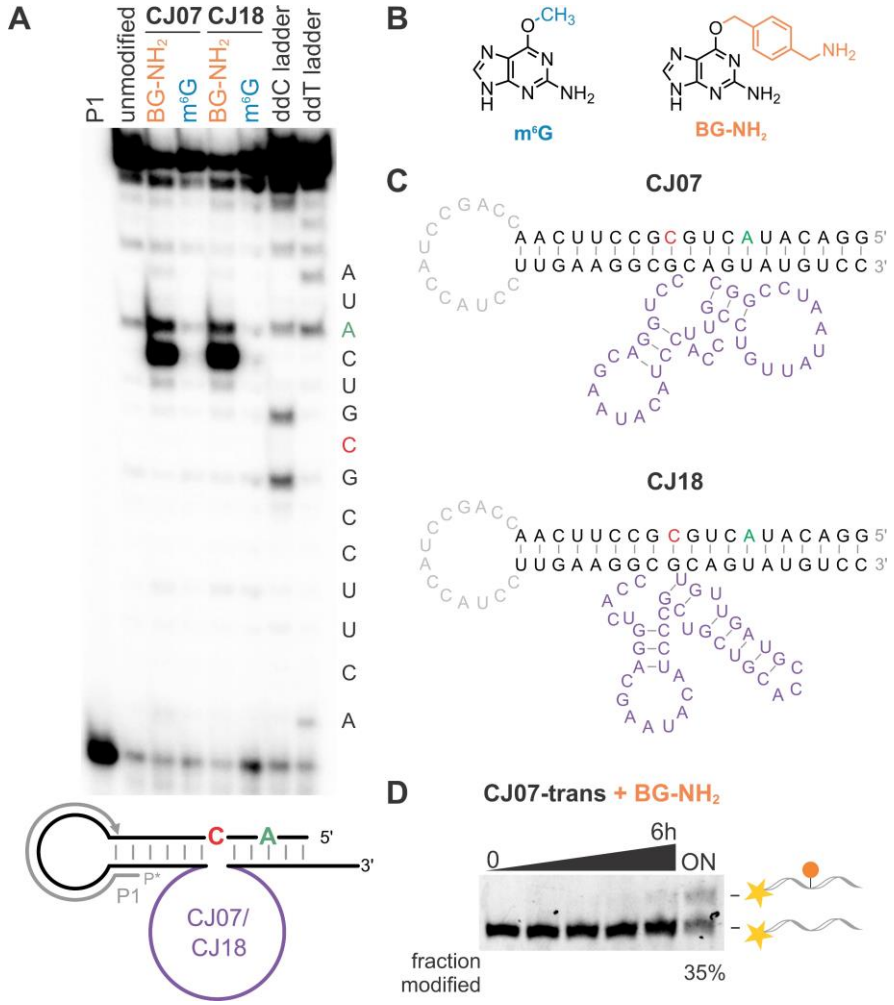



**Figure 6.11:** *Cis*- and *trans*-activity assay of CJ07 and CJ18 ribozymes. **A** RT primer extension of *cis*-active CJ07 and CJ18 that were reacted ON with $m^6G$ or BG-$NH_2$ using 5'-$^{32}$P-labeled primer P1. **B** Molecular structures of $m^6G$ and BG-$NH_2$. **C** Predicted secondary structure of *cis*-active CJ07 and CJ18. The catalytic core is depicted in purple, the desired modification site C is marked red, the actual target site is shown in green. The secondary structures were predicted using the RNAfold web server.[391] **D** Test of CK03 for *trans*-activity under single turnover conditions. The 5′-fluorescein-labeled 21-mer RNA substrate (R3) was reacted with CJ07-trans generated from the respective plasmid using primer D1 and D5 under single turnover conditions. Timepoints: 0, 30min, 1h, 2h, 6h, ON.

According to secondary structure predictions of the full-length *cis*-active CJ07 and CJ18 sequences (Figure 6.11C), both ribozymes contain two stem-loop structures within their catalytic domain and anneal completely to the substrate RNA. In this fold, no bulged nucleoside is generated. Whereas the catalytic core of CJ07 is positioned between the actual and expected modification sites A7 and C11, CJ18 is placed directly next to the targeted cytosine. From these secondary structure predictions, it is unclear how the ribozymes recognize and modify adenosine A7. Most likely not all nucleotides, which are predicted to be part of the binding arms actually anneal to the target RNA but may rather play a crucial structural role in cofactor binding and catalysis. However, further investigations are needed to characterize the recognition mode of the ribozymes binding arms and their programmability in closer detail. This may also guide the future design of new starting pools in order to prevent a modification site shift as observed in these selections.

In conclusion, RT primer extension experiments show that both CJ ribozymes obtained from *in vitro* selection modify an adenosine, initially designed to be part of a binding arm, instead of the bulged cytosine. Furthermore, no methyltransferase activity was detectable for either ribozyme. Only CJ07 was tested in *trans* with BG-$NH_2$ as the substrate, revealing low catalytic activity. Therefore, no further characterization of the CJ ribozymes was pursued. Instead, it was continued with the analysis of the CK selection outcome.

#### 6.2.1.4 Characterization of the CK03 ribozyme

Regarding the selection CK, which targeted a G within a GGG RNA substrate context, the ribozyme variant CK03 (Figure 9.31B) that shows high sequence similarities to the CJ18 ribozyme was selected for further analysis. It was first tested for activity in *cis* using a streptavidin gel-shift assay, which revealed around 70 % biotinylated RNA after 4 h incubation under selection conditions (Figure 6.12A). Encouraged by this result, the acceptance of cofactors other than BG-biotin was explored using RT primer extension. Analogously to the experiment conducted with cis-active CJ07 and CJ18 (chapter 6.2.1.3), the ribozyme was reacted overnight using either BG-biotin, BG or $m^6G$ as small molecule cofactor followed by RT readout with primer P1 (Figure 6.12D). This experiment revealed a high self-alkylation activity of CK03 with both, BG and BG-biotin, as indicated by the strong termination product band. Thus, biotin is not essential for the reactivity of CK03. With $m^6G$, however, only a faint abortive band is visible on gel indicating only a slight, almost undetectable methyltransferase activity in *cis*. When CK03 was challenged to modify an external 21-mer substrate RNA R2 (*trans*-activity), the ribozyme retained its high reactivity with BG as the alkyl donor (Figure 6.12B): Approximately 65 % benzylated substrate RNA could be obtained after 7 h of reaction time.

With regard to the modification site of CK03, the RT primer extension assay shows that the CK03 ribozyme again targets adenosine A7 instead of the intended modification site G11 (Figure 6.12D). Thus, although three consecutive guanosines were positioned directly opposite the $N_{40}$ random region, *in vitro* selection yielded a ribozyme, which modifies an adenosine instead. Additionally, an RT primer extension was also performed on the pool of the last selection round (15CK), which was incubated with BG-biotin beforehand (Figure 9.32A). This experiment tested whether the enriched pool contained ribozyme variants that modify positions other than A7 but were missed by Sanger sequencing. However, only one abortive band indicating successful modification of A7 could be detected.

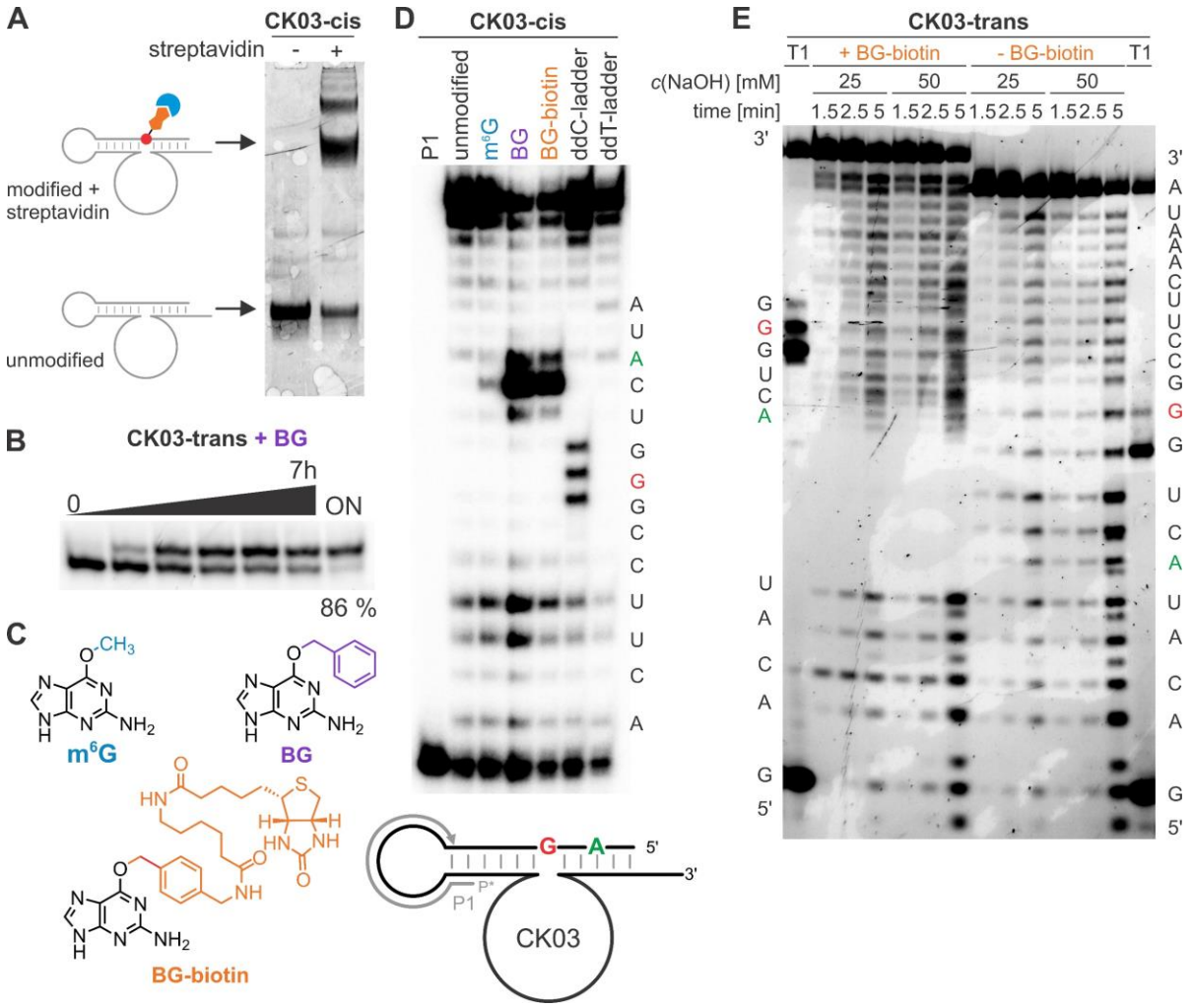



**Figure 6.12:** Initial characterization of the CK03 ribozyme and determination of its modification site. **A** Streptavidin gel-shift assay of CK03. The RNA was incubated for 4 h in the presence of 100 µM BG-biotin and 40 mM $MgCl_2$ at 37 °C before coupling to streptavidin (blue). The desired modification site is depicted in red; biotin is shown in orange. The gel was stained with SYBR gold. **B** Test of CK03 for *trans*-activity under single turnover conditions. The 5′-fluorescein-labeled 21-mer RNA substrate (R2) was reacted with CK03 and 100 µM BG at 37 °C. Timepoints: 0, 15min, 1h, 2h, 4h, 7h, ON. **C** Chemical structure of cofactors used for RT primer extension. **D** RT primer extension of *cis*-active CK03 that was reacted overnight with $m^6G$, BG or BG-biotin using 5'-$^{32}$P-labeled primer P1 for readout. **E** RNase T1 digestion and alkaline hydrolysis of unmodified RNA R2 in comparison to the isolated CK03 reaction product of R2 reacted with Rz1 using BG-biotin as alkyl-donor.

To gain deeper insight into the modification site of CK03, RNase T1 digestion and alkaline hydrolysis of the isolated biotinylated reaction product was performed (Figure 6.12E). These experiments further corroborate A7 as the modification site, as they reveal up-shifted digestion products starting from this position. Moreover, additional hydrolysis bands (double bands) for fragments containing the modified A can be observed, which are not present in the unmodified reference sample. Such additional bands were also observed for RNA modified by the MTR1 ribozyme, which generates *N*1-modified adenosine (Figure 3.2). Upon incubation under alkaline conditions and high temperatures, this modification easily undergoes Dimroth rearrangement, generating $N^6$-modified adenosine. As the resulting RNA product is depleted of one positive charge, it exhibits an altered electrophoretic mobility resulting in the formation of double bands. Thus, the alkaline hydrolysis result hints towards *N*1 of A7 as the exact modification site of the CK03 ribozyme. However, as the alkaline hydrolysis samples of the modified RNA are not very well resolved (especially in the region around A7), no definitive conclusion can be drawn from this single experiment. Further experiments, such as digestion of the modified RNA followed by LC-MS analysis and comparison with reference samples, is required to ultimately determine the exact constitution of the CK03 modification product.

Due to the promising preliminary results regarding acceptance of cofactors other than BG-biotin and the high activity in *trans* (Figure 6.12), the labeling characteristics of CK03 were investigated in more detail. Different kinetic assays were conducted under single turnover conditions using 1 µM 5'-fluorescein labeled substrate RNA (R2) together with a ten-fold excess of the trans-CK03 ribozyme. Unless otherwise mentioned, all kinetic assays were performed at 37 °C and pH 7.5 in the presence of 100 µM of the cofactor and 40 mM $MgCl_2$. The results are presented in Figure 6.13 and Figure 9.33. Table 9.23 provides a detailed summary of the rate constants obtained from for all kinetic assays**.**

First, the ribozyme was tested for acceptance of different $O^6$-modified guanine cofactors. While no reactivity could be detected with $O^6$-methylguanine ($m^6G$) or $O^6$-propargylguanine (O6PG; Figure 9.33A), fast and efficient alkylation of the target RNA was observed with all benzylated cofactors. The highest yield (nearly 80 % after 7 h incubation) was achieved with the smallest substrate BG followed by BG-$NH_2$ and BG-biotin (Figure 6.13A). Regarding the alkyl donor concentration requirement, the kinetic studies revealed a moderate labeling efficiency for a BG-$NH_2$ concentration as low as 25 µM (25 % benzylated RNA after 7 h incubation). A significant higher reactivity was observed for concentrations above 50 µM (Figure 6.13B, Figure 9.33B). However, when the BG-$NH_2$ concentration was increased to 200 µM, no further acceleration was observed, resulting in a maximum yield of 80 % after 7 h reaction time. Compared to the CA13 ribozyme, however, CK03 is less robust to a reduction in alkyl donor concentration: While for CA13, similar rate constants were obtained at 100 µM and 25 µM BG-$NH_2$, the

reaction rate decreases by more than a factor of four when CK03 was tested under the same conditions (Figure 9.33, Table 9.23). Regarding the kinetic characterization, it should further be noted that the CJ18 ribozyme may show comparable labeling characteristics, due to its high sequence similarity to CK03.

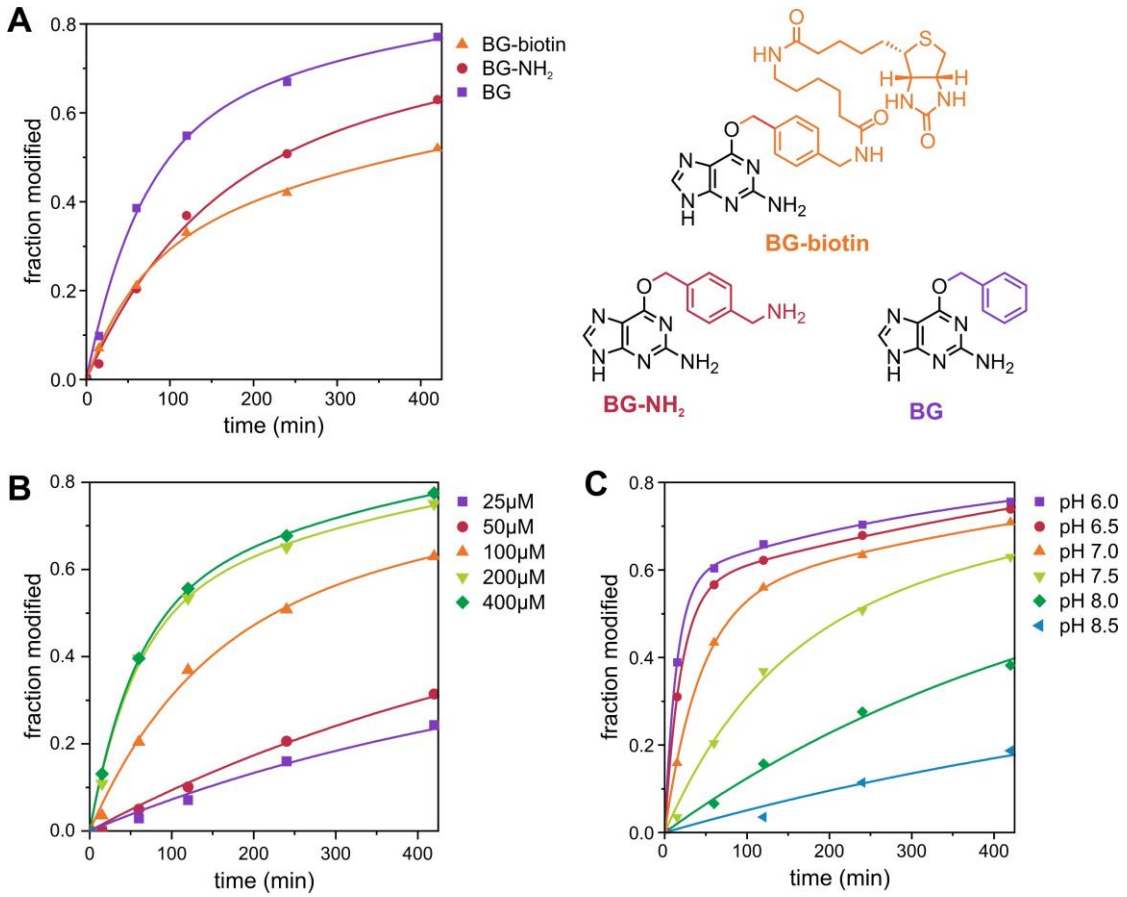



**Figure 6.13:** Characterization of the CK03 ribozyme in *trans*. All reactions were performed under single turnover conditions using Rz2 with substrate RNA R2. Unless otherwise mentioned, all experiments were conducted at pH 7.5 and 37 °C in the presence of 100 µM BG-$NH_2$ and 40 mM $MgCl_2$. The data was fitted using pseudo-first order kinetics (for slow reactions) or a biexponential model (for fast kinetic). The obtained rate constants are summarized in Table 9.23. The corresponding gel images are depicted in Figure 9.33. **A** Substrate scope of CK03. No reactivity was determined for O6PG and $m^6G$. **B** BG-$NH_2$-dependency of CK03. **C** pH-dependency of CK03.

Another important observation is the pronounced pH-dependence of the CK03-catalyzed reaction. Although the ribozyme was initially selected at pH 7.5, the reaction rate was nearly nine times higher at pH 6.0 (Figure 6.13C, Figure 9.33B). Under alkaline conditions (pH 8.0 and pH 8.5), the reaction rate experiences a sudden drop, indicating deprotonation of residues, which are crucial for catalysis. In general, the pH rate profile shows high similarities to that of MTR1, where a protonated cytidine next to the small molecule substrate acts as a general acid in the methyl transfer reaction. It is tempting to speculate that a similar mechanism is responsible for the alkyl-transfer by the CK03 ribozyme. However, more detailed kinetic studies are required to obtain conclusive answers regarding the mechanism of the CK03-catalyzed reaction. These may include the investigation of multiple ribozyme mutants in combination with structural investigations.

To gain first insight into the folding of CK03 and its substrate RNA requirements, the activity of a point mutant version of the ribozyme (CK03 C14G; Rz4) was analyzed in more detail. This mutant was designed based on the secondary structure prediction of CK03: Residue G53 of the ribozyme is predicted to be part of the 3' binding arm. It is expected to anneal to C7 of the target RNA, which is located directly 3' of the modification site A6 (Figure 6.14A). However, in this arrangement the catalytic core of the ribozyme is positioned at least one nucleotide away from the target adenosine. Within this constitution, it is difficult to imagine how CK03 might be able to access and alkylate the modification site. In order to introduce a shift of the catalytic core and reposition it one nucleotide towards the 5'-end of the RNA target, C14 was mutated to a G. This guanosine can then theoretically form a Watson-Crick base pair with C7 (Figure 6.14A). However, the resulting ribozyme mutant did not show any detectable activity when tested with BG-$NH_2$, neither at pH 7.5, nor at pH 6.5. This indicates that C14 may be an essential nucleotide of the catalytic core that cannot be mutated. In addition, it is conceivable that G33 may not be engaged in a Watson-Crick base pair with the substrate RNA as suggested by the secondary structure prediction. Instead, it may be an important component of the catalytic center as well. Moreover, the introduced G14 mutation may also interact with C7 in the substrate RNA, resulting in a misfolded state of the ribozyme that lacks catalytic activity. However, kinetic assays with additional mutants, in particular alternative substrate sequences, are required to investigate this issue in closer detail.

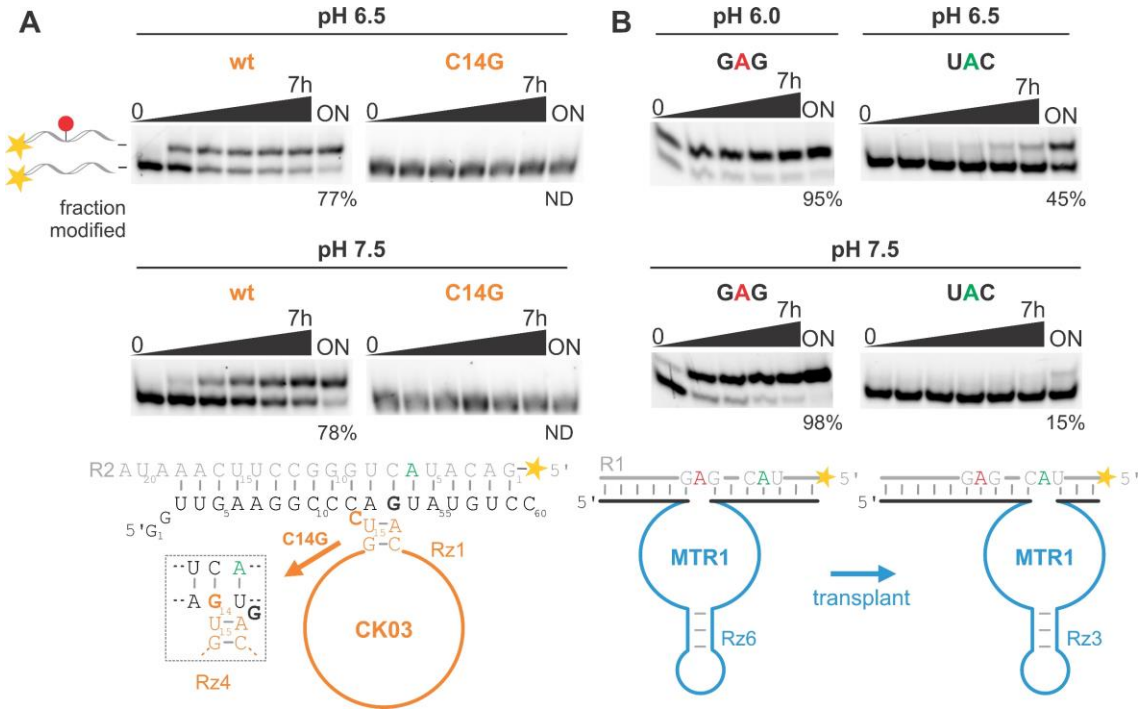



**Figure 6.14:** Further investigation of the CK03 modification site and comparison of CK03 to MTR1. **A** *Trans*-activity assay of the CK03 C14G mutant (Rz4) at pH 6.5 and pH 7.5 under single turnover conditions with RNA R2 and BG-$NH_2$ as small-molecule substrate. Timepoints: 0, 15min, 1h, 2h, 4h, 7h, ON. **B** Testing MTR1 for targeting the modification site of CK03 in RNA R1 using Rz3 with BG-$NH_2$. Timepoints: 0, 45min, 2h, 4h, 7h, ON (for GAG context) and 0, 15min, 1h, 2h, 4h, 7h, ON (for UAC context).

Even though CK03, as well as CJ07 and CJ18, are incapable of modifying the desired G or C, but instead, just like MTR1, target an adenosine, this A is still of particular interest: It is located in an UAC sequence context, whereas MTR1 has been selected to modify an A embedded between two guanosines (GAG). To test whether MTR1 functions as efficiently as CK03 when targeting the same UAC RNA context, a ribozyme variant with appropriate binding arms was designed and subsequently tested for activity (Figure 6.14B). Under the same conditions that were applied for CK03 with BG-$NH_2$ as the alkyl donor, MTR1 yielded only 45 % benzylated RNA after overnight incubation at pH 6.5, whereas the reaction with CK03 resulted in 77 % RNA product. In comparison, with almost 80 % modified RNA after only 15 min incubation at pH 7.5, very high reactivity was observed for MTR1 targeting A within the GAG context. This establishes CK03 as a valuable addition to the family of RNA alkylating ribozymes by expanding the scope of possible modification sites.

#### 6.2.1.5 Discussion

In conclusion, both, the CJ and CK selection failed to yield an RNA alkylating ribozyme, with a modification site other than adenosine. Although the starting pools were designed to comprise a bulged C or G to serve as the modification site, all ribozymes evolved to modify an adenosine, originally designed to anneal to the 3' binding arm. However, the ribozymes obtained from both selections are highly active alkyltransferases that are functional in *cis* as well as in *trans*. The exact modification position within the adenosine, however, still requires further investigation. The ribozymes may generate reaction products different from MTR1, which would be a valuable addition to the toolbox of RNA modifying ribozymes. Moreover, further in-depth characterization regarding the substrate sequence requirements is needed. So far, the ribozymes were only tested for modification of one particular RNA oligonucleotide, but nothing is known about the sequence requirements. Substrate RNA mutants covering all possible nucleosides as modification sites have been tested for reactivity with CK03 (Figure 9.32B). However, this analysis is of low significance as only the originally desired modification site (G11) was mutated, which now, however, anneals to the binding arm of CK03. Mutants of the actual modification site A7 remain to be analyzed. Other aspects such as the $MgCl_2$-dependency or the possibility to minimize the catalytic core of CK03 are also of high interest, as these are important factors considering prospective applications of the ribozyme.

Regarding future projects, the CJ and CK selections provided valuable information regarding the starting pool design. To obtain C- or G-modifying ribozymes using BG-biotin, alternative random regions may be an option: For instance, a 3-4 nt long sequence stretch inserted at the 3'- and 5'-ends of the randomized region, which contains only three instead of all four possible

nucleotides may be advantageous. This setup may suppress an annealing of the random region to the RNA target, thereby preventing a modification site shift. Alternatively, as all six ribozymes obtained with BG-biotin modify adenosine, other cofactors, which are not based on guanine might result in C-, G- or U-alkylating ribozymes. In this regard, biotinylated $O^2$-benzylcytidine (BC-biotin) can be an option: Using a pyrimidine- instead of a purine-based cofactor, may likely yield ribozymes with highly different binding pockets to efficiently accommodate the small molecule substrate. Such ribozymes may be able to alkylate positions other than A. In addition, BC-biotin can conveniently be used in an *in vitro* evolution experiment that employes the same overall pool design and protocol as the previous selections, which renders time and labor-intensive optimization redundant.

## 6.2.2 Selecting an improved alkyltransferase ribozyme (CL selection)

Especially for *in vivo* applications, ribozymes that require low concentrations of bivalent metal ions and cofactor are highly desirable as the concentration of $Mg^{2+}$ in cells is only 0.25 - 1 mM[251]. Furthermore, the intracellular delivery of small molecule substrates is limited, and some cofactors can also be lethal to cells at higher concentrations. The previously evolved CA13 ribozyme was selected in the presence of 40 mM $MgCl_2$ and 100 µM BG-biotin, which was reduced to 50 µM in the last round. Analysis of the labeling characteristics of CA13 revealed that the modification yield drops drastically at a $MgCl_2$ concentration below 5 mM.[95] When the cofactor concentration is reduced in parallel, hardly any reactivity can be observed. To evolve a ribozyme that catalyzes the same reaction as CA13, but requires much lower $Mg^{2+}$ and substrate concentrations, a reselection of the initial CA ribozyme selection was performed, named CL selection.

### 6.2.2.1 Selection progress

The original CA library, that was generated after the first selection round (1CA), served as the starting pool. The selection cycles were performed as described for selections CJ and CK using streptavidin or neutravidin coated magnetic beads for capturing of biotinylated and therefore active sequences. In the first round, the pool was incubated overnight at 37 °C and pH 7.5 together with 100 µM BG-biotin in the presence of 40 mM $MgCl_2$. Compared to the original CA selection, however, much more stringent conditions were applied in subsequent rounds: Despite the lack of detectable enrichment, the BG-biotin and $MgCl_2$ concentrations were consistently reduced starting from the second round. The selection cycle was repeated a total of eleven times. An overview of the reaction conditions and the enrichment observed in each round is provided in Figure 6.15A as well as in Table 9.22.

Already in the third round, 1.8 % active species were observed, which dropped to 0.4 % as the stringency was further increased (Figure 6.15A). Despite a continuous reduction of the BG-biotin and $MgCl_2$ concentrations, the activity then increased again to nearly 5 % and remained on a consistently high level for the sixth and seventh round. In addition to fluorescence measurements, regular streptavidin gel shift assays were performed during the selection (Figure 6.15B) to monitoring the enrichment level. In these assays, the enriched CL libraries were always compared to the *cis*-active CA13 ribozyme as a positive reference. In round seven, the CL pool generated around 50 % self-modification after 5 h in the presence of 25 µM BG-biotin and 4 mM $MgCl_2$, which was comparable to the activity of CA13 under identical conditions. Thus, in order to obtain only variants with labeling characteristics superior to those of CA13, the incubation time was reduced from overnight to only 4 h in round eight. This resulted in a

sudden drop of the enrichment level to just 0.1 %, rising again to 1.7 % in the subsequent round. This low activity was consistent with the results obtained by streptavidin gel-shift assays, in which only a small fraction of the round ten pool was shifted after 2 h. CA13, however, showed the same low activity under the more stringent reaction conditions. As a further reduction in incubation time led to a continuous decrease in pool activity, the selection was stopped after eleven rounds. The enriched library was then cloned and a total of ten randomly selected colonies were sent for Sanger sequencing.

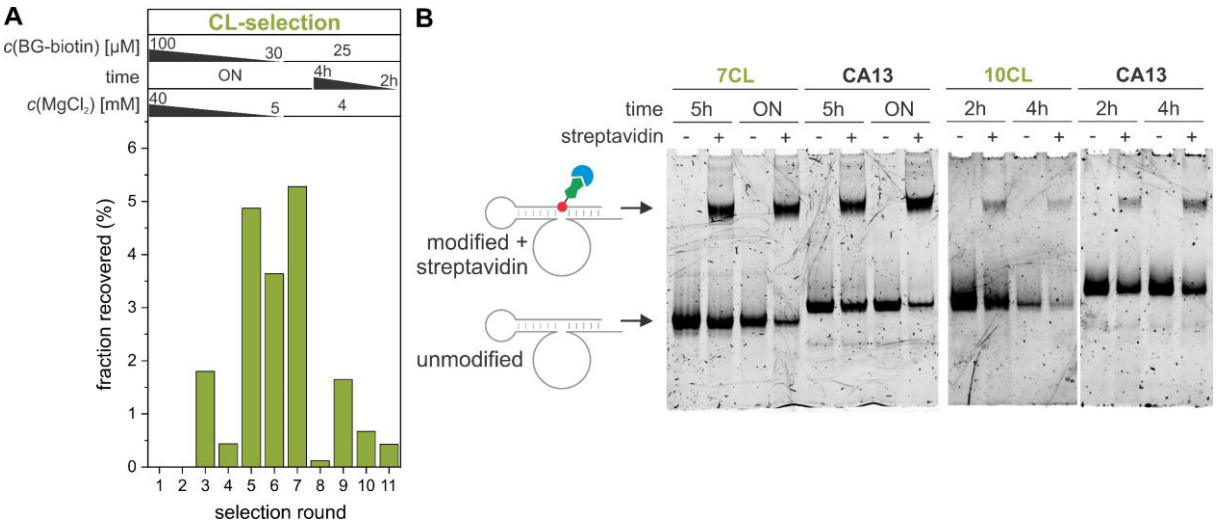



**Figure 6.15:** Progress of selection CL. **A** Overview of the enrichment level and selection parameters. The fraction recovered in each round was determined using fluorescence measurements of the initial and eluted library. A detailed summary of the reaction conditions and enrichment levels of each round is provided in Table 9.22. **B** Streptavidin gel shift assays of the round 7 (7CL) and round 10 pools (10CL) in comparison to the *cis* reactive CA13 ribozyme. The RNAs were incubated overnight in the presence of 25 µM BG-biotin and 4 mM $MgCl_2$ at 37 °C before coupling to streptavidin (blue). The desired modification site is depicted in red; biotin is shown in dark green. The gels were stained with SYBR gold.

### 6.2.2.2 Sanger sequencing results of CL-selection

Sequencing of the round eleven CL library revealed three distinct ribozyme families, which are all composed of several highly similar or identical sequences (Table 6.5). The first family includes ribozymes CL01, CL06, CL10 and CL19, which show great similarities to the CA13 ribozyme. Except for the sequence of the stem-loop, their catalytic core is identical to that of CA13 (Figure 6.17D). Only CL01 contains an A-to-U point mutation. Unlike CA13, the new ribozymes comprise only a 3 bp long stem-loop with an AACCAA hexaloop, which is a pentaloop in CL10. As a result, the total length of the catalytic core was reduced from 40 nt to 34 nt (or 33 nt in CL10) during the selection process. In contrast, sequence CL24 is unique with a 40 nt catalytic core and no similarities to any previously selected ribozyme.

Finally, the third group of CL ribozymes consists of five highly similar RNA sequences (CL4, CL12, CL17, CL20 and CL26). Most noticeable, these variants consist of only the 40 nt long

catalytic region and the adjacent 3' binding arm, which also acts as primer binding site. The binding arm 5' of the catalytic region, however, was lost during the selection, while the connecting loop was maintained. The catalytic cores of this third family are very similar to the sequence of CL24 and largely identical but contain up to three point mutations in the centre.

**Table 6.5:** Sequencing results obtained from the CL selection. Shown are the ribozymes binding arms (green) and the randomized region, which evolved into the catalytic core (black). Base paired nucleotides forming the ribozymes stem-loop structure are underlined. Point mutations are marked in blue. Secondary structure predictions of CL06, CL12 and CL24 are provided in Figure 6.16C and Figure 6.17B, respectively.

| ribozyme | 5'-sequence-3' |
|---|---|
| pool | UUGAAGGC NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN CAGUAUGUCC |
| CA13 | UUGAAGGC UGACCGACCCCCCGACCCUUCUCUGGGACAACUAGACAUA CAGUAUGUCC |
| MTR1 | UUGAAGGC UGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUA CAGUAUGUCC |
| CL01 | UUGAAGGC UGACCGACCCGC AACCAA GCGACUACUAGACAUA CAGUAUGUCC |
| CL06 | UUGAAGGC UGACCGACCCGC AACCAA GCGACAACUAGACAUA CAGUAUGUCC |
| CL10 | UUGAAGGC UGACCGACCCGC AACCA GCGACAACUAGACAUA CAGUAUGUCC |
| CL19 | UUGAAGGC UGACCGAUCCGC AACCAA GCGACAACUAGACAUA CAGUAUGUCC |
| CL24 | UUGAAGGC CAUCUUGUGUCCUGGACAGCACCGUUCUCCCCCCGGAGGG CAGUAUGUCC |
| CL04 | CAUCUUGUGUCCUGGACAGCACCGUCCUCCCCCCGGAGGG CAGUAUGUCC |
| CL12 | CCAUCUUGUGUCCUGGACAUUACCGUUCUCCCCCCGGAGGG CAGUAUGUCC |
| CL17 | CCAUCUUGUGUCCUGGACAUUACCGUUCUCCCCCCGGAGGG CAGUAUGUCC |
| CL20 | CCAUCUUGUGUCCUGGACAAUGCCGUUCUCCCCCCGGAGGG CAGUAUGUCC |
| CL26 | CCAUCCUGUGUCCUGGACAUCACCGUUCUCCCCCCGGAGGG CAGUAUGUCC |

### 6.2.2.3 Activity assay of CL ribozymes

One ribozyme from each family (CL06, CL12 and CL24) were then selected for further analysis. To first assess their self-biotinylation activity, the ribozymes were 3'-fluorescently labeled and tested in *cis* using a streptavidin gel shift assay. As comparison, the CA13 and MTR1 ribozymes were incubated under identical conditions and included in the analysis. With approximately 30 % shifted RNA after 4 h incubation, CL06 was identified as the only CL ribozyme capable of catalyzing the self-biotinylation reaction (Figure 6.17A). Although most of the sequences obtained belong to the family of CL12, this variant as well as CL24 exhibited no detectable activity even after overnight incubation (Figure 6.16A). However, this is not too surprising, given their predicted secondary structures: Both ribozymes adopt a tight fold, in which the catalytic core forms intense base pairs with the substrate RNA and the 3'-binding arm (Figure 6.16C). It is uncertain how these inactive species survived the selection process and even make up such a large proportion of the round eleven library. One possibility is that they

could act as helper oligonucleotides to improve the catalytic activity of other ribozymes, as reported by Attwater *et al.* for an RNA replicase ribozyme.[277] However, when CL12 was tested for such an activity in combination with the CL06 ribozyme, no accelerating effect was observed (Figure 6.16C). Potentially, the catalytic activity of other variants, which were not identified by Sanger sequencing may be positively influenced by CL12 or CL24.

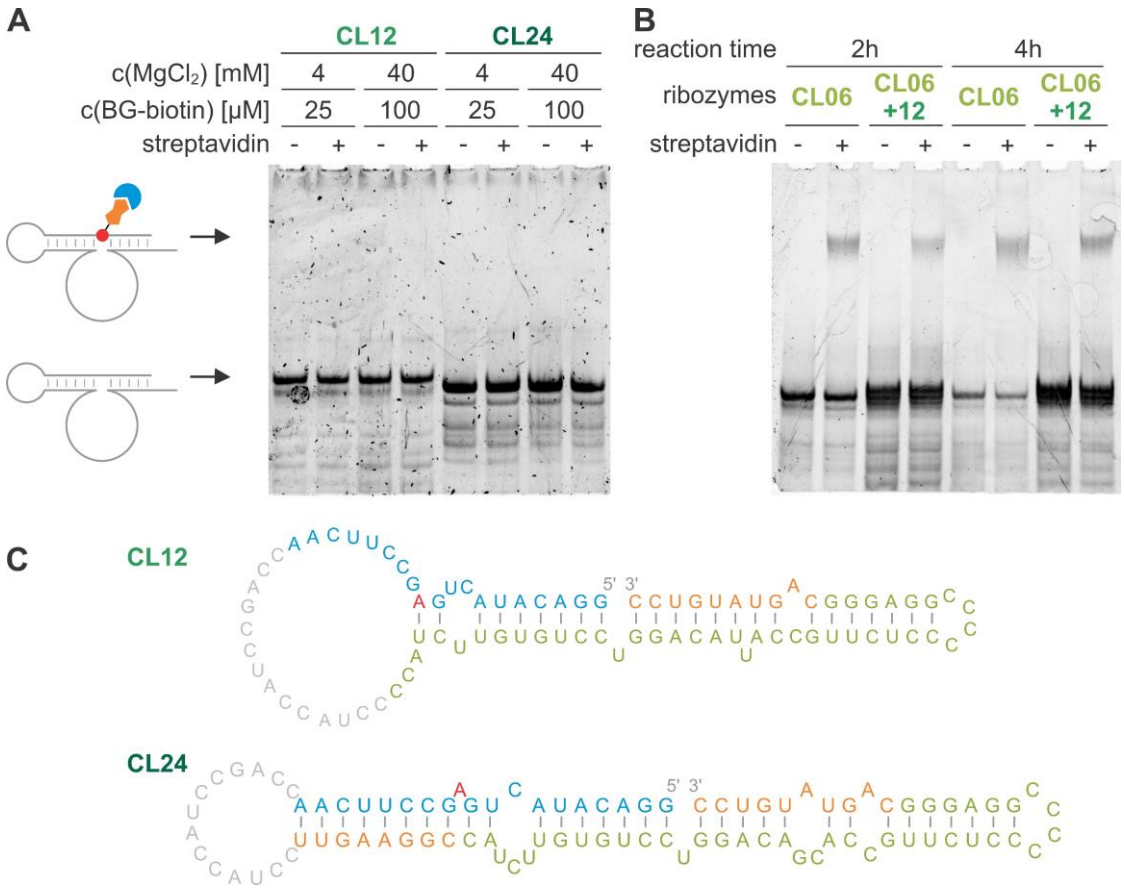



**Figure 6.16:** Activity assays of CL12 and CL24 ribozymes. **A** Streptavidin gel-shift assay of CL12 and CL24. The RNAs were incubated overnight at pH 7.5 and 37 °C before coupling to streptavidin (blue). The desired modification site is depicted in red; biotin is shown in orange. The gel was stained with SYBR gold. **B** Streptavidin gel-shift assay of CL06 by itself or in combination with an equal amount of CL12 to test if CL12 can act as a helper oligonucleotide. RNAs were incubated for 2 h or 4 h at pH 7.5 and 37 °C in the presence of 25 µM BG-biotin and 4 mM $MgCl_2$ before coupling to streptavidin. The gel was stained with SYBR gold **C** Secondary structure predictions of CL12 and CL24. The sequence is colored according to the original starting pool design: The catalytic core is marked green flanked by the binding arms in orange. The connecting loop is colored grey, the substrate sequence is shown in blue with the intended modification site A in red. For CL12, the original 5' binding arm was lost during the selection process. The secondary structures were predicted using the RNAfold web server.[391]

CL06 by its own, however, showed good *cis*-activity with BG-biotin. After 2 h incubation in the presence of 25 µM BG-biotin and 4 mM $MgCl_2$, around 20 % self-biotinylation could be detected (Figure 6.17A). This is comparable to the yields obtained with MTR1, which is also consistent with the overall highly similar secondary structure of both ribozymes (Figure 6.17D). Yet, when CL06 was then tested for modification of an external 17-mer RNA substrate in *trans* using 25 µM BG-$NH_2$ and 4 mM $MgCl_2$, only moderate activity was observed: While near quantitative conversion was observed for MTR1 after 9 h incubation, CL06 only yielded about 50 % modified RNA and showed slower kinetics.

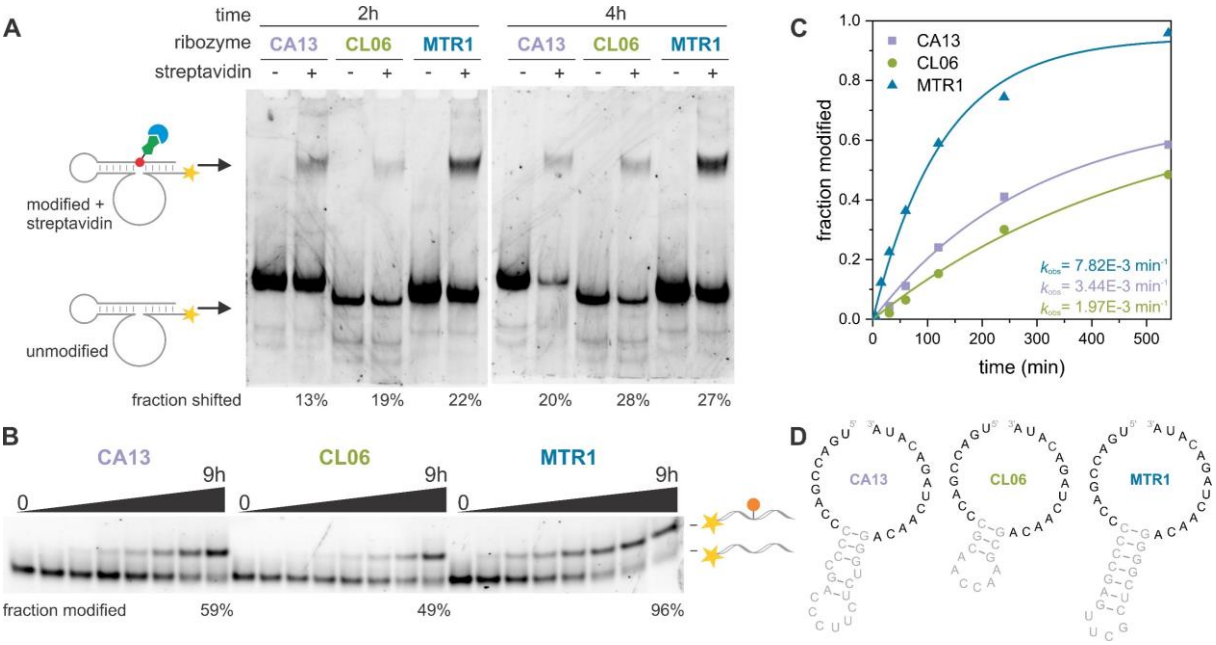



**Figure 6.17:** Characterization of the CL06 ribozyme. **A** Streptavidin gel-shift assay of 3'-fluorescein labeled *cis* active ribozyme CL06 in comparison to CA13 and MTR1. The RNAs were incubated for 2 h or 4 h at pH 7.5 in the presence of 25 µM BG-biotin and 4 mM $MgCl_2$ at 37 °C before coupling to streptavidin (blue). The desired modification site is depicted in red; biotin is shown in dark green. **B** Test of CL06 for activity in *trans* under single turnover conditions. The 5′-fluorescein-labeled 17-mer RNA substrate was reacted with CL06 at pH 7.5 in the presence of 25 µM BG-$NH_2$ and 4 mM $MgCl_2$ at 37 °C. CA13 and MTR1 reacted under identical conditions were included as references. Timepoints: 0, 15min, 30min, 1h, 2h, 4h, 9h. **C** Kinetic data obtained from **B** fitted to pseudo-first order kinetics. **D** Predicted secondary structure of the CL06 catalytic core in comparison to CA13 and MTR1. Identical sequences in the catalytic core are depicted black, the stem-loop structure is sown in gray. The secondary structures were predicted using the RNAfold web server.[391]

### 6.2.2.4 Discussion

In summary, the alkyltransferase reselection CL failed to evolve ribozymes that function under more stringent conditions, mimicking the intracellular environment. The only active ribozyme obtained was CL06, a variant of MTR1 with a shorter stem loop structure. Although it possesses similar labeling characteristics in *cis*, it shows a much lower activity and slower kinetics, when applied in *trans*. Due to the high similarity regarding sequence and secondary structure, CL06 may also be an active methyltransferase ribozyme with the same modification site as MTR1. However, further research is needed to confirm this hypothesis.

Moreover, evolution and survival of the inactive variants CL12 and CL24 that dominated the sequencing results is unexpected. It is further unclear how the CL12 variants lost the 5' binding arm. Neither CL12 nor CL24 were identified as a helper oligonucleotides accelerating the reaction of the active CL06 ribozymes. By altering the substrate RNA and the corresponding binding arms, CL24 may be engineered to show some alkyltransferase activity, as it may then adopt an alternative fold.

In general, the CL selection did not result in a ribozyme with labeling characteristics superior to that of MTR1. To obtain variants that perform well under lo $Mg^{2+}$ and substrate concentrations, even more stringent selection conditions need to be applied. Initiating the CL selection from the second round of selection CA may not have been the best choice. At this stage, the pool was already pre-enriched and had lost some of its complexity. As the first selection round was carried out using high concentrations of $Mg^{2+}$ and BG-biotin, some variants that are capable to work under near-cellular conditions may have already been eliminated. Starting with much more stringent conditions in the first round can therefore be a promising approach for future selections, as it was recently reported for aptamer selections.[393] In addition, selections with alternative small molecule substrates may also result in new alkyl- or methyltransferase ribozymes that function under more stringent conditions.

# 7 Conclusion and Outlook

In biological systems, multiple classes of RNA can be found that play various important roles. Although RNA is only composed of four different nucleosides, its functional diversity is enhanced by post-transcriptional modifications, which generates a second layer of genetic information, often referred to as the 'RNA epitranscriptome'. These chemical alterations, especially methylations strongly influence the function, stability or localization of RNA molecules and are therefore crucial regulators in biological systems.[386-387] Nowadays, methyl groups are installed by protein enzymes using the universal nucleoside derived methyl donor SAM.[54] However, it is believed that in early stages of life, RNA rather than DNA and proteins, was responsible for both, storage of genetic information as well as catalysis.[223-224] In such an RNA-world, post-transcriptional RNA methylation may have also been enabled by RNA enzymes, so-called ribozymes. Catalytically active RNAs can still be found in contemporary biology today where they are mainly responsible for the transformation of phosphodiester bonds.[261-262] However, no methyltransferase ribozymes have been identified until launch of this thesis, neither in nature, nor in the laboratory using *in vitro* evolution techniques. Its discovery would have enormous implications concerning the RNA-world hypothesis as well as early life research in general.

The focus of this work was therefore the development and characterization of efficient RNA catalysts for site-specific RNA modification with special focus on methylations. In the course of this thesis, the first methyltransferase ribozyme (MTR1), which uses $m^6G$ as methyl group donor was developed and further characterized. To identify the chemical structure of the methylated product, a combination of biochemical experiments and analytical methods such as kinetic evaluation of mutated RNAs, alkaline hydrolysis, HPLC, HR-ESI-MS and LC-MS analysis was used. It was discovered that MTR1 targets the *N*1 of a specific adenosine within the target RNA thereby generating $m^1A$. Further characterization of the ribozyme also revealed its broad RNA substrate scope: MTR1 can even be used for the site-specific installation of $m^1A$ in long and highly structured oligonucleotides as demonstrated by the modification of several tRNAs or the 658 nt long tetR RNA transcript. In this context, MTR1 was also shown to target multiple adenosines simultaneously without interfering crosstalk between ribozyme variants with different binding arms. Additionally, the ribozyme was further investigated for future cellular applications: Although direct *in vivo* methylation was limited by the availability of $m^6G$ and the insufficiently low cellular $Mg^{2+}$ concentration, the intact ribozyme was isolated from *E. coli* cells and successfully modified *in vitro*.

In order to investigate the underlying reaction mechanism of MTR1, the structure of the ribozyme was determined by X-ray crystallography. Using a trimolecular construct consisting of a

split ribozyme variant in complex with the target RNA, highly reproducible, diffraction quality crystals were obtained that allowed solving the MTR1 crystal structure to a resolution of 2.8 Å. The obtained structure represents the post-catalytic state of the ribozyme, which contains the guanine leaving group bound to the catalytic centre as well as the methylated RNA substrate. Structurally, the guanine binding site is of particular interest, as it shows great similarities to that of natural purine riboswitches. This finding is impressive, considering that MTR1 was evolved in the laboratory for catalytically activity rather than for tight binding of cofactors in nature. In combination with the experimentally determined pH-dependency of the methylation reaction, the crystal structure also guided the proposal of a reaction mechanism, in which a protonated cytidine within the active site acts as a general acid catalyst. This was further supported by extensive mutagenic biochemical experiments, which also confirm the overall fold of the ribozyme. Additionally, the structure helped to identify two pyrimidine nucleosides within the guanine binding site, which are crucial for the catalysis: 2'-*O*-methylation of these residues resulted in a positive synergistic effect that severely accelerated the reaction rate of the ribozyme, demonstrating that small modifications can have a profound impact on RNA catalysis.

Besides its application for RNA methylation, MTR1 was further repurposed for site-specific aldehyde generation in RNAs. Using MTR1 as an alkyltransferase with BG as the small molecule cofactor in combination with nebularine as a new target nucleoside, a 5-amino-4-formylimidazole residue was generated. Detailed HR-ESI-MS and HPLC analyses revealed that the reaction proceeds *via* an instable *N*1-benzylated purine nucleobase, followed by a ring opening reaction and subsequent hydrolysis. It was then demonstrated that the obtained internal formyl group can be further conjugated with various aldehyde reactive probes to obtain different affinity tagged or fluorescently labeled RNAs. Furthermore, condensation of the aldehyde containing moiety with a 2,2,3-trimethylindole resulted in the formation of a novel and environmentally sensitive hemicyanine fluorophore directly on the RNA in high yields. Although nebularine-containing RNA can easily be produced using solid-phase synthesis, this approach is limited to short oligonucleotides. For easy access to longer RNAs with internal purine ribosides, the corresponding triphosphate was synthesized and enzymatically incorporated into RNA using *in vitro* transcription with T7 RNA polymerase. This approach enabled straight forward synthesis of a nebularine containing full-length tRNA in high yields. Subsequent treatment with MTR1 resulted in highly site-specific aldehyde unmasking even in the presence of multiple nebularines, demonstrating the versatility of this novel MTR1-mediated RNA-labeling approach.

Even though the *in vitro* selected ribozyme CA13 and its engineered variant MTR1 already exhibit high methyltransferase activity, it was attempted to further improve their catalytic properties by conducting multiple reselections. For this purpose, a DNAzyme-based enrichment strategy was developed that allows the direct selection of active methyltransferases. The first

reselections (CD-CG) used pre-structured starting pools that were designed based on the CA13 ribozyme interspersed with sequence stretches of random nucleotides. Although some *trans*-active methyltransferase ribozymes were obtained with this setup, all variants showed lower activity than their ancestor CA13. However, analysis of all active sequences that were obtained revealed that the stability of the stem-loop structure was positively correlated with the catalytic activity. This observation may also guide the rational optimization of other ribozymes generated by alternative *in vitro* selection projects. Consequently, fully, or partially randomized starting pools containing a constant stable stem-loop structure were used in the second reselection approach. Despite regular rounds of negative selection to eliminate non-truly active variants from the library, only three *cis*-active methyltransferases were obtained. Although one of the ribozymes (6CH18) showed high similarities to MTR1 regarding structure and sequence, it was barely active in *trans*.

To create additional RNA-based tools for RNA methylation and alkylation, two different *in vitro* selections were performed: Using BG-biotin as the cofactor and starting pools that were designed to guide the desired modification site to either a cytosine or a guanosine, several active alkyltransferases were obtained. Further analysis revealed that the ribozymes were not biotin dependent and could efficiently work in *trans* to alkylate an external substrate RNA. However, no methyltransferase activity was detectable for any of the tested variants. Moreover, the selections failed to identify C- or G-modifying ribozymes, since all candidates once more targeted an adenosine, initially designed to anneal to the binding arm of the ribozyme. The most promising candidate, CK03, was further characterized regarding its dependency on the pH-value of the medium, the alkyl donor concentration or the substrate scope. Compared to CA13, similar labeling characteristics were observed. However, the overall activity of CK03 was lower and the ribozyme was less robust regarding a reduced cofactor concentration. Concerning the $Mg^{2+}$ dependency, the exact chemical composition of the modified product and the substrate RNA requirements of CK03, further research is needed. This may for example be achieved by additional kinetic studies, atomic mutagenesis of the original substrate sequence and digestion of the modified RNA product in combination with LC-MS analysis.

Finally, another reselection of CA13 was performed to obtain a ribozyme variant suitable for future *in vivo* applications. Thus, the medium of the self-biotinylation reaction contained only low $Mg^{2+}$ and BG-biotin concentrations. The evolved ribozyme CL06 showed strong structural similarities to CA13 and MTR1. Moreover, it is equally active in *cis*, when tested under stringent reaction conditions. However, when challenged to modify an RNA substrate in *trans*, no improved activity was observed. Therefore, especially in the first rounds, even more stringent reaction conditions should be applied in future reselections. In addition, to prevent evolution of another CA13 variant, alternative new starting pool designs or alkyl donors can be used.

In summary, this work reports on the *in vitro* selection, characterization and application of highly active methyl- and alkyltransferase ribozymes. The findings presented herein demonstrate that RNA, just like proteins, can form defined three dimensional structures, which enables efficient and specific catalysis. Natural ribozymes as well as riboswitches are thought to be remnants of an ancient RNA-world. The fact that the scope of RNA-catalyzed reactions extends far beyond phosphodiester cleavage and ligations but also include fundamental chemical transformations such as the installation of a simple methyl group, provides strong support for this hypothesis. Furthermore, the discovery, that 2'-*O*-methylation at defined positions of the MTR1 ribozyme can greatly enhance its catalytic activity suggests, that modified nucleotides may have also been highly advantageous for ancient RNA catalysts.

In this regard, starting pools that are spiked with xeno nucleic acid residues may likely yield ribozymes not only with enhanced stability, but also with superior catalytic performance. However, the selection of such hybrid RNA-XNAzymes is challenging due to the need for site-specific incorporation of XNA building blocks. Nevertheless, with the steady development of new engineered polymerases for XNA synthesis and amplification, this may be a future exciting direction in the field of ribozyme selection. Alternatively, the development of a method for screening the effect of single XNA incorporation into existing ribozymes is also in high demand.

Finally, as until today three independent groups have demonstrated that RNA can catalyze the transfer of a one-carbon unit to another RNA[95, 185-186], the next step would be to develop a catalytically active RNA for the reverse reaction (e.g., the removal a methyl group from RNA). Such a ribozyme would act as an RNA-repair enzyme and could have also been greatly beneficial at the early stages of life. The DNAzyme based reselection developed in this work can potentially be repurposed to achieve this goal, when performed in the opposite direction: The cleaved and therefore unmethylated RNA needs to be isolated and amplified. This may allow the selection of $m^1A$ or $m^6A$ demethylating ribozymes.

Overall, this work establishes RNA in general, and MTR1 in particular, as a valuable tool for efficient and highly site-specific post-transcriptional methylation of RNA. Future research on RNA-methylating ribozymes will further expand our understanding of RNA biology, including ways of substrate recognition as well as the underlying catalytic mechanisms. Given the large number of natural modifications found in different types of RNA, numerous ribozymes that are able to catalyze their installation remain to be discovered. They may be hidden somewhere in a large RNA sequence space, but with the development of an appropriate selection method, they can likely be found.

# 8 Zusammenfassung und Ausblick

In biologischen Systemen lassen sich unterschiedlichste Klassen an RNA finden, welche alle verschiedene, essenzielle Funktionen besitzen. Obwohl sich alle RNA-Moleküle nur aus vier verschiedenen Nukleosiden zusammensetzen, wird deren funktionelle Vielfalt durch eine Vielzahl posttranskriptioneller Modifikationen erweitert. Dies lässt eine zweite Ebene genetischer Informationen entstehen: das „RNA-Epitranskriptom". Chemische Veränderungen der RNA, insbesondere Methylierungen, haben einen immensen Einfluss auf die Funktionalität, Stabilität oder Lokalisation einzelner RNA-Molekülen innerhalb eines Organismus.[386-387] In unserer heutigen, modernen Biologie werden Methylgruppen durch Protein-Enzyme unter Zuhilfenahme des universalen Methylgruppendonors SAM angebracht.[54] Es wird jedoch davon ausgegangen, dass sehr frühe Stadien des Lebens vollkommen auf RNA basierten.[223-224] Somit waren es zu dieser Zeit auch RNAs und nicht Proteine, die für die Katalyse chemischer Reaktionen sowie Transformationen verantwortlich waren. In einer solchen RNA-Welt wurden konsequenterweise wahrscheinlich auch posttranskriptionelle RNA-Methylierungen durch RNA-Enzyme, sogenannte Ribozyme, angebracht. Katalytisch aktive RNA ist auch noch in heutigen, modernen Lebensformen zu finden. Jedoch ist sie dort hauptsächlich für die Bildung und Spaltung von Phosphodiesterbindungen verantwortlich.[261-262] Zu Beginn dieser Arbeit, war weder ein natürliches noch ein im Labor entwickeltes Ribozym mit Methyltransferase-Aktivität bekannt. Dessen Entdeckung wäre jedoch von enormer Bedeutung sowohl für die RNA-Welt-Hypothese als auch für Forschungen zu ersten, primitiven Lebensformen.

Der Fokus dieser Doktorarbeit lag daher auf der Entwicklung and Charakterisierung effizienter RNA-Katalysatoren für die spezifische Methylierung von RNA. Hierbei wurde das erste Methyltransferase-Ribozym namens MTR1 entwickelt und im Detail charakterisiert. Das Ribozym überträgt die Methylgruppe des Kofaktors $O^6$-Methylguanin ($m^6G$) auf ein definiertes Adenosin innerhalb einer Ziel-RNA. Um jedoch die exakte Konstitution des methylierten Reaktionsproduktes herauszufinden, wurden unterschiedliche biochemische Experimente mit verschiedenen analytischen Verfahren kombiniert. Diese beinhalteten unter anderem kinetische Untersuchungen von Ribozym- und Substrat-Mutanten, alkalische Hydrolysen sowie HPLC, HR-ESI-MS and LC-MS basierte Analysen des RNA-Produkts. Dies ermöglichte die Identifizierung der Modifikationsstelle von MTR1 als die *N*1-Position eines Adenosins innerhalb der Ziel-RNA. Das Ribozym generiert somit die natürlich vorkommende Modifikation $m^1A$. Zudem konnte gezeigt werden, dass MTR1 ein sehr breites Spektrum an RNA-Substraten akzeptiert. Das Ribozym kann selbst stark strukturierte und sehr lange RNA-Stränge, wie beispielsweise tRNAs oder das 658 nt lange tetR RNA-Transkript methylieren. In diesem Zusammenhang konnte

auch gezeigt werden, dass MTR1 in der Lage ist innerhalb einer RNA eine Vielzahl verschiedener Andenosine zu methylieren, ohne dass sich dabei Varianten mit unterschiedlichen Bindungsarmen gegenseitig inhibieren. Auch für einen möglichen Einsatz von MTR1 in Zellen wurden bereits erste Schritte unternommen: Eine direkte *in vivo* Anwendung war zwar durch die geringe Verfügbarkeit von $m^6G$ in der Zelle sowie der unzureichenden intrazellulären $Mg^{2+}$-Konzentration noch nicht möglich. Allerdings konnte das Ribozym erfolgreich von *E. coli* Bakterien produziert, daraus isoliert und anschließend für *in vitro* Methylierungsreaktionen verwendet werden.

Um einen Einblick in den zugrundeliegenden Reaktionsmechanismus von MTR1 zu bekommen, wurde anschließend mit Hilfe von Röntgenstrukturanalyse die dreidimensionale Struktur des Ribozyms untersucht. Ein Konstrukt bestehend aus einer zweigeteilten Variante von MTR1 sowie der Substrat-RNA ermöglichte die reproduzierbare Erzeugung von gut streuenden Kristallen hoher Qualität. Mit deren Hilfe konnte die Kristallstruktur von MTR1 mit einer Auflösung von 2.8 Å gelöst werden. Diese stellt den post-katalytischen Zustand des Ribozyms dar, d.h. die Übertragung der Methylgruppe hat bereits stattgefunden. Insbesondere das katalytische Zentrum, in dem sich das demethylierte Guanin befindet, ist von besonderem Interesse: Es zeigt große strukturelle Ähnlichkeiten zur Bindungstasche, die man auch in natürlich vorkommenden Purin-Riboschaltern findet. Diese entstanden in der Natur mit dem Ziel möglichst effizient kleine Moleküle zu binden. MTR1 jedoch stammt aus dem Labor und wurde evolviert, um katalytisch aktiv zu sein, was dessen Ähnlichkeit zu Riboschaltern noch erstaunlicher macht. Des Weiteren ermöglichte die Kristallstruktur zusammen mit der experimentell bestimmten pH-Abhängigkeit die Ableitung eines plausiblen Reaktionsmechanismus: Ein protoniertes Cytidin, das Bestandteil des aktiven Zentrums und in direktem Kontakt mit dem Kofaktor ist, kann als Säure fungieren. Es hilft dadurch die Methylgruppe auf das Adenin zu übertragen und dabei Guanin als Abgangsgruppe zu generieren. Weitere Belege für diesen Mechanismus stammten aus zahlreichen zusätzlichen biochemischen Experimenten, die zudem die erhaltene dreidimensionale Struktur des Ribozyms zusätzlich untermauern. Diese Erkenntnisse trugen auch dazu bei, zwei Pyrimidine innerhalb der Guanin-Bindungstasche zu identifizieren, welche einen starken Einfluss auf die katalytische Aktivität von MTR1 ausüben: Eine Methylierung dieser beiden 2'OH Positionen hat einen positiven synergistischer Effekt zur Folge. Dieser führt zu einer massiven Steigerung der Reaktionsrate des Ribozyms. Somit konnte gezeigt werden, dass bereits kleinste Modifikationen einen enormen Einfluss auf die Aktivität katalytischer RNA haben können.

Neben der Verwendung von MTR1 für die Methylierung von RNA-Molekülen, wurde es auch für die positions-spezifische Einführung von Aldehyd-Gruppen in RNA umfunktioniert. In seiner Funktion als Alkyltransferase, unter Verwendung von $O^6$-Benzylguanin (BG) als Kofaktor,

wurde es erfolgreich dazu verwendet 5-Amino-4-formylimidazol aus dem Nukleosid Nebularin zu generieren. Durch zahlreiche intensive Analysen unter anderem mittels HR-ESI-MS und HPLC konnte gezeigt werden, dass die Reaktion über die *N*1-benzylierte Purin Nukleobase als instabile Zwischenstufe verläuft. Diese wird sofort einer nukleophilen Ringöffnungsreaktion unterzogen und abschließend zum Endprodukt hydrolysiert. Der entstandene Aldehyd kann daraufhin weiter funktionalisiert werden, was beispielsweise das Anbringen von Biotin oder verschiedener Fluorophore an die RNA ermöglicht. Zudem erlaubte das neu entstandene 5-Amino-4-formylimidazol durch die Kondensation mit 2,2,3-Trimethylindol die effiziente Synthese eines neuartigen Hemicyanin-Farbstoffs direkt an der intakten RNA. Zwar kann RNA, die Nebularin beinhaltet, problemlos über Festphasensynthese hergestellt werden, allerdings ist dies nur für kurze Oligonukleotide effizient. Um auf einem einfachen Weg auch längere modifizierte RNAs erhalten zu können, wurde das entsprechende Nebularin-Triphosphat synthetisiert. Dieses konnte anschließend mit hohen Ausbeuten enzymatisch über eine *in vitro* Transkription mittels T7 RNA-Polymerase eingebaut werden, was die einfache und effiziente Herstellung einer tRNA mit internen Nebularin-Modifikationen in hohen Ausbeuten ermöglichte. Als diese anschließend mit MTR1 und BG umgesetzt wurde, konnte die spezifische Entstehung des entsprechenden Aldehyds beobachtet werden. Dies war selbst dann der Fall, wenn mehrere Nebularin-Nukleoside als mögliche Modifikationsstellen vorhanden waren. Somit konnte nochmals eindeutig das Potential und die Vielseitigkeit dieser neuartige MTR1-katalysierten Methode zur RNA-Markierung aufgezeigt werden.

Zwar besitzen das *in vitro* selektierte Ribozym CA13 sowie seine weiterentwickelte Variante MTR1 bereits eine hohe Methyltransferase-Aktivität. Dennoch wurde versucht ihre katalytischen Eigenschaften durch unterschiedliche Reselektionen noch weiter zu verbessern. Hierzu wurde eine Anreicherungsstrategie für methylierte RNA entwickelt, die auf der katalytischen Aktivität eines Desoxyribozyms beruht. Dieses ermöglicht die direkte Selektion aktiver Methyltransferasen. Der Umweg über die Selektion von Alkyltranferasen, welche erst anschließend in Methyltransferasen umgewandelt werden, ist somit nicht mehr notwendig. Die ersten Reselektionen (CD-CG) verwendeten ein Pool-Design, welches auf der Struktur und Sequenz von CA13 basierte. Generell wurde das ursprüngliche Ribozym beibehalten, lediglich einzelne randomisierte Sequenzabschnitte wurden an unterschiedlichen Positionen eingefügt. Obwohl eine dieser Selektionen einige *trans*-aktive Methyltransferasen hervorbrachte, zeigten diese lediglich eine geringere Aktivität als ihr Vorläufer-Ribozym CA13. Eine genauere Analyse bezüglich der Struktur aller aktiven Ribozyme führte jedoch zur wichtigen Erkenntnis, dass die Stabilität ihrer Haarnadelstruktur eine positive Korrelation mit ihrer Methyltransferase-Aktivität aufwies. Dieser Befund ist vor allem auch für zukünftige Projekte von Bedeutung, in denen durch Selektion evolvierte Ribozyme weiter optimiert werden sollen.

Infolgedessen wurden für die zweite Runde an Reselektionen ein Pooldesign verwendet, welches eine konstante, stabile Haarnadelstruktur beinhaltete. Die verbliebenen Nukleoside wurden hingegen vollständig oder teilweise randomisiert. Trotz regelmäßig durchgeführter Runden zur Negativ-Selektion, die dabei helfen sollten alle Varianten zu eliminieren, die nicht aktiv waren, konnten nur drei *cis*-aktive Methyltransferase erhalten werden. Auch wenn eines dieser Ribozyme (6CH18) große Ähnlichkeiten zu MTR1 hinsichtlich Sequenz und Struktur aufwies, besaß es jedoch kaum *trans*-Aktivität.

Um weitere RNA-basierte Werkzeuge zur Methylierung oder Alkylierung von RNA zu erhalten, wurden zwei weitere *in vitro* Selektionen durchgeführt. Hierbei wurde wie auch schon in der CA Selektion BG-Biotin als Kofaktor sowie ein Pooldesign, das 40 randomisierte Nukleotide enthielt, verwendet. Allerdings sollte dieses Mal anstelle von Adenosin ein Cytosin (CJ) oder ein Guanosin (CK) modifiziert werden. Daher befand sich gegenüber der randomisierten Region jeweils ein C oder ein G, welches keine Basenpaarung mit den Ribozym-Bindungsarmen aufwies. Insgesamt gingen aus diesen Selektionen drei Alkyltransferasen hervor, welche jeweils *trans*-aktiv waren und zudem BG sowie BG-$NH_2$, jedoch kein $m^6G$ als Kofaktoren akzeptierten. Eine genauere Analyse der exakten Modifikationsstelle zeigte allerdings, dass alle Ribozyme erneut ein Adenosin und nicht wie gewünscht ein Cytosin oder Guanosin modifizierten. Das Ribozym CK03 wurde anschließend zur näheren Charakterisierung ausgewählt. Bezüglich der Abhängigkeit vom pH-Wert und der Kofaktor-Konzentration zeigte es ähnliche Trends wie CA13. Allerdings wies es insgesamt eine deutlich geringere Aktivität als CA13 auf und war zudem weniger robust gegenüber einer Verringerung der Kofaktor-Konzentration. Wie flexibel die Substrat-RNA für CK03 gewählt werden kann, bedarf allerdings noch weiterer Analysen. Dasselbe gilt für die exakte chemische Zusammensetzung des alkylierten RNA-Produkts sowie der $Mg^{2+}$-Abhängigkeit. Dies kann beispielsweise mit Hilfe weiterer kinetischer Analysen oder durch die Verwendung verschiedener Substrat-RNA Mutanten näher untersucht werden. Auch ein enzymatischer Verdau des Produktes gefolgt von einer anschließenden Analyse mittels LC-MS kann weitere Erkenntnisse bringen.

Schließlich wurde noch eine weitere Reselektion von CA13 durchgeführt (CL). Hierbei war das Ziel ein Ribozym zu erhalten, das für zukünftige Anwendungen *in vivo* geeignet ist. Folglich wurde die Biotinylierungsreaktion in Gegenwart von nur geringen Konzentrationen an $Mg^{2+}$ and BG-biotin durchgeführt. Die Selektion resultierte in einem aktiven Ribozym namens CL06, welches starke Ähnlichkeiten zu CA13 und MTR1 bezüglich Struktur und Sequenz ausweist. Aktivitätstests unter stringenten Reaktionsbedingungen ergaben, dass CL06 in *cis* genauso aktiv ist wie seine Vorläufer-Ribozyme. Als es jedoch für Alkylierung in *trans* getestet wurde, konnte nur eine geringere Aktivität im Vergleich zu MTR1 beobachtet werden. Folglich sollte für zukünftige Reslektionen solcher Art bereits in den ersten Runden noch stringentere

Bedingungen verwendet werden. Um zudem zu verhindern, dass in Zukunft immer wieder lediglich CA13-Varianten evolvieren, könnte außerdem die Verwendung eines neuen Pool-Designs sowie alternativer Alkyldonor-Substrate von Nutzen sein.

Zusammenfassend zeigt diese Arbeit das große Potential katalytisch aktiver RNA insbesondere für die zielgenaue RNA-Methylierung auf. Die hier präsentierten Ergebnisse veranschaulichen unter anderem, dass RNA, genau wie Proteine, definierte dreidimensionale Strukturen ausbilden kann, was eine effiziente und spezifische Katalyse von Reaktionen ermöglicht. Sowohl Riboschalter als auch natürliche Ribozyme können als Relikte einer präbiotischen RNA-Welt angesehen werden, welche der heutigen modernen DNA- und Protein-basierten Biologie vorausgegangen ist. Diese Hypothese wird auch dadurch unterstützt, dass die Bandbreite RNA-katalysierter Reaktionen weit über die Knüpfung und Spaltung von Phosphodiester-Bindungen hinausgeht. Mit Hilfe von *in vitro* Selektionen konnte gezeigt werden, dass RNA auch in der Lage ist fundamentale chemische Reaktionen wie die Methylierung von RNA zu katalysieren. Zudem ist in diesem Zusammenhang außerdem von großer Bedeutung, dass die katalytische Aktivität von MTR1 allein durch das Anbringen von zwei kleinen Methylgruppen erheblich gesteigert werden konnte. Diese legt nahe, dass modifizierte Nukleotide auch in einer früheren RNA-Welt die Katalyse-Aktivität einzelner Ribozyme erheblich verbessert haben könnte.

Diesbezüglich ist es auch wahrscheinlich, dass mit Selektions-Bibliotheken, die Xeno-Nukleoside enthalten, neuartige Ribozyme erhalten werden können, welche nicht nur eine höhere Stabilität, sondern auch eine bessere katalytische Aktivität aufweisen. Eine Selektion mit solchen RNA-XNA-Hybridsequenzen durchzuführen ist jedoch herausfordernd und nicht trivial, da der Einbau der XNA-Bausteine positionsspezifisch erfolgen muss. Mit der stetigen Neu- und Weiterentwicklung von Polymerasen für die XNA-Synthese und -Amplifikation könnte dies jedoch in Zukunft eine spannende und vielversprechende Richtung auf dem Gebiet der Ribozymselektionen sein. Des Weiteren ist auch die Entwicklung eines Verfahrens, mit dem der Effekt einzelner XNA-Substitutionen auf die Aktivität bereits bestehender Ribozyme schnell und einfach analysiert werden kann, von großem Interesse.

Da inzwischen unabhängig voneinander bereits drei unterschiedliche Gruppen zeigen konnten, dass RNA in der Lage ist die Übertragung einer Methylgruppe auf eine Ziel-RNA zu katalysieren[95, 185-186], besteht folglich der nächste Schritt in der Entwicklung eines Ribozyms für die Umgekehrte Reaktion: die RNA-Demethylierung. Ein solches Ribozym würde als RNA-Reparaturenzym fungieren und könnte bereits in einer RNA-basierten Welt von großem Nutzen gewesen sein. Für dessen Evolution könnte unter anderem die hier entwickelte Desoxyribozym-basierte Selektionsstrategie Verwendung finden. Sie muss lediglich in umgekehrter Richtung eingesetzt werden: Durch die Isolation und Amplifikation der geschnittenen und damit

unmethylierte RNA könnte diese Methode dazu beitragen $m^1A$- oder $m^6A$-demethylierende Ribozyme zu selektieren.

Zusammenfassend zeigt diese Arbeit, dass RNAs im Allgemeinen sowie MTR1 im Besonderen wertvolle Werkzeuge sind, die unter anderem dabei helfen hochgradig effizient und spezifisch Methylgruppen an RNA anzubringen. Künftige Forschungen an RNA-methylierenden Ribozymen können unter anderem dabei helfen das Wissen auf dem Gebiet der RNA-Biologie weiter zu vertiefen. Dies schließt vor allem die Art und Weise der Substraterkennung sowie die zugrunde liegenden katalytischen Mechanismen ein. Angesichts der enormen Anzahl natürlicher Modifikationen, die in unterschiedlichen Arten von RNA zu finden sind, warten noch unzählige Ribozyme darauf entdeckt zu werden, die diese Modifikationen generieren können. Sie mögen im großen RNA-Sequenzraum versteckt sein, mit einer geeigneten Selektionsmethode können sie jedoch vermutlich gefunden werden.

# 9 Appendix

## 9.1 Supporting information for chapter 3

This Supporting Information has been published in ref.[95]. For the sake of unity of this thesis, several editorial changes have been made, which, however, do not affect the contents of the thesis.



### 9.1.1 Material and Methods

#### RNA synthesis

RNA oligonucleotides were prepared by solid-phase synthesis using phosphoramidite chemistry (2'-*O*-TOM-protected) on controlled-pore glass solid supports.[394] RNA/DNA sequences are given in Table 9.1 and Table 9.2. Modified phosphoramidites for atomic mutagenesis and synthesis of reference oligonucleotides were purchased or prepared in house, following published procedures.[395-397] RNA oligonucleotides were deprotected with ammonia/methyl amine (AMA), followed by 1 M tetrabutylammonium fluoride in THF, desalted and purified by denaturing polyacrylamide gel electrophoresis. Mild deprotection conditions were used for $m^1A$ RNA (3.5 M $NH_3$ in MeOH, at 25 °C for 72 h) to avoid Dimroth rearrangement during deprotection, and for methylphosphate-modified RNA (0.05 M $K_2CO_3$ in MeOH at 25 °C for 7 h) to avoid loss of the phosphotriester. Quality of RNAs (purity and identity) was analyzed by anion exchange HPLC (Dionex DNAPac PA200, 2x250 mm, at 60 °C. Solvent A: 25 mM Tris-HCl (pH 8.0), 6 M Urea. Solvent B: 25 mM Tris-HCl (pH 8.0), 6 M Urea, 0.5 M $NaClO_4$. Gradient: linear, 0-40 % solvent B, 4 % solvent B per 1 CV) and HR-ESI-MS (micrOTOF-Q III, negative mode, direct injection). Measured and calculated masses are listed in Table 9.4.

Unmodified RNA substrates and tRNAs were prepared by *in vitro* transcription with T7 RNA polymerase (prepared according to ref[398] with minor modifications) from synthetic DNA templates (purchased from Microsynth), following standard procedures with 4 mM NTPs and 30 mM $MgCl_2$.[184]

### *In vitro* selection

The DNA template for *in vitro* transcription of the initial RNA library was assembled from two DNA oligonucleotides (D2+D3, $N_{40}$: A:C:G:T=1:1:1:1) by overlap extension using Klenow fragment with the sequence of the connecting loop acting as the overlapping region. The dsDNA template (450 pmol) was used for *in vitro* transcription with T7 RNA polymerase in a final volume of 450 µl. For the first selection round, 3.3 nmol RNA pool (containing 10 % 3'-fluorescently labeled RNA, obtained by sodium periodate oxidation and reaction with Lucifer yellow carbohydrazide, according to ref.[25]) were folded in selection buffer (120 mM KCl, 5 mM NaCl, 50 mM HEPES, pH 7.5; 3 minutes at 95 °C, then 10 minutes at 25 °C). Biotinylated *O*$^6$-benzylguanine (BG-biotin New England Biolabs) and $MgCl_2$ were added (100 µM and 40 mM final concentrations, respectively) to a final reaction volume of 60 µL and the reaction mixture was incubated at 37 °C for 16 h. In subsequent rounds, the incubation time, the amount of RNA, and the concentration of the biotinylated substrate were reduced in order to increase the selection pressure. After precipitation by ethanol, the biotinylated RNAs were captured using either neutravidin- or streptavidin-coated magnetic beads (Dynabeads, Thermo Fisher Scientific, ca. 1 nmol RNA per mg of beads), eluted with formamide, and amplified by RT-PCR, following established procedures.[184, 285] *In vitro* transcription was performed (total volume of 100 µL), followed by PAGE purification to prepare the enriched RNA library for the next selection round. After 11 rounds of selection, the library was cloned (TOPO-TA cloning), and ribozymes generated from randomly picked colonies were examined for catalytic activity (by streptavidin gel shift assay on native PAGE) and sequenced. Three sequence families were identified (Table 9.5), two of which retained catalytic activity *in trans* (i.e., in an intermolecular setup upon removing the connecting loop between binding arm and substrate sequence), named CA13 and CA21.

### Kinetic assays of RNA-catalyzed RNA methylation reactions

Single turnover assays were performed as described previously with a 10-fold excess of ribozyme over the target RNA.[184] Briefly, 10 pmol ($^{32}$P- or fluorescein-labeled) RNA target were mixed with 100 pmol of the corresponding ribozyme in 10 µL of selection buffer (120 mM KCl, 5 mM NaCl, 50 mM HEPES, pH 7.5) including 100 µM of substrate (BG-$NH_2$, BG or m$^6$G) and 40 mM $MgCl_2$. To ensure proper folding and formation of the ribozyme-substrate RNA complex, an annealing step (3 min at 95°C, 10 min at 25 °C) was performed prior to addition of $MgCl_2$ and the small molecule substrate. The mixture was incubated at 37 °C and 1 µL aliquots were taken at desired timepoints and quenched immediately by adding 4 µL of stop solution. Half of each timepoint sample was analysed by PAGE (20 %), and band intensities were quantified

by phosphorimaging or by fluorescence imaging using blue epi illumination and 530/28 nm emission filter. The yield versus time data were fit to (fraction reacted) = $Y(1 - e^{-kt})$, where k = $k_{obs}$ and Y = final yield using KaleidaGraph (4.3) or Origin (2019). All kinetic assays were carried out as three independent replicates, and representative gel images are shown.

**Analysis of the RNA methylation products**

From a 20 µL methylation reaction with 1 nmol target RNA, 1.2 nmol ribozyme, 100 µM $m^6G$ and 40 mM $MgCl_2$ at pH 7.5 (120 mM KCl, 5 mM NaCl, 50 mM HEPES) for 21 h at 37 °C, the methylated RNA product was isolated by PAGE, and subjected to HR-ESI-MS (Bruker microOTOF-Q III, direct injection), RNase T1 digestion (150 IPS of 5'-$^{32}$P-RNA were digested with 0.5 U RNase T1 in 5 µL 50 mM Tris (pH 7.5) for 30 sec at 37 °C), and alkaline hydrolysis (250 IPS of 5'-$^{32}$P-RNA in 5 µL 25 mM NaOH were incubated at 95 °C for 5 min). Dimroth rearrangement was examined in a volume of 5 µL with 90 IPS of 5'-$^{32}$P-RNA in 25 mM $Na_2CO_3$ buffer (pH 10) with 1 mM EDTA at 65 °C or 1 h. After quenching with high dye gel loading buffer, the samples were resolved on denaturing PAGE and visualized by autoradiography.

**tRNA methylation and primer extension assays**.

*In vitro* transcribed tRNA (10 pmol) was annealed with the corresponding ribozyme (100 pmol) and optional disruptor oligo (25 pmol), and then incubated in a final volume of 10 µL of 1x selection buffer (120 mM KCl, 5 mM NaCl, 50 mM HEPES, pH 7.5) including 100 µM of $m^6G$ and 40 mM $MgCl_2$, at 25°C for 22 h. Disruptor oligos were used for *B. subtilis* tRNA-Ser and *R. norvegicus* tRNA-Lys. The unmodified tRNA reference samples were prepared analogously, but without addition of $m^6G$. RT primer extension stop experiments were carried out with 4 pmol of the methylated or the unmodified tRNA, and the appropriate 5'-$^{32}$P-labeled primer (100 IPS, ca. 4 pmol). After annealing in 5 mM Tris-HCl (pH 7.5) and 0.1 mM EDTA, the sample was combined with 5 mM DTT, 0.5 mM of each dNTP and 50 U of SuperScript III RT (ThermoFisher Scientific) in 1x first strand buffer (50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$) to yield a final reaction volume of 10 µL. After incubation at 55 °C for 1 h, the reaction was stopped by adding 1 µL of 2 N NaOH and incubation at 95 °C for 5 min. RT primer extension on total *E. coli* tRNA was carried out for 105 min at 42 °C using 1 µg total *E. coli* tRNA, followed by workup as above. The primer extension products were recovered by ethanol precipitation, dissolved in high dye solution, and resolved on 15 % or 20 % denaturing PAGE. Sequencing ladders were prepared in analogy with suitable dNTP/ddNTP mixtures (0.5 mM

ddNTP, 0.05 mM corresponding dNTP, 0.5 mM each of the other three dNTPs), and analyzed in parallel.

**RNA structure probing by DMS and SHAPE**

MTR1 (Rz3) hybridized to unreactive 17-nt RNA (R6) in 10 µL selection buffer with $MgCl_2$ (40 mM) was treated with DMS or 1M7 in absence or presence of $m^6G$ (100 µM). For DMS probing, 0.5 µL DMS solution (5 % in EtOH) was added and incubated for 1 h at 25 °C. The reaction was quenched by the addition of 10 µL of 1 M 2-mercaptoethanol and 1.5 M NaCl. SHAPE probing was performed by addition of 1 µL 1M7 solution (130 mM in dry DMSO; synthesized according to ref.[399]) and 50 min incubation at 37 °C. After ethanol precipitation, the modification pattern was analyzed by primer extension as described above, using 5'-$^{32}$P-labeled primer (D4).

**Construction of F30-Broccoli-MTR1-containing plasmids and expression in *E. coli***

The F30-Broccoli-MTR1 constructs were prepared by overlap extension of synthetic DNA oligonucleotides, amplified by PCR and inserted into a pET14 vector using restriction enzymes BglII and BlpI. The sequence of the insert and successful ligation into the plasmid was confirmed by Sanger sequencing. The F30-Broccoli-MTR1 plasmid was transformed into *E. coli* BL21(DE3) cells, and expression was induced by addition of 1 mM IPTG. After 1 h incubation at 37 °C, total *E.coli* RNA was extracted as previously reported.[285] A fraction (0.5 µg) was analyzed by 10 % denaturing PAGE, that was stained with a solution of 20 µM DFHBI in 100 mM KCl, 5 mM $Mg^{2+}$, 40 mM HEPES, pH 7.5, for 15 min, and imaged on a ChemiDoc imager. Afterwards the gel was stained with SYBR gold and imaged again to visualize all RNA and the size marker.

For testing the activity of the F30-Broccoli-MTR1 constructs, 200 ng of total cellular RNA was incubated *in vitro* at 37 °C for 4 h in 10 µL of selection buffer (120 mM KCl, 5 mM NaCl, 50 mM HEPES, pH 7.5) including 100 µM $m^6G$ or BG and 40 mM $MgCl_2$. RT primer extension experiments were then performed as described above for probing of the modification site.

**LC-MS analysis of MTR1-catalyzed methylation**

For LC-MS analysis, 30 µg total *E. coli* tRNA were mixed with 5.4 µg (300 pmol) of tRNA-Asp-A58-specific ribozyme in 10 µL of selection buffer (120 mM KCl, 5 mM NaCl, 50 mM HEPES,

pH 7.5) including 100 µM $m^6G$ and 40 mM $MgCl_2$. An annealing step (3 min at 95°C, 10 min at 25 °C) was performed prior to addition of $m^6G$ and $MgCl_2$. After 22 h incubation at 25 °C the RNA was digested for 18 h at 37 °C using 7.5 U bacterial alkaline phosphatase and 2.0 U snake venom phosphodiesterase in 40 mM Tris-HCl, pH 7.5 in the presence of 20 mM $MgCl_2$. The unmodified reference was generated by digestion of 30 µg unmodified *E. coli* tRNA. After extracting the sample twice with chloroform, the aqueous layer was concentrated, and an aliquot was analysed by LC-MS, using an RP-18 column (Synergi, 4 µm Fusion-RP C18 80 Å, 250 x 2 mm; Phenomenex) at 25 °C with aqueous mobile phase A (5 mM $NH_4OAc$, pH 5.3) and organic mobile phase B (100 % acetonitrile). The flow rate was 0.2 mL/min with a gradient of 0-5 % B in 15 min, followed by 5-70 % B in 30 min. The micrOTOF-Q III with an ESI ion source was operated in positive ion mode, with capillary voltage of 4.5 kV, end plate offset of 500 V, nitrogen nebulizer pressure 1.4 bar, dry gas flow 9 L/min, and dry temperature 200 °C. Data were analyzed with Data Analysis software DA 4.2 (Bruker Daltonics).

Analysis of F30-Broccoli-*cis*-MTR1 and F30-Broccoli-*trans*-tRNA-Asp-MTR1 were performed analogously using 200 pmol of *in vitro* transcribed constructs that were incubated at 37 °C for 22 h in the presence of 100 µM $m^6G$ (*cis*) or BG (*trans*). Synthetic reference nucleosides $m^1A$, $m^6A$, $bn^1A$ and $bn^6A$ (synthesized in analogy to literature-known procedures),[397, 400] were injected at a concentration of 50 nM.

**Transcription and MTR1-catalyzed methylation of tetR and blaR RNAs**

The template for *in vitro* transcription was generated from the respective plasmid using the primer pairs P19 and P20 for tetR or P21 and P22 for blaR. For the PCR, 150 ng plasmid were mixed with 50 pmol of each primer, 0.2 mM dNTP-mix, and 1.25 U DreamTaq polymerase in a total reaction volume of 50 µL. Amplification was carried out under the following cycling conditions:

| 95 °C | 95 °C | 50 °C | 72 °C | 72 °C | 4 °C |
|---|---|---|---|---|---|
| 4:00 min | 0:30 min | 0:30 min | 0:50 min | 5:00 min | hold |
| | x28 | | | | |

The crude PCR product was then purified using the PCR clean-up kit (Macherey-Nagel). For *in vitro* transcription, 700 ng dsDNDA template were used in a 100 µL reaction containing 40 mM Tris (pH 8.0), 30 mM $MgCl_2$, 10 mM DTT, 4 mM of each NTP, 2 mM spermidine and 4 µL T7 RNA-Pol. (4 mg/mL). After incubation at 37 °C for 3 h, 5 µL EDTA (0.5 M, pH 8.0) and 100 µL LiCl (7.5 M) were added and the RNA was precipitated through centrifugation at 4 °C for 30 min followed by washing with 70 % EtOH for 10 min.

Methylation was performed at 25 °C for 4.5 h with 200 ng tetR (Tr17) or blaR RNA (Tr18) and 50 pmol of each MTR1 variant (Rz24 - Rz29 for tetR, Rz30 - Rz31 for blaR) in 10 µL selection buffer (120 mM KCl, 5 mM NaCl, 50 mM MES, pH 6.0) including 100 µM $m^6G$ and 40 mM $MgCl_2$. For proper folding and formation of the MTR1-substrate RNA complex, an annealing step (3 min at 95°C, 10 min at 25 °C) was performed prior to addition of $MgCl_2$ and $m^6G$. The reaction was quenched by addition of 1 µL EDTA (0.5 M, pH 8.0) followed by EtOH-precipitation. Methylated RNA was then separated from the ribozymes on 5% denaturing PAGE (35W, 30 cm, 2:30 h). Extraction from the gel was carried out twice with overnight incubation at 4°C to minimize RNA degradation. After the following EtOH-precipitation of the RNA, its purity was confirmed on 1 % agarose gel (20 ng, 90 V, 70 min). Readout via RT primer extension was performed as described above for tRNA methylation using primer P11 - P16 for tetR and P17 - P18 for blaR.

**Statistics and reproducibility statement**

Kinetic experiments for characterization of ribozyme core sequence requirements, to determine $k_{obs}$, $m^6G$ and $Mg^{2+}$ concentration dependence were run as three independent experiments. Kinetic experiments for atomic mutagenesis of RNA substrates were repeated twice. All primer extension experiments with *in vitro* transcribed tRNA were repeated three times with similar results. Experiments with isolated *E. coli* tRNA and total *E. coli* RNA were performed two times with freshly extracted RNA from independent cultures and gave similar results.

## 9.1.2 Supplementary tables

**Table 9.1:** DNA oligonucleotides prepared by solid-phase synthesis.

| No | description | 5'-sequence-3' |
|---|---|---|
| D1 | T7 promotor | CTGTAATACGACTCACTATA |
| D2 | DNA Pool | GGTAAGGTGGACATACTG-N40-GCCTTCAAGGATGGTAGGCTGG |
| D3 | Forward primer | CTGTAATACGACTCACTATAGGACATACTGAGCCTTCAACCAGCCTACCATCC |
| D4 | Reverse primer 1st & 2nd PCR | GGTAAGGTGGACATACTG |
| D5 | Forward primer 1st PCR | CTTCAACCAGCCTACCATCC |
| D6 | Cloning f. primer | TAAATAAAATAACTGTAATACGACTCACTATAGGACATACTGAGC |
| D7 | P1 (*R. norveg.* tRNA-Lys; 17-34) | AACTTAAAAGGTTAACGC |
| D8 | P2 (*B. subtilis* tRNA-Ser; 30-47) | CCTACACGATTTCCAATC |
| D9 | P3 (*T. thermo*. tRNA-Asp; 66-72) | TGGCGGCCCCG |
| D10 | Disruptor 1 (*R. norveg.* tRNA-Lys) | TGGTCATTGTGGAGATTTGTTGTCTCTA |
| D11 | Disruptor 2 (*B. subtilis* tRNA-Ser) | ACTCGGACAGCTCTCC |
| D12 | P4 (*E. coli.* tRNA-Asp GTC) | TGGCGGAACGG |
| D13 | P5 (*E. coli.* tRNA-Glu TTC) | TGGCGTCCCCT |
| D14 | P6 (*E. coli.* tRNA-Gly CCC) | TGGAGCGGGCG |
| D15 | P7 (*E. coli.* tRNA-Gly TCC) | TGGAGCGGGCA |
| D16 | P8 (*E. coli.* tRNA-Ser GCT, GGA) | TGGCGGTGAGG |
| D17 | P9 (*E. coli.* tRNA-His GTG) | TGGGGTGGCTA |
| D18 | P10 (cis-MTR1; 110-128) | CAAGGATGGTAGGCTGGT |
| D19 | P11 (tetR RNA A44) | AACCTTCGATTCCGACCT |
| D20 | P12 (tetR RNA A110) | TTTTTACATGCCAATACA |
| D21 | P13 (tetR RNA A173) | AAAAGTGAGTATGGTGCC |
| D22 | P14 (tetR RNA A288) | TGTAGGCCGTGTACCTAA |
| D23 | P15 (tetR RNA A449) | GTTTCCCTTTCTTCTTTA |
| D24 | P16 (tetR RNA A548) | ATCAATTCAAGGCCGAAT |
| D25 | P17 (blaR RNA A45) | ATGGGGATGCTGTTGATT |
| D26 | P18 (blaR RNA A245) | CGTCCGCAGGGGCTCAAG |
| D27 | P19 (T7-tetR fw primer) | TAATACGACTCACTATAGATGTCTAGATTAGATAAAAGTAAAG |
| D28 | P20 (tetR rv primer) | TTTAATAAGATCTGAATTCCCGGGATC |
| D29 | P21 (T7-blaR fw primer) | TAATACGACTCACTATAGATGGCCAAGCCTTTGTCTCAAGAAG |
| D30 | P22 (blaR rv primer) | TTAGCCCTCCCACACATAACCAGAGGG |

**Table 9.2:** RNA oligonucleotides prepared by solid-phase synthesis. Target nucleosides are shown in bold.

| No | description | 5'-sequence-3' |
|---|---|---|
| R1 | Parent 5'-Alkyne-RNA (22 nt) | Hexyne-GACAUACUG**A**GCCUUCAAAUA |
| R2 | Parent RNA (17 nt) 3'-$NH_2$ | ACAUACUG(**A**)GCCUUCAA-C6-$NH_2$ |
| R3 | Inosine 3'-$NH_2$ | ACAUACUG(**I**)GCCUUCAA-C6-$NH_2$ |
| R4 | 2'dA 3'-$NH_2$ | ACAUACUG(**dA**)GCCUUCAA-C6-$NH_2$ |
| R5 | 2'OMe A-3'-$NH_2$ | ACAUACUG(**Am**)GCCUUCAA-C6-$NH_2$ |
| R6 | Purine 3'-$NH_2$ | ACAUACUG(**P**)GCCUUCAA-C6-$NH_2$ |
| R7 | 2-Aminopurine-3'-$NH_2$ | ACAUACUG(**2AP**)GCCUUCAA-C6-$NH_2$ |
| R8 | 7-deazaA-3'-$NH_2$ | ACAUACUG(**c7A**)GCCUUCAA-C6-$NH_2$ |
| R9 | 2'-dA-methylphosphonate-3'-$NH_2$ | ACAUACUG(**dA-PCH$_3$**)GCCUUCAA-C6-$NH_2$ |
| R10a | Parent Short RNA (13 nt) (S*) | AUACUG(**A**)GCCUUC-C6-$NH_2$ |
| R10 | Parent Short RNA (13 nt) (S) | AUACUG(**A**)GCCUUC |
| R11 | 2'dA | AUACUG(**dA**)GCCUUC |

| | | |
|---|---|---|
| R12 | 2'dA-methylphosphate (dApm) | AUACUG(**dA-POCH$_3$**)GCCUUC |
| R13 | 3-deaza-2'dA | AUACUG(**c3dA**)GCCUUC |
| R14 | m$^6$A | AUACUG(**m6A**)GCCUUC |
| R15 | m$^1$A | AUACUG(**m1A**)GCCUUC |
| R16 | unmodified 16 nt RNA (GAC) | GGAUAAUACG**A**CUCAC |
| R17 | 1,3-dideaza-A (c$^1$c$^3$A) var.of R16 | GGAUAAUACG(**c$^1$c$^3$A**)CUCAC |
| R18 | *R. norvegicus* fragment A9 | UUGCGA**A**GCUUAG |
| R19 | *T. thermophil.* fragment A58 | GGUUCG**A**GUCCCG |
| R20 | *B. subtilis* fragment A22 | GGUCGA**A**GGAGCA |

**Table 9.3:** RNA oligonucleotides prepared by *in vitro* transcription. Target nucleosides are shown in bold.

| No | description | 5'-sequence-3' |
|---|---|---|
| Tr1 | Selection Library | GGACAUACUGAGCCUUCAACCAGCCUACCAUCCUUGAAGGC-N$_{40}$-CAGUAUGUCCACCUUACC |
| Tr2 | Parent RNA (19 nt) | GGACAUACUG**A**GCCUUCAA |
| Tr3 | A11G | GGACAUACUG**G**GCCUUCAA |
| Tr4 | A11C | GGACAUACUG**C**GCCUUCAA |
| Tr5 | TM | GGGUGCGUCG**A**GUUCCUGG |
| Tr6 | TV1 | GGUGUAUGAG**A**GGGAAGUU |
| Tr7 | TV2 | GGCACGCAGG**A**GAAGGACC |
| Tr8 | AAG | GGACAUACUA**A**GCCUUCAA |
| Tr9 | CAG | GGACAUACUC**A**GCCUUCAA |
| Tr10 | UAG | GGACAUACUU**A**GCCUUCAA |
| Tr11 | GAA | GGACAUACUG**A**ACCUUCAA |
| Tr12 | GAC | GGACAUACUG**A**CCCUUCAA |
| Tr13 | GAU | GGACAUACUG**A**UCCUUCAA |
| Tr14 | *R. norvegicus* tRNA-Lys A9 | GGCAUUGCGA**A**GCUUAGAGCGUUAACCUUUUAAGUUAAAGUUAGAGACAACAAAUCUCCACAAU GACCA (69nt) |
| Tr15 | *B. subtills* tRNA-Ser A22 | GGAGAGCUGUCCGAGUGGUCGA**A**GGAGCACGAUUGGAAAUCGUGUAGGCGGUCAACUCCGUCUC AAGGGUUCGAAUCCCUUGCUCUCCGCCA (92nt) |
| Tr16 | *T. thermophilus* RNA-Asp A58 | GGCCCCGUGGUGUAGUUGGUUAACACACCCGCCUGUCACGUGGGAGAUCGCGGGUUCG**A**GUCCC GUCGGGGCCGCCA (77nt) |
| Tr17 | tetR RNA | AUGUCUAGAUUAGAUAAAAGUAAAGUGAUUAACAGCGCAUUAG**A**GCUGCUUAAUGAGGUCGGAA UCGAAGGUUUAACAACCCGUAAACUCGCCCAGAAGCUAGGUGUAG**A**GCAGCCUACAUUGUAUUG GCAUGUAAAAAAUAAGCGGGCUUUGCUCGACGCCUUAGCCAUUG**A**GAUGUUAGAUAGGCACCAU ACUCACUUUUGCCCUUUAGAAGGGGAAAGCUGGCAAGAUUUUUUACGUAAUAACGCUAAAAGUU UUAGAUGUGCUUUACUAAGUCAUCGCGAUGG**A**GCAAAAGUACAUUUAGGUACACGGCCUACAGA AAAACAGUAUGAAACUCUCGAAAAUCAAUUAGCCUUUUUAUGCCAACAAGGUUUUUCACUAGAG AAUGCAUUAUAUGCACUCAGCGCUGUGGGGCAUUUUACUUUAGGUUGCGUAUUGGAAGAUCAAG **A**GCAUCAAGUCGCUAAAGAAGAAAGGGAAACACCUACUACUGAUAGUAUGCCGCCAUUAUUACG ACAAGCUAUCGAAUUAUUUGAUCACCAAGGUGCAG**A**GCCAGCCUUCUUAUUCGGCCUUGAAUUG AUCAUAUGCGGAUUAGAAAAACAACUUAAAUGUGAAAGUGGGUCCGCGUACAGCGGAUCCCGGG AAUUCAGAUCUUAUUAAA(658nt) |
| Tr18 | blaR RNA | AUGGCCAAGCCUUUGUCUCAAGAAGAAUCCACCCUCAUUGAAAG**A**GCAACGGCUACAAUCAACA GCAUCCCCAUCUCUGAAGACUACAGCGUCGCCAGCGCAGCUCUCUCUAGCGACGGCCGCAUCUU CACUGGUGUCAAUGUAUAUCAUUUUACUGGGGGACCUUGUGCAGAACUCGUGGUGCUGGGCACU GCUGCUGCUGCGGCAGCUGGCAACCUGACUUGUAUCGUCGCGAUCGGAAAUG**A**GAACAGGGGCA UCUUGAGCCCCUGCGGACGGUGCCGACAGGUGCUUCUCGAUCUGCAUCCUGGGAUCAAAGCCAU AGUGAAGGACAGUGAUGGACAGCCGACGGCAGUUGGGAUUCGUGAAUUGCUGCCCUCUGGUUAU GUGUGGGAGGGCUAA (399nt) |

| | | **Ribozymes (prepared by *in vitro* transcription, binding arms underlined)** |
|---|---|---|
| Rz1 | CA13 | GGUUGAAGGCUGACCGACCCCCCGACCCUUCUCUGGGACAACUAGACAUACAGUAUGUCC |
| Rz1s | CA13 short binding arms (sh) | GGAGGCUGACCGACCCCCCGACCCUUCUCUGGGACAACUAGACAUACAGU |
| Rz1a | CA13 sh for R16/17 | GGUGAGUGACCGACCCCCCGACCCUUCUCUGGGACAACUAGACAUACGUAUUAUCC |
| Rz2 | CA21 | GGUUGAAGGCUGACCGACCCGCUACAAUACAACAGCGACAACUAGAAAUACAGUAUGUCC |
| Rz2s | CA21 sh | GGAAGGCUGACCGACCCGCUACAAUACAACAGCGACAACUAGACAUACAGU |
| Rz3 | MTR1 | GGUUGAAGGCUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUACAGUAUGUCC |
| Rz3s | MTR1 sh | GGAGGCUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUACAGU |
| Rz4 | CA13 U24G | GGAGGCUGACCGACCCCCCGACCCUUCUCGGGGACAACUAGACAUACAGU |
| Rz5 | Rz4 C10G27→G10C27 | GGAGGCUGACCGACCGCCCGACCCUUCUCGGGCACAACUAGACAUACAGU |
| Rz6 | Rz4 C11G26→G11C26 | GGAGGCUGACCGACCCGCCGACCCUUCUCGGCGACAACUAGACAUACAGU |
| Rz7 | Rz4 C13G24→G13C24 | GGAGGCUGACCGACCCCCGGACCCUUCUCCGGGACAACUAGACAUACAGU |
| Rz8 | Rz3 5bp | GGAGGCUGACCGACCCCCCGUUCGCGGGGACAACUAGACAUACAGU |
| Rz9 | Rz2 Δstem | GGUUGAAGGCUGACCGACCCGACAACUAGAAAUACAGUAUGUCC |
| Rz10 | 4bp stem tetraloop | GGUUGAAGGCUGACCGACCCGCUAACCAGCGACAACUAGAAAUACAGUAUGUCC |
| Rz11 | 4bp stem hexaloop | GGUUGAAGGCUGACCGACCCGCUACAUACAGCGACAACUAGAAAUACAGUAUGUCC |
| Rz12 | CA13 TM | GGAACUUCCCUGACCGACCCCCCGACCCUUCUCUGGGACAACUAGACAUACUCAUACACC |
| Rz13 | CA13 TV1 | GGGGUCCUUCUGACCGACCCCCCGACCCUUCUCUGGGACAACUAGACAUACCUGCGUGCC |
| Rz14 | CA13 TV2 | GGUUGAAGGCUGACCGACCCGCUACAAUACAACAGCGACAACUAGAAAUACAGUAUGUCC |
| Rz15 | CA13 AAG | GGUUGAAGGCUGACCGACCCCCCGACCCUUCUCUGGGACAACUAGACAUAUAGUAUGUCC |
| Rz16 | CA13 CAG | GGUUGAAGGCUGACCGACCCCCCGACCCUUCUCUGGGACAACUAGACAUAGAGUAUGUCC |
| Rz17 | CA13 UAG | GGUUGAAGGCUGACCGACCCCCCGACCCUUCUCUGGGACAACUAGACAUAAAGUAUGUCC |
| Rz18 | CA13 GAA | GGUUGAAGGUUGACCGACCCCCCGACCCUUCUCUGGGACAACUAGACAUACAGUAUGUCC |
| Rz19 | CA13 GAC | GGUUGAAGGGUGACCGACCCCCCGACCCUUCUCUGGGACAACUAGACAUACAGUAUGUCC |
| Rz20 | CA13 GAU | GGUUGAAGGAUGACCGACCCCCCGACCCUUCUCUGGGACAACUAGACAUACAGUAUGUCC |
| Rz21 | MTR1 *R. norvegicus* $tRNA^{Lys}$ A9 | GGCUAAGCUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUAUCGCAA |
| Rz22 | MTR1 *B. subtilis* $tRNA^{Ser}$ A22 | GGUGCUCCUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUAUCGACC |
| Rz23 | MTR1 *T. thermophilus* $tRNA^{Asp}$ A58 = MTR1 *E. coli* $tRNA^{Asp}$ A58 | GGACGGGACUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUACGAACCC |
| Rz24 | MTR1 tetR A44 | GGUUAAGCAGCUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUACUAAUGCGC |
| Rz25 | MTR1 tetR A110 | GGUAGGCUGCUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUACUACACCUAG |
| Rz26 | MTR1 tetR A173 | GGUCUAACAUCUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUACAAUGGCUA |
| Rz27 | MTR1 tetR A288 | GGUACUUUUGCUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUACCAUCGCGA |
| Rz28 | MTR1 tetR A449 | GGACUUGAUGCUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUACUUGAUCUU |
| Rz29 | MTR1 tetR A548 | GGAAGGCUGGCUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUACUGCACCUU |
| Rz30 | MTR1 blaR A45 | GGAGCCGUUGCUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUACUUUCAAUG |
| Rz31 | MTR1 blaR A245 | GGCCCCUGUUCUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUACAUUUCCGA |
| | | **F30-Broccoli MTR1 constructs (binding arms underlined)** |
| C1 | F30-Broccoli-MTR1*cis* (Broccoli green, substrate red, MTR1 blue) | GGGUUGCCAUGUGUAUGUGGGAGACGGUCGGGUCCAGAUAUUCGUAUCUGUCGAGUAGA-GUGUGGGCUCCCAUAUACUCUGAUGAUCCUCGAGGACAUACUGAGCCUUCAACCAGCCUACCAU CCUUGAAGGCUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGA-CAUACAGUAUGUCCAUGGAUCAUUCAUGGCAAGC |
| C2 | F30-Broccoli-MTR1*trans* $tRNA^{Asp}$ (Broccoli green, MTR1 blue) | GGGUUGCCAUGUGUAUGUGGGAGACGGUCGGGUCCAGAUAUUCGUAUCUGUCGAGUAGA-GUGUGGGCUCCCAUAUACUCUGAUGAUCCUCGAGCCACCAACGGUACCGACGGGACUGACCGAC CCCCCGAGUUCGCUCGGGGACAACUAGACAUACGAACCCUAAACCAUGGAUCAU-UCAUGGCAAGC |

**Table 9.4:** ESI-MS of synthetic RNAs and methylated products.

| No | description | Length (nt) | Mass calculated | Mass found |
|---|---|---|---|---|
| R1 | Parent 5'-Alkyne-RNA (22 nt) | 22 | 6821.09 Da | 6821.03 Da |
| R2 | Parent RNA (17 nt) 3'-$NH_2$ | 17 | 5530.83 Da | 5530.86 Da |
| R3 | Inosine 3'-$NH_2$ | 17 | 5531.82 Da | 5531.87 Da |
| R4 | 2'dA 3'-$NH_2$ | 17 | 5514.84 Da | 5514.86 Da |
| R5 | 2'OMe A-3'-$NH_2$ | 17 | 5544.85 Da | 5544.87 Da |
| R6 | Purine 3'-$NH_2$ | 17 | 5515.82 Da | 5515.85 Da |
| R7 | 2-Aminopurine-3'-$NH_2$ | 17 | 5530.83 Da | 5530.88 Da |
| R8 | 7-deazaA-3'-$NH_2$ | 17 | 5529.84 Da | 5529.88 Da |
| R9 | 2'-dA-methylphosphonate-3'-$NH_2$ | 17 | 5512.96Da | 5512.90 Da |
| R10a | Parent Short RNA (13 nt) (S*) | 13 | 4238.63 Da | 4238.69 Da |
| R10 | Parent Short RNA (13 nt) (S) | 13 | 4059.56 Da | 4059.62 Da |
| R11 | 2'dA | 13 | 4043.57 Da | 4043.62 Da |
| R12 | 2'dA-methylphosphate (dApm) | 13 | 4057.58 Da | 4057.63 Da |
| R13 | 3-deaza-2'dA | 13 | 4042.57 Da | 4042.62 Da |
| R14 | $m^6A$ | 13 | 4073.58 Da | 4073.62 Da |
| R15 | $m^1A$ | 13 | 4073.58 Da | 4073.63 Da |
| R16 | unmodified 16 nt RNA (GAC) | 16 | 5085.83 Da | 5085.75 Da |
| R17 | 1,3-dideaza-A ($c^1c^3A$) var.of R16 | 16 | 5083.75 Da | 5083.77 Da |
| R18 | *R. norvegicus* fragment A9 | 13 | 4139.58 da | 4139.61 Da |
| R19 | *T. thermophil.* fragment A58 | 13 | 4130.57 Da | 4130.60 Da |
| R20 | *B. subtilis* fragment A22 | 13 | 4240.65 Da | 4240.68 Da |
| R1-BG-Bio | isolated biotinylated RNA by Rz1 with BG-biotin | | 7279.33 Da | 7279.26 Da |
| R9-BG-$NH_2$ | isolated alkylated RNA by Rz1 with BG-$NH_2$ | | 5631.93 Da | 5631.99 Da |
| R10a-Me | isolated methylation product by MTR1 (Fig 2b) | | 4252.65 Da | 4252.70 Da |
| R18-Me | isolated *R. norv.* RNA after methylation by Rz21 | | 4153.59 Da | 4153.58 Da |
| R19-Me | isolated *T. therm*.RNA after methylation by Rz23 | | 4144.59 Da | 4144.56 Da |
| R20-Me | isolated *B. subtilis*.RNA after methylation by Rz22 | | 4254.66 Da | 4254.65 Da |

**Table 9.5:** Ribozyme sequences obtained from selection and their *cis*- and *trans*-activity.

| No | # clones | 5'-binding arm – core (40) – binding arm-3' | *cis* | *trans* |
|---|---|---|---|---|
| CA7 | 2 | CUGAAGGC GCCCGAGAAACAAUGAUGCCUUGUUAGAUACAGUAGUG CAGUAUGUCC | – | |
| **CA9** | **2** | **UUGAAGGC GCCCGAGAAACAAUGAUGCCUUGUUAGAUACAGUAGUG CAGUAUGUCC** | **+** | **–** |
| CA28 | 1 | UUGAAGGC GCCCGAGAAACAAUGAUGCCUUGUUAGAUACAGGAGUG CAGUAUGUCC | – | |
| **CA13** | **1** | **UUGAAGGC UGACCGACCCCCCGACCCUUCUCUGGGACAACUAGACAUA CAGUAUGUCC** | **+** | **+** |
| CA23 | 1 | UUGAAGGC UGACCGACCCCCCGACCAUUCUCUGGGACAACUAGACAUA CAGUAUGUCC | + | |
| CA38 | 1 | UUGAGGGC UGACCGACCCCCCGACCAUUCUCUGGGACAACUAGACAUA CAGUAUGUCC | + | |
| **CA21** | **1** | **UUGAAGGC UGACCGACCCGCUACAAUACAACAGCGACAACUAGAAAUA CAGUAUGUCC** | **+** | **+** |

### 9.1.3 Extended data figures

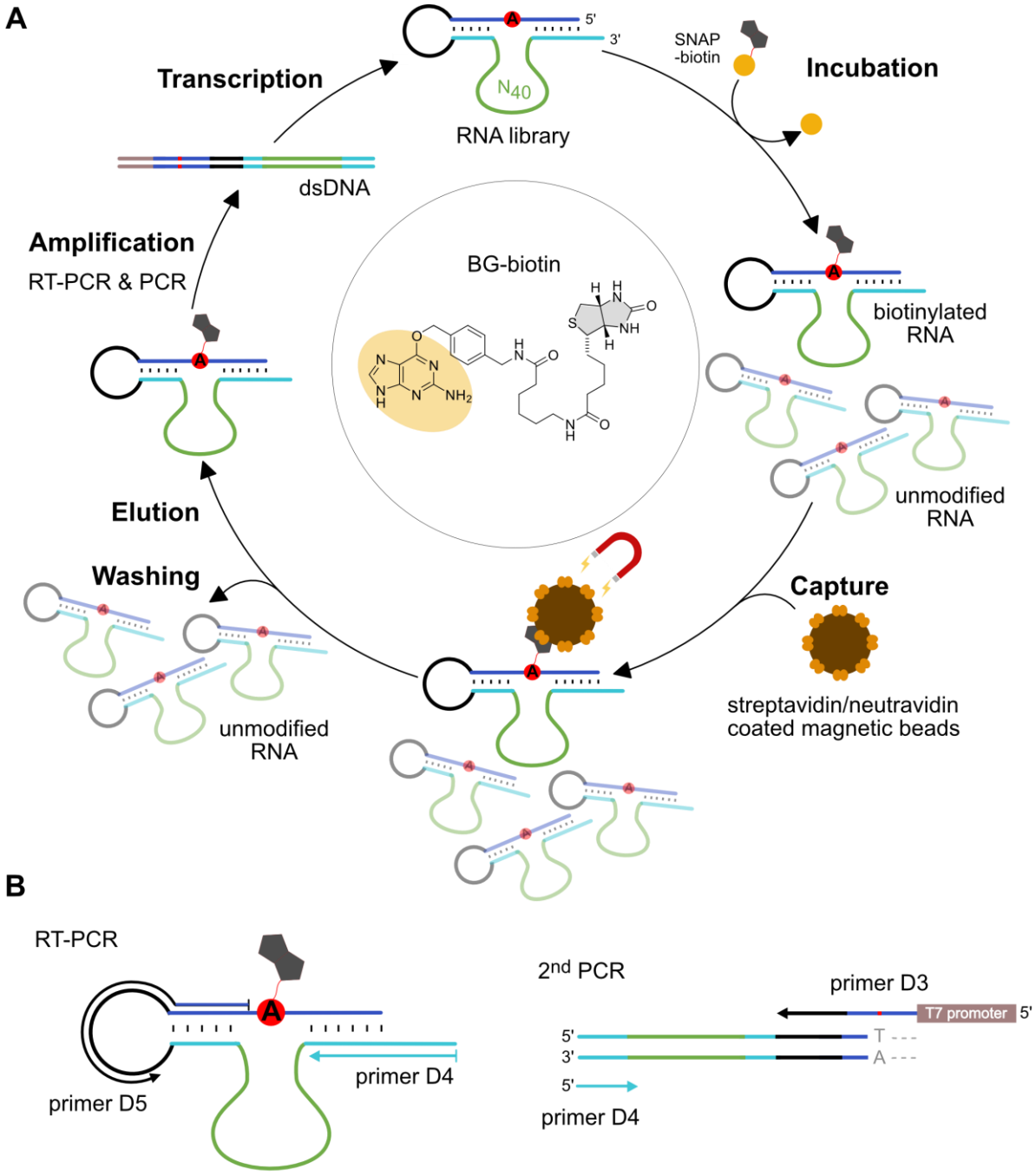



**Figure 9.1:** *In vitro* selection of methyltransferase ribozymes. **A** *In vitro* selection scheme consisting of incubation, capture, wash, elution, amplification, and transcription steps. The RNA substrate (blue) contains an unpaired adenosine (red, A) and is connected to the RNA library via the single-stranded loop (black). The library contains 40 random nucleotides (green) and two constant binding arms (cyan) complementary to the RNA substrate upstream and downstream of the bulged A. **B** Schematic depiction of primer binding sites for RT-PCR and PCR.

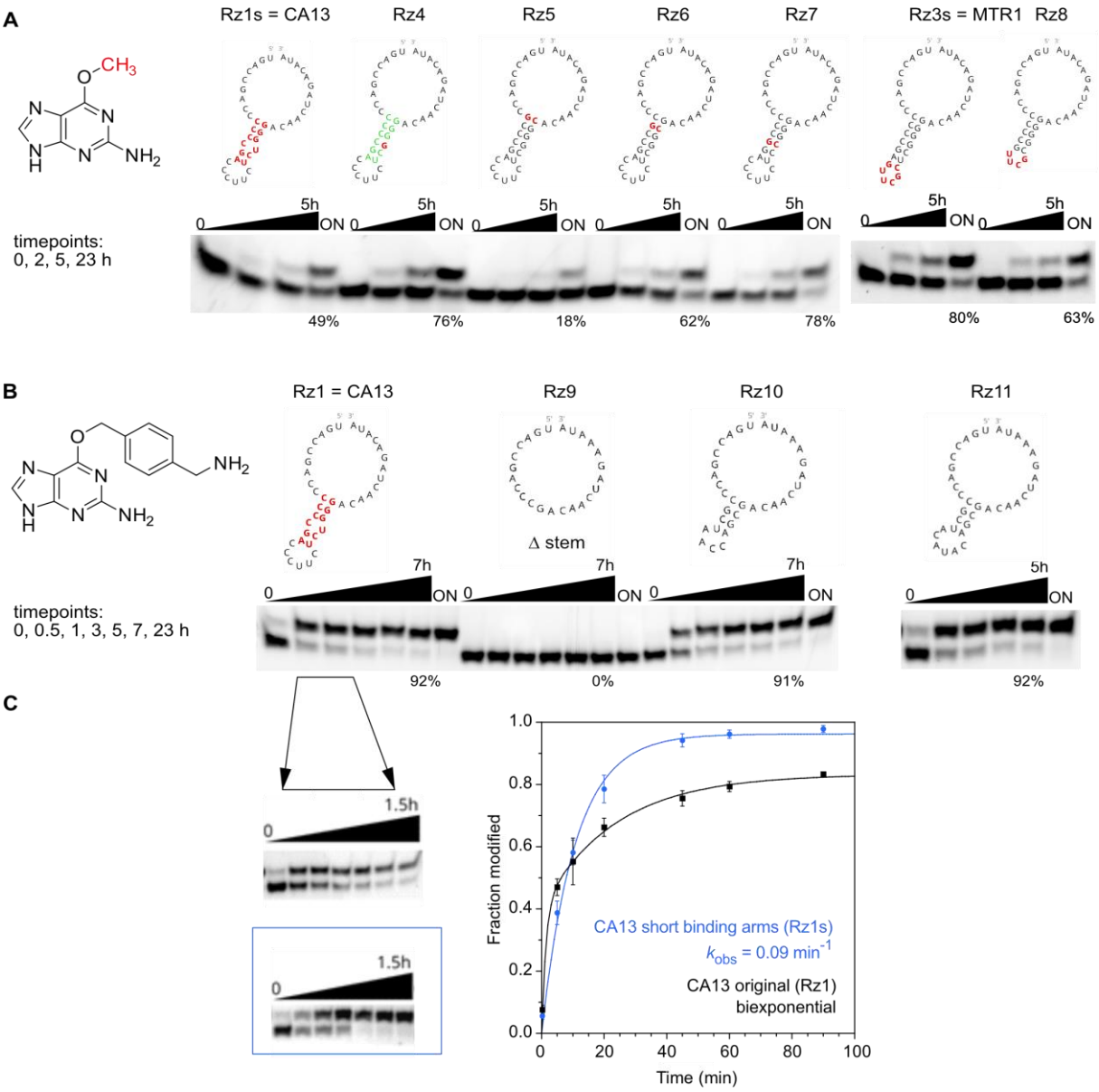



**Figure 9.2:** Activity of methyltransferase ribozymes: Examination of mutations in the stem-loop. **A** 3'-fluorescein-labeled R10a tested with 100 µM m$^6$G. **B** 5'-fluorecein-labeled R1 tested with 100 µM BG-NH$_2$. (ON = 23 h). **C** Kinetics of CA13 (Rz1/Rz1s)-catalyzed alkylation of R1 using BG-NH$_2$ as cofactor. Fraction modified is shown as mean ± SD ($n$ = 3), and fit to a monexponential model (Y = $Y_{max}(1 - e^{-kt})$, blue), or a biexponential model (Y = $Y_{max}(a (1 - e^{-k1t}) + (1 - a) (1 - e^{-k2t})$), black). **D** Dependence of MTR1 methylation yield on $Mg^{2+}$ concentration, reactions performed with 100 µM m$^6$G (on R2 with Rz3) at 37 °C. Individual data points are shown as empty symbols ($n$ = 2 for 5 h and 23 h, $n$ = 3 for 7 h timepoints), and mean is depicted as filled symbol.

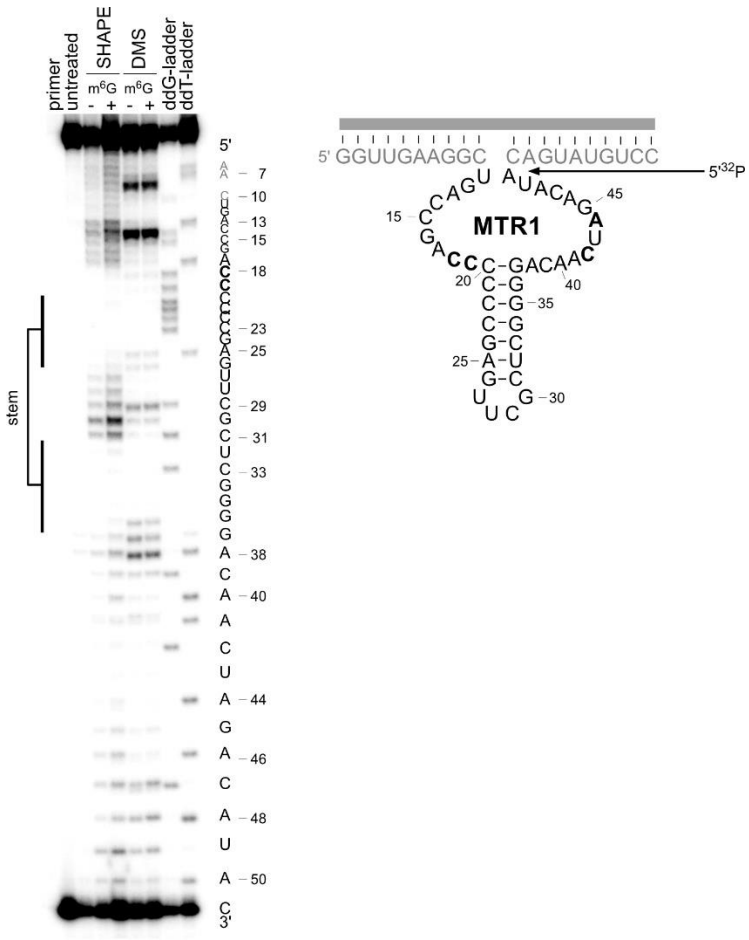



**Figure 9.3:** RNA structure probing by DMS and SHAPE. MTR1 (Rz3) was annealed with 17-nt RNA (R6), treated with dimethyl sulfide (DMS) or 1-methyl-7-nitroisatoic anhydride (1M7), in presence (+) or absence (-) of $m^6G$, and the modification pattern was analyzed by primer extension (5'-$^{32}$P-labeled D4) with Superscript III. Both probing methods confirm the central base paired stem and reveal the protection of several additional nucleotides (bold). The experiment was repeated three times with similar results.

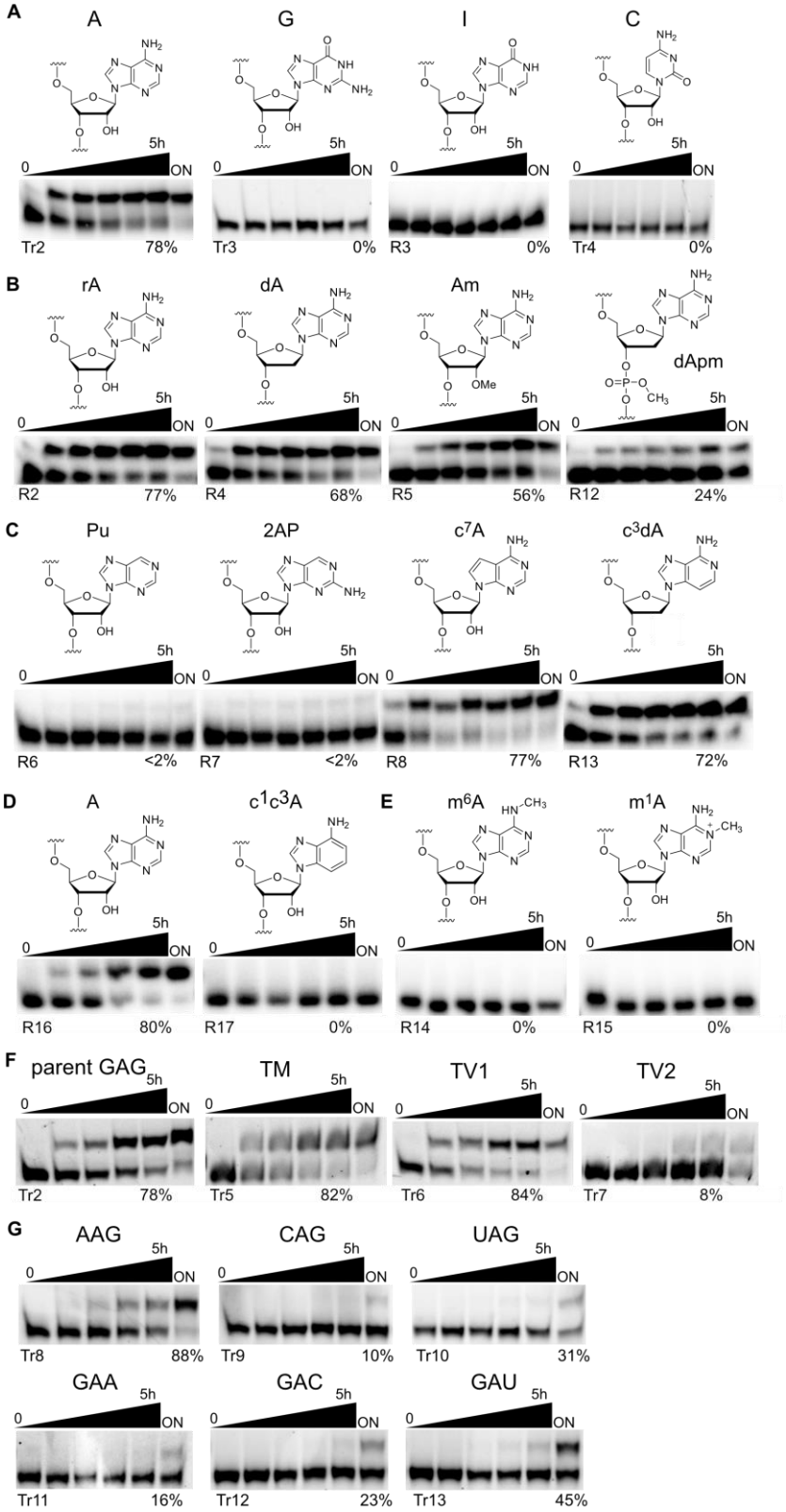



**Figure 9.4:** Representative gel images of RNA-catalyzed alkylation reactions of RNA substrate mutants by their corresponding ribozymes with complementary binding arms as listed in Table 9.3. **A** Adenosine point mutations. **B** Atomic mutagenesis of the backbone. **C, D** Atomic mutagenesis of adenosine. **E** Reaction sites blocked by methylation. **F** Binding arm mutations outside of GAG. **G** Point mutations next to target nucleoside A. Reactions were performed with 100 µM BG-$NH_2$ (**A** - **E**) or BG-biotin (**F** - **G**) at pH 7.5, 40 mM $MgCl_2$, 37°C and repeated two times for each substrate. The parent reaction with adenosine was performed with fluorescein- and $^{32}$P-labeled RNA independently for six times.

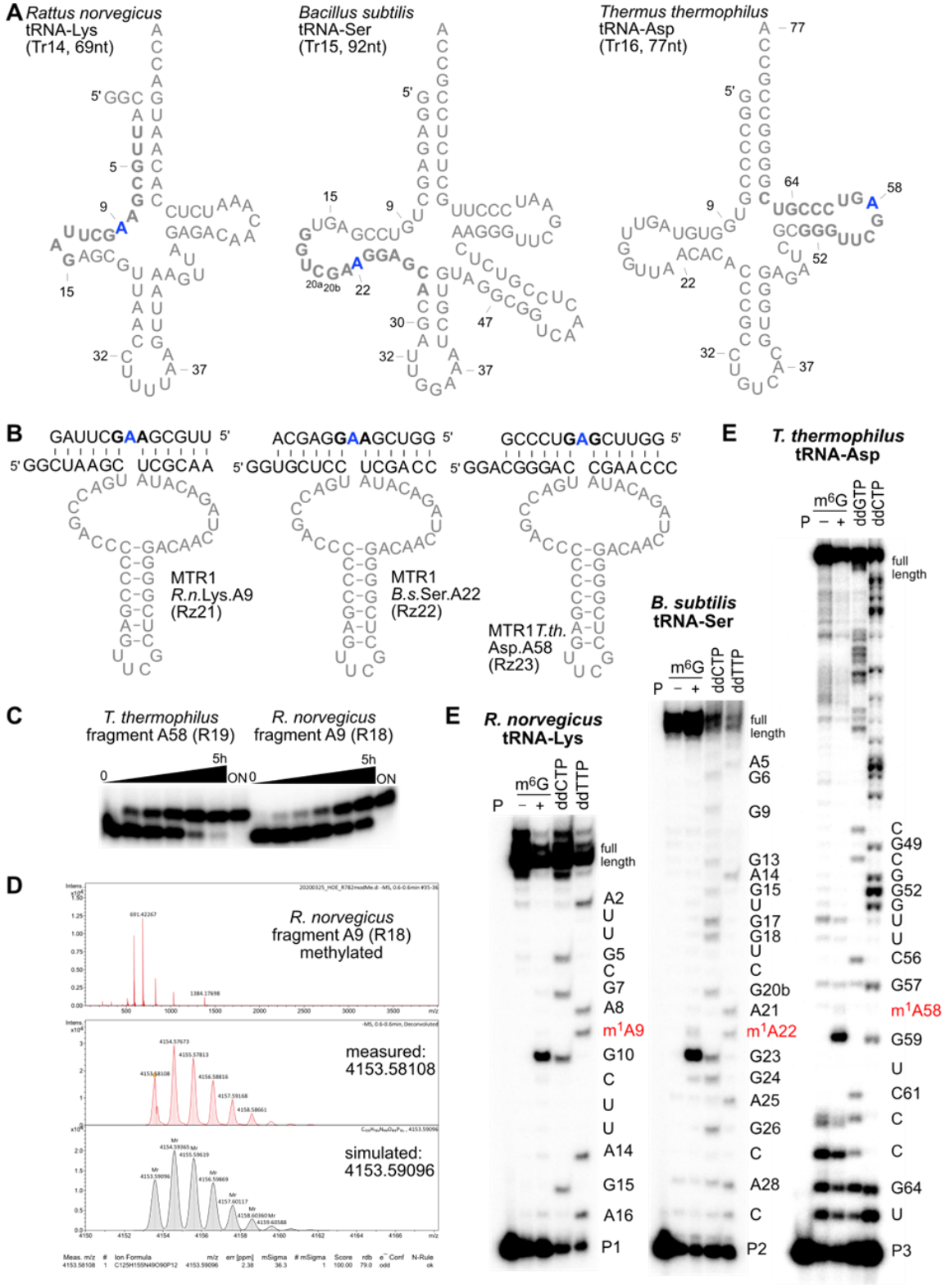



**Figure 9.5:** RNA-catalyzed methylation of tRNAs. **A** tRNA sequences studied. **B** Synthetic fragments and corresponding ribozymes. **C** Exemplary gel images for kinetic analysis of MTR1-catalyzed fragment methylation showing quantitative formation of $m^1A$. **D** Exemplary HR-ESI MS of isolated methylated *R. norvegicus* RNA fragment. **E** Full gel images of primer extension analysis shown in Figure 3.3. Representative gel images of three independent experiments with similar results.

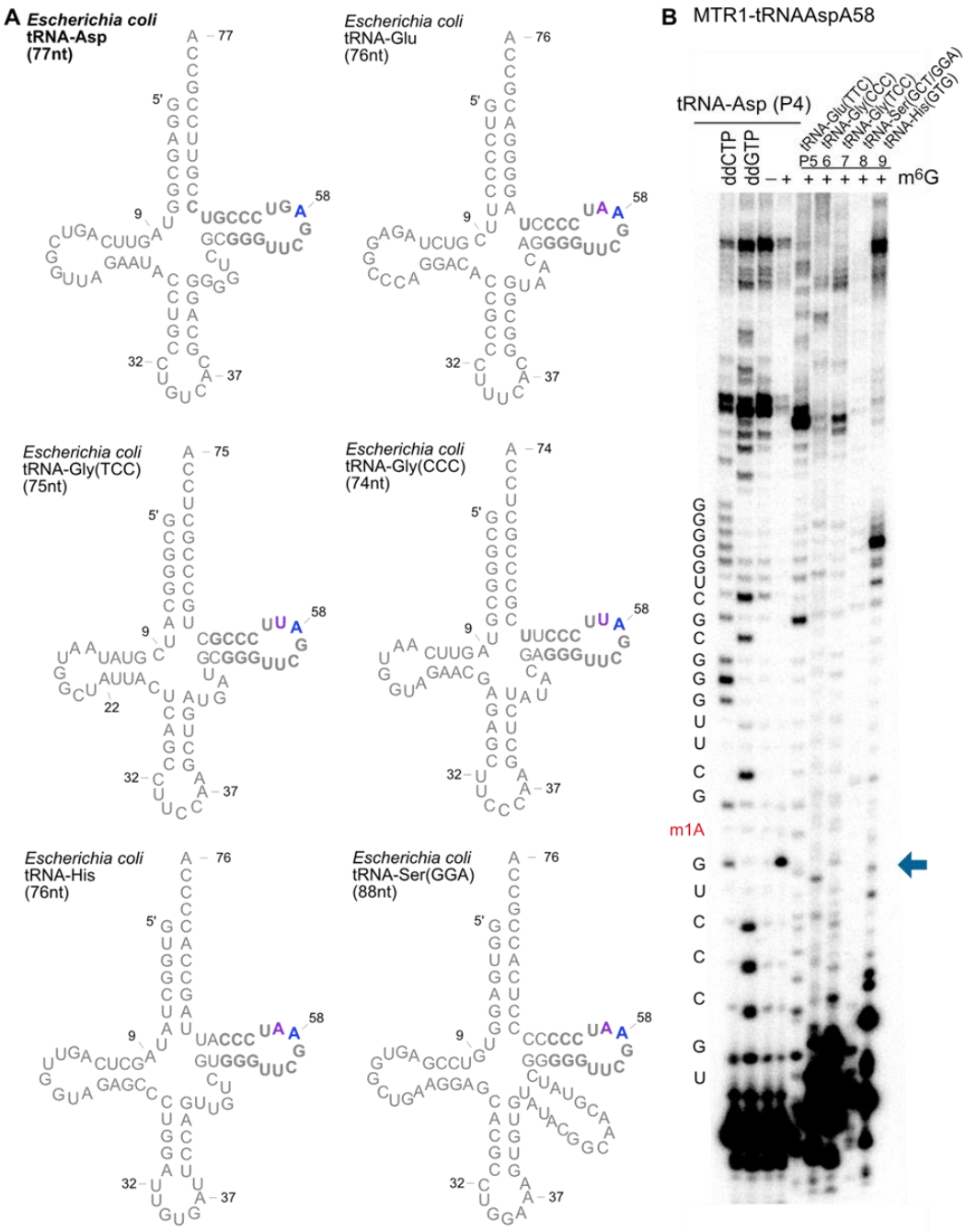



**Figure 9.6: A** Secondary structure schemes of six *E. coli* tRNAs, with very similar TΨC-stem-loop sequences (drawn without natural modifications). Target tRNA-Asp in top left corner, with A58 indicated in blue. The nucleotides complementary to the binding arms of MTR1-tRNA-Asp-A58 (Rz23) are shown in bold. The purple nucleotides indicate mismatched positions with the binding arms. **B** Full gel image of RT primer extension analysis on total *E. coli* tRNA with six different primers, shown in Figure 3.3C. Primer extension analysis were repeated twice.

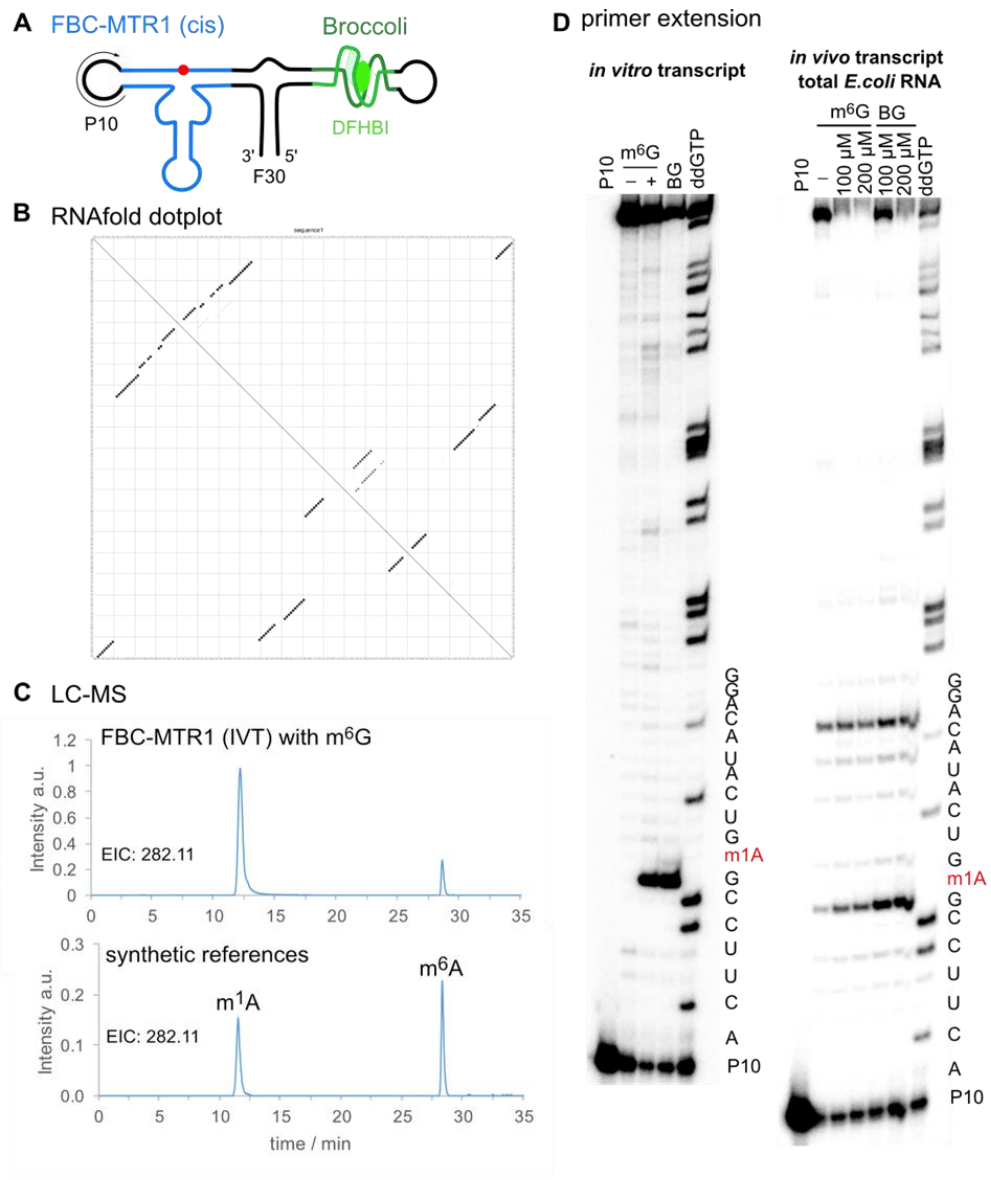



**Figure 9.7:** Plasmid-encoded *cis*-active ribozyme. **A** Schematic depiction of F30-Broccoli-cis (FBC)-MTR1 construct. **B** Dot plot for FBC-MTR1 generated by the RNAfold web server[391], indicates high probability of folding into the designed structure. **C** LC-MS analysis of SVPD/BAP-digested FBC-MTR1 *in vitro* transcript after reaction with $m^6G$. Extracted ion chromatogram (EIC) for detection of $MH^+$ 282.11±0.05 (methylated adenosines) shows production of $m^1A$, and $m^6A$ to a small extent (due to partial Dimroth rearrangement during digestion). Bottom trace for synthetic references $m^1A$ and $m^6A$ (50 nM each), is same as shown in Figure 3.3D. **D** Primer extension stop assays also confirm activity of FBC-MTR1 transcribed *in vitro* and *in vivo*, in the presence of total *E.coli* RNA. Left: full gel image shown in Figure 3.4A for *in vitro* transcribed FBC-MTR1. Right: primer extension on total *E. coli* RNA, isolated 1 h after IPTG induction, and incubated with indicated $m^6G$ or BG concentration *in vitro*. These experiments were independently repeated two times with similar results.

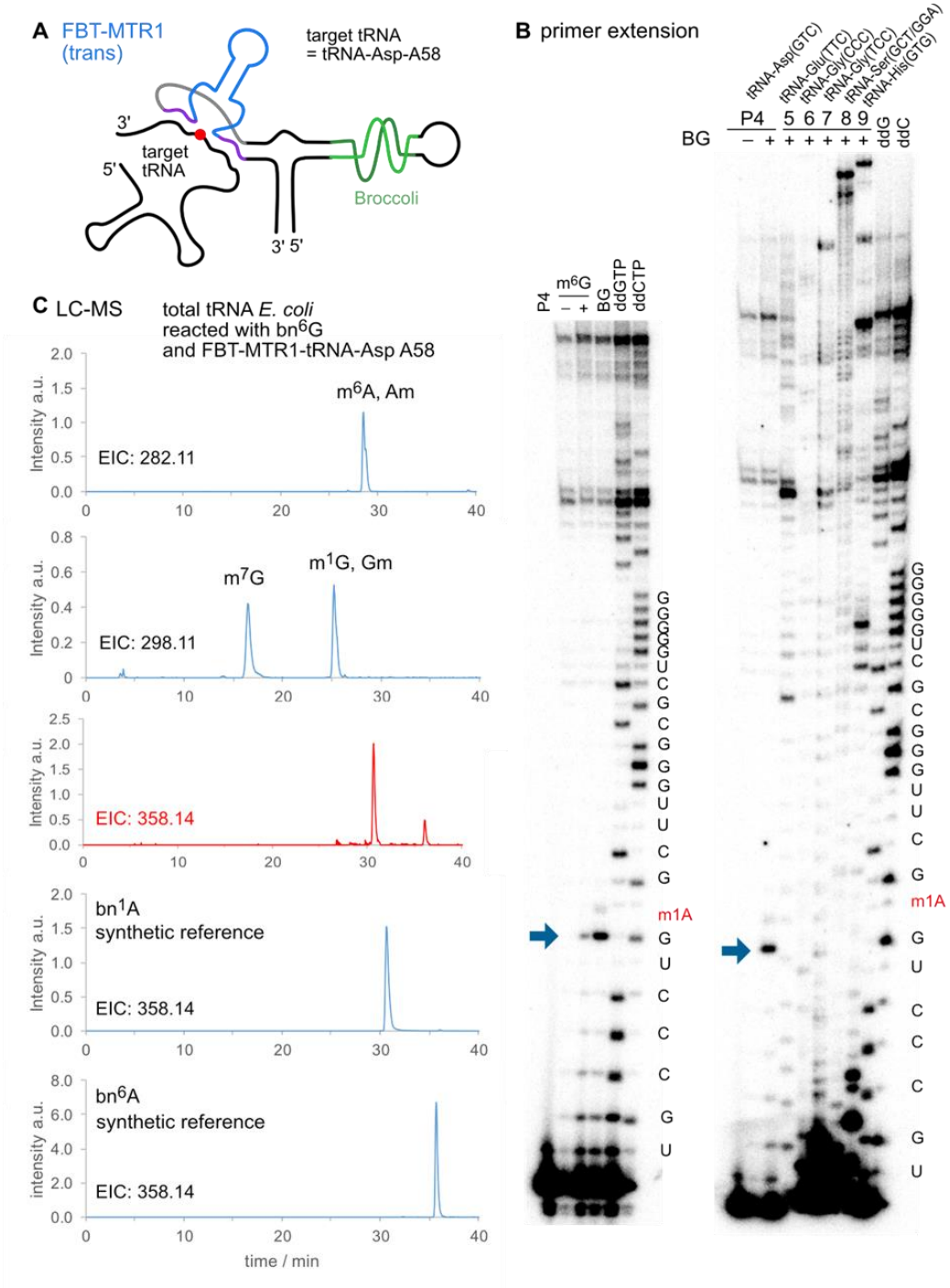



**Figure 9.8:** Plasmid-encoded *trans*-active ribozyme. **A** Schematic depiction of F30-Broccoli-*trans* (FBT)-MTR1 with specific binding arms for *E. coli* tRNA-Asp (A58). **B** Primer extension stop assays confirm the activity and specificity of the FBT-MTR1 *in vitro* transcript. Left: full gel image shown in Figure 3.4B for FBT-MTR1 reacted with $m^6G$ and BG on total *E. coli* tRNA. Right: primer extension on BG-treated sample with six different *E. coli* tRNA-specific primers (P4-P9), repeated twice. **C** LC-MS analysis after digestion of total *E. coli* tRNA treated with FBT-MTR1 and BG (same sample as for right gel image). EIC for $MH^+$ 282.11 (methylated adenosines) and 298.11 (methylated guanosines) demonstrate the presence of natural tRNA modifications. EIC 358.11 in comparison to reference nucleosides $bn^1A$ and $bn^6A$ shows bn-modified adenosines produced by FBT-MTR1.

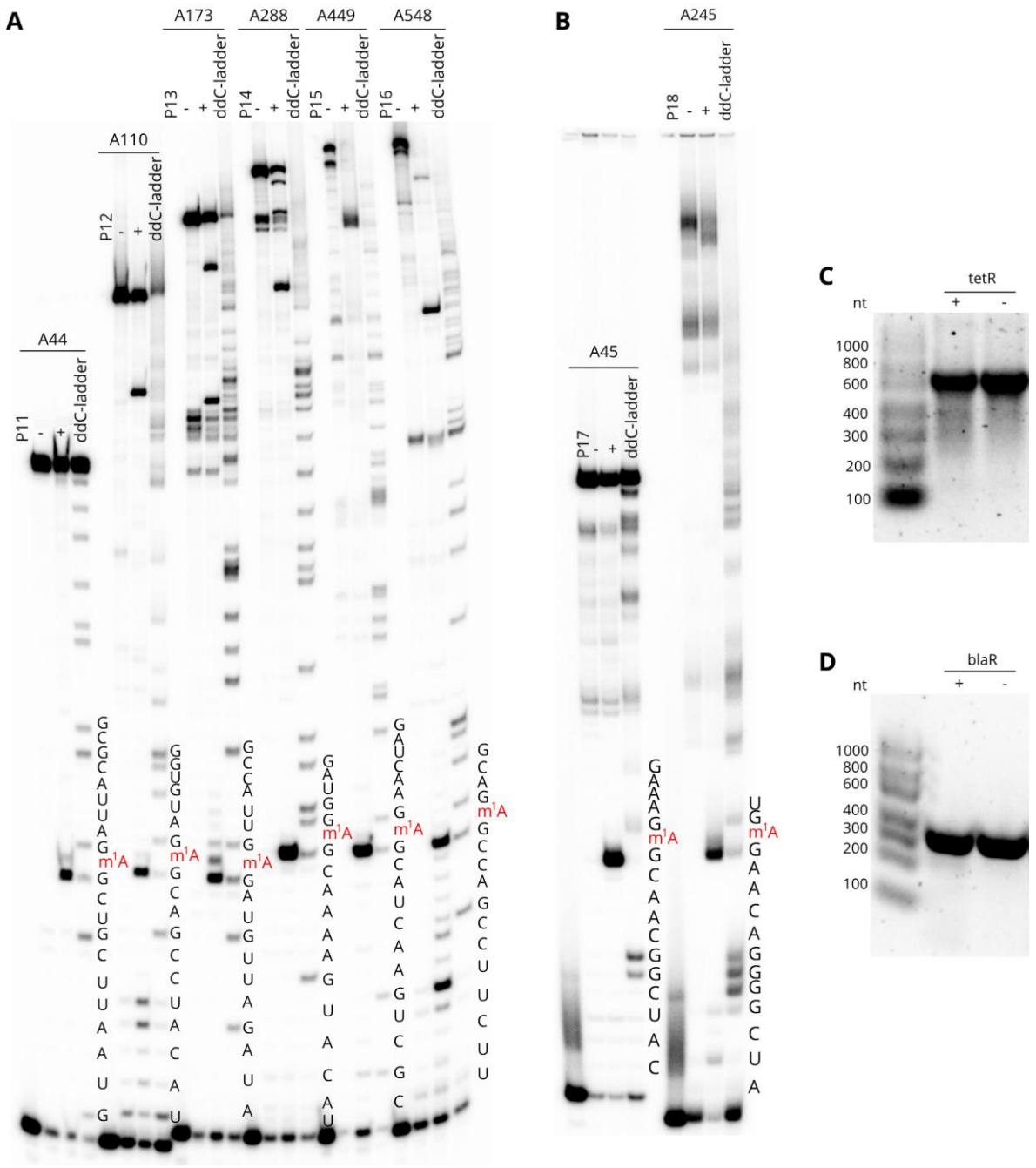



**Figure 9.9:** MTR1-catalyzed methylation of long RNA transcripts. RT primer extension confirms successful simultaneous installation of six different $m^1A$ sites in tetR (**A**) and two $m^1A$ sites in blaR (**B**) transcripts. Full gel images shown in Figure 3.5. Analysis of PAGE-purified methylated (+) and unmethylated (-) tetR (**C**) and blaR (**D**) transcripts on agarose gel next to an RNA-ladder confirm the purity of isolated RNAs.

## 9.2 Supporting information for chapter 4

This Supporting Information has been published in ref.[299]. For the sake of unity of this thesis, several editorial changes have been made, which, however, do not affect the contents of the thesis.



### 9.2.1 Material and Methods

#### RNA synthesis

The MTR1 ribozyme and point mutants were prepared by *in vitro* transcription with T7 RNA polymerase from synthetic DNA templates (purchased from Microsynth), as described previously.[184] Typical reaction conditions included 1 µM DNA template, 4 mM NTPs, 30 mM $MgCl_2$, 2 mM spermidine and 10 mM DTT. RNA oligonucleotides for the split MTR1 complexes were prepared by solid-phase synthesis using 2'-*O*-TOM-protected ribonucleotide phosphoramidites. Modified phosphoramidites were purchased (2'-OMe-U, 2'-OMe-C, 2'-F-U, 2'-F-C, $m^5C$) or prepared in-house following known procedures (2'-SeMe-U,[317] $m^4C$[401]). Deprotection was performed with a 1:1 (v/v) mixture of 28 % aqueous ammonia and 40 % aqueous methylamine (known as AMA) at 37 °C for 4-6 h, followed by 1 M tetrabutylammonium fluoride in THF at 25 °C for 12-16 h. 2'-SeMe-U containing RNA was treated with 2 mM DTT in water for 5 min at 37 °C to reduce partially oxidized 2'-methylselenoxide groups back to the desired 2'-selenomethyl moiety.[402] RNAs were purified by denaturing PAGE, followed by extraction and ethanol precipitation. Purity was confirmed by anion exchange HPLC (Dionex DNAPac PA200, 2 × 250 mm, at 60 °C; solvent A: 25 mM Tris-HCl (pH 8.0) and 6 M urea; solvent B: 25 mM Tris-HCl (pH 8.0), 6 M urea and 0.5 M $NaClO_4$; linear gradient, 0-40 % solvent B, with a slope of 4 % solvent B per column volume), and the identity of the RNAs was confirmed by HR-ESI-MS (negative ion mode).

#### Complex formation and crystallization

Crystallization complexes were formed by mixing RNA strands in a 1:1:1 molar ratio in 50 mM KCl and 10 mM HEPES (pH 7.5). The sample was heated to 95 °C for 3 min and cooled at 20 °C for 15 min. 5 mM $MgCl_2$ and 240 µM substrate ($m^6G$, BG or BG-$NH_2$) were added to a

final complex concentration of 200 µM. $MgCl_2$ was omitted for crystallization in the absence of magnesium ions.

Crystals were grown at 20 °C using the hanging drop vapor-diffusion method by mixing the RNA complex in a 1:1 ratio with solutions containing 100 mM NaCl, 100 mM LiCl, 10 mM $MgCl_2$, 50 mM MES (pH 6.4-6.7) and 36-42 % MPD.

Drops with volumes of 1 µL produced crystals after 3-4 days (5-6 days when no $MgCl_2$ was present) and were harvested after 14 days. Crystals were flash frozen in liquid nitrogen without any additional cryoprotection.

Incorporation of thallium ions was achieved by co-crystallization. The crystallization complex was prepared as described before, but with 50 mM KOAc instead of KCl to prevent precipitation of TlCl. The reservoir buffer contained 100 mM NaOAc, 100 mM LiOAc, 50 mM MES (pH 6.4-6.7) and 36-42 % MPD. Crystals appeared after 5-6 days and were harvested after 14 days, and flash frozen in liquid nitrogen without any additional cryoprotection.

**Data collection and structure determination**

X-ray diffraction data were collected at 100K on PILATUS 6M and EIGER 2X 16 M detectors at the ID23-01 (European Synchrotron Radiation Facility) or P11 (DESY) beamlines. Diffraction data were indexed, integrated, scaled with XDS,[403] and reduced with POINTLESS, AIMLESS, and CTRUNCATE within the CCP4 package.[404] Initial phases for the Crystal I data (wavelength 0.9198 Å) were determined by molecular replacement (MR) using Phaser in Phenix.[405] Search models were double-stranded RNA A-form helix fragments as described previously.[316] Here we used the 7-bp bottom stem from the Chili RNA aptamer[406] (PDB: 7OAX, chain A nt 2-8/45-51), searched for three copies in the ensemble and obtained a MR solution with a Z-score (TFZ) of 14.4 (TFZ >8 indicates a definite solution). The initial electron density map obtained from Phaser was used for automated model building with the AutoBuild module in Phenix.[407] The model from AutoBuild was completed by iterative rebuilding in COOT[408] and refinement with Phenix.refine.[409] The final structures of Crystal II ($bn^6A$), Crystal III ($ab^6A$), and Crystal IV (no $Mg^{2+}$) were solved by MR using the model built with the data of Crystal I ($m^1A$). The $N^1$-modified adenosine residue restraints were prepared using JLigand in the CCP4 suite[410] or eLBOW in Phenix.[411]

The final structures of Crystal V (with SeMeU) and Crystal VI (with thallium) were determined by MR-single anomalous dispersion (SAD) using the anomalous data collected at the Selenium K-edge (0.9786 Å) and Thallium L-III edge (0.9751 Å). Data collection and refinement statistics

are summarized in Table 9.6. Figures displaying molecular models were prepared using PyMol (Schrödinger).

### HPLC analysis of the RNA methylation reaction and the crystal content

For analysis of the methylation status and the composition of the crystals, a crystal was harvested from the drop, briefly washed in buffer, then dissolved in 15 µL $H_2O$, and analyzed by anion exchange HPLC on Dionex DNAPac PA200 column, 2 × 250 mm at 60 °C with UV detection at 260 nm. Solvent A was 25 mM Tris-HCl (pH 8.0), 6 M urea and solvent B was 25 mM Tris-HCl (pH 8.0), 6 M urea and 0.5 M $NaClO_4$ with a linear gradient 0-48 % solvent B in 12 column volumes.

### Kinetic assays of RNA-catalyzed methyltransfer reaction

Single turnover kinetic assays were carried out as described in chapter 9.1.1. Briefly, in a standard reaction, 100 pmol ribozyme were mixed with 10 pmol of trace-labeled 5'-$^{32}$P-RNA R1 in 10 µl of reaction buffer (120 mM KCl, 5 mM NaCl and 50 mM Bis-Tris (pH 6.0) or HEPES (pH 7.5)) and annealed (3 min at 95 °C and 10 min at 25 °C), followed by addition of 100 µM $m^6G$ and 40 mM $MgCl_2$. The reaction mixture was incubated at 25 °C and aliquots (1 µl) were taken at different timepoints, quenched immediately with loading dye (4 µl) and half of each timepoint sample was resolved on 20 % denaturing PAGE. Methylation yields were determined by measuring band intensities using phosphorimaging. The data was fit with the pseudo-first order kinetic equation $Y = Y_{max}(1 - e^{-kobs*t})$ to obtain the observed reaction rate $k_{obs}$ and the final yield $Y_{max}$ using Origin (2021). For slow ribozymes with ≤ 20% methylated RNA product after overnight reaction, $k_{obs}$ values were approximated linearly using the 7 h methylation yield. For measuring the pH-dependence of the MTR1-catalyzed methylation reaction between pH 4.5-8.5, the following buffers were used: 50 mM NaOAc pH 4.5, 5.0, 5.5; 50 mM Bis-Tris pH 6.0, 6.5; 50 mM MES pH 6.0; 50 mM sodium cacodylate pH 6.0; 50 mM HEPES pH 7.0, 7.5, 8.0; 50 mM MOPS pH 7.5, 50 mM PIPES pH 7.5; 50 mM Tris-HCl pH 7.7, 8.0; pH determined at 25°C. Additional kinetic experiments were performed at pH 6.0 (50 mM Bis-Tris) using only 0.5, 1, or 5 mM $MgCl_2$. All kinetic assays were carried out with at least two independent replicates.

**Preparative scale methylation of RNA**

The methylated RNA $m^1$A-R1 was prepared in a preparative scale ribozyme-mediated methylation reaction. 1 nmol target RNA was mixed with 1.1 nmol MTR1 in 20 µl of reaction buffer (120 mM KCl, 5 mM NaCl and 50 mM Bis-Tris, pH 6.0) including 100 µM $m^6$G as small molecule substrate and 40 mM $MgCl_2$. Before addition of $m^6$G and $MgCl_2$, an annealing step (3 min at 95 °C and 10 min at 25 °C) was performed. After overnight incubation at 25 °C, the methylated RNA was purified by denaturing PAGE yielding ca. 75 % isolated product.

**In-line probing**

In-line probing was performed using the 5'-$^{32}$P-labeled full-length MTR1 ribozyme in complex with the substrate RNA R1 (unmodified or methylated). In general, 125 IPS ribozyme were annealed to 25 pmol R1 or $m^1$A-R1 (3 min at 95 °C and 10 min at 25 °C). 100 mM KCl and 100 mM Tris-HCl (pH 8.0) were added along with 20 mM $MgCl_2$ and $m^6$G or guanine in various concentrations to yield a final volume of 5 µL. After incubation at 20 °C for 36 h, the reactions were quenched by adding loading dye (5 µL), resolved on 20 % denaturing PAGE, and visualized by autoradiography. As a reference, an RNase T1 digestion (150 IPS of 5′-$^{32}$P-RNA in 10 µl 50 mM Tris (pH 7.5) digested with 1 U RNase T1 at 37 °C for 45 s) and an alkaline hydrolysis lane (250 IPS of 5′-$^{32}$P-RNA in 10 µl 20 mM NaOH incubated for 4 min at 95 °C) were used. Band intensities of cleaved RNA at three selected positions (U36, A37, G38) were quantified and normalized, plotted against ligand concentration, and fitted with a one-site binding model using Origin (2021).

**Data availability**

Structural data obtained by X-ray crystallography were deposited in the Protein Data Bank (PDB) and are available with the following accession codes: 7G7X, 7Q7Y, 7Q7Z, 7Q80, 7Q81, 7Q82.

## 9.2.2 Supplementary tables

**Table 9.6:** Data collection and refinement statistics.

| | MTR1-m$^1$A (Crystal I) | MTR1-bn$^1$A (Crystal II) | MTR1-ab$^1$A (Crystal III) | MTR1-m$^1$A; no $Mg^{2+}$ (Crystal IV) | MTR1-m$^1$A; 2'-SeMeU (Crystal V) | MTR1-m$^1$A; $Tl^+$ (Crystal VI) |
|---|---|---|---|---|---|---|
| **PDB code** | **7Q7X** | **7Q7Y** | **7Q7Z** | **7Q80** | **7Q81** | **7Q82** |
| **Data collection** | | | | | | |
| Space group | P $4_1$ $2_1$ 2 | P $4_1$ $2_1$ 2 | P $4_1$ 2 2 | P $4_1$ 2 2 | P $4_1$ $2_1$ 2 | P $4_1$ $2_1$ 2 |
| Cell dimensions | | | | | | |
| a, b, c (Å) | 71.0 71.0 82.8 | 71.3 71.3 83.1 | 117.9 117.9 83.3 | 120.9 120.9 82.2 | 71.1 71.1 83.6 | 72.0 72.0 84.5 |
| α, β, γ (˚) | 90 90 90 | 90 90 90 | 90 90 90 | 90 90 90 | 90 90 90 | 90 90 90 |
| Resolution (Å) | 42.94 - 2.80 (2.90 - 2.80)[1] | 43.09 - 2.90 (3.08 - 2.90) | 44.54 - 3.26 (3.38 - 3.26) | 45.17 - 3.15 (3.26 - 3.15) | 43.07 - 2.85 (2.95 - 2.85) | 43.59 - 2.95 (3.06 - 2.95) |
| $R_{merge}$ | 0.11 (>1.0) | 0.14 (>1.0) | 0.14 (>1.0) | 0.12 (>1.0) | 0.07 (>1.0) | 0.10 (>1.0) |
| $I/\sigma I$ | 18.74 (1.13) | 11.4 (1.20) | 11.92 (0.60) | 13.52 (0.56) | 27.07 (1.38) | 18.01 (0.81) |
| Completeness (%) | 99.9 (99.8) | 99.9 (100) | 99.71 (98.83) | 99.1 (91.8) | 99.9 (99.6) | 99.8 (100) |
| Redundancy | 24.8 (25.5) | 11.0 (11.4) | 15.4 (15.7) | 25.3 (23.1) | 24.5 (25.7) | 24.9 (26.7) |
| CC1/2 (%) | 99.6 (66.4) | 99.5 (56.1) | 99.9 (23.6) | 99.4 (25.8) | 99.9 (69.3) | 99.9 (36.6) |
| Unique reflections | 5592 (533) | 5134 (799) | 9584 (937) | 11036 (1054) | 5365 (515) | 5037 (487) |
| Wilson *B*-factor (Å$^2$) | 90.4 | 111.1 | 162.3 | 138.3 | 122.5 | 145.5 |
| Wavelength (Å) | 0.9198 | 0.9716 | 1.0332 | 1.0332 | 0.9786 | 0.9751 |
| No. mol/AU | 1 | 1 | 2 | 2 | 1 | 1 |
| **Refinement** | | | | | | |
| Resolution | | | | | | |
| No. reflections | 5587 (531) | 5091 (492) | 9568 (928) | 10973 (993) | 5363 (515) | 5037 (487) |
| $R_{work}$ / $R_{free}$ (%) | 21.3/25.5 | 19.8/24.4 | 21.0/26.4 | 17.6/20.3 | 18.8/21.4 | 20.9/22.5 |
| No. of non-H atoms | 1340 | 1346 | 2671 | 2647 | 1347 | 1334 |
| Macromolecules | 1312 | 1289 | 2578 | 2624 | 1313 | 1312 |
| Ligands/ions | 28 | 57 | 93 | 23 | 34 | 22 |
| Average *B*-factors (Å$^2$) | 104.4 | 98.4 | 166.2 | 151.9 | 127.1 | 123.9 |
| Macromolecules (Å$^2$) | 121.3 | 106.9 | 186.4 | 189.5 | 125.1 | 114.7 |
| Ligands/ions (Å$^2$) | 87.5 | 89.8 | 146.0 | 114.4 | 129.0 | 133.2 |
| r.m.s deviations | | | | | | |
| Bond lengths (Å) | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 |
| Bond angles (°) | 0.20 | 0.41 | 0.23 | 0.27 | 0.29 | 0.31 |
| Coord. Precis.[2] (Å) | 0.33 | 0.34 | 0.46 | 0.37 | 0.16 | 0.37 |

[1] Statistics for the highest-resolution shell are shown in parentheses. [2] Calculated by Maximum-Likelihood Estimation.

**Table 9.7:** Description, sequence, calculated and measured masses of RNA oligos prepared by solid phase synthesis.

| description | 5'-sequence-3' | Mass calculated | Mass found |
|---|---|---|---|
| R1 (unmodified substrate) | CCACUGAGAGCUUC | 4403.62 | 4403.63 |
| $m^1$A R1 (methylated product) | CCACUG$m^1$AGAGCUUC | 4418.65 | 4418.65 |
| R1 with SeMeU12 | CCACUGAGAGCUSemUC | 4481.56 | 4481.57 |
| MTR1 5' fragment (R2) | GGAAGCUCUGACCGACCCCCAGCC | 7621.13 | 7621.14 |
| MTR1 3' fragment (R3) | GCUGGGACAACUAGACAUACAGUG | 7735.10 | 7735.13 |
| MTR1 5' fragment C12U | GGAAGCUCUGAUCGACCCCCAGCC | 7622.08 | 7622.11 |
| MTR1 3' fragment U42C | GCUGGGACAACUAGACACACAGUG | 7734.12 | 7734.17 |
| MTR1 5' fragment $m^5$C12 | GGAAGCUCUGA$m^5$CCGACCCCCAGCC | 7635.11 | 7635.14 |
| MTR1 5' fragment $m^4$C12 | GGAAGCUCUGA$m^4$CCGACCCCCAGCC | 7635.11 | 7635.14 |
| MTR1 5' fragment dC12 | GGAAGCUCUGAdCCGACCCCCAGCC | 7605.10 | 7605.13 |
| MTR1 3' fragment dU42 | GCUGGGACAACUAGACAdUACAGUG | 7719.11 | 7719.13 |
| MTR1 5' fragment Cf12 | GGAAGCUCUGACfCGACCCCCAGCC | 7623.09 | 7623.13 |
| MTR1 3' fragment Uf42 | GCUGGGACAACUAGACAUfACAGUG | 7737.10 | 7737.11 |
| MTR1 5' fragment Cm12 | GGAAGCUCUGACmCGACCCCCAGCC | 7635.11 | 7635.17 |
| MTR1 3' fragment Um42 | GCUGGGACAACUAGACAUmACAGUG | 7749.12 | 7749.13 |
| MTR1 3' fragment USem42 | GCUGGGACAACUAGACAUSemACAGUG | 7813.04 | 7816.09 |
| MTR1 3' fragment Am41 | GCUGGGACAACUAGACAmUACAGUG | 7749.12 | 7749.13 |

**Table 9.8:** Ribozymes prepared by *in vitro* transcription.

| No | description | 5'-sequence-3' (binding arms underlined) |
|---|---|---|
| | wt MTR1 ribozyme | GGAAGCUCUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUACAGUGG |
| M1 | C17U | GGAAGCUCUGACCGACUCCCCGAGUUCGCUCGGGGACAACUAGACAUACAGUGG |
| M2 | A15G | GGAAGCUCUGACCGGCCCCCCGAGUUCGCUCGGGGACAACUAGACAUACAGUGG |
| M3 | △C17A31 | GGAAGCUCUGACCGAC-CCCCGAGUUCGCUCGGGG-CAACUAGACAUACAGUGG |
| M4 | A15U+C16G+C17U | GGAAGCUCUGACCGUGUCCCCGAGUUCGCUCGGGGACAACUAGACAUACAGUGG |
| M5 | A15G+ C16U+ C32U+A33G | GGAAGCUCUGACCGGUCCCCCGAGUUCGCUCGGGGAUGACUAGACAUACAGUGG |
| M6 | G14A+C35U | GGAAGCUCUGACCAACCCCCCGAGUUCGCUCGGGGACAAUUAGACAUACAGUGG |
| M7 | C13U+G38A | GGAAGCUCUGACUGACCCCCCGAGUUCGCUCGGGGACAACUAAACAUACAGUGG |
| M8 | U9C | GGAAGCUCCGACCGACUCCCCGAGUUCGCUCGGGGACAACUAGACAUACAGUGG |
| M9 | U42C | GGAAGCUCUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACACACAGUGG |

### 9.2.3 Extended data figures

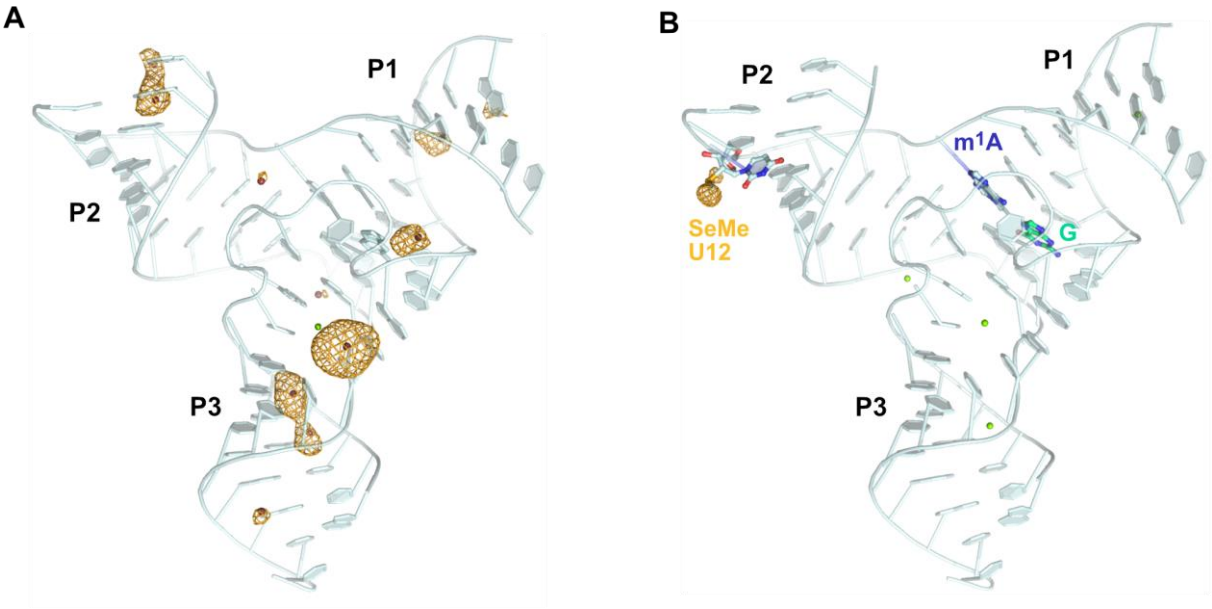



**Figure 9.10:** Heavy atom derivatives of the MTR1 crystal structure. **A** Overall structure of MTR1 co-crystallized with $Tl^+$. **B** MTR1 containing 2′-Selenomethyl-uridine modified residue in the RNA substrate. Yellow mesh indicates anomalous difference Fourier map contoured at **A** 3σ and **B** 5σ. The difference maps were computed from data collected at the Thallium L-III edge and Selenium K-edge, respectively.

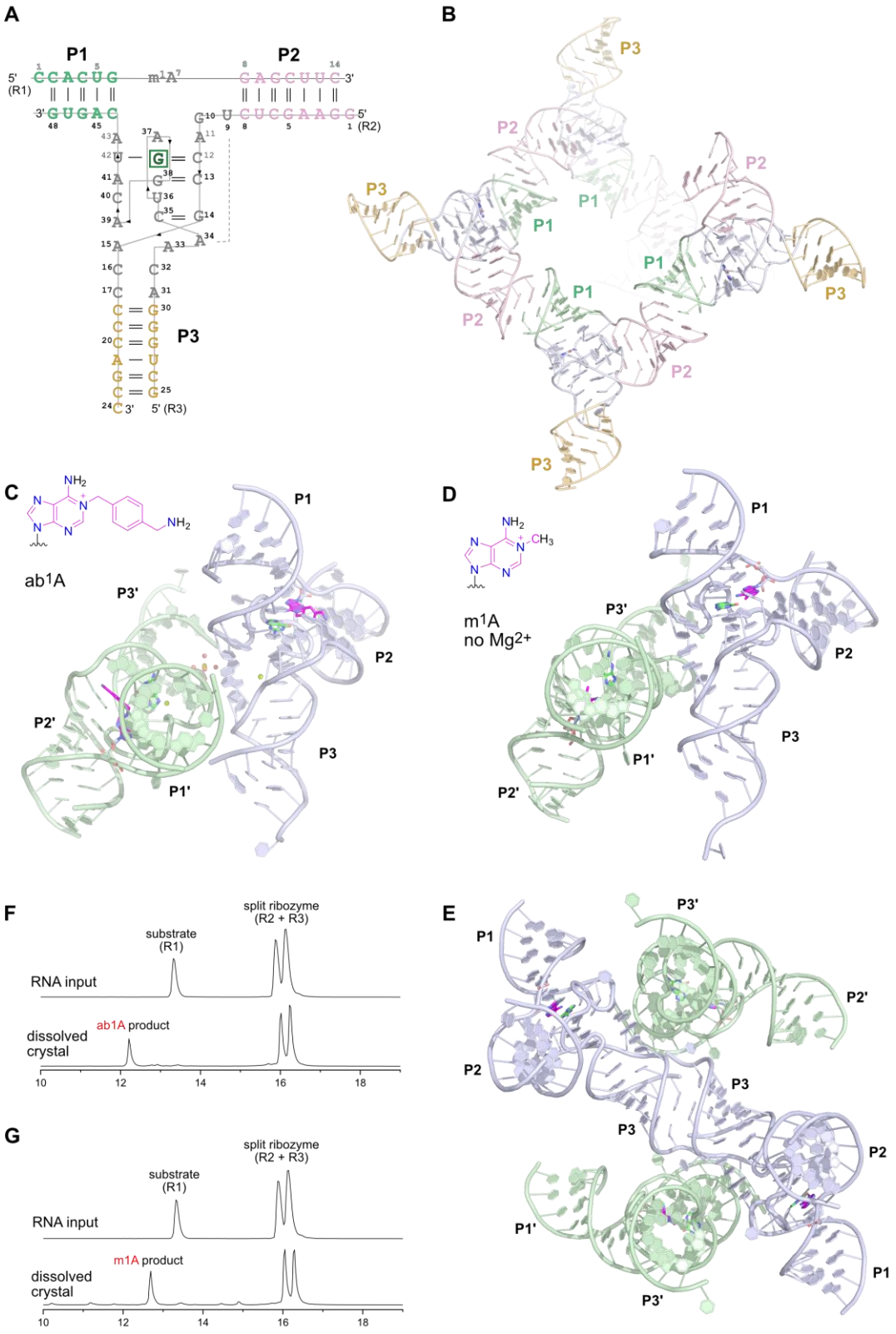



**Figure 9.11:** Overall structures of MTR1 and crystal contacts. **A** Secondary structure scheme of trimolecular MTR1. **B** Packing of overhangs of P1 and P2 to form a semi-continuous double helix. Arrangement of two copies in the asymmetric unit for crystals grown with **C** BG-$NH_2$ ($ab^1A$ in the product) and **D** with $m^6G$ ($m^1A$ in the crystal) but no added $Mg^{2+}$. **E** Crystal contact via stacking of P3 in the structure shown in **D**. **F,G** Anion exchange HPLC analysis of dissolved crystals corresponding to **C** and **D**.

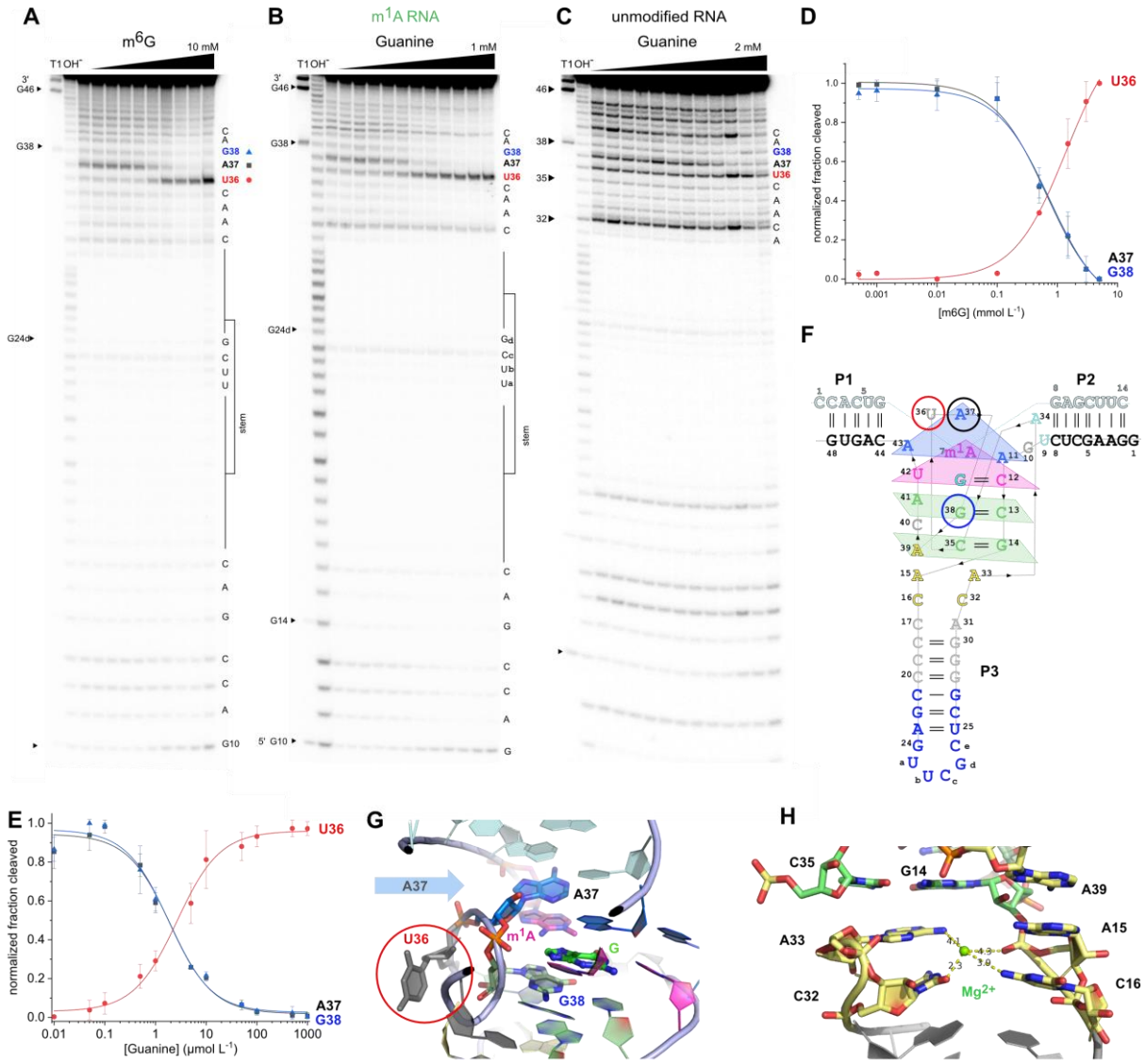



**Figure 9.12:** In-line probing of MTR1. **A-C** Full gel images for the excerpts shown in Figure 4.2 **A** In-line probing of MTR1 hybridized to unmethylated RNA R1, with increasing concentrations of $m^6G$. **B** In-line probing of guanine binding to the MTR1 product complex containing $m^1A$. **C** In-line probing of guanine binding to the MTR1 starting complex hybridized to unmethylated RNA R1. **A-C** Incubation at pH 8.0, 20 °C, 36 h. **D** Normalized band intensities seen in **A**, shown for U36 (red), A37 (black), G38 (blue). The $[m^6G]_{1/2}$ value is ca 800 µM, however, this value cannot be interpreted as a $K_d$ or $K_m$ because of multiple overlapping equilibria (with $m^6G$ and G since partial/slow methylation occurred during incubation). **E** Normalized in-line probing band intensities seen in **B** for U36 (red), A37 (black), G38 (blue). $K_{d,app}$ = 2.0 ± 0.3 µM. Data in **D** and **E** are fitted to a one-site binding model. Error bars denote ± s.d. of the mean for $n$ = 3 (**D**) or 4 (**E**) independent replicates. **F** Secondary structure scheme of MTR1 used in in-line probing experiments (with connecting loop at P3; numbers of nucleotides correspond to the split version). **G** Excerpt of the catalytic core showing solvent exposed location of U36 and stacking of A37 on $m^1A$. **H** Excerpt of the metal ion binding site in the transition of the catalytic core to P3.

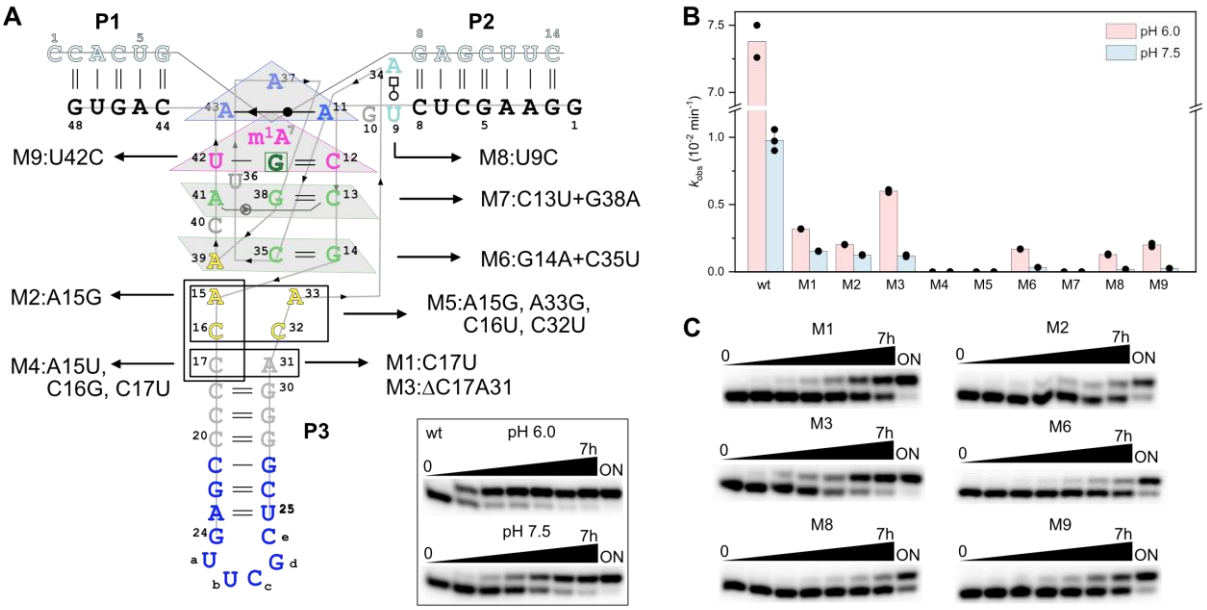



**Figure 9.13:** MTR1 mutagenesis of the core nucleotides and pH-dependence. **A** Scheme of MTR1 with mutated nucleotides and base pairs indicated. **B** Summary of rate constants determined at pH 6.0 and 7.5. Individual data points of two independent replicates are shown as black dots. Reaction of MTR1-wt at pH 7.5 war repeated three times. **C** Representative gel images of kinetic experiments of two independent replicates. Reaction conditions: 5′-$^{32}$P-labeled R1, 10 µM MTR1 or mutant, 100 µM $m^6G$, 40 mM $MgCl_2$, 25 °C. Timepoints: 0, 0.2, 0.5, 1, 2, 4, 7, 23 h (ON).

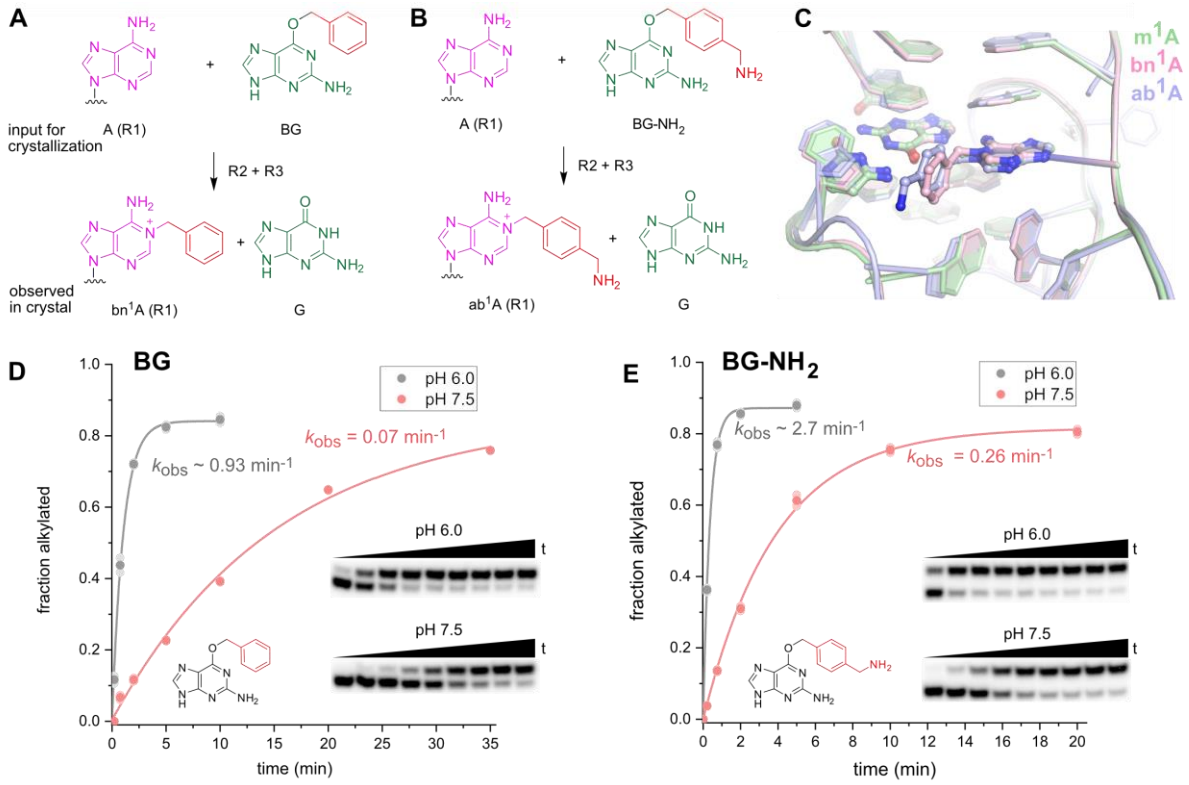



**Figure 9.14:** MTR1 as alkyltransferase ribozyme and pH dependence. **A,B** Chemical structures of substrates (adenosine and benzyl guanine cofactor) before the RNA-catalyzed reaction, and after transfer to the target adenosine in R1 and release of guanine. **C** Overlay of active sites containing $m^1A$, $bn^1A$ or $ab^1A$ in stick representation. **D** Kinetics of benzyl group transfer at pH 6.0 and pH 7.5, with exemplary gel images as inset. **E** Same as **D**, but reaction with BG-$NH_2$. The fastest reaction was observed with BG-$NH_2$ at pH 6.0, yielding ca 90 % $ab^1A$-RNA within 2 min. All reactions were performed in duplicate, individual data points and representative gel images are shown (timepoints shown in gels up to 60 min).

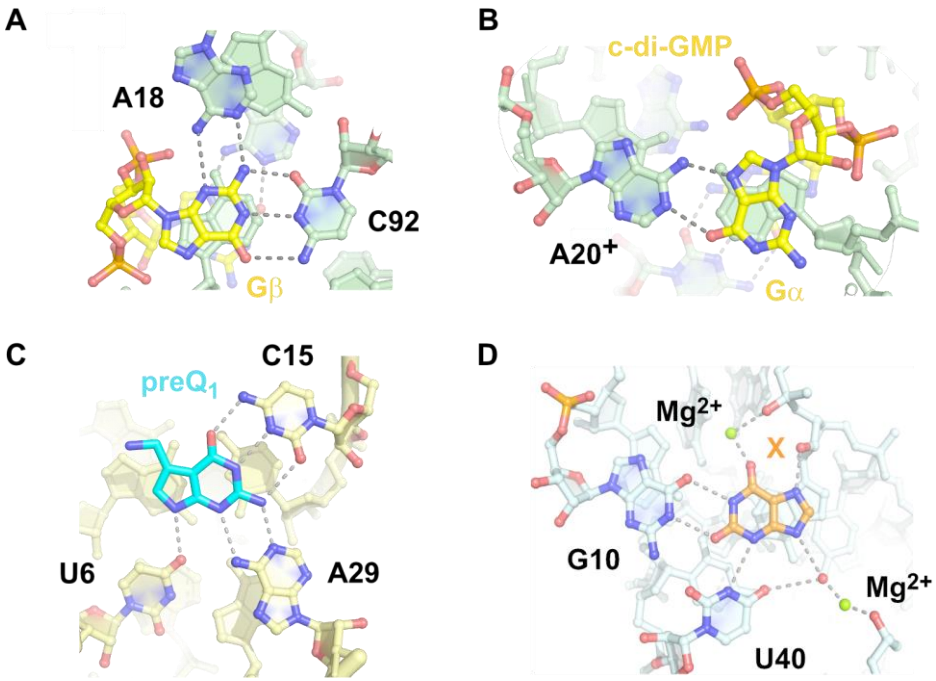



**Figure 9.15:** Comparison to binding sites of purine riboswitches also involving protonated nucleobases or metal ion binding. **A,B** G20A mutant of Vc2 riboswitch with c-di-GMP (PDB: 3MUM), showing **A** $G_\beta$ and **B** $G_\alpha$ bound via stably protonated A20. **C** *T. tengcongensis* $preQ_1$ riboswitch with bound preQ1 (PDB: 3Q50). **D** NMT1 riboswitch in complex with xanthine (PDB 7ELR).

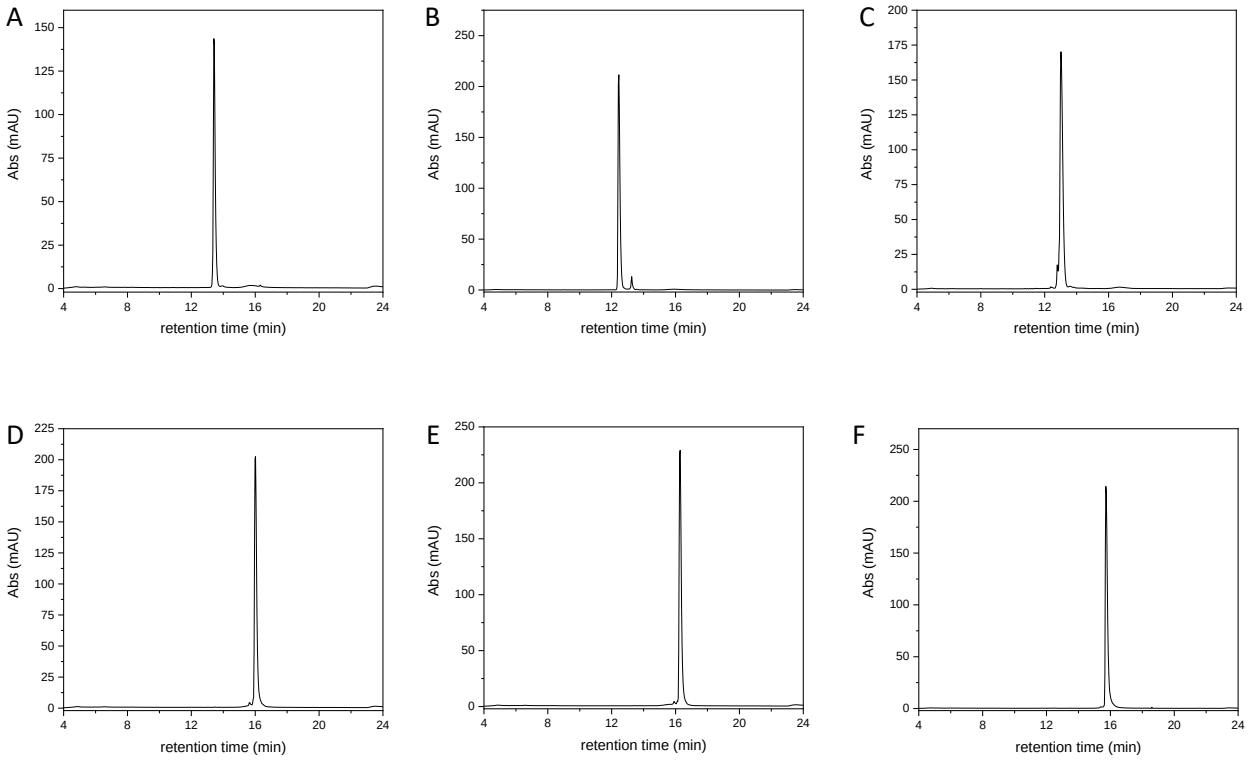



**Figure 8.16:** Continued on the next page.

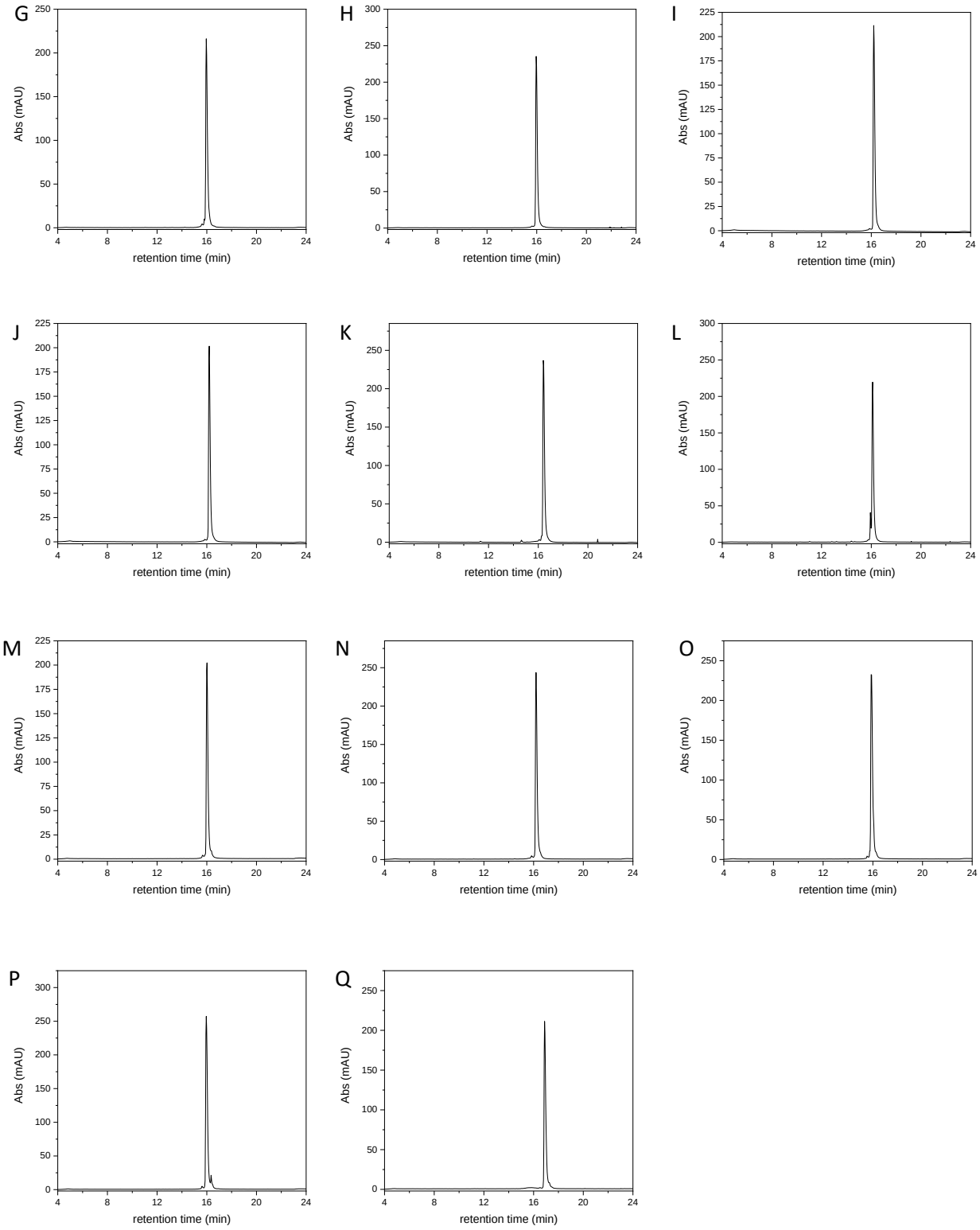



**Figure 9.16:** Anion exchange HPLC chromatograms to confirm purity of RNA oligonucleotides prepared by solid-phase synthesis. **A** R1, **B** $m^1$A R1, **C** R1_USem12 (USem = 2'-SeMe-U), **D** R2, **E** R3, **F** R2_C12U, **G** R3_U42C, **H** R3_dU42, **I** R2_dC12, **J** R2_Cm12, **K** R3_Um42, **L** R3_USem42, **M** R2_Cf12, **N** R3_Uf42, **O** R2_$m^5$C12, **P** R2_$m^4$C12, **Q** R3_Am43. Dionex DNAPAc PA200, 2x250 mm, 60 °C, monitored at 260 nm.

## 9.3 Supporting information for chapter 5

This Supporting Information has been published in ref.[335]. For the sake of unity of this thesis, several editorial changes have been made, which, however, do not affect the contents of the thesis.



### 9.3.1 Material and Methods

#### Materials and Equipment

All standard chemicals and solvents including urea, acrylamide/bisacrylamide stock solution, buffer reagents and phenol/chloroform/isoamyl alcohol were purchased from commercial suppliers and used without further purification. Water used for all experiments was purified with a Milli-Q-unit from Millipore. Snake venom phosphodiesterase (SVPD) was purchased from Merck. Bacterial alkaline phosphatase (BAP) was from Invitrogen. RNase T1, T4 Polynucleotide kinase (PNK), Calf Intestinal Alkaline Phosphatase (CIAP) and Klenow exo$^-$ were from ThermoFisher Scientific. Lucifer Yellow-carbohydrazide, Fluorescein-thiosemicarbazide and Biotin-hydrazide were purchased from Sigma. 1-Ethyl-2,3,3-trimethylindolenium-5-sulfonate (**4**) was from Toronto Research Chemicals. Nebularine nucleoside was purchased from Carbosynth. γ-$^{32}$P-ATP was from Hartmann Analytic GmbH.

Unmodified DNA oligonucleotides were purchased from Microsynth and purified on denaturing PAGE (10-20 % acrylamide) prior to use. Modified and unmodified short RNA oligonucleotides were synthesized in house using solid phase synthesis with standard phosphoramidite chemistry. Phosphoramidites for incorporation of unmodified and modified nucleosides were purchased from ChemGenes. NTPs for transcription were purchased from Jena biosciences. T7 RNA polymerase was prepared as previously described.[398]

Fluorescence gel images were recorded using a ChemiDoc MP with epi illumination using blue LEDs (emission filters 530/28) from BioRad. Gels containing radioactive oligonucleotides were exposed to phosphor storage screens and imaged using a Typhoon Phosphorimager from Cytiva. Fluorescent measurements were performed on a Jasco FP-8300 spectrofluorometer. UV-vis spectra were recorded on an Agilent Cary 3500 spectrophotometer with an Agilent Cary UV-Vis Multicell Peltier.

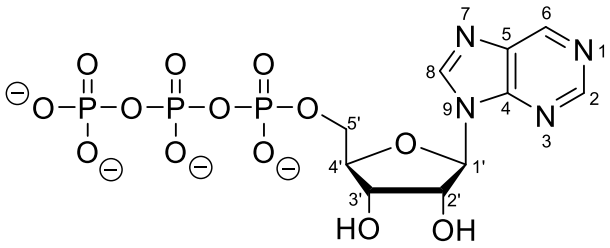
NMR spectra were recorded on a Bruker Avance III HD 400 spectrometer. Chemical shifts were measured relative to the solvent signal ($^1$H-NMR: 4.79 ppm for $D_2O$). Signal multiplicities are denoted as s (singlet), d (doublet), t (triplet) and m (multiplet). Processing of the raw data was performed with Topspin 3.5.

HR-ESI-MS spectra of the oligonucleotides were recorded on a Bruker micrOTOF-Q III spectrometer in negative ion mode upon direct injection of a 3 µM RNA solution in 44 % ACN (v/v) containing 44 mM hexafluoroisopropanol (HFIP) and 3.8 mM triethylamine (TEA), using capillary voltage 3 kV, end plate voltage 2.5 kV, nitrogen nebulizer pressure 0.4 bar, dry gas flow 4 L/min, dry temperature 200 °C, and recorded in the range m/z 200 - 4000. Data evaluation and ion deconvolution were performed by Data analysis software DA 4.2 (Bruker Daltonics).

**RNA solid phase synthesis**

Short RNA oligonucleotides were prepared by solid-phase synthesis using phosphoramidite chemistry with 2′-*O*-TOM-protecting groups. The RNAs were deprotected with a 1:1 mixture of aqueous ammonia and methylamine (AMA) at 37 °C for 4-6 h, followed by 1 M tetrabutylammonium fluoride in THF at 25 °C for 12-16 h. The RNAs were desalted and purified by denaturing PAGE, extracted, and EtOH-precipitated. The RNA purity was confirmed by anion exchange HPLC (Dionex DNAPac PA200, 2 × 250 mm at 60 °C; solvent A: 25 mM Tris-HCl (pH 8.0) and 6 M urea; solvent B: 25 mM Tris-HCl (pH 8.0), 6 M urea and 0.5 M $NaClO_4$; linear gradient: 0-40 % solvent B, with a slope of 4 % solvent B per column volume). The identity of the RNAs was confirmed by HR-ESI-MS (micrOTOF-Q III, negative ion mode, direct injection). The oligonucleotide sequences are summarized in Table 9.9 together with their respective measured and calculated molecular weights and anion exchange HPLC chromatograms.

**Synthesis of Nebularine-triphosphate (NeTP)**



**1) Synthesis of bis(tributylammonium)pyrophosphate**

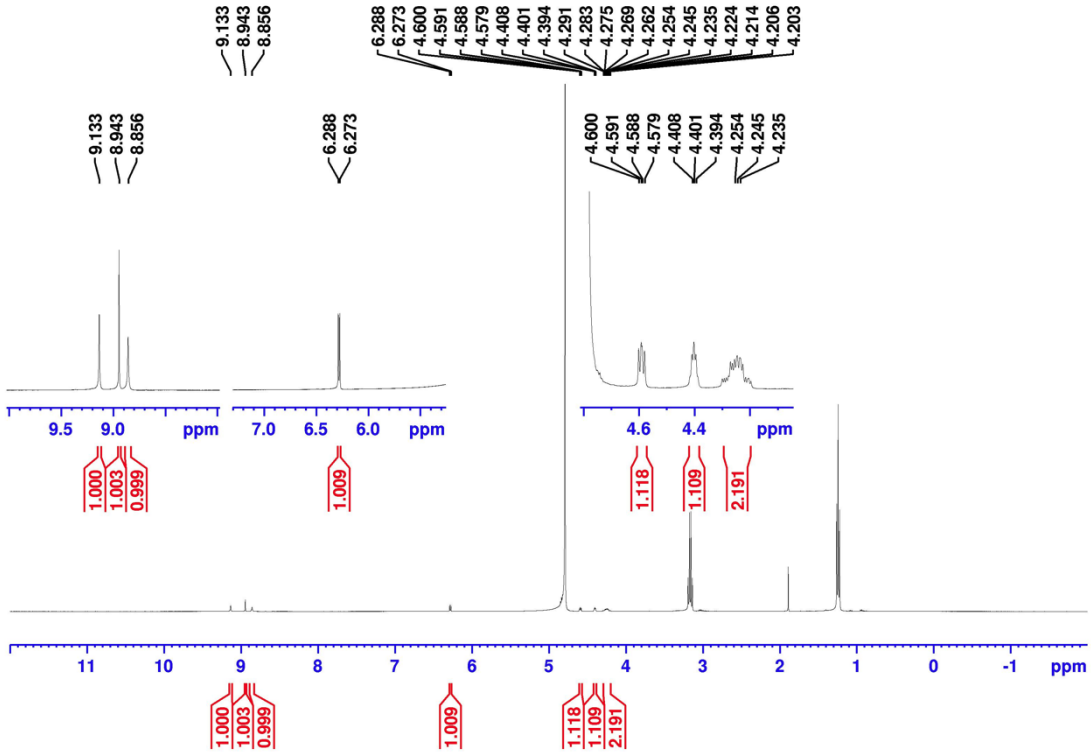
Ion exchange resin Dowex ($H^+$) (15 g) was filled in a column and washed multiple times with methanol until the eluate was colorless followed by washing with nanopure water. Tetrasodium pyrophosphate·10 $H_2O$ (1.12 g, 2.51 mmol, 1.0 eq.) was dissolved in 10 mL nanopure water, added to the resin, and incubated at ambient temperature for 10 min. The pyrophosphate was eluted with nanopure water until a pH of 7.0 was reached. The solution was then adjusted to pH 3.5 using 40 % aq. tetrabutylammonium hydroxide. The clear liquid was lyophilized to yield a colorless solid.

**2) Synthesis of nebularine-5'-triphospate (NeTP) sodium salt**

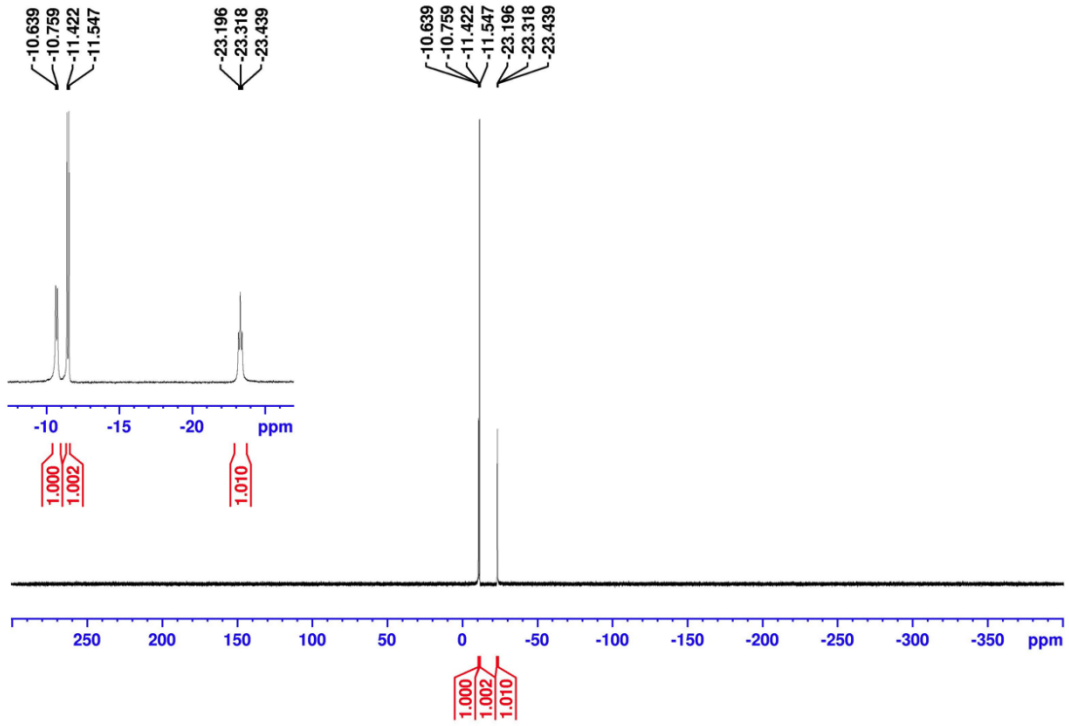
Synthesis of nebularine-5'-triphosphate (NeTP) was previously reported, and here performed following a general established procedure for nucleoside triphosphates.[412] Nebularine (25 mg, 0.10 mmol) and proton sponge (28.4 mg, 0.13 mmol, 1.3 eq.) were dissolved in trimethyl phosphate (1.5 mL). After cooling to -20 °C, freshly distilled $POCl_3$ (14 µL, 0.15 mmol, 1.5 eq.) was added and the reaction mixture was stirred at -15 °C to -20 °C for 3 h. A solution of bis(tributylammonium)pyrophosphate in dry DMF (1 M, 0.60 mL) and tri-*n*-butylamine (0.15 mL, 0.63 mmol, 6.3 eq.) were added. After stirring for 1.5 h at -15 °C to -20 °C, the reaction was quenched by the addition of triethylammonium bicarbonate (TEAB) buffer (1 M, pH 7.5, 12 mL), warmed to room temperature and washed three times with ethyl acetate. The aqueous layer was lyophilized, and the crude product was purified using anion exchange chromatography (DEAE Sephadex A-25 column, solvent A: 0.1 M TEAB, pH 7.5; solvent B: 1 M TEAB, pH 7.5; linear gradient: 0-100 % solvent B in 40 min)

The product-containing fraction was lyophilized and further purified by reversed-phase HPLC using triethylammonium acetate (TEAA) buffer (Nucleosil, 8x250 mm, C18 column at 30 °C; solvent A: 100 mM TEAA (pH = 7.0); solvent B: 100 mM TEAA (pH 7.0) in 90 % ACN; linear gradient: 0.1-10 % solvent B in 5 CV, 2 mL/min).

The obtained triethylammonium salt was converted to the sodium salt by precipitation from 0.3 ml NaOAc (3 M) and 0.9 mL ice cold abs. EtOH. After 30 min centrifugation at 4 °C, the product was obtained as a white solid, dissolved in nanopure water and the concentration was determined with UV absorption spectroscopy using $\varepsilon^{245}$ = 7000 $M^{-1}cm^{-1}$.[413] The yield was 7.0 µmol (7 %).




**$^{1}$H** NMR (400 MHz, $D_2O$) δ(ppm) = 9.13 (s, 1H, 6-H), 8.94 (s, 1H, 2-H), 8.86 (s, 1H, 8-H), 6.28 (d, *J* = 5.9 Hz, 1H, 1'-H), 4.59 (dd, *J* = 5.1, 3.5 Hz, 1H, 3'-H), 4.40 (p, *J* = 2.9 Hz, 1H, 4'-H), 4.25 (qdd, *J* = 11.7, 5.3, 3.0 Hz, 2H, 5'-H).




**$^{31}$P** NMR (162 MHz, $D_2O$) δ(ppm) = -10.70 (d, *J* = 19.7 Hz), -11.48 (d, *J* = 20.2 Hz), -23.32 (t, *J* = 20.0 Hz).

### *In vitro* transcription and enzymatic incorporation of nebularine

MTR1 ribozymes with varying binding arms, short unmodified und Ne containing substrate RNAs as well as the tRNA were prepared by *in vitro* transcription with T7 RNA polymerase from synthetic DNA templates (Microsynth), as described previously.[184] The reactions contained 1 µM DNA template, 4 mM NTPs, 30 mM $MgCl_2$, 2 mM spermidine and 10 mM DTT. The template for tRNA transcription was assembled from two ssDNA oligonucleotides using Klenow exo$^-$. For Ne incorporation, ATP was substituted by an equal amount of NeTP. The sequences of the transcripts are summarized in Table 9.10.

### 5'-$^{32}$P-labeling of oligonucleotides

5'-Radiolabeling reactions were performed by incubation of 100 pmol RNA in 10 µL 1x PNK buffer (50 mM Tris, pH 7.6, 10 mM $MgCl_2$, 5 mM DTT, 0.1 mM spermidine) in the presence of 5 µCi γ-$^{32}$P-ATP and 5 U T4 PNK for 1h at 37 °C. The reaction was quenched by addition of loading dye followed by PAGE-purification.

*In vitro* transcripts were dephosphorylated beforehand by incubation of 100 pmol RNA transcript in 10 µL 1x CIAP buffer (50 mM Tris-HCl (pH 8.5), 0.1 mM EDTA) and 1 U CIAP for 1 h at 37 °C followed by PCI-extraction and EtOH-precipitation.

### Analytical scale aldehyde generation and functionalization

Generation of the aminoimidazole aldehyde **3** for further functionalization was performed using 10 pmol 5′-$^{32}$P-labeled Ne-containing RNA and 100 pmol MTR1 ribozyme, that were folded (3 min at 95 °C then 10 min at 25 °C) in reaction buffer (120 mM KCl, 5 mM NaCl and 50 mM MES (pH 6.0)). 100 µM substrate BG and 40 mM $MgCl_2$ were added to a final concentration of 10 µL. After overnight incubation at 25 °C, the reaction was EtOH-precipitated and 1/8 of the reaction (1.25 pmol modified RNA substrate with 12.5 pmol MTR1) was submitted to the aldehyde functionalization reaction.

To attach different probes, the MTR1-reacted RNA was mixed with 25 mM *N,N*-Dimethylethylenediamine (*N,N*-DMED, adjusted to pH 7.5 with HCl) and 500 µM reagent (Fluorescein-thiosemicarbazide, Biotin-hydrazide or Lucifer Yellow-carbohydrazide), followed by 37 °C incubation for 3 h.

Generation of the chromophore **5** was performed by mixing the MTR1-treated RNA with 25 mM Tris (pH 8.5), 100 mM KCl and 75 mM of commercially available 1-ethyl-2,3,3-trimethyl-

indoleninium-5-sulfonate (**4**, dissolved in 0.1 M NaOH). The reaction mixture was incubated at 45 °C for 2 h.

For both functionalization reactions, aliquots of 1 µL were taken at timepoint 0 and after completion of the reaction and quenched by immediate addition of loading dye (4 µl) and freezing in liquid nitrogen. Half of each timepoint sample was resolved on denaturing PAGE (20 %) and the band intensities were quantified using phosphorimaging.

For conversion of chromophore **5a** to isomer **5b**, 125 pmol PAGE purified RNA modified with Indole **4** were incubated in 15 µL 25 mM Tris (pH 8.0) at 37 °C or 50 °C for 4 h or 23 h followed by immediate analysis using anion exchange HPLC.

**One-pot aldehyde generation and functionalization reactions**

For one-pot modification reactions, 5 pmol of 5′-$^{32}$P-labeled Ne-containing RNA and 50 pmol of the respective MTR1 ribozyme were annealed (3 min at 95 °C then 10 min at 25 °C). 100 µM BG and 40 mM $MgCl_2$ were added. For fluorescein modification, the sample was mixed with 25 mM *N,N*-DMED (pH 6.0/7.5) and 500 µM Flu-TSC. For generation of chromophore **5**, 25 mM MES (pH 6.0) or Tris (pH 7.5), 100 mM KCl and 75 mM **4** were used instead. The reactions had a final volume of 5 µL and were incubated at 25 °C or 37 °C.

1µL aliquots were taken at timepoint 0 and after ON incubation and quenched by immediate addition of loading dye (4 µl). Half of each timepoint sample was resolved on denaturing PAGE (20 %) and the band intensities were quantified using phosphorimaging.

**Preparative scale generation of the aldehyde (3)**

For generation of the aldehyde **3**, 1 nmol Ne-containing RNA were mixed with 1.2 nmol of the respective MTR1 ribozyme in reaction buffer (120 mM KCl, 5 mM NaCl and 50 mM MES (pH 6.0)). After annealing (3 min at 95 °C then 10 min at 25 °C) 100 µM substrate BG and 40 mM $MgCl_2$ were added to a final reaction volume of 20 µL. After overnight incubation at 25 °C the RNA was EtOH-precipitated and separated from the ribozyme and $N^1$-modified RNA on denaturing PAGE. The isolated RNA was then submitted for ESI-MS measurements. The calculated and observed masses are summarized in Table 9.11.

**Two-step preparative scale generation of chromophore (5)**

Generation of the aldehyde **3** was performed by mixing 3 nmol Ne-containing RNA and 3.4 nmol of the respective MTR1 ribozyme in reaction buffer (120 mM KCl, 5 mM NaCl and 50 mM MES (pH 6.0)). After annealing (3 min at 95 °C then 10 min at 25 °C) 100 µM substrate BG and 40 mM $MgCl_2$ were added to a final reaction volume of 60 µL. After overnight incubation at 25 °C the RNA was EtOH-precipitated and dissolved in 25 mM Tris (pH 8.5), 100 mM KCl and 75 mM **4** (in 0.1 M NaOH) to a final reaction volume of 40 µL. The reaction was incubated at 45 °C for 2 h followed by EtOH-precipitation and PAGE purification. The isolated RNA was submitted for ESI-MS measurements. The calculated and observed masses are summarized in Table 9.11.

**Analysis of modified RNA using alkaline hydrolysis and RNase T1 digestion**

RNase T1 digestion was performed by incubating 90 IPS 5'-$^{32}$P-RNA at 37 °C for 30 s in 5 µL 50 mM Tris-HCl (pH 7.5) using 5 U RNase T1. For alkaline hydrolysis, 300 IPS $^{32}$P-RNA were incubated in 10 µL NaOH (25 mM) at 95 °C for 5 min. The reactions were quenched by immediate addition of loading dye and placing on ice.

**Anion exchange HPLC analysis of modified RNA**

90 pmol substrate RNA were annealed with 100 pmol of the respective MTR1 ribozyme (3 min at 95 °C the 10 min at 25 °C) in reaction buffer (120 mM KCl, 5 mM NaCl, 50 mM MES, pH 6.0). 100 µM substrate BG and 40 mM $MgCl_2$ were added to a final volume of 10 µL and the reaction was incubated overnight at 25 °C. The sample was analyzed by anion exchange HPLC on a Dionex DNAPac PA200 column, 2 × 250 mm, at 60 °C with UV detection at 260 nm. Solvent A was 25 mM Tris-HCl (pH 8.0) and 6 M urea, and solvent B was 25 mM Tris-HCl (pH 8.0), 6 M urea and 0.5 M $NaClO_4$, with a linear gradient of 0-48% solvent B in 12 column volumes.

For anion exchange HPLC-analysis of the generation of chromophore **5**, the sample was prepared in a two-step preparative scale reaction as described above and 100 pmol of the PAGE purified RNA were injected.

**Digestion and LC-MS analysis of modified RNA**

For LC-MS analysis, 450 pmol of the RNA were digested at 37 °C for 20 h using 12 U bacterial alkaline phosphatase (BAP) and 1 U snake venom phosphodiesterase (SVPD) in 50 µL 40 mM Tris-HCl (pH 7.5) containing 20 mM $MgCl_2$. The sample was then brought to 100 µL with $H_2O$

and extracted twice using an equal volume of chloroform. The aqueous layer was lyophilized and analyzed by LC-MS using an RP-18 column (Synergi, 4-µm Fusion-RP C18 80 Å, 250 × 2 mm; Phenomenex) at 25 °C with aqueous mobile phase A (5 mM $NH_4OAc$, pH 5.3) and organic mobile phase B (100 % ACN). The flow rate was 0.2 ml/min with a gradient of 0-5 % B in 15 min, followed by 5-70 % B in 30 min. The micrOTOF-Q III with an ESI ion source was operated in positive-ion mode, with capillary voltage of 4.5 kV, end plate offset of 500 V, nitrogen nebulizer pressure 1.4 bar, dry gas flow 9 l/min and dry temperature 200 °C. The data were analyzed with Data Analysis software DA 4.2 (Bruker Daltonics).

**UV-vis measurements of modified RNA**

UV-vis measurements were conducted in a total volume of 100 µL $H_2O$ containing 500 pmol RNA modified with **4** that was freshly prepared in a two-step preparative scale reaction as described above. The spectra were measured in Hellma Quartz Suprasil cuvettes (10 mm path length) at ambient temperature using the following conditions: 200-550 nm measurement range, 2 nm spectral bandwidth, 3000 nm/min scan speed, 1 nm data interval, 20 ms averaging time.

**Measurements of fluorescence emission spectra of modified RNA**

Fluorescence emission spectra were recorded using 30 pmol RNA in 12 µL 10 mM Tris (pH 7.5). The RNA was freshly modified with **4** in a two-step preparative scale reaction as described above. For measurement with the complementary RNAs, 30 pmol of the RNA modified with **4** were annealed to 30 pmol of its complement (3 min at 95 °C then 10 min at 25 °C). Measurements were performed in Hellma Quartz Suprasil cuvettes (1.5x1.5 mm) at 25 °C using the following conditions: 500-700 nm measurement range, excitation at 486 nm, 5 nm ex/em bandwidth, 50 ms response, 0.2 nm data interval, 1000 nm/min scan speed, 680 V PMT voltage. The spectra were manually background corrected by subtraction of a spectrum of the respective untreated RNA in pure buffer.

## 9.3.2 Supplementary tables

**Table 9.9:** RNA oligonucleotides prepared by solid phase synthesis together with calculated and measured masses using HR-ESI-MS. Red denotes the targeted modification site. Anion exchange HPLC traces to confirm purity of the RNAs are summarized in Figure 9.24.

| description | 5'-sequence-3' | Mass (calculated) | Mass (found) |
|---|---|---|---|
| R1_un | CCACUGAGAGCUUC | 4403.62 | 4403.63 |
| R2_un | GGAUAAUACGACUCAC | 5085.74 | 5085.75 |
| R1_Ne | CCACUGNeGAGCUUC | 4388.62 | 4388.63 |
| R2_Ne | GGAUAAUACGNeCUCAC | 5070.73 | 5070.72 |
| R2_DAP | GGAUAAUACGDAPCUCAC | 5100.75 | 5100.74 |
| R2_2AP | GGAUAAUACG2APCUCAC | 5085.74 | 5085.73 |

**Table 9.10:** RNA oligonucleotides prepared by *in vitro* transcription. Red denotes the targeted modification site, blue denotes incorporation of A or Ne, green denotes nucleotides placed opposite chromophore **5**.

| description | 5'-sequence-3' |
|---|---|
| Tr1 | GGGUGCGUCGAGUUCCUGG |
| Tr1_G5A | GGGUACGUCGAGUUCCUGG |
| Tr1_C16A | GGGUGCGUCGAGUUCAUGG |
| Tr1_G10A = Tr2 | GGGUGCGUCAAGUUCCUGG |
| Tr1_G5A+C16A | GGGUACGUCGAGUUCAUGG |
| MTR1_R1 | GGAAGCUCUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUA CAGUGG |
| MTR1_R2 | GGUGAGUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUACG UAUUAUC |
| MTR1_Tr1 | GGCCAGGAACUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACA UACGACGCACCC |
| MTR1_Tr1_G5A+C16A | GGCCAUGAACUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACA TACGACGUACCC |
| MTR1_Tr1_G10A | GGCCAGGAACUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACA UAUGACGCACCC |
| U-complement of R1 | GGAAGCUCUCAGUGG |
| A-complement of R1 | GGAAGCUCACAGUGG |
| C-complement of R1 | GGAAGCUCCCAGUGG |
| G-complement of R1 | GGAAGCUCGCAGUGG |
| *T.therm.* tRNA-Asp A58 | GGCCCCGUGGUGUAGUUGGUUAACACACCCGCCUGUCACGUGGGAGAU CGCGGGUUCGAGUCCCGUCGGGGCCGCCA |
| MTR1 *T.therm.* tRNA-Asp A58 | GGACGGGACUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAU ACGAACCC |

**Table 9.11:** Isolated RNA modification products together with calculated and measured masses using HR-ESI-MS. Red denotes the modification (Ald = 5-aminoimidazole-4-carbaldehyde; Hcy = hemicyanine).

| description | 5'-sequence-3' | Mass (calculated) | Mass (found) |
|---|---|---|---|
| R1_Ne_Ald | CCACUGAldGAGCUUC | 4379.62 | 4379.60 |
| R1_Ne_Hcy | CCACUGHcyGAGCUUC | 4628.70 | 4628.71 |

## 9.3.3 Extended data figures

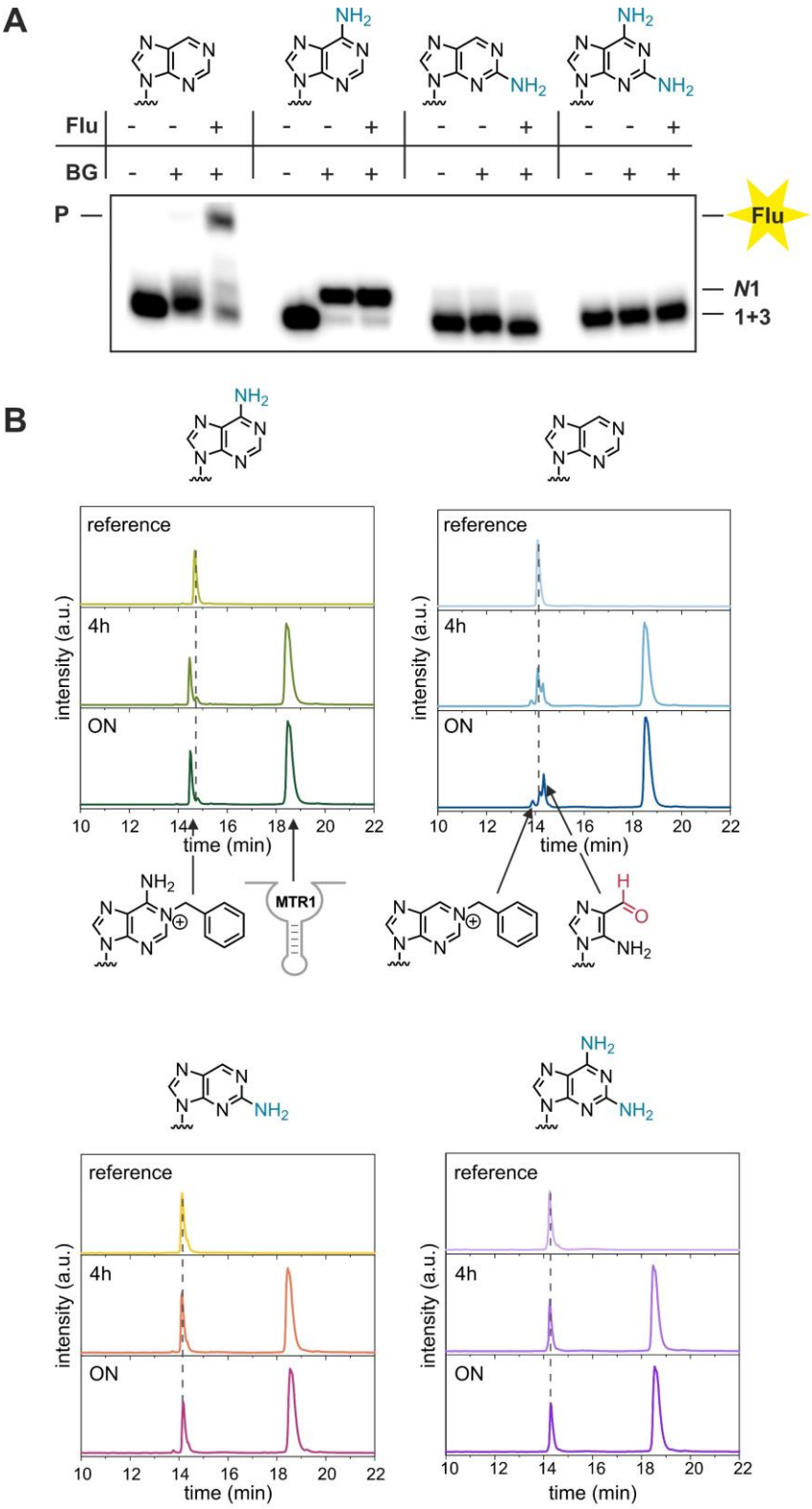



**Figure 9.17:** Reactivity of atomic nebularine mutants. **A** Atomic mutagenesis of the RNA substrate R2. The RNA was incubated with BG at 25 °C for 23 h in the presence (+) or absence (-) of BG followed by labeling with Flu-TSC (+). P = fluorescently labeled reaction product **B** Anion exchange HPLC analysis of MTR1-catalyzed reaction of RNA R2 atomic mutants with A, Ne, 2AP or DAP as target sites.

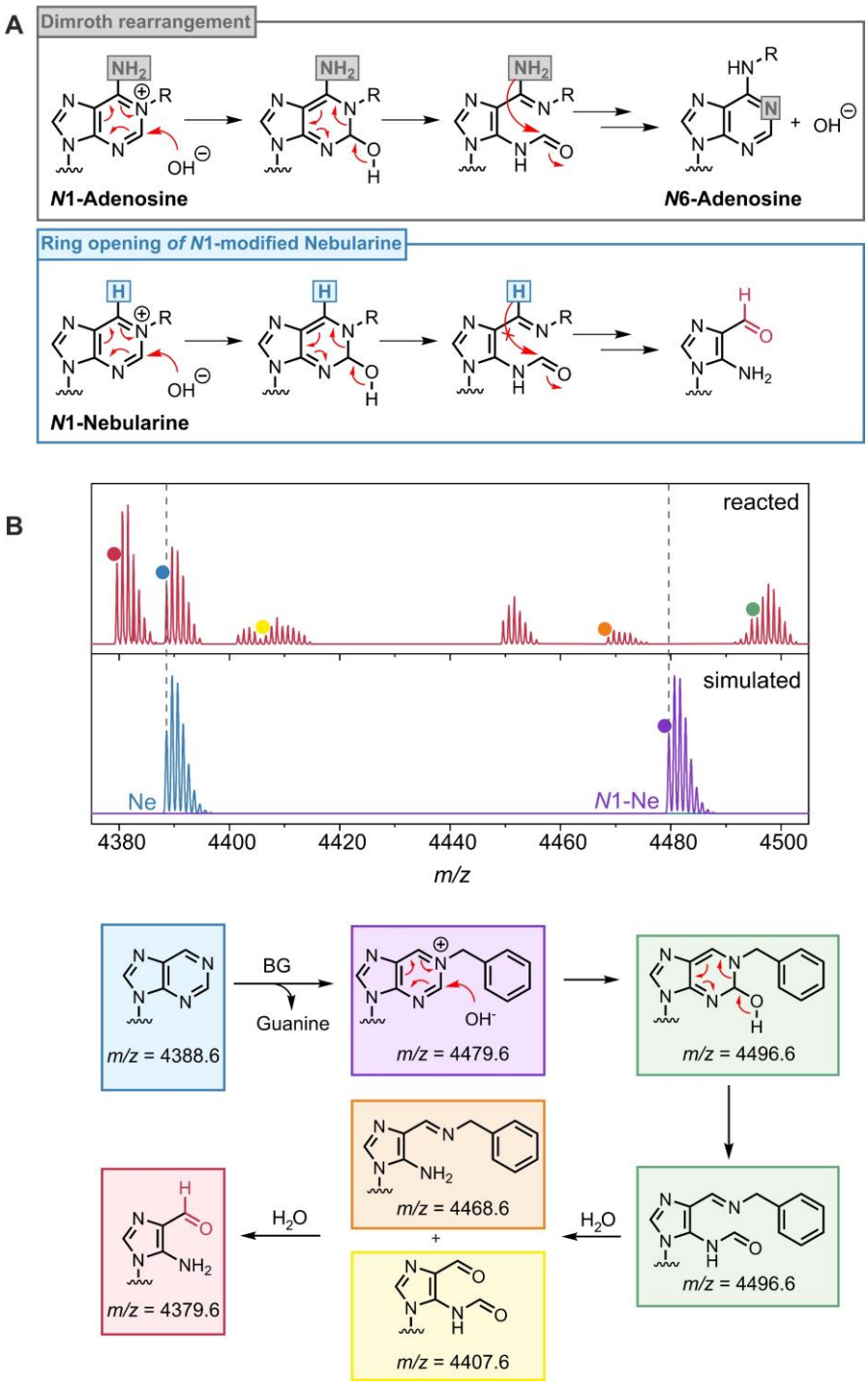



**Figure 9.18:** Mechanism of the MTR1 ribozyme-catalyzed aminoimidazole-aldehyde generation. **A** Reaction mechanism of the Dimroth rearrangement of $N^1$- to $N^6$-modified adenosine and analogous reaction on $N^1$-modified nebularine. **B** Deconvoluted HR-ESI-MS spectrum of the isolated MTR1 + BG treated Ne-RNA R1 (top) and the simulated references of respective unmodified or $N^1$-benzylated Ne-RNA R1 (bottom) together with the proposed reaction mechanism derived thereof. Note: C2 is expected to be more electrophilic than C6. However, in an alternative mechanism, the first attack of $H_2O$ (or $OH^-$) could also occur at C6 instead of C2, which would lead to isomeric intermediates with same mass, and result in the same final product.

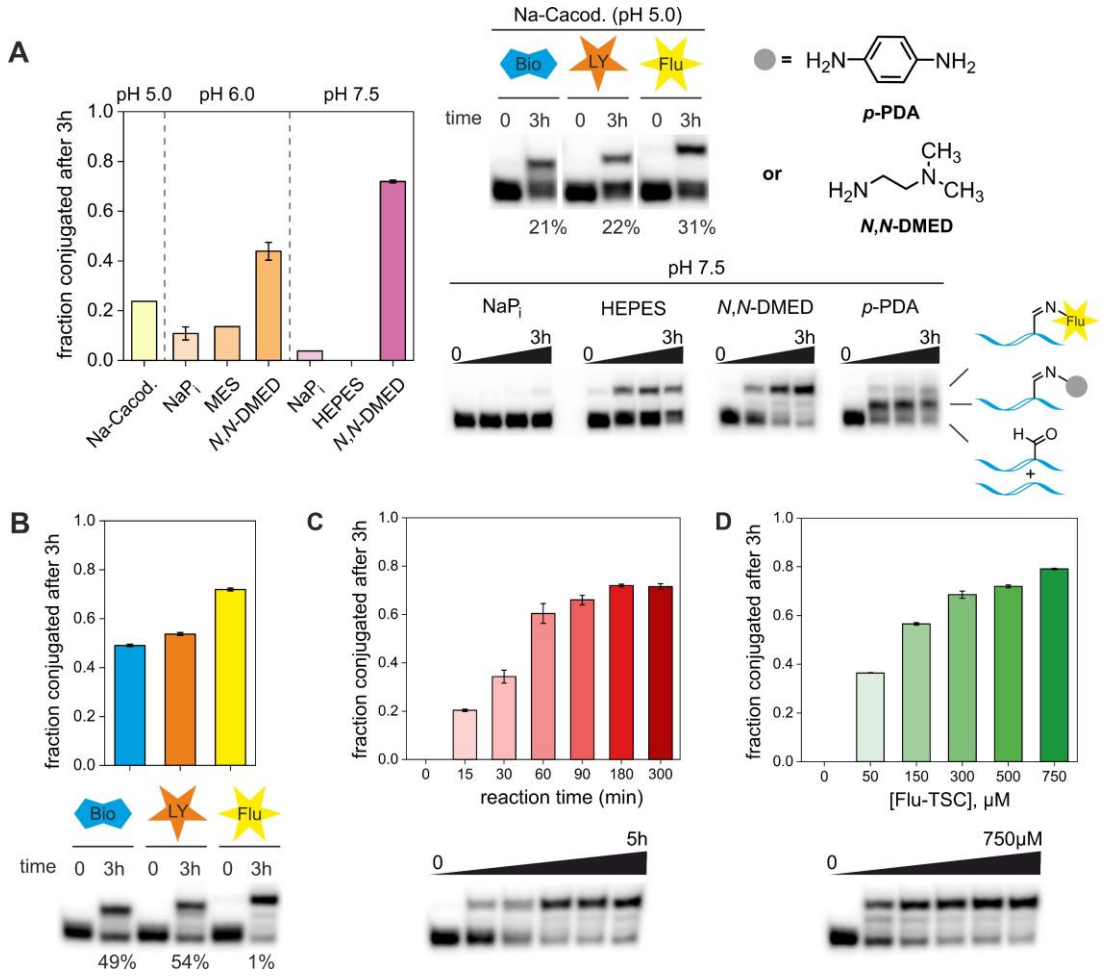



**Figure 9.19:** Optimization of the aldehyde labeling reaction. **A** Aldehyde functionalization using various buffers with different pH. Ne-RNA R1 was treated with MTR1 + BG followed by reaction with Flu-TSC for 3 h at 37°C. **B** Aldehyde functionalization on Ne-RNA R1 treated with MTR1 + BG using Flu-TSC in *N,N*-DMED (pH 7.5) for 3 h at 37 °C. **C** Time dependency of the aldehyde functionalization in Ne-RNA R1 in *N,N*-DMED (pH 7.5) using 500 µM Flu-TSC with incubation at 37 °C. **D** Flu-TSC concentration dependency of the aldehyde functionalization in Ne-RNA R1 in *N,N*-DMED (pH 7.5) with 3 h incubation at 37 °C.

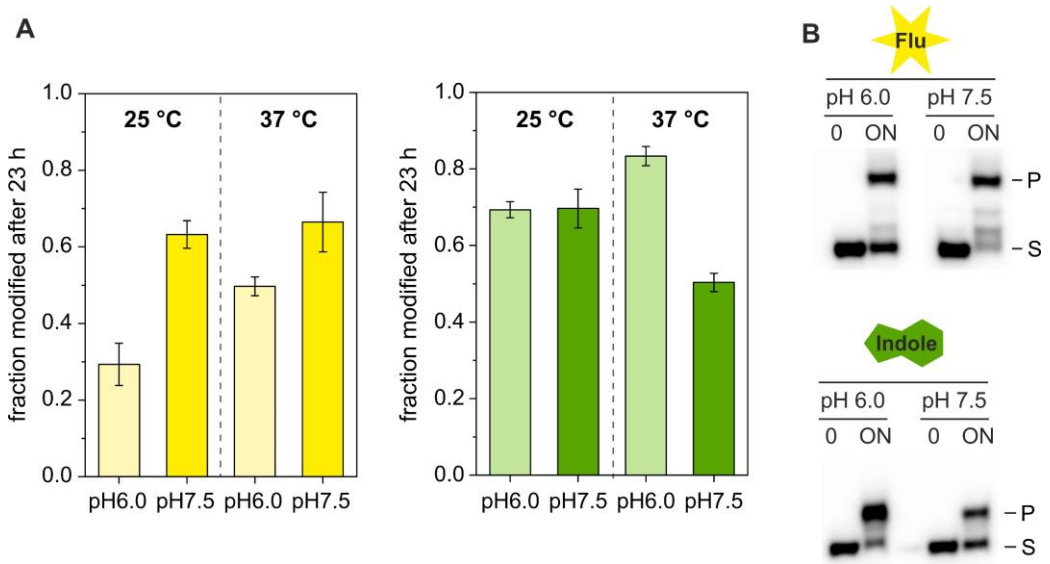



**Figure 9.20: A** Analysis of one-pot Ne-modification reactions in Ne-RNA R1 using Flu-TSC (yellow) or 4 (green). Reactions were conducted for 23 h at 25 °C or 37 °C using *N,N*-DMED (pH 6.0/pH7.5) for Flu-TSC or MES (pH 6.0)/Tris (pH 7.5) for **4**. **B** Representative gel images from one-pot reactions performed at 37 °C. P, product; S, substrate. ON = over night, here 23 h.

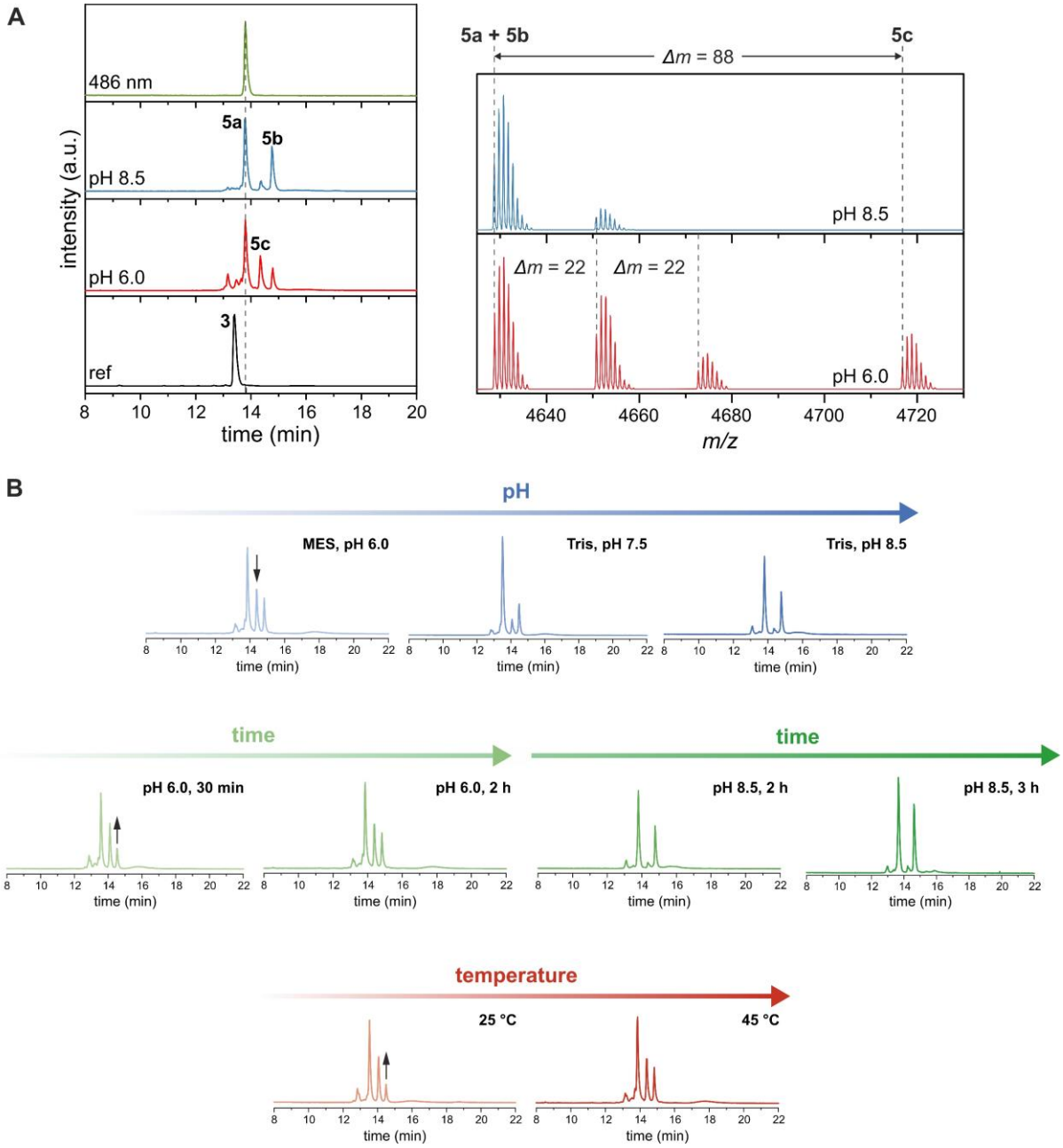



**Figure 9.21:** Optimization of the synthesis of chromophore **5** on RNA. **A** Anion exchange HPLC analysis of the generation of chromophore **5a** in Ne-RNA R1 with formation of side products **5b** and **5c** conducted at pH 6.0 (monitoring at 260 nm) or at pH 8.5 (monitoring at 260 nm and 486 nm). **B** Deconvoluted HR-ESI-MS spectra of the isolated products for the Ne-RNA R1 reaction that was conducted at pH 8.5 (top) or pH 6.0 (bottom). Δm = 22 denotes $Na^+$-adducts. **C** Anion exchange HPLC analysis of PAGE-purified modification products. Ne-RNA R1 was first reacted overnight with MTR1 in the presence of BG followed by reaction with **4** under different conditions, as indicated by variation of pH, reaction time or temperature. The standard set of conditions were (unless otherwise mentioned): 85 µM substrate RNA in 25 mM MES (pH 6.0), 100 mM KCl and 75 mM 4 at 45 °C for 2 h.

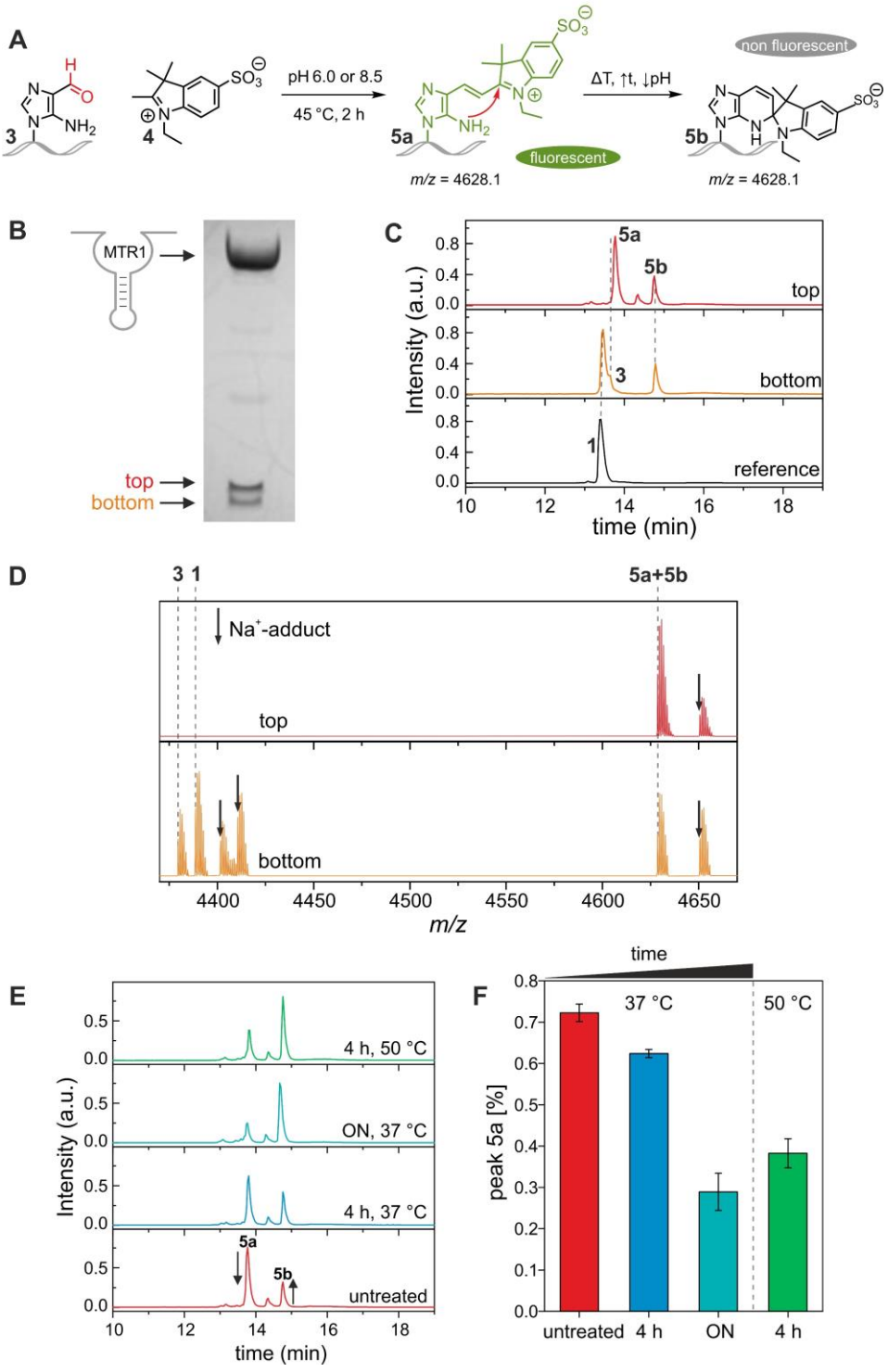



**Figure 9.22:** Conversion of chromophore **5a** to the isomeric spiro compound **5b**. **A** Proposed reaction mechanism for the generation of the isomeric non-fluorescent spiro compound **5b** that is favored under elevated temperatures (ΔT), longer reaction times (↑t,) and lower pH (↓pH). **B** Purification of MTR1 + BG treated Ne-RNA R1 reacted with **4** on denaturing PAGE. **C** Anion exchange HPLC analysis of RNA fractions isolated from **B**. **D** Deconvoluted HR-ESI-MS spectrum of RNA fractions isolated from **B**. Arrows denote sodium adducts. **E** Monitoring the conversion of **5a** to **5b** in the "top" RNA fraction from **B** using anion exchange HPLC. The RNA was incubated at pH 8 under elevated temperatures followed by analysis with anion exchange HPLC. **F** Analysis of the peak areas corresponding to **5a** and **5b** from **E**.

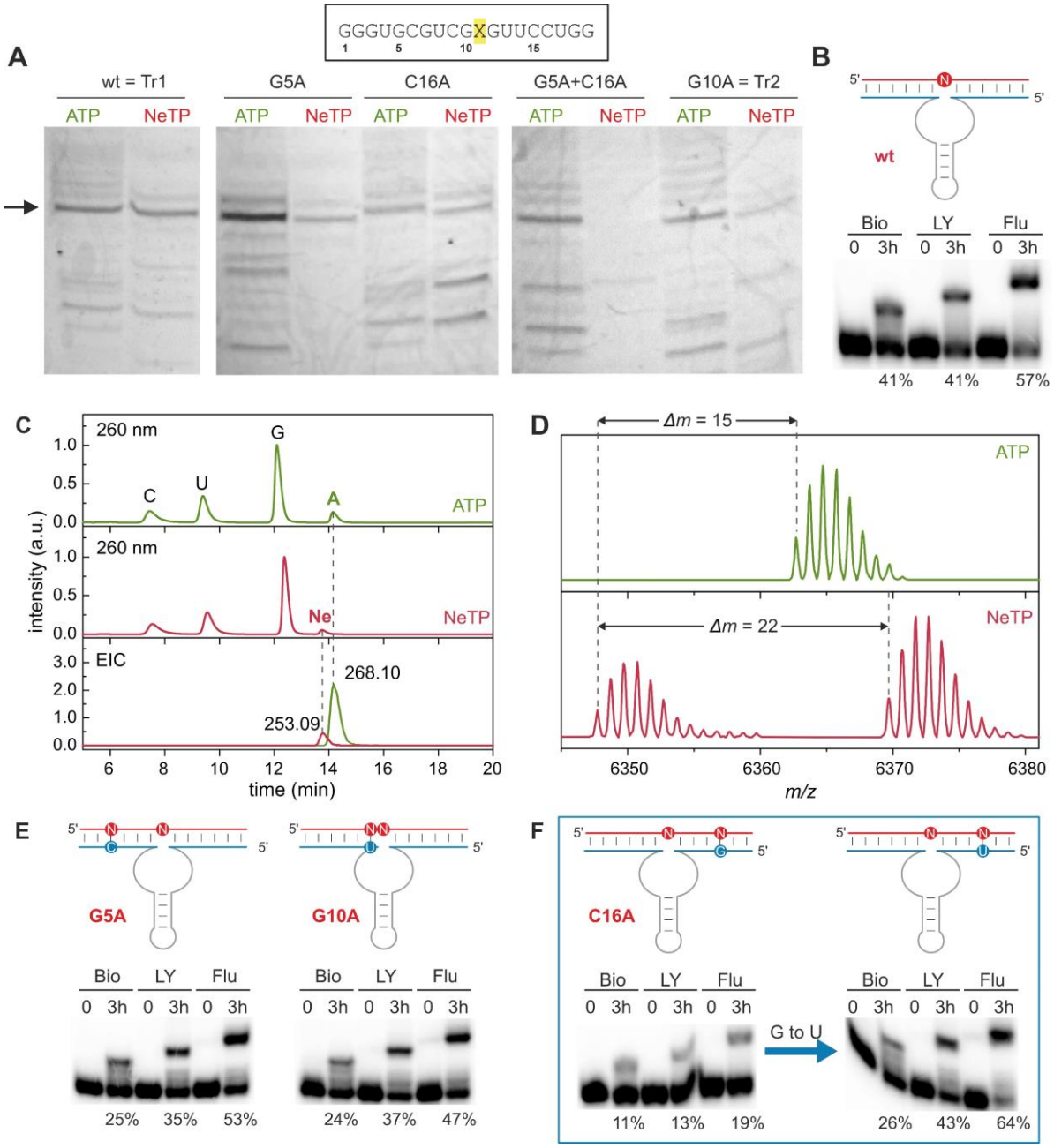



**Figure 9.23:** *In vitro* transcription using NeTP and modification of the transcripts. **A** Purification of a 19-mer substrate RNA Tr1 and its point mutants transcribed using ATP or NeTP. X denotes incorporation site of A or Ne in the wild type (wt) RNA Tr1. Arrow indicates the desired product. **B** Aldehyde functionalization of MTR1 + BG Ne-transcript Tr1 using Bio-hyd (Bio), LY-CH (LY) or Flu-TSC (Flu). **C** LC-MS analysis of digested unreacted ATP and NeTP wt transcript Tr1. UV traces at 260 nm and EICs (m/z 253.09 ± 0.01; corresponding to Ne and m/z 268.10 ± 0.01 corresponding to A) are shown. **D** Deconvoluted HR-ESI-MS spectra of wt ATP and NeTP Tr1 transcripts, containing a single adenosine or nebularine, respectively. As expected, the Ne-Tr1 is lighter by 15 mass units. Δm = 22 denotes $Na^+$ adduct. **E** Aldehyde functionalization of MTR1 + BG treated Ne-Tr1 and its point mutants with Bio-hyd, LY-CH or Flu-TSC. Circles denote Ne incorporation sites (N; red) and the respective base pairing partner in the ribozymes binding arm (blue). **F** Aldehyde functionalization of MTR1 + BG treated Ne-Tr2 (= G10A mutant) with Bio-hyd, LY-CH or Flu-TSC. The G to U point mutation in the ribozymes binding arm restored the labeling efficiency.

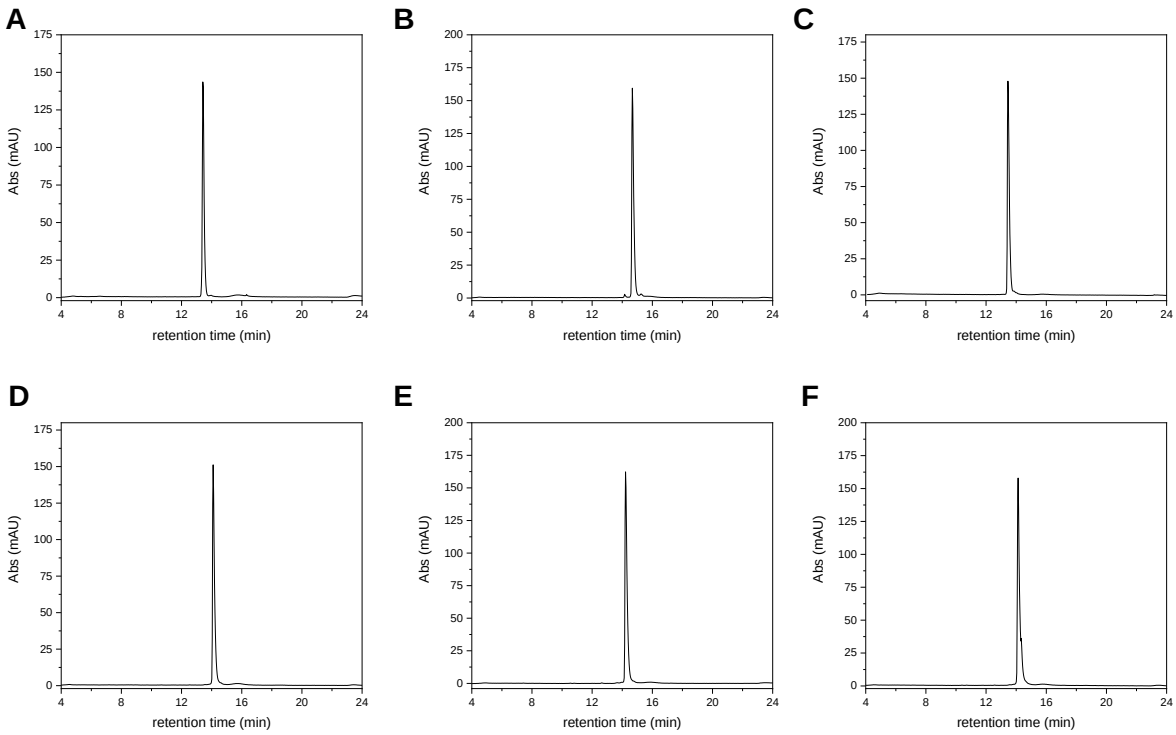



**Figure 9.24:** Anion exchange HPLC chromatograms to confirm purity of RNA oligonucleotides prepared by solid-phase synthesis. **A** R1_un, **B** R2_un, **C** R1_Ne, **D** R2_Ne, **E** R2_DAP, **F** R2_2AP. Dionex DNAPAc PA200, 2 x 250 mm, 60 °C, monitored at 260 nm.

## 9.4 Supporting information for chapter 6.1

### 9.4.1 Material and methods

#### Oligonucleotide synthesis

Unmodified DNA oligonucleotides (Table 9.14) and ribozyme templates were purchased from Microsynth and purified on denaturing PAGE (10-20 % acrylamide) prior to use.

Ribozymes (Table 9.13) were prepared by *in vitro* transcription from the corresponding synthetic DNA templates using T7 RNA polymerase, as described previously.[184] The T7 RNA polymerase was prepared as reported in literature.[398]

Short RNA oligonucleotides were synthesized in house using solid-phase synthesis with standard phosphoramidite chemistry (2′-*O*-TOM protected) on controlled pore glass solid support.[394] Deprotection of RNA oligonucleotides was performed in two steps using a 1:1 mixture of aqueous ammonia and methylamine (AMA) for 6 h at 37 °C, followed by treatment with 1 M tetrabutylamoonium fluoride (TBAF) in THF at 25 °C for 12-16 h. Deprotected oligonucleotides were first desalted using size exclusion chromatography and then purified via denaturing PAGE. The purity and identity of the RNAs was analyzed using anion exchange HPLC (Dionex DNAPac PA200, 2x250 mm, at 60 °C; solvent A: 25 mM Tris-HCl (pH 8.0), 6 M Urea; solvent B: 25 mM Tris-HCl (pH 8.0), 6 M Urea, 0.5 M $NaClO_4$. Gradient: linear, 0-40 % solvent B with a slope of 4 % solvent B per 1 CV) and HR-ESI-MS (micrOTOF-Q III, negative ion mode, direct injection). Measured and calculated masses are listed Table 9.12 together with the respective RNA sequence. The anion-exchange HPLC-chromatograms are summarized in Figure 9.29.

#### VMC10 cleavage assay

For single turnover assays of the VMC10-mediated cleavage reaction of ssRNA, 10 pmol A/m$^1$A-containing fluorescently- or $^{32}$P-labeled substrate RNA was annealed (5 min at 95°C, then 10 min at 25 °C) to 100 pmol of the DNAzyme in reaction buffer (150 mM NaCl, 50 mM Tris-HCl, pH 7.5). 20 mM $MgCl_2$ were added to a final reaction volume of 10 µL and the reaction was incubated at 37 °C. 1 µL aliquots were taken after defined timepoints and immediately quenched by the addition of 4 µL gel loading dye. Half of each timepoint sample was analyzed on PAGE (20 %, 20 cm, 25 W, 45 min) and then submitted to phosphorimaging or fluorescence imaging using blue epi illumination and a 530/28 nm emission filter.

For cleavage of the full-length *cis*-active selection pool, the same reaction conditions were used, but 25 pmol of the reverse primer D2 acting as a disruptor oligo were added in addition

to ensure efficient annealing of VMC10 to the target RNA. Analysis of the reaction war performed on 10 % PAGE (30 cm, 35 W, 90 min).

### Calculation of pool randomization

The extend of pool randomization was calculated according to ref.[392] using the following equation: $P(n) = x^{m-n} \cdot (1-x)^{n} \cdot {}_{m}C_{n}$, where $P(n)$ is the probability of a total of $n$ changes with respect to the parent sequence, $m$ is the length of the partially randomized sequence, $x$ denotes the fraction of parent nucleosides incorporated at each position and ${}_{m}C_{n}$ is the combinatorial function of $m$ objects taken $n$ at a time. In case of the CH starting pool, $m$ = 22 and $x$ = 0.76. An overview of the calculation results is provided in Figure 9.27 and Table 9.15.

### *In vitro* selection

The DNA templates for *in vitro* transcription of the RNA starting pools were assembled from two DNA oligonucleotides (fw primer D1 + rv primer D5 - D8 or D11) using Klenow polymerase overlap extension. The reaction was performed at 37 °C for 2 h using 300 pmol of each primer together with 0.2 mM of each dNTP and 5 U Klenow Polymerase exo$^-$ in 100 µL 1x Klenow buffer. The dsDNA template for the pool of selection CH was assembled from three parts in two subsequent Klenow extension steps (first D9 + D10, then addition of D1) with PC/I-extraction and EtOH-precipitation in between the Klenow primer extension reactions. Following PC/I-extraction and EtOH-precipitation of the fully assembled template, 100 pmol of the dsDNA was used per 100 µL *in vitro* transcription with T7 RNA polymerase.

For the first selection round, 400 pmol of each RNA pool were used. Each pool contained ~100 pmol 3'-fluorescently labeled RNA, obtained by sodium periodate labeling with Lucifer yellow carbohydrazide, performed according to ref.[25]. The RNA was annealed (3 minutes at 95 °C, then 10 minutes at 25 °C) in reaction buffer (120 mM KCl, 5 mM NaCl, 50 mM HEPES, pH 7.5). Then 100 µM $m^6G$ as well as 40 mM $MgCl_2$ were added to a total reaction volume of 20 µL (CD-CG) or 10 µL (CH-CI). After overnight incubation at 37 °C, the RNA was EtOH-precipitated and submitted to the VMC10 reaction.

For the VMC10 cleavage reaction, 600 pmol of the disruptor DNA (D2) and 1.2 nmol VMC10 (D12) were annealed (3 minutes at 95 °C, then 10 minutes at 25 °C) in reaction buffer (150 mM NaCl, 50 mM Tris-HCl, pH 7.5). 20 mM $MgCl_2$ were added to a final reaction volume of 10 µL and the reaction was incubated overnight at 37 °C. Following, the pools were resolved on

PAGE (10 %, 0.4 mm, 40 cm, 45 W, 1:30 h) next to fluorescently labeled size markers indicating the full-length RNA pools.

The uncleaved fractions were isolated from the gel and reverse transcribed in a total reaction volume of 50 µL using 0.5 µM forward primer D3 and 1 µM reverse primer D2 as well as one Ready-To-Go RT-PCR Bead. RT-PCR amplification was performed under the following cycling conditions:

| 42 °C | 95 °C | 95 °C | 65 °C | 72 °C | 72 °C | 4 °C |
|---|---|---|---|---|---|---|
| 30 min | 1:00 min | 0:30 min | 0:30 min | 0:25 min | 5:00 min | hold |
| | | x6 | | | | |

The cycle number was gradually decreased down to three cycles in the last round according to the selection progress.

Amplification of the DNA thus generated was performed in a 200 µL scale using 10 µL RT-PCR product as template. The DNA was mixed with 0.5 µM forward primer D1 and 1.0 µM reverse primer D2, 0.2 mM dNTP-mix, 20 µL DMSO, 20 µL 10x Dream Taq buffer and 1 U Dream Taq polymerase. Amplification was performed under the following conditions:

| 95 °C | 95 °C | 63 °C | 72 °C | 72 °C | 4 °C |
|---|---|---|---|---|---|
| 4:00 min | 0:30 min | 0:30 min | 0:25 min | 5:00 min | hold |
| | x32 | | | | |

The cycle number was gradually decreased according to the selection progress to 26 in the last selection rounds. The PCR product was then EtOH-precipitated and *in vitro* transcribed using T7 RNA polymerase, which yielded the enriched selection pool that was send through the next round of selection.

After six rounds (CH selection) or eight rounds (CD and CH selection), the PCR product obtained from the last selection round cloned (TOPO-TA cloning) and randomly picked clones were submitted for Sanger sequencing.

**Activity assay in *cis* using RT primer extension**

The dsDNA templates for transcription of the full-length *cis*-active ribozymes were obtained by PCR using the respective plasmids as templates. The reaction was performed in a 200 µL scale under the same conditions as the second PCR during the selection using primer D1 and D2 with 30 amplification cycles. The dsDNA thus obtained was EtOH-precipitated and submitted to the *in vitro* transcription reaction using T7 RNA polymerase.

For the self-methylation reaction, 10 pmol of each ribozyme was annealed (3 minutes at 95 °C, then 10 minutes at 25 °C) in reaction buffer (120 mM KCl, 5 mM NaCl, 50 mM MES, pH 7.5). 100 µM $m^6G$ and 40 mM $MgCl_2$ were added, and the reaction was incubated overnight at 37 °C followed by EtOH-precipitation. Readout via RT primer extension was performed as described in chapter 9.1.1 for tRNA methylation using 5'-$^{32}$P-labeled primer P1. The reactions were resolved on 10 % PAGE (40 cm, 0,4 mm, 45 W, 2:15 h) and visualized by autoradiography.

**3'-end labeling of RNAs using aminolysis of NHS-ester**

RNA substrates R1 and R2 were 3'-end labeled with fluorescein using NHS-ester chemistry. Up to 10 nmol of RNA modified with an amino functionality at the 3'-end were mixed with 7.5 mM NHS-fluorescein (Thermo Scientific; in DMF) in 25 µL 75 mM carbonate buffer (pH 9.0). The reaction was incubated at 37 °C in the dark followed by EtOH-precipitation. The RNA was then dissolved in equal volumes of $H_2O$ and gel loading buffer followed by purification on denaturing PAGE (20 %, 0.7 mm, 30 cm, 35W, 2:15 h).

**5'-$^{32}$P-labeling of DNA**

The primer P1 that was used for RT primer extension experiments was 5'-radiolabeled using PNK and [γ-$^{32}$P] ATP. The labeling reaction was performed as described in chapter 9.3.1.

**Activity assay in *trans***

Single turnover kinetic assays to test for *trans*-activity were performed as described in chapter 9.1.1 at pH 7.5 with incubation at 37 °C using ribozymes Rz1-Rz6 together with 3'-fluorescein labeled substrate RNA R1. The reactions were resolved on 20 % PAGE (30 cm, 0.4 mm, 35 W, 2:15 h) followed by fluorescence imaging using blue epi illumination and a 530/28 nm emission filter. The data was fit with the pseudo-first order kinetic equation $Y = Y_{max}(1 - e^{-kobs*t})$ to obtain the observed reaction rate $k_{obs}$ and the final yield $Y_{max}$ using Origin (2021).

## 9.4.2 Supplementary tables

**Table 9.12:** RNA oligonucleotides prepared by solid phase synthesis together with calculated and measured masses using HR-ESI-MS. Red marks denote the targeted modification site. Anion exchange HPLC traces to confirm purity of the RNAs are summarized in Figure 9.29.

| No. | description | 5'-sequence-3' | Mass (calculated) | Mass (found) |
|---|---|---|---|---|
| R1 | Sub. Seq. (short) | AUACUGAGCCUUC-C6-$NH_2$ | 4059.56 Da | 4259.62 Da |
| R1_m1A | Sub. Seq. (short)_$m^1$A | AUACUG($m^1$A)GCCUUC | 4073.58 Da | 4073.63 Da |
| R2 | Sub. Seq. (long) | ACAUACUGAGCCUUCAA-C6-$NH_2$ | 5530.83 Da | 5530.86 Da |

**Table 9.13:** RNA oligonucleotides prepared by *in vitro* transcription. Binding arms are underlined.

| No. | description | 5'-sequence-3' |
|---|---|---|
| Rz1 | CA13 trans (short) | GGAGGCTGACCGACCCCCCGACCCTTCTCTGGGACAACTAGACATACAGT |
| Rz2 | CA13 trans (long) | GGTTGAAGGCTGACCGACCCCCCGACCCTTCTCTGGGACAACTAGACATACAGTATGTCC |
| Rz3 | CA21 trans (short) | GGAGGCTGACCGACCCGCTACAATACAACAGCGACAACTAGAAATACAGT |
| Rz4 | MTR1 trans (long) | GGTTGAAGGCTGACCGACCCCCCGAGTTCGCTCGGGGACAACTAGACATACAGTATGTCC |
| Rz5 | CD2 trans | GGAGGCTGACCGACCCGCCACCAATATAAAAAACGCGACAACTAGAAATACAGT |
| Rz6 | CD5 trans | GGAGGCTGACCGACCCGGCACTACAGCTCGTTCCGACAACTAGACATACAGT |
| Rz7 | CD8 trans | GGAGGCTGACCGACCCGCCCGCGAGACAAAGCGCGACAACTAGACATACAGT |
| Rz8 | CD40 trans | GGAGGCTGACCGACCCGGACCAAACTCACAGACCGACAACTAGAAATACAGT |
| Rz9 | 6CH18 trans | GGTTGAAGGCTGACCGATCCGCAGCCAAGCGACAACTAGACATACAGTATGTCC |
| Rz10 | 8CH36 trans | GGTTGAAGGCTTAGCTGTCCGCAACCAAGCGACAACTGGGTCTCCAGTATGTCC |

**Table 9.14:** DNA oligonucleotides prepared by solid phase synthesis. Binding arms and primer binding sites are underlined.

| No. | description | 5'-sequence-3' |
|---|---|---|
| D1 | fw primer | CTGTAATACGACTCACTATAGGACATACTGAGCCTTCAACCAGCCTACCATCC |
| D2 | P1; rv primer 1st & 2nd PCR; disruptor oligo | GGTAAGGTGGACATACTG |
| D3 | fw primer 1st PCR | CTTCAACCAGCCTACCATCC |
| D4 | cloning fw primer | TAAATAAAATAACTGTAATACGACTCACTATAGGACATACTGAGC |
| D5 | rv primer CD pool | GTAAGGTGGACATACTGTATTTCTAGTTGTCG-$N_{16}$-CGGGTCGGTCAGCCTTCAAGGATGGTAGGCTG |
| D6 | rv primer CE pool | GTAAGGTGGACATACTGTATTTCTAGTTGT-$N_{10}$-GGTCGGTCAGCCTTCAAGGATGGTAGGCTG |
| D7 | rv primer CF pool | GTAAGGTGGACATACTGTATTTCTAGTTGT-$N_{20}$-GGTCGGTCAGCCTTCAAGGATGGTAGGCTG |
| D8 | rv primer CG pool | GTAAGGTGGACATACTGTATTTCT-$N_8$-AGTTGTCGCTGGTTAGCGGGTCGGTCAGCCTTCAAGGATGGTAGGCTG |
| D9 | rv primer 1 CH pool | GTAAGGTGGACATACTGTATGTCTAGTTGTCGCTTGGTTGCG |
| D10 | rv primer 2 CH pool | CCAGCCTACCATCCTTGAAGGCTGACCGACCCGCAACCAAGCG |
| D11 | rv primer CI pool | GTAAGGTGGACATACTG-$N_{13}$-CGCTTGGTTGCG-$N_9$-GCCTTCAAGGATGGTAGGCTGG |
| D12 | VMC10 for cutting the substrate RNA | CTGGTTGAAGGCGGGTCTCCAGCTGGACGTTAGTATGTCC |

**Table 9.15:** Calculated probability distribution for the partially randomized starting pool of selection CH. *n* = number of mutated nucleosides relative to the parent sequence; P(*n*) = probability of *n* changes in the starting pool. The calculations were performed according to ref.[392] and as described in the material and methods section.

| ***n*** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **P(*n*)** | 0.002 | 0.017 | 0.055 | 0.116 | 0.174 | 0.197 | 0.177 | 0.128 | 0.075 | 0.037 | 0.015 |

| 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.005 | 0.002 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |

**Table 9.16:** Enrichment for each selection round of selections CD-CG. Fraction uncleaved denotes sequences resistant to VMC10 cleavage for standard selection rounds. In negative selections (NS), the uncleaved fraction was discarded and the cleaved fraction was isolated and amplified.

| round | fraction uncleaved | round | fraction uncleaved | round | fraction uncleaved | round | fraction uncleaved |
|---|---|---|---|---|---|---|---|
| 1CD | 0.127 | 1CE | 0.075 | 1CF | 0.058 | 1CG | 0.127 |
| 2CD | 0.200 | 2CE | 0.148 | 2CF | 0.087 | 2CG | 0.163 |
| 3CD | 0.306 | 3CE | 0.281 | 3CF | 0.298 | 3CG | 0.334 |
| 4CD | 0.488 | 4CE | 0.459 | 4CF | 0.405 | 4CG | 0.471 |
| 4CD (NS) | 0.554 | 4CE (NS) | 0.410 | 4CF (NS) | 0.264 | 4CG (NS) | 0.483 |
| 5CD | 0.296 | 5CE | 0.079 | 5CF | 0.078 | 5CG | 0.208 |
| 6CD | 0.274 | 6CE | 0.252 | 6CF | 0.178 | 6CG | 0.115 |
| 7CD | 0.360 | 7CE | 0.485 | 7CF | 0.551 | 7CG | 0.297 |
| 8CD | 0.443 | 8CE | 0.620 | 8CF | 0.571 | 8CG | 0.444 |

**Table 9.17:** Enrichment and incubation times for each selection round of CH (left) and CI (right) selections. ON = overnight (i.e., 15-18 h). Fraction uncleaved represents sequences resistant to VMC10 cleavage for standard selection rounds. In negative selections (NS), the uncleaved fraction was discarded and the cleaved fraction was isolated and amplified.

| round | fraction uncleaved | incubation time |
|---|---|---|
| 1CH | 0.017 | ON |
| 1CH (NS) | 0.121 | |
| 2CH | 0.073 | |
| 2CH (NS) | 0.018 | |
| 3CH | 0.068 | |
| 3CH (NS) | 0.108 | |
| 4CH | 0.011 | |
| 5CH | 0.183 | |
| 6CH | 0.299 | |
| 7CH | 0.527 | 6 h |
| 7CH (NS) | 0.535 | |
| 8CH | 0.259 | |
| 9CH | 0.431 | 4 h |
| 9CH (NS) | 0.579 | |
| 10CH | 0.520 | |
| 10CH (NS) | 0.627 | |

| round | fraction uncleaved | incubation time |
|---|---|---|
| 1CI | 0.013 | ON |
| 1CI (NS) | 0.091 | |
| 2CI | 0.009 | |
| 2CI (NS) | 0.099 | |
| 3CI | 0.049 | |
| 3CI (NS) | 0.014 | |
| 4CI | 0.059 | |
| 5CI | 0.020 | |
| 6CI | 0.430 | |
| 6CI (NS) | 0.483 | |
| 7CI | 0.301 | |
| 7CI (NS) | 0.521 | |
| 8CI | 0.421 | 6 h |
| 8CI (NS) | 0.470 | |

## 9.4.3 Extended data figures

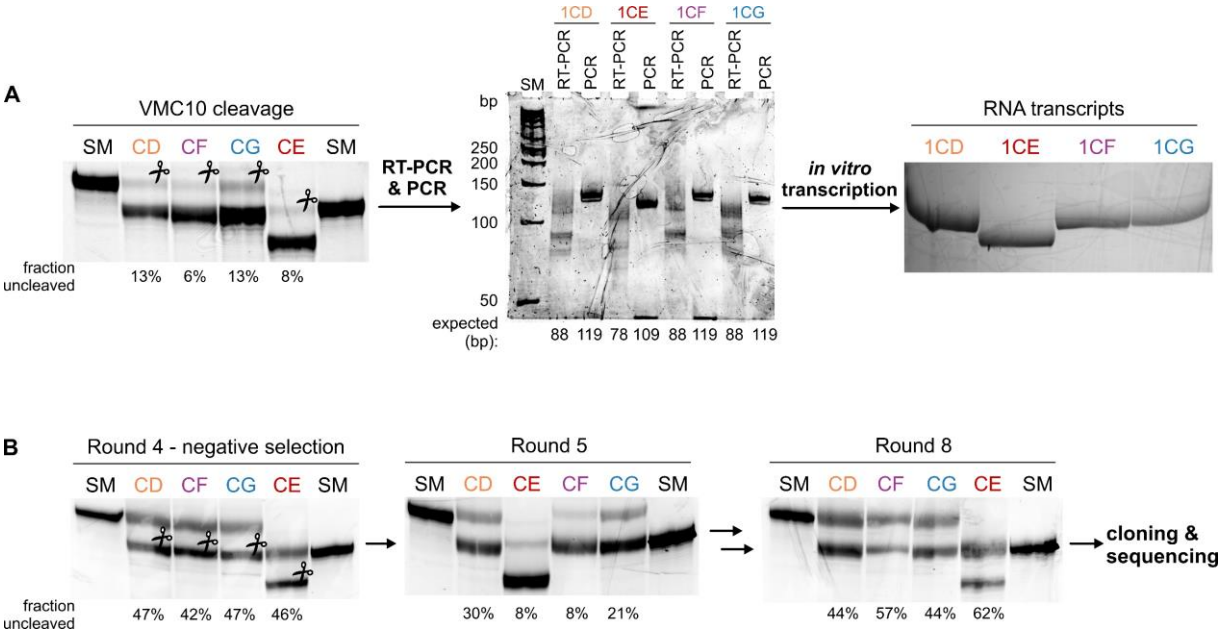



**Figure 9.25:** Representative gel images from methyltransferase reselections CD-CG. **A** Gel images from the first selection round. The uncleaved fractions resulting from the VMC10 reaction were extracted and amplified via RT-PCR and PCR. The DNA thus obtained was then *in vitro* transcribed yielding the input RNA for the next selection round. **B** In the negative selection that was performed after round four, the cleaved RNA was isolated from the gel and amplified. This results in a sudden drop in the fraction of uncleaved RNA. After eight selection rounds, the cDNA was submitted to cloning and sequencing.

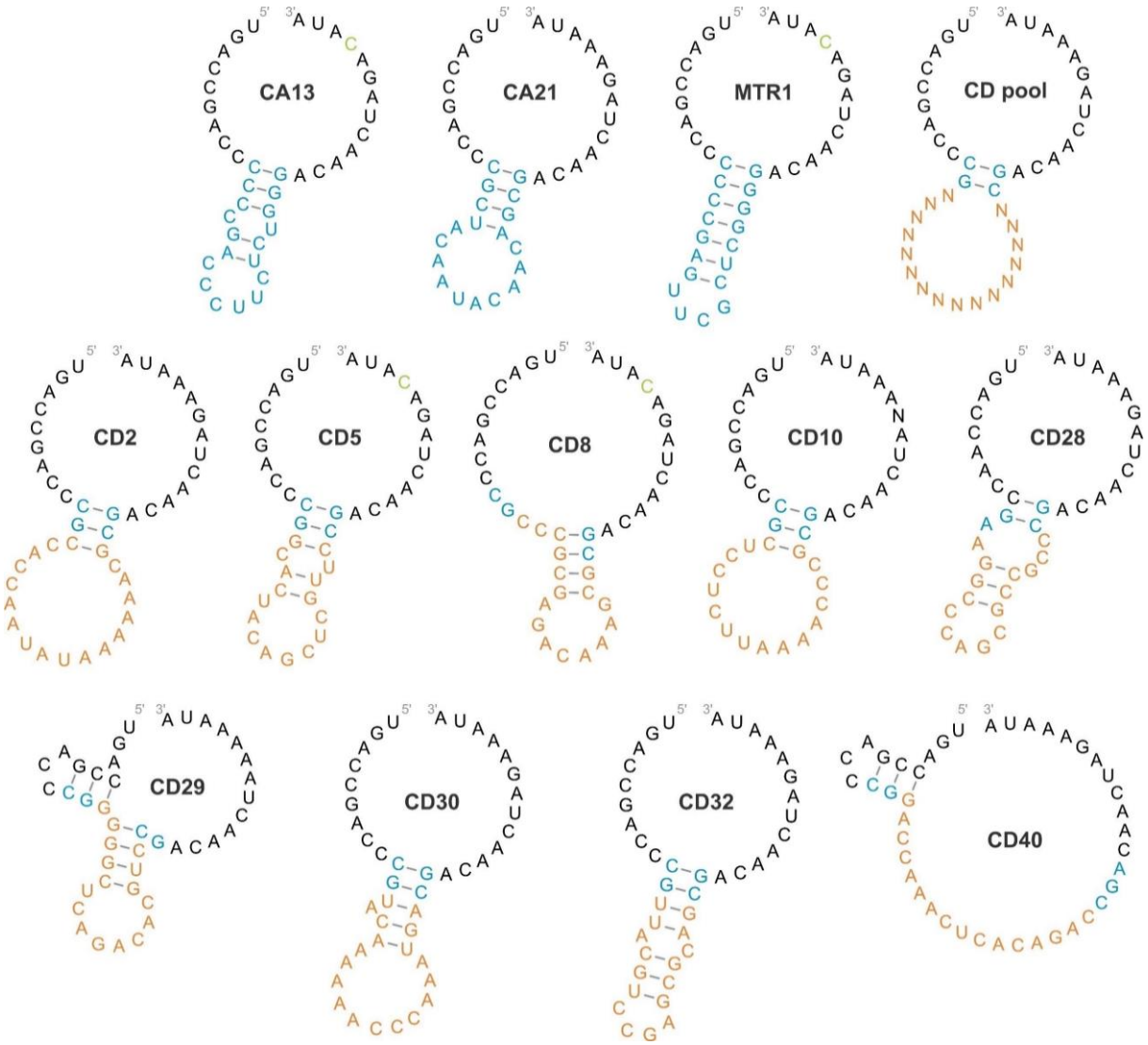



**Figure 9.26:** Comparison of Sequences and secondary structures of CA13, CA21, MTR1 and the CD ribozymes obtained from Sanger sequencing together with the starting pool design. The stem-loop structure present in CA13, CA21, MTR1 and the starting pool is marked blue. The randomized region is shown in orange. The A-to-C mutation in the catalytic core of CA13, MTR1, CD5 and CD8 is depicted in light green. The secondary structures were predicted using the RNAfold web server.[391]

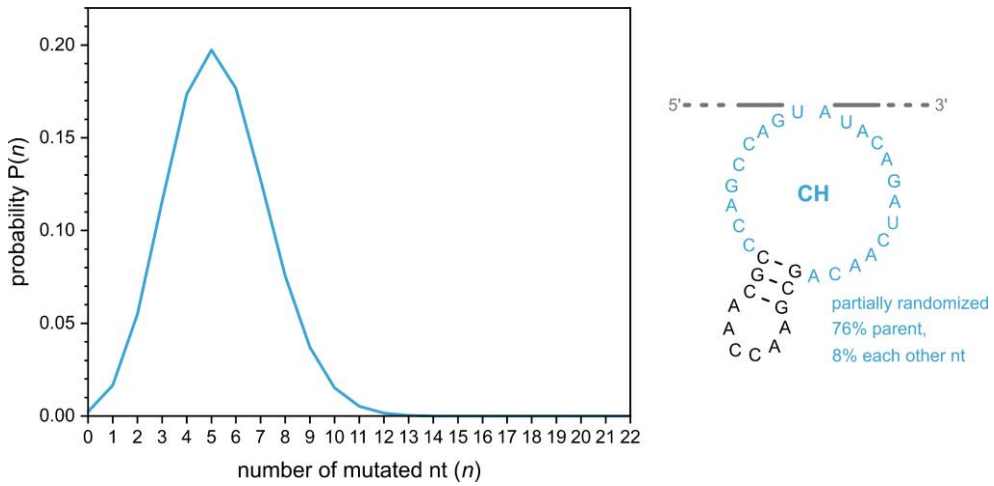



**Figure 9.27:** Calculated probability distribution of nucleotide mutations *n* for the CH starting pool containing 22 partially randomized nucleotides (76 % wt sequence, 8 % each other nucleotide). Calculated according to ref.[392] and as described in the material and methods section. The exact probabilities are summarized in Table 9.15.

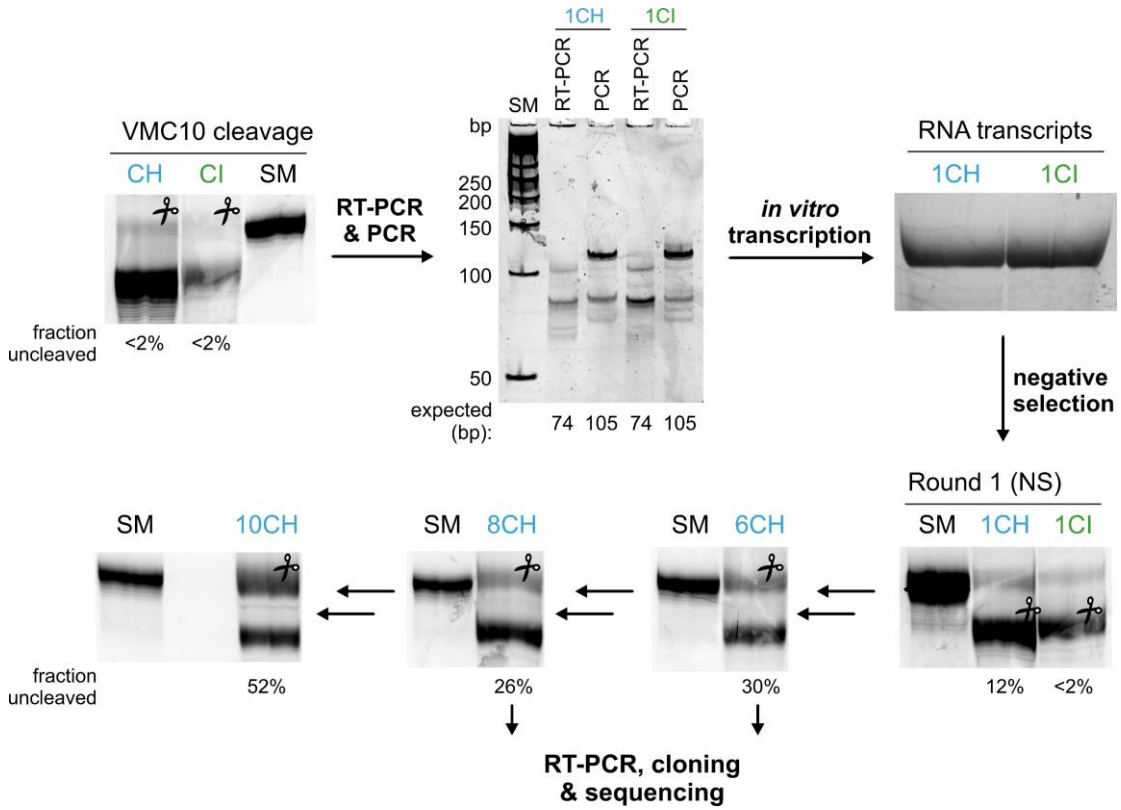



**Figure 9.28:** Representative gel images from methyltransferase reselections CH and CI. Shown are gel images of the VMC10 cleavage reaction, analysis of the RT-PCR and PCR products as well as for the purification of the *in vitro* transcripts and the following VMC10 cleavage in the negative selection round. Selection CH was performed for a total of ten rounds. The cDNA obtained after the sixth and eighth round was submitted to cloning and sequencing.

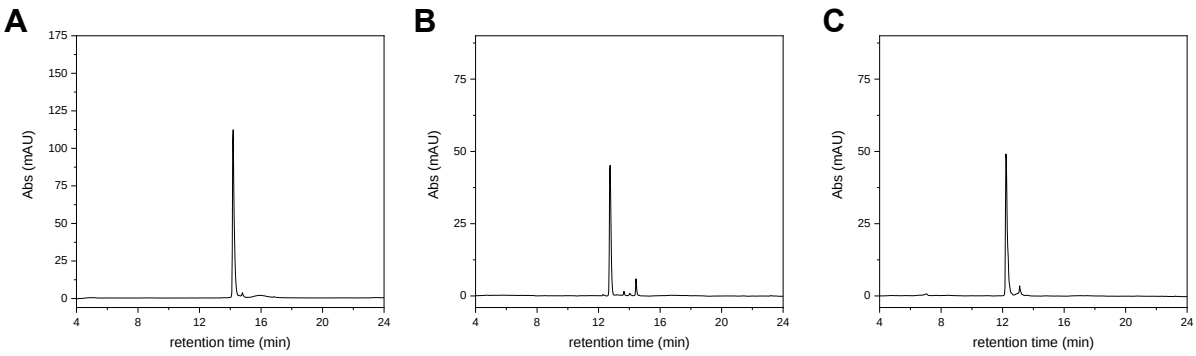



**Figure 9.29:** Anion exchange HPLC chromatograms to confirm purity of RNA oligonucleotides prepared by solid-phase synthesis. **A** R1, **B** R1_m$^1$A, **C** R2. Dionex DNAPAc PA200, 2x250 mm, 60°C, monitored at 260 nm.

## 9.5 Supporting information for chapter 6.2

### 9.5.1 Material and methods

#### Oligonucleotide synthesis

Unmodified DNA oligonucleotides (Table 9.20) and ribozyme templates were purchased from Microsynth and purified on denaturing PAGE (10-20 % acrylamide) prior to use. Ribozymes (Table 9.19) were prepared by *in vitro* transcription from the corresponding synthetic DNA templates using T7 RNA polymerase, as described previously.[184] The T7 RNA polymerase was prepared in house as reported in ref.[398].

Short RNA oligonucleotides (Table 9.18) were synthesized in house using solid-phase synthesis with standard phosphoramidite chemistry (2′-*O*-TOM protected) on controlled pore glass solid support as described in chapter 9.4.1.[394] The anion-exchange HPLC-chromatograms are summarized in Figure 9.34.

#### *In vitro* selection

The CJ, CK and CL *in vitro* selections were performed analogously to the CA selection as described in chapter 9.1.1 and ref.[184]. The dsDNA templates for *in vitro* transcription of the RNA starting pools were assembled using fw primer D1 (CJ), D2 (CK) or D3 (CL) together with rv primer D4 by Klenow polymerase overlap extension. In a total reaction volume of 100 µL 1x Klenow buffer, 500 pmol of each primer were mixed with 0.2 mM of each dNTP and 5 U Klenow polymerase exo$^-$. The reaction was incubated for 2 h at 37 °C followed by P/CI extraction and EtOH-precipitation. The dsDNA template thus obtained was submitted to *in vitro* transcription with T7 RNA polymerase using 100 pmol assembled template per 100 µL transcription reaction.

In the first selectin round, 3.15 nmol of each RNA pool were submitted to the self-modification reaction, each containing ~200 pmol 3'-Lucifer yellow labeled RNA, which was obtained by sodium periodate glycol oxidation, performed according to ref.[25]. After annealing of the RNA pool (3 minutes at 95 °C, 10 minutes at 25 °C) in reaction buffer (120 mM KCl, 5 mM NaCl, 50 mM HEPES, pH 7.5), 100 µM BG-biotin (New England Biolabs) as well as 40 mM $MgCl_2$ were added to a total volume of 60 µL. The reactions were incubated ON (i.e., 16 h) at 37 °C, followed by EtOH-precipitation of the RNA. The incubation time, cofactor and $MgCl_2$ concentrations as well as the amount of input RNA were reduced in subsequent rounds according to the enrichment level to increase the selection pressure. An overview of the selection parameters is provided in Table 9.21 and Table 9.22. The capturing of active species was performed with

streptavidin- or neutravidin-coated magnetic beads (Dynabeads, Thermo Fisher Scientific, ca. 1 nmol RNA/mg beads), followed by elution with formamide. The RNA was recovered by EtOH-precipitation and then amplified by RT-PCR, PCR and *in vitro* transcription as described in the literature.[184, 285] After 14 (CJ), 15 (CK) or eleven rounds (CL) of selection, the library was cloned using TOPO-TA cloning and randomly picked colonies were sent for Sanger sequencing.

### *In vitro* transcription of ribozymes using plasmids as templates

The dsDNA templates for *in vitro* transcription of the full-length *cis*-active ribozymes were obtained by PCR using approx. 150-200 ng of the respective plasmid together with 100 pmol of the rv primer D4 and 100 pmol of the appropriate fw primer D1-D3 to obtain the templates for the *cis*-active ribozymes. To remove the substrate RNA and generate *trans*-active ribozymes, fw primer D7 was used together with rv primer D4. The PCR was carried out under the same conditions as the second PCR during *in vitro* selection with an annealing temperature of 65 °C and 30 amplification cycles. The PCR product was EtOH-precipitated followed by *in vitro* transcription with T7 RNA polymerase and purification of the RNA on denaturing PAGE.

### Streptavidin gel shift assay

To test the enriched libraries or single ribozymes for activity in *cis*, streptavidin gel shift assays were performed. In a total reaction volume of 5 µL reaction buffer (120 mM KCl, 5 mM NaCl, 50 mM HEPES, pH 7.5), 75 pmol of the RNA pool/individual ribozyme were reacted with 100 µM BG-biotin in the presence of 40 mM $MgCl_2$. An annealing step (3 minutes at 95 °C, then 10 minutes at 25 °C) was performed before addition of the cofactor and $MgCl_2$. The reaction mixture was incubated at 37 °C for an appropriate amount of time and subsequently precipitated with EtOH. The RNA was dissolved in 25 µL $H_2O$ and 1 µL was gently mixed with 1 µg streptavidin in 1x TBS buffer (5 µL total reaction volume). After 5 min incubation at 25 °C, 1 µL 6x loading dye was added and the reaction was analyzed on 10 % native PAGE (200 V, 50 min). Following staining with SYBR gold, the RNA was imaged using UV transillumination.

### RT primer extension for determination of modification site

RT primer extension reactions were performed as described in chapters 9.1.1 and 9.4.1. In short, 10 pmol *cis*-active ribozyme were submitted to a standard self-modification reaction containing 100 µM small molecule substrate and 40 mM $MgCl_2$ that was incubated at 37 °C for an appropriate amount of time. After EtOH-precipitation, the samples were submitted to the RT

primer extension reaction using 5'-$^{32}$P-labeled primer P1. The reactions were resolved on 15 % denaturing PAGE (40 cm, 0.4 mm, 45 W, 2:15 h) and visualized by autoradiography.

### 3'-end labeling of RNAs using aminolysis of NHS-ester

RNA substrate R5 was 3'-fluorescein labeled using NHS-ester chemistry as described in chapter 9.4.1.

### 5'-end labeling of RNAs using copper(I)-catalyzed azide-alkyne cycloaddition

RNA substrates R1 - R4 were 5'-fluorescein labeled using Click-chemistry. Up to 10 nmol RNA carrying a 5'-terminal alkyne functionality were mixed with 3 µL of a freshly prepared DMSO:*t*BuOH (3:1) solution and 0.5 µL of a 50 mM 6-FAM-PEG$_3$-azide (Thermo Scientific) solution in DMSO:*t*BuOH (3:1). Then, 0.5 µL of freshly prepared CuBr solution (100 mM in DMSO:*t*BuOH, 3:1) was premixed with 1 µL of a freshly prepared TBTA solution (100 mM in DMSO:*t*BuOH, 3:1) and added to the reaction mixture to a final reaction volume of 10 µL. The reaction mixture was incubated at 37 °C for 2-3 h followed by EtOH-precipitation of the labeled oligonucleotide. The RNA was dissolved in loading buffer and purified on denaturing PAGE.

### 5'-$^{32}$P-labeling of DNA

The primer P1 that was used for RT primer extension experiments was 5'-radiolabeled using PNK and [γ-$^{32}$P] ATP. The labeling reaction was performed as described in chapter 9.3.1.

### Activity assay in *trans*

To test the ribozymes for activity in *trans*, single turnover kinetic assays were performed as described in chapter 9.1.1 at pH 7.5 and 37 °C using ribozymes Rz1-Rz6 together with 5'- or 3'-fluorescently labeled substrate RNAs R1-R5. The timepoint samples were resolved on 20 % denaturing PAGE (BG-biotin: 0.4 mm, 20 cm, 25W, 1 h; BG-NH$_2$/BG: 0.4 mm, 30 cm, 35W, 2:15 h; m$^6$G: 0.4 mm 40 cm, 45 W, 3:45 h) followed by autoradiography or fluorescence imaging using blue epi illumination and a 530/28 nm emission filter. For slow reactions, the data was fit with the pseudo-first order kinetic equation $Y = Y_{max}(1 - e^{-kobs*t})$. For fast reactions, a biexponential model with $Y = Y_{max}(a\ (1 - e^{-k1t}) + (1 - a)\ (1 - e^{-k2t}))$ was used, to obtain the observed reaction rates $k_1$ and $k_2$/$k_{obs}$ and the final yield $Y_{max}$ using Origin (2021).

## 9.5.2 Supplementary tables

**Table 9.18:** RNA oligonucleotides prepared by solid phase synthesis together with calculated and measured masses using HR-ESI-MS. Red marks denote the targeted modification site. Alk = alkyne. Anion exchange HPLC traces to confirm purity of the RNAs are summarized in Figure 9.34.

| No. | description | 5'-sequence-3' | Mass calculated | Mass found |
|---|---|---|---|---|
| R1 | GAG RNA substrate (21 nt) | Alk-GACAUACUGAGCCUUCAAAUA | 6821.09 Da | 6821.03 Da |
| R2 | GGG RNA substrate (21 nt) | Alk-GACAUACUGGGCCUUCAAAUA | 6836.96 Da | 6836.94 Da |
| R3 | GCG RNA substrate (21 nt) | Alk-GACAUACUGCGCCUUCAAAUA | 6796.96 Da | 6797.00 Da |
| R4 | GUG RNA substrate (21 nt) | Alk-GACAUACUGUGCCUUCAAAUA | 9797.94 Da | 6797.93 Da |
| R5 | GAG RNA substrate (17 nt) | ACAUACUGAGCCUUCAA-C6-$NH_2$ | 5530.83 Da | 5530.86 Da |

**Table 9.19:** RNA oligonucleotides prepared by *in vitro* transcription. Binding arms are underlined.

| No. | description | 5'-sequence-3' |
|---|---|---|
| Rz1 | CK03 trans | GGUUGAAGGCCCACUGGACGAAGACAUCCCGCUGCUGCGCCGUAGUUGUGCAGUAUGUCC |
| Rz2 | CL06 trans | GGUUGAAGGCUGACCGACCCGCAACCAAGCGACAACUAGACAUACAGUAUGUCC |
| Rz3 | MTR1 for UAC | GGAAGCCCAGUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUAAUGUCC |
| Rz4 | CK03 C14G | GGUUGAAGGCCCAGTGGACGAAGACAUCCCGCUGCUGCGCCGUAGUUGUGCAGUAUGUCC |
| Rz5 | CA13 trans | GGUUGAAGGCUGACCGACCCCCCGACCCUUCUCUGGGACAACUAGACAUACAGUAUGUCC |
| Rz6 | MTR1 trans | GGUUGAAGGCUGACCGACCCCCCGAGUUCGCUCGGGGACAACUAGACAUACAGUAUGUCC |

**Table 9.20:** DNA oligonucleotides prepared by solid phase synthesis.

| No. | description | 5'-sequence-3' |
|---|---|---|
| D1 | fw primer CJ (GCG) | CTGTAATACGACTCACTATAGGACATACTGCGCCTTCAACCAGCCTACCATCC |
| D2 | fw primer CK (GGG) | CTGTAATACGACTCACTATAGGACATACTGGGCCTTCAACCAGCCTACCATCC |
| D3 | fw primer CL selection | CTGTAATACGACTCACTATAGGACATACTGAGCCTTCAACCAGCCTACCATCC |
| D4 | rv primer 1st & 2nd PCR | GGTAAGGTGGACATACTG |
| D5 | fw primer 1st PCR | CTTCAACCAGCCTACCATCC |
| D6 | cloning fw primer | TAAATAAAATAACTGTAATACGACTCACTATAGGACATACTGAGC |
| D7 | Primer to obtain *trans* ribozymes | CTGTAATACGACTCACTATAGGCCAGCCTACCATCC |
| D8 | Primer (P1) for RT readout | CAAGGATGGTAGGCTGGT |

**Table 9.21:** Enrichment and incubation conditions for each selection round of CJ (left) and CK (right) selections. ON = overnight (i.e., 15-18 h); (-) denotes below detection limit.

| round | activity | incubation time | $[MgCl_2]$ |
|---|---|---|---|
| 1CJ | - | ON | 40 mM |
| 2CJ | - | | |
| 3CJ | - | | |
| 4CJ | - | | |
| 5CJ | - | | |
| 6CJ | - | | |
| 7CJ | - | | |
| 8CJ | 0.002 | | |
| 9CJ | 0.046 | | |
| 10CJ | 0.015 | | |
| 11CJ | 0.061 | 6 h | |
| 12CJ | 0.009 | 3 h | |
| 13CJ | 0.031 | 1.5 h | |
| 14CJ | 0.035 | 1 h | 20 mM |

| round | activity | incubation time | $[MgCl_2]$ |
|---|---|---|---|
| 1CK | - | ON | 40 mM |
| 2CK | - | | |
| 3CK | - | | |
| 4CK | - | | |
| 5CK | - | | |
| 6CK | - | | |
| 7CK | - | | |
| 8CK | 0.030 | | |
| 9CK | 0.008 | | |
| 10CK | 0.029 | | |
| 11CK | 0.045 | | |
| 12CK | 0.036 | 6 h | |
| 13CK | 0.056 | 4 h | |
| 14CK | 0.040 | 2 h | |
| 15CK | 0.031 | 1.5 h | 20 mM |

**Table 9.22:** Enrichment and incubation conditions for each selection round of selection CL. ON = overnight (i.e., 15-18 h); (-) denotes below detection limit.

| round | activity | $[MgCl_2]$ | [SNAP-biotin] | incubation time |
|---|---|---|---|---|
| 1CL | - | 40 mM | 100 µM | ON |
| 2CL | - | 30 mM | | |
| 3CL | 0.018 | 20 mM | 75 µM | |
| 4CL | 0.004 | 10 mM | 50 µM | |
| 5CL | 0.049 | 7.5 mM | 40 µM | |
| 6CL | 0.036 | 5 mM | 30 µM | |
| 7CL | 0.053 | 4 mM | 25 µM | |
| 8CL | 0.001 | | | 4 h |
| 9CL | 0.017 | | | |
| 10CL | 0.007 | | | 2 h |
| 11CL | 0.004 | | | |

**Table 9.23:** Kinetic parameters of CK03 measured with different substrates, BG-$NH_2$ concentrations or at varying pH-values as well as the comparison to CA13. For kinetics fitted using pseudo-first order kinetics, $k_{obs}$ values are listed. For biexponential fitting, rate constants $k_1$ and $k_2$ are shown. The respective graphs are shown in Figure 6.13 and Figure 9.33.

| substrate | $k_1$ ($min^{-1}$) | $k_2/k_{obs}$ ($min^{-1}$) |
|---|---|---|
| BG-biotin | 1.86E-3 | 1.54E-2 |
| BG-$NH_2$ | 8.68E-4 | 7.80E-3 |
| BG | 2.07E-3 | 1.68E-3 |

| [BG-$NH_2$] | $k_1$ ($min^{-1}$) | $k_2/k_{obs}$ ($min^{-1}$) |
|---|---|---|
| 25 µM | - | 1.77E-3 |
| 50 µM | - | 1.63E-3 |
| 100 µM | 8.68E-4 | 7.80E-3 |
| 200 µM | 1.92E-3 | 1.83E-2 |
| 400 µM | 2.16E-3 | 1.84E-2 |

| pH | $k_1$ ($min^{-1}$) | $k_2/k_{obs}$ ($min^{-1}$) |
|---|---|---|
| 6.0 | 2.49E-3 | 7.01E-2 |
| 6.5 | 1.30E-3 | 5.12E-2 |
| 7.0 | 1.28E-3 | 2.34E-2 |
| 7.5 | 8.68E-4 | 7.80E-3 |
| 8.0 | - | 1.88E-3 |
| 8.5 | - | 1.27E-3 |

| Ribozyme | [BG-$NH_2$] | $k_1$ ($min^{-1}$) | $k_2/k_{obs}$ ($min^{-1}$) |
|---|---|---|---|
| CK03 | 25 µM | - | 1.77E-3 |
| CK03 | 100 µM | 8.68E-4 | 7.80E-3 |
| CA13 | 25 µM | 1.80E-3 | 6.73E-2 |
| CA13 | 100 µM | 2.90E-3 | 7.20E-2 |

## 9.5.3 Extended data figures

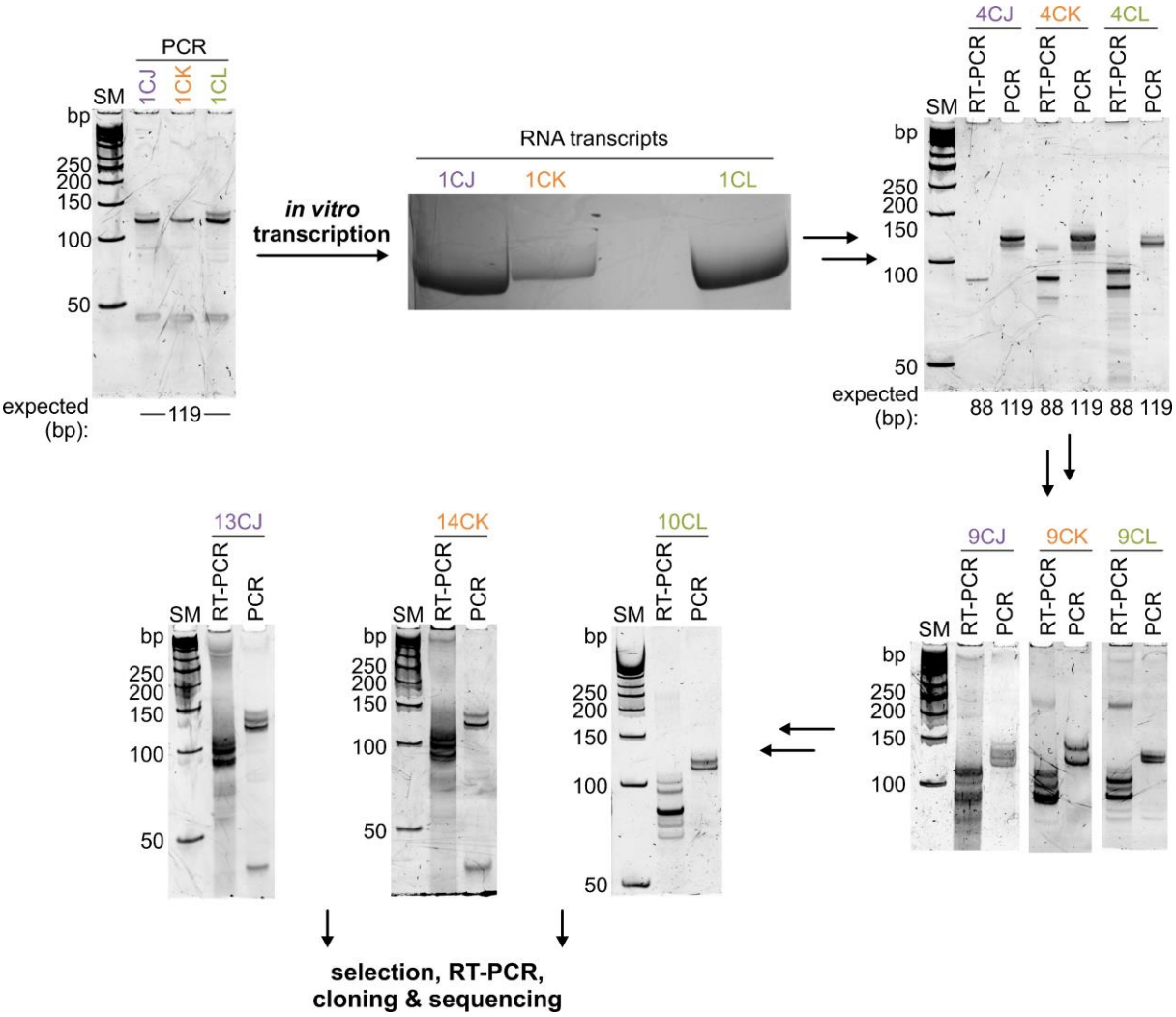



**Figure 9.30:** Representative gel images from selections CJ-CL. Shown are gel images of the analysis of the RT-PCR and PCR products after specific selection rounds and for the purification of the *in vitro* transcripts of the first round. First signs of enrichment were detected after round 8 (CJ and CK) or round 3 (CL). The RT-PCR product of selection rounds 14 (CJ), 15 (CK) and 11 (CL) were cloned and submitted for Sanger sequencing.

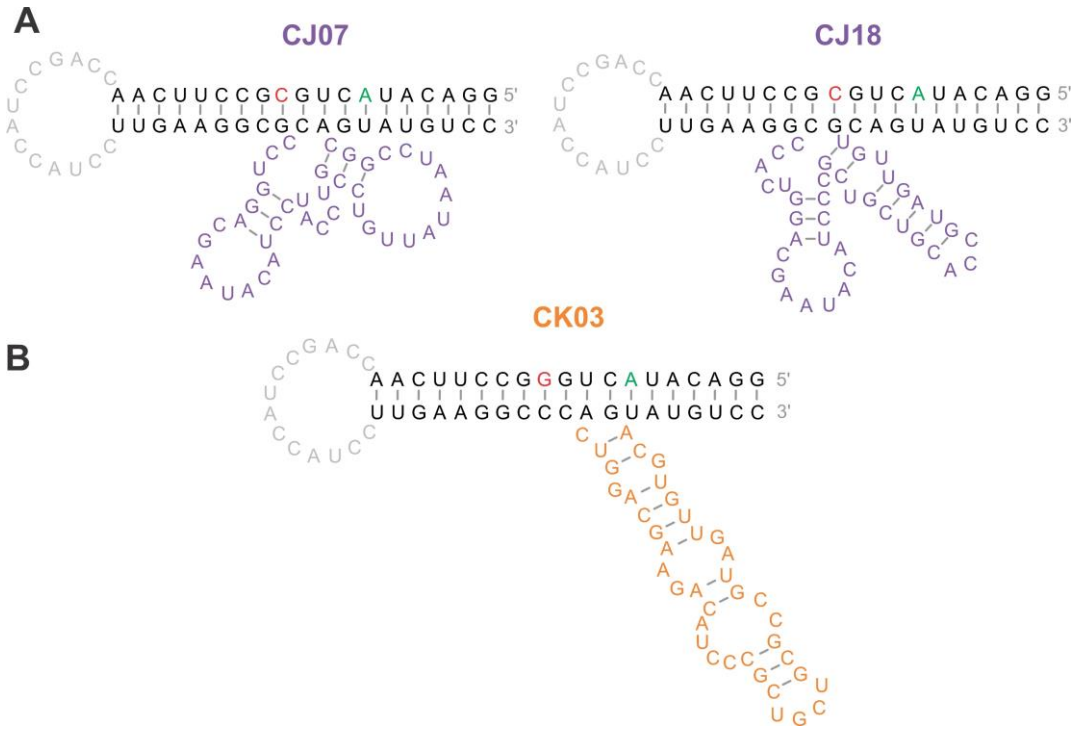



**Figure 9.31:** Comparison of sequences and secondary structures of *cis*-active CJ07, CJ18 and CK03 ribozymes The catalytic core is marked in purple (CJ ribozymes) or orange (CK03 ribozyme), the connecting loop is colored grey. The intended modification site is shown in red and the adenosine modified by the ribozymes is marked green. The secondary structures were predicted using the RNAfold web server.[391]

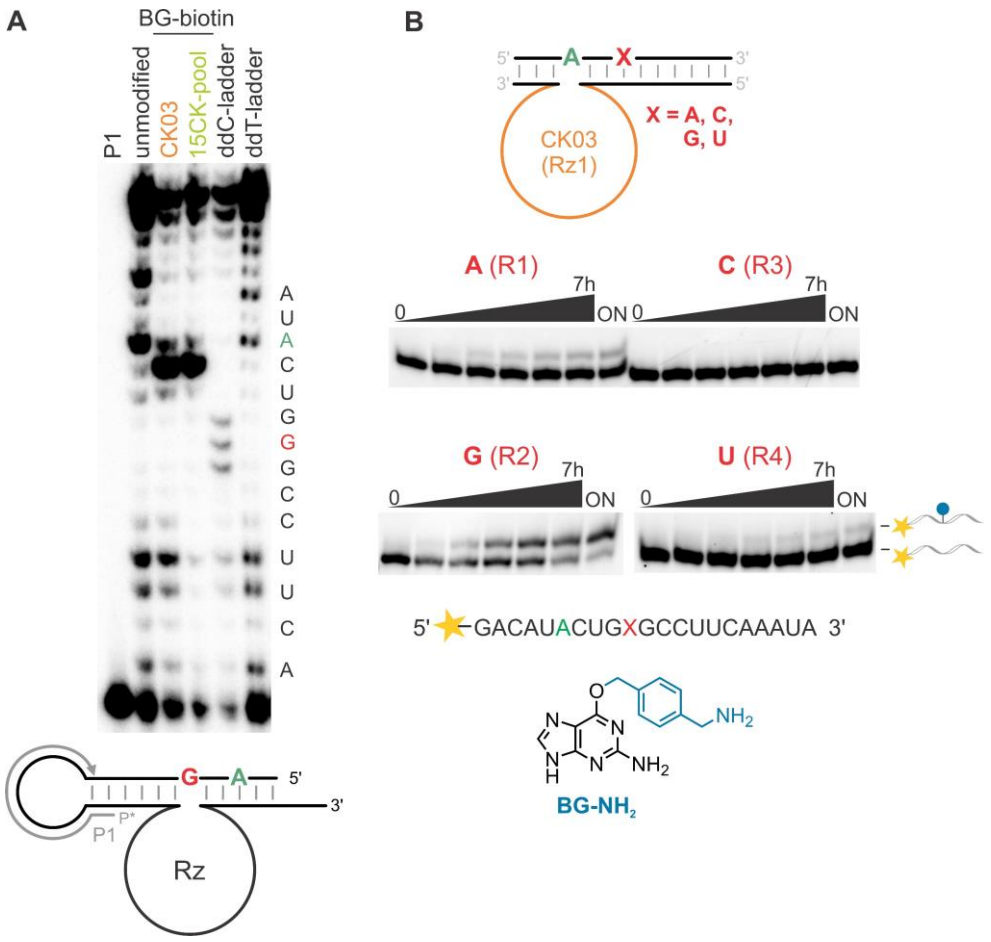



**Figure 9.32:** Characterization of the 15CK pool and the CK03 ribozyme. **A** RT primer extension of *cis*-active CK03 and the round 15 CK selection pool (15CK) that were reacted with 100 µM BG-biotin for 4 h in the presence of 40 mM $MgCl_2$ at 37 °C using 5'-$^{32}$P-labeled primer P1. **C** *Trans*-activity test of CK03 (Rz1) under single turnover conditions with substrate RNAs R1-R4 using BG-$NH_2$ as alkyl donor. Timepoints: 0, 15min, 1h, 2h, 4h, 7h, ON.

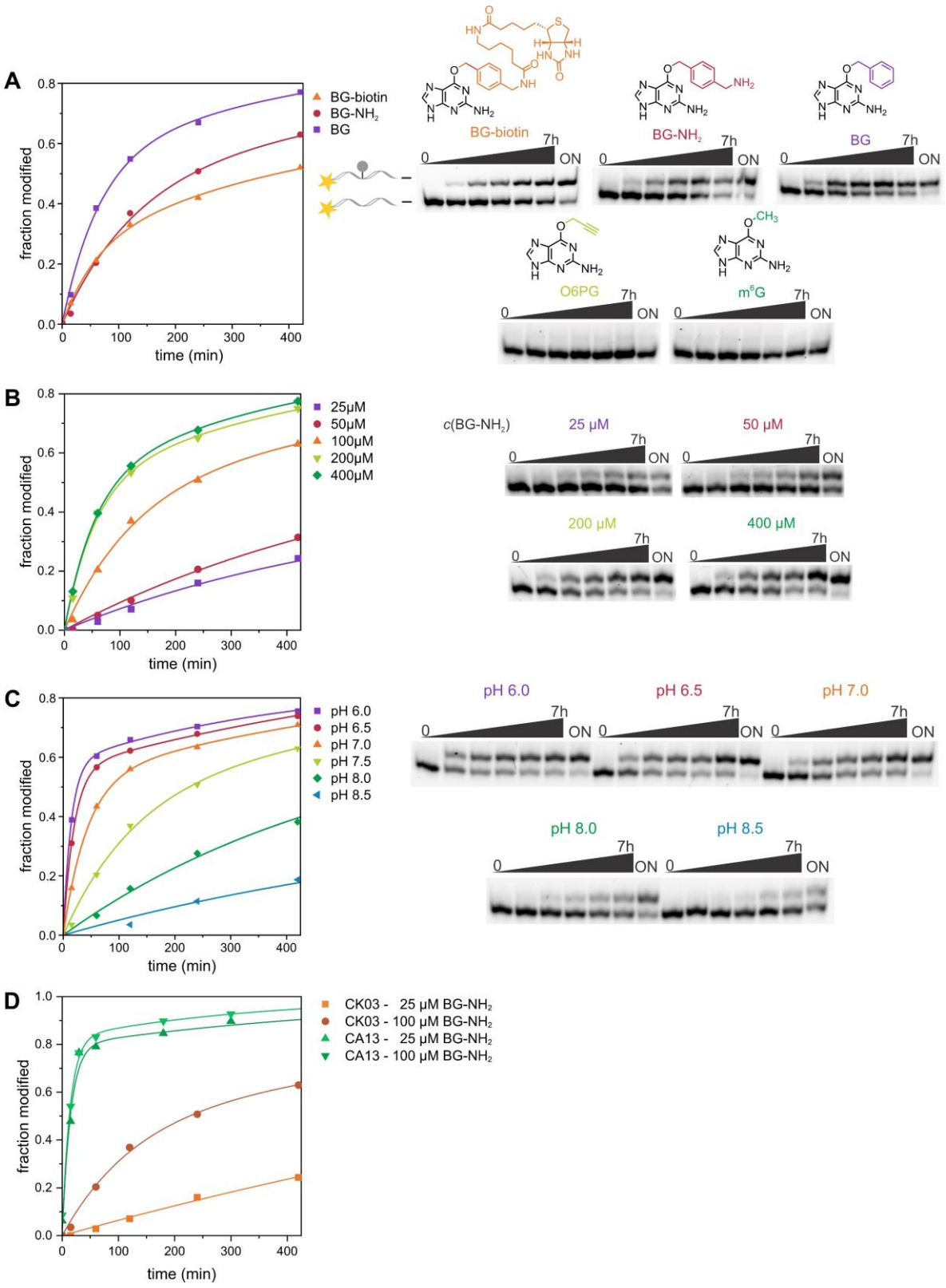



**Figure 9.33:** Characterization of the CK03 ribozyme. All reactions were performed under single turnover conditions using the *trans*-active CK03 ribozyme Rz2 with substrate RNA R2. Unless otherwise noted, all reactions were conducted at pH 7.5 and 37 °C in the presence of 100 µM BG-$NH_2$ and 40 mM $MgCl_2$. The data was fitted using pseudo-first order kinetics (for slow reactions) or a biexponential model (for fast kinetics). **A** Investigating the substrate scope of CK03. No reactivity was determined for O6PG and $m^6G$. **B** BG-$NH_2$-dependency of CK03. **C** pH-dependency of CK03. **D** Comparison of CK03 and CA13 reacted in the presence of 25 µM or 100 µM BG-$NH_2$. The $k_{obs}$ values obtained from all experiments are summarized in Table 9.23.

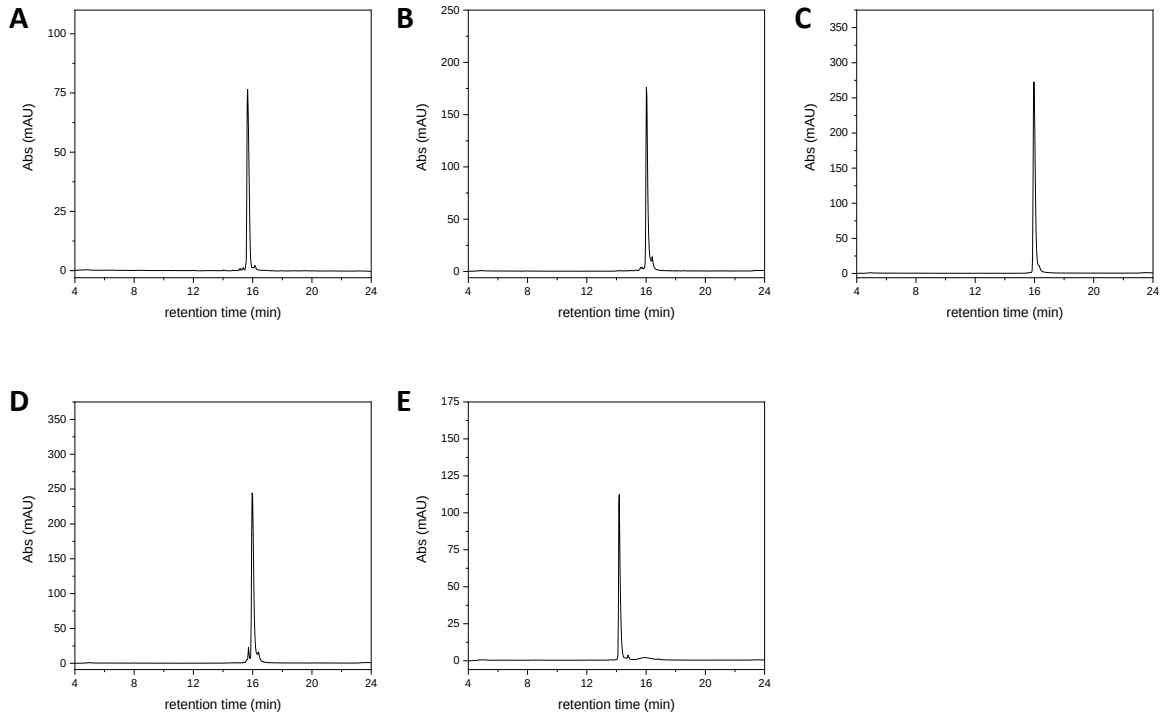



**Figure 9.34:** Anion exchange HPLC chromatograms to confirm purity of RNA oligonucleotides prepared by solid-phase synthesis. **A** R1, **B** R2, **C** R3, **D** R4, **E** R5. Dionex DNAPAc PA200, 2x250 mm, 60 °C, monitored at 260 nm.

# Individual contributions

The coauthors of the publications included in this semi-cumulative thesis are informed and agree with the reprint and the individual contributions as stated below.

**Site-specific RNA methylation by a methyltransferase ribozyme**

Scheitl, C.P.M.; Ghaem Maghami, M.; Lenz, A.-K.; Höbartner, C., *Nature* **2020**, *587*, 663-667.

Autoren/innen (ggf. Haupt- / Ko- / korrespondierende/r Autor/in) mit Vorname Nachname (Initialen): Carolin P.M. Scheitl (Hauptautorin, C.P.M.S.), Mohammad Ghaem Maghami (Koautor, M.G.M.) Ann-Kathrin Lenz (Koautorin, A.-K.L.), Claudia Höbartner (korrespondierende Autorin, C.H.)

Detaillierte Darstellung der Anteile an der Veröffentlichung (in %):

| **Autor** | **C.P.M.S.** | **M.G.M.** | **A.-K.L.** | **C.H.** | **∑ in Prozent** |
|---|---|---|---|---|---|
| Research conception | 1 % | 2 % | - | 2 % | 5 % |
| *In vitro* selection | 20 % | 2 % | - | - | 22 % |
| RNA solid-phase synthesis | - | - | 3 % | 2 % | 5 % |
| Ribozyme characterization | 12 % | 1 % | 2 % | - | 15 % |
| Plasmid construction | 2 % | 5 % | 3 % | - | 10 % |
| RNA structure probing | 10 % | - | - | - | 10 % |
| RT & LC-MS analysis | 13 % | - | - | 2 % | 15 % |
| Publication writing | 3 % | - | - | 7 % | 10 % |
| Publication correction | 1 % | - | - | 3 % | 4 % |
| Publication coordination | - | - | - | 4 % | 4 % |
| **Summe** | **62 %** | **10 %** | **8 %** | **20 %** | **100 %** |

**Structure and mechanism of the methyltransferase ribozyme MTR1**

C.P.M. Scheitl, M. Mieczkowski, H. Schindelin, C. Höbartner, *Nat. Chem. Biol.* **2022**, *18* (5), 547-555.

Autoren/innen (ggf. Haupt- / Ko- / korrespondierende/r Autor/in) mit Vorname Nachname (Initialen): Carolin P.M. Scheitl (Hauptautorin, C.P.M.S.), Mateusz Mieczkowski (Hauptautor, M.M.) Hermann Schindelin (Koautor, H.S.), Claudia Höbartner (korrespondierende Autorin, C.H.)

Detaillierte Darstellung der Anteile an der Veröffentlichung (in %):

| **Autor** | **C.P.M.S.** | **M.M.** | **H.S.** | **C.H.** | **∑ in Prozent** |
|---|---|---|---|---|---|
| Research conception | 1 % | 1 % | - | 3 % | 5 % |
| Ribozyme crystallization | 12 % | 3 % | - | - | 15 % |
| Data collection | 2 % | 2 % | 3 % | 2 % | 9 % |
| Data analysis and model building | - | 21 % | 1 % | 1 % | 23 % |
| Ribozyme kinetics | 20 % | - | - | - | 20 % |
| RNA structure probing | 10 % | - | - | - | 10 % |
| Publication writing | 2 % | 2 % | 1 % | 5 % | 10 % |
| Publication correction | 1 % | - | - | 3 % | 4 % |
| Publication coordination | - | - | - | 4 % | 4 % |
| **Summe** | **47 %** | **29 %** | **5 %** | **18 %** | **100 %** |

**Ribozyme-catalyzed Late-stage Functionalization and Fluorogenic Labeling of RNA**

C.P.M. Scheitl, T. Okuda, J. Adelmann, C. Höbartner, *Angew. Chem. Int. Ed.* **2023**, e202305463.

Autoren/innen (ggf. Haupt- / Ko- / korrespondierende/r Autor/in) mit Vorname Nachname (Initialen): Carolin P.M. Scheitl (Hauptautorin, C.P.M.S.), Takumi Okuda (Koautor, T.O.), Juliane Adelman (Koautor, J.A.), Claudia Höbartner (korrespondierende Autorin, C.H.)

Detaillierte Darstellung der Anteile an der Veröffentlichung (in %):

| **Autor** | **C.P.M.S.** | **T.O.** | **J.A.** | **C.H.** | **∑ in Prozent** |
|---|---|---|---|---|---|
| Research conception | 2 % | - | - | 3 % | 5 % |
| RNA synthesis (solid-phase and enzymatic) | 20 % | 4 % | - | - | 24 % |
| Ribozyme and product characterization (biochemical and spectroscopic) | 37 % | - | - | - | 37 % |
| LC/ESI-MS analysis | 3 % | 5 % | 5 % | 3 % | 16 % |
| Publication writing | 7 % | - | - | 3 % | 10 % |
| Publication correction | 1 % | - | - | 3 % | 4 % |
| Publication coordination | - | - | - | 4 % | 4 % |
| **Summe** | **70 %** | **9 %** | **5 %** | **16 %** | **100 %** |

# References


1. Fu, X. D., Non-coding RNA: a new frontier in regulatory biology. *Natl Sci Rev* **2014,** *1* (2), 190-204.
2. Moosavi, A.; Motevalizadeh Ardekani, A., Role of Epigenetics in Biology and Human Diseases. *Iran Biomed J* **2016,** *20* (5), 246-58.
3. Boccaletto, P.; Machnicka, M. A.; Purta, E., et al., MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Research* **2018,** *46* (D1), D303-D307.
4. Song, J.; Yi, C., Chemical Modifications to RNA: A New Layer of Gene Expression Regulation. *ACS Chem Biol* **2017,** *12* (2), 316-325.
5. Carell, T.; Brandmayr, C.; Hienzsch, A., et al., Structure and function of noncanonical nucleobases. *Angew Chem Int Ed Engl* **2012,** *51* (29), 7110-31.
6. Charette, M.; Gray, M. W., Pseudouridine in RNA: What, where, how, and why. *Iubmb Life* **2000,** *49* (5), 341-351.
7. Wu, Y.; Tang, Y.; Dong, X., et al., RNA Phosphorothioate Modification in Prokaryotes and Eukaryotes. *ACS Chem Biol* **2020,** *15* (6), 1301-1305.
8. Coghill, T. B. J. R. D., Researches on Pyrimidines. C111. The Discovery of 5-Methyl-Cytosine in Tuberculinic Acid, the Nucleic Acid of the Tubercle Bacillus 1. *J Am Chem Soc* **1925,** *47* (11), 2838-2844.
9. Pan, T., Modifications and functional genomics of human transfer RNA. *Cell Res* **2018,** *28* (4), 395-404.
10. Yarian, C.; Townsend, H.; Czestkowski, W., et al., Accurate translation of the genetic code depends on tRNA modified nucleosides. *J Biol Chem* **2002,** *277* (19), 16391-16395.
11. Haag, S.; Sloan, K. E.; Ranjan, N., et al., NSUN3 and ABH1 modify the wobble position of mt-tRNA(Met) to expand codon recognition in mitochondrial translation. *Embo J* **2016,** *35* (19), 2104-2119.
12. Helm, M.; Brule, H.; Degoul, F., et al., The presence of modified nucleotides is required for cloverleaf folding of a human mitochondrial tRNA. *Nucleic Acids Research* **1998,** *26* (7), 1636-1643.
13. Helm, M.; Giege, R.; Florentz, C., A Watson-Crick base-pair-disrupting methyl group (m(1)A9) is sufficient for cloverleaf folding of human mitochondrial tRNA(Lys). *Biochemistry-Us* **1999,** *38* (40), 13338-13346.
14. Nobles, K. N.; Yarian, C. S.; Liu, G., et al., Highly conserved modified nucleosides influence Mg2+-dependent tRNA folding. *Nucleic Acids Res* **2002,** *30* (21), 4751-60.
15. Davis, D. R.; Veltri, C. A.; Nielsen, L., An RNA model system for investigation of pseudouridine stabilization sf the codon-anticodon interaction in tRNA(Lys), tRNA(His) and tRNA(Tyr). *J Biomol Struct Dyn* **1998,** *15* (6), 1121-1132.
16. Incarnato, D.; Anselmi, F.; Morandi, E., et al., High-throughput single-base resolution mapping of RNA 2 '-O-methylated residues. *Nucleic Acids Research* **2017,** *45* (3), 1433-1441.
17. Bakin, A.; Lane, B. G.; Ofengand, J., Clustering of pseudouridine residues around the peptidyltransferase center of yeast cytoplasmic and mitochondrial ribosomes. *Biochemistry-Us* **1994,** *33* (45), 13475-83.
18. Rackham, O.; Wang, K.; Chin, J. W., Functional epitopes at the ribosome subunit interface. *Nat Chem Biol* **2006,** *2* (5), 254-258.

19. Zhang, W.; Dunkle, J. A.; Cate, J. H. D., Structures of the Ribosome in Intermediate States of Ratcheting. *Science* **2009,** *325* (5943), 1014-1017.
20. Sloan, K. E.; Warda, A. S.; Sharma, S., et al., Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *Rna Biol* **2017,** *14* (9), 1138-1152.
21. Liang, X. H.; Liu, Q.; Fournier, M. J., rRNA modifications in an intersubunit bridge of the ribosome strongly affect both ribosome biogenesis and activity. *Molecular Cell* **2007,** *28* (6), 965-977.
22. Shatkin, A. J., Capping of eucaryotic mRNAs. *Cell* **1976,** *9* (4 PT 2), 645-53.
23. Mauer, J.; Luo, X.; Blanjoie, A., et al., Reversible methylation of m(6)A(m) in the 5' cap controls mRNA stability. *Nature* **2017,** *541* (7637), 371-375.
24. Wei, C.; Gershowitz, A.; Moss, B., N6, O2'-dimethyladenosine a novel methylated ribonucleoside next to the 5' terminal of animal cell and virus mRNAs. *Nature* **1975,** *257* (5523), 251-3.
25. Cowling, V. H., Regulation of mRNA cap methylation. *Biochem J* **2009,** *425* (2), 295-302.
26. Zhao, B. X. S.; Roundtree, I. A.; He, C., Post-transcriptional gene regulation by mRNA modifications. *Nat Rev Mol Cell Bio* **2017,** *18* (1), 31-42.
27. Liu, N.; Pan, T., RNA epigenetics. *Transl Res* **2015,** *165* (1), 28-35.
28. Desrosiers, R.; Friderici, K.; Rottman, F., Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc Natl Acad Sci U S A* **1974,** *71* (10), 3971-5.
29. Deng, X.; Chen, K.; Luo, G. Z., et al., Widespread occurrence of N6-methyladenosine in bacterial mRNA. *Nucleic Acids Res* **2015,** *43* (13), 6557-67.
30. Krug, R. M.; Morgan, M. A.; Shatkin, A. J., Influenza viral mRNA contains internal N6-methyladenosine and 5'-terminal 7-methylguanosine in cap structures. *J Virol* **1976,** *20* (1), 45-53.
31. Harper, J. E.; Miceli, S. M.; Roberts, R. J., et al., Sequence specificity of the human mRNA N6-adenosine methylase in vitro. *Nucleic Acids Res* **1990,** *18* (19), 5735-41.
32. Meyer, K. D.; Saletore, Y.; Zumbo, P., et al., Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **2012,** *149* (7), 1635-46.
33. Liu, N.; Dai, Q.; Zheng, G. Q., et al., N-6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **2015,** *518* (7540), 560-564.
34. Zheng, G. Q.; Dahl, J. A.; Niu, Y. M., et al., ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Molecular Cell* **2013,** *49* (1), 18-29.
35. Wang, X.; Lu, Z.; Gomez, A., et al., N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **2014,** *505* (7481), 117-20.
36. Geula, S.; Moshitch-Moshkovitz, S.; Dominissini, D., et al., Stem cells. m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science* **2015,** *347* (6225), 1002-6.
37. Wang, X.; Zhao, B. S.; Roundtree, I. A., et al., N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **2015,** *161* (6), 1388-99.
38. Zou, S.; Toh, J. D. W.; Wong, K. H. Q., et al., N-6-Methyladenosine: a conformational marker that regulates the substrate specificity of human demethylases FTO and ALKBH5. *Sci Rep-Uk* **2016,** *6*.
39. Schwartz, S.; Mumbach, M. R.; Jovanovic, M., et al., Perturbation of m6A Writers Reveals Two Distinct Classes of mRNA Methylation at Internal and 5 ' Sites. *Cell Rep* **2014,** *8* (1), 284-296.
40. Huang, H.; Weng, H.; Chen, J., m(6)A Modification in Coding and Non-coding RNAs: Roles and Therapeutic Implications in Cancer. *Cancer Cell* **2020,** *37* (3), 270-288.

41. Zhang, B. J.; Jiang, H.; Dong, Z., et al., The critical roles of m6A modification in metabolic abnormality and cardiovascular diseases. *Genes Dis* **2021,** *8* (6), 746-758.
42. Li, X. Y.; Xiong, X. S.; Wang, K., et al., Transcriptome-wide mapping reveals reversible and dynamic N-1-methyladenosine methylome. *Nat Chem Biol* **2016,** *12* (5), 311-+.
43. Dominissini, D.; Nachtergaele, S.; Moshitch-Moshkovitz, S., et al., The dynamic N-1-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **2016,** *530* (7591), 441-+.
44. Li, X.; Xiong, X.; Zhang, M., et al., Base-Resolution Mapping Reveals Distinct m(1)A Methylome in Nuclear- and Mitochondrial-Encoded Transcripts. *Mol Cell* **2017,** *68* (5), 993-1005 e9.
45. Safra, M.; Sas-Chen, A.; Nir, R., et al., The m(1)A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* **2017,** *551* (7679), 251-+.
46. Schwartz, S., m(1)A within cytoplasmic mRNAs at single nucleotide resolution: a reconciled transcriptome-wide map. *Rna* **2018,** *24* (11), 1427-1436.
47. Carlile, T. M.; Rojas-Duran, M. F.; Zinshteyn, B., et al., Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **2014,** *515* (7525), 143-+.
48. Li, X. Y.; Zhu, P.; Ma, S. Q., et al., Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat Chem Biol* **2015,** *11* (8), 592-U93.
49. Karijolich, J.; Yu, Y. T., Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature* **2011,** *474* (7351), 395-+.
50. Shimizu-Motohashi, Y.; Miyatake, S.; Komaki, H., et al., Recent advances in innovative therapeutic approaches for Duchenne muscular dystrophy: from discovery to clinical trials. *Am J Transl Res* **2016,** *8* (6), 2471-89.
51. Pranke, I.; Golec, A.; Hinzpeter, A., et al., Emerging Therapeutic Approaches for Cystic Fibrosis. From Gene Editing to Personalized Medicine. *Front Pharmacol* **2019,** *10*, 121.
52. Linde, L.; Kerem, B., Introducing sense into nonsense in treatments of human genetic diseases. *Trends Genet* **2008,** *24* (11), 552-63.
53. Srinivasan, S.; Torres, A. G.; Ribas de Pouplana, L., Inosine in Biology and Disease. *Genes (Basel)* **2021,** *12* (4).
54. Swinehart, W. E.; Jackman, J. E., Diversity in mechanism and function of tRNA methyltransferases. *Rna Biol* **2015,** *12* (4), 398-411.
55. Ping, X. L.; Sun, B. F.; Wang, L., et al., Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Research* **2014,** *24* (2), 177-189.
56. Sledz, P.; Jinek, M., Structural insights into the molecular mechanism of the m(6)A writer complex. *Elife* **2016,** *5*.
57. Wang, X.; Feng, J.; Xue, Y., et al., Structural basis of N6-adenosine methylation by the METTL3-METTL14 complex (vol 534, pg 575, 2016). *Nature* **2017,** *542* (7640), 260-260.
58. Wang, P.; Doxtader, K. A.; Nam, Y., Structural Basis for Cooperative Function of Mettl3 and Mettl14 Methyltransferases. *Molecular Cell* **2016,** *63* (2), 306-317.
59. Wang, X.; Feng, J.; Xue, Y., et al., Structural basis of N-6-adenosine methylation by the METTL3-METTL14 complex. *Nature* **2016,** *534* (7608), 575-+.
60. Bawankar, P.; Lence, T.; Paolantoni, C., et al., Hakai is required for stabilization of core components of the m(6)A mRNA methylation machinery. *Nature Communications* **2021,** *12* (1).
61. Knuckles, P.; Lence, T.; Haussmann, I. U., et al., Zc3h13/Flacc is required for adenosine methylation by bridging the mRNA-binding factor Rbm15/Spenito to the m(6)A machinery component Wtap/Fl(2)d. *Gene Dev* **2018,** *32* (5-6), 415-429.

62. Jia, G. F.; Fu, Y.; Zhao, X., et al., N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol* **2011,** *7* (12), 885-887.
63. Gerken, T.; Girard, C. A.; Tung, Y. C., et al., The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **2007,** *318* (5855), 1469-72.
64. Mauer, J.; Jaffrey, S. R., FTO, m(6) A(m) , and the hypothesis of reversible epitranscriptomic mRNA modifications. *FEBS Lett* **2018,** *592* (12), 2012-2022.
65. Wei, J. B.; Liu, F. G.; Lu, Z. K., et al., Differential m(6)A, m(6)A(m), and m(1)A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm. *Molecular Cell* **2018,** *71* (6), 973-+.
66. Gulati, P.; Avezov, E.; Ma, M., et al., Fat mass and obesity-related (FTO) shuttles between the nucleus and cytoplasm. *Bioscience Rep* **2014,** *34*, 621-628.
67. Li, Z. J.; Weng, H. Y.; Su, R., et al., FTO Plays an Oncogenic Role in Acute Myeloid Leukemia as a N-6-Methyladenosine RNA Demethylase. *Cancer Cell* **2017,** *31* (1), 127-141.
68. Fu, Y.; Jia, G.; Pang, X., et al., FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. *Nat Commun* **2013,** *4*, 1798.
69. Zheng, G.; Dahl, J. A.; Niu, Y., et al., Sprouts of RNA epigenetics: the discovery of mammalian RNA demethylases. *Rna Biol* **2013,** *10* (6), 915-8.
70. Kurowski, M. A.; Bhagwat, A. S.; Papaj, G., et al., Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB. *BMC Genomics* **2003,** *4* (1), 48.
71. Kawarada, L.; Suzuki, T.; Ohira, T., et al., ALKBH1 is an RNA dioxygenase responsible for cytoplasmic and mitochondrial tRNA modifications. *Nucleic Acids Research* **2017,** *45* (12), 7401-7415.
72. Ma, C. J.; Ding, J. H.; Ye, T. T., et al., AlkB Homologue 1 Demethylates N-3-Methylcytidine in mRNA of Mammals. *Acs Chemical Biology* **2019,** *14* (7), 1418-1425.
73. Ougland, R.; Lando, D.; Jonson, I., et al., ALKBH1 Is a Histone H2A Dioxygenase Involved in Neural Differentiation. *Stem Cells* **2012,** *30* (12), 2672-2682.
74. Chen, B. E.; Liu, H. C.; Sun, X. X., et al., Mechanistic insight into the recognition of single-stranded and double-stranded DNA substrates by ABH2 and ABH3. *Mol Biosyst* **2010,** *6* (11), 2143-2149.
75. Li, P.; Gao, S.; Wang, L., et al., ABH2 couples regulation of ribosomal DNA transcription with DNA alkylation repair. *Cell Rep* **2013,** *4* (4), 817-29.
76. Ougland, R.; Zhang, C. M.; Liiv, A., et al., AlkB restores the biological function of mRNA and tRNA inactivated by chemical methylation. *Mol Cell* **2004,** *16* (1), 107-16.
77. Fu, Y.; Dai, Q.; Zhang, W., et al., The AlkB Domain of Mammalian ABH8 Catalyzes Hydroxylation of 5-Methoxycarbonylmethyluridine at the Wobble Position of tRNA. *Angew Chem Int Edit* **2010,** *49* (47), 8885-8888.
78. Songe-Moller, L.; van den Born, E.; Leihne, V., et al., Mammalian ALKBH8 possesses tRNA methyltransferase activity required for the biogenesis of multiple wobble uridine modifications implicated in translational decoding. *Mol Cell Biol* **2010,** *30* (7), 1814-27.
79. van den Born, E.; Vagbo, C. B.; Songe-Moller, L., et al., ALKBH8-mediated formation of a novel diastereomeric pair of wobble nucleosides in mammalian tRNA. *Nat Commun* **2011,** *2*, 172.
80. Li, M. M.; Nilsen, A.; Shi, Y., et al., ALKBH4-dependent demethylation of actin regulates actomyosin dynamics. *Nature Communications* **2013,** *4*.
81. Dango, S.; Mosammaparast, N.; Sowa, M. E., et al., DNA Unwinding by ASCC3 Helicase Is Coupled to ALKBH3-Dependent DNA Alkylation Repair and Cancer Cell Proliferation. *Molecular Cell* **2011,** *44* (3), 373-384.

82. Koike, K.; Ueda, Y.; Hase, H., et al., Anti-Tumor Effect of AlkB Homolog 3 Knockdown in Hormone-Independent Prostate Cancer Cells. *Curr Cancer Drug Tar* **2012,** *12* (7), 847-856.

83. Ozanick, S.; Krecic, A.; Andersland, J., et al., The bipartite structure of the tRNA m(1)A58 methyltransferase from S-cerevisiae is conserved in humans. *Rna* **2005,** *11* (8), 1281-1290.

84. Kiss-Laszlo, Z.; Henry, Y.; Bachellerie, J. P., et al., Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* **1996,** *85* (7), 1077-88.

85. Tollervey, D.; Lehtonen, H.; Jansen, R., et al., Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. *Cell* **1993,** *72* (3), 443-57.

86. Ni, J.; Tien, A. L.; Fournier, M. J., Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell* **1997,** *89* (4), 565-73.

87. Lafontaine, D. L.; Bousquet-Antonelli, C.; Henry, Y., et al., The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev* **1998,** *12* (4), 527-37.

88. Keith, G., Mobilities of Modified Ribonucleotides on 2-Dimensional Cellulose Thin-Layer Chromatography. *Biochimie* **1995,** *77* (1-2), 142-144.

89. Grosjean, H.; Droogmans, L.; Roovers, M., et al., Detection of enzymatic activity of transfer RNA modification enzymes using radiolabeled TRNA substrates. *Rna Modification* **2007,** *425*, 57-101.

90. Nees, G.; Kaufmann, A.; Bauer, S., Detection of RNA modifications by HPLC analysis and competitive ELISA. *Methods Mol Biol* **2014,** *1169*, 3-14.

91. Jora, M.; Lobue, P. A.; Ross, R. L., et al., Detection of ribonucleoside modifications by liquid chromatography coupled with mass spectrometry. *Biochim Biophys Acta Gene Regul Mech* **2019,** *1862* (3), 280-290.

92. Kellner, S.; Ochel, A.; Thuring, K., et al., Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. *Nucleic Acids Research* **2014,** *42* (18).

93. Globisch, D.; Pearson, D.; Hienzsch, A., et al., Systems-based analysis of modified tRNA bases. *Angew Chem Int Ed Engl* **2011,** *50* (41), 9739-42.

94. Liu, N.; Parisien, M.; Dai, Q., et al., Probing N-6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *Rna* **2013,** *19* (12), 1848-1856.

95. Scheitl, C. P. M.; Ghaem Maghami, M.; Lenz, A. K., et al., Site-specific RNA methylation by a methyltransferase ribozyme. *Nature* **2020,** *587* (7835), 663-667.

96. Sharma, S.; Hartmann, J. D.; Watzinger, P., et al., A single N-1-methyladenosine on the large ribosomal subunit rRNA impacts locally its structure and the translation of key metabolic enzymes. *Sci Rep-Uk* **2018,** *8*.

97. Swinehart, W. E.; Henderson, J. C.; Jackman, J. E., Unexpected expansion of tRNA substrate recognition by the yeast m(1)G(9) methyltransferase Trm10. *Rna* **2013,** *19* (8), 1137-1146.

98. Xu, L.; Liu, X. Y.; Sheng, N., 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.

99. Sharma, S.; Yang, J.; Duttmann, S., et al., Identification of novel methyltransferases, Bmt5 and Bmt6, responsible for the m3U methylations of 25S rRNA in Saccharomyces cerevisiae. *Nucleic Acids Res* **2014,** *42* (5), 3246-60.

100. Dominissini, D.; Moshitch-Moshkovitz, S.; Schwartz, S., et al., Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **2012,** *485* (7397), 201-6.

101. Li, X.; Xiong, X.; Wang, K., et al., Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome. *Nat Chem Biol* **2016,** *12* (5), 311-6.
102. Gu, X.; Liang, Z., Transcriptome-Wide Mapping 5-Methylcytosine by m(5)C RNA Immunoprecipitation Followed by Deep Sequencing in Plant. *Methods Mol Biol* **2019,** *1933*, 389-394.
103. Chen, K.; Lu, Z.; Wang, X., et al., High-resolution N(6) -methyladenosine (m(6) A) map using photo-crosslinking-assisted m(6) A sequencing. *Angew Chem Int Ed Engl* **2015,** *54* (5), 1587-90.
104. Linder, B.; Grozhik, A. V.; Olarerin-George, A. O., et al., Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods* **2015,** *12* (8), 767-72.
105. Feederle, R.; Schepers, A., Antibodies specific for nucleic acid modifications. *Rna Biol* **2017,** *14* (9), 1089-1098.
106. Meyer, K. D., DART-seq: an antibody-free method for global m(6)A detection. *Nat Methods* **2019,** *16* (12), 1275-1280.
107. Tegowski, M.; Flamand, M. N.; Meyer, K. D., scDART-seq reveals distinct m(6)A signatures and mRNA methylation heterogeneity in single cells. *Mol Cell* **2022,** *82* (4), 868-878 e10.
108. Wang, Y.; Xiao, Y.; Dong, S., et al., Antibody-free enzyme-assisted chemical approach for detection of N(6)-methyladenosine. *Nat Chem Biol* **2020,** *16* (8), 896-903.
109. Garcia-Campos, M. A.; Edelheit, S.; Toth, U., et al., Deciphering the "m(6)A Code" via Antibody-Independent Quantitative Profiling. *Cell* **2019,** *178* (3), 731-+.
110. Zhang, Z.; Chen, L. Q.; Zhao, Y. L., et al., Single-base mapping of m(6)A by an antibody-independent method. *Sci Adv* **2019,** *5* (7).
111. Birkedal, U.; Christensen-Dalsgaard, M.; Krogh, N., et al., Profiling of ribose methylations in RNA by high-throughput sequencing. *Angew Chem Int Ed Engl* **2015,** *54* (2), 451-5.
112. Marchand, V.; Blanloeil-Oillo, F.; Helm, M., et al., Illumina-based RiboMethSeq approach for mapping of 2'-O-Me residues in RNA. *Nucleic Acids Res* **2016,** *44* (16), e135.
113. Zhu, Y.; Pirnie, S. P.; Carmichael, G. G., High-throughput and site-specific identification of 2'-O-methylation sites using ribose oxidation sequencing (RibOxi-seq). *Rna* **2017,** *23* (8), 1303-1314.
114. Dai, Q.; Moshitch-Moshkovitz, S.; Han, D., et al., Nm-seq maps 2'-O-methylation sites in human mRNA with base precision. *Nat Methods* **2017,** *14* (7), 695-698.
115. Marchand, V.; Ayadi, L.; Ernst, F. G. M., et al., AlkAniline-Seq: Profiling of m(7) G and m(3) C RNA Modifications at Single Nucleotide Resolution. *Angew Chem Int Ed Engl* **2018,** *57* (51), 16785-16790.
116. Cui, J.; Liu, Q.; Sendinc, E., et al., Nucleotide resolution profiling of m3C RNA modification by HAC-seq. *Nucleic Acids Res* **2021,** *49* (5), e27.
117. Jain, M.; Olsen, H. E.; Paten, B., et al., The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* **2016,** *17* (1), 239.
118. Wang, Y.; Zhao, Y.; Bollas, A., et al., Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol* **2021,** *39* (11), 1348-1365.
119. Branton, D.; Deamer, D. W.; Marziali, A., et al., The potential and challenges of nanopore sequencing. *Nat Biotechnol* **2008,** *26* (10), 1146-53.
120. Garalde, D. R.; Snell, E. A.; Jachimowicz, D., et al., Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* **2018,** *15* (3), 201-206.
121. Dong, X.; Tian, L.; Gouil, Q., et al., The long and the short of it: unlocking nanopore long-read RNA sequencing data with short-read differential expression analysis tools. *NAR Genom Bioinform* **2021,** *3* (2), lqab028.

122. Furlan, M.; Delgado-Tejedor, A.; Mulroney, L., et al., Computational methods for RNA modification detection from nanopore direct RNA sequencing data. *Rna Biol* **2021,** *18* (sup1), 31-40.
123. Thomas, N. K.; Poodari, V. C.; Jain, M., et al., Direct Nanopore Sequencing of Individual Full Length tRNA Strands. *ACS Nano* **2021,** *15* (10), 16642-16653.
124. Smith, A. M.; Jain, M.; Mulroney, L., et al., Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. *PLoS One* **2019,** *14* (5), e0216709.
125. Jenjaroenpun, P.; Wongsurawat, T.; Wadley, T. D., et al., Decoding the epitranscriptional landscape from native RNA sequences. *Nucleic Acids Res* **2021,** *49* (2), e7.
126. Liu, H.; Begik, O.; Lucas, M. C., et al., Accurate detection of m(6)A RNA modifications in native RNA sequences. *Nat Commun* **2019,** *10* (1), 4079.
127. Wan, Y. K.; Hendra, C.; Pratanwanich, P. N., et al., Beyond sequencing: machine learning algorithms extract biology hidden in Nanopore signal data. *Trends Genet* **2022,** *38* (3), 246-257.
128. White, L. K.; Hesselberth, J. R., Modification mapping by nanopore sequencing. *Front Genet* **2022,** *13*.
129. Buchhaupt, M.; Peifer, C.; Entian, K. D., Analysis of 2'-O-methylated nucleosides and pseudouridines in ribosomal RNAs using DNAzymes. *Anal Biochem* **2007,** *361* (1), 102-8.
130. Sednev, M. V.; Mykhailiuk, V.; Choudhury, P., et al., N(6) -Methyladenosine-Sensitive RNA-Cleaving Deoxyribozymes. *Angew Chem Int Ed Engl* **2018,** *57* (46), 15117-15121.
131. Bujnowska, M.; Zhang, J. C.; Dai, Q., et al., Deoxyribozyme-based method for absolute quantification of N-6-methyladenosine fractions at specific sites of RNA. *J Biol Chem* **2020,** *295* (20), 6992-7000.
132. Liaqat, A.; Sednev, M. V.; Stiller, C., et al., RNA-Cleaving Deoxyribozymes Differentiate Methylated Cytidine Isomers in RNA. *Angew Chem Int Ed Engl* **2021,** *60* (35), 19058-19062.
133. Klein, A. M.; Mazutis, L.; Akartuna, I., et al., Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell* **2015,** *161* (5), 1187-1201.
134. Macosko, E. Z.; Basu, A.; Satija, R., et al., Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **2015,** *161* (5), 1202-1214.
135. Lee, J.; Hyeon, D. Y.; Hwang, D., Single-cell multiomics: technologies and data analysis methods. *Exp Mol Med* **2020,** *52* (9), 1428-1442.
136. Kruger, K.; Grabowski, P. J.; Zaug, A. J., et al., Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **1982,** *31* (1), 147-57.
137. Guerriertakada, C.; Gardiner, K.; Marsh, T., et al., The Rna Moiety of Ribonuclease-P Is the Catalytic Subunit of the Enzyme. *Cell* **1983,** *35* (3), 849-857.
138. The Nobel Prize in Chemistry 1989. https://www.nobelprize.org/prizes/chemistry/1989/summary/ (accessed 14.06.2023).
139. Cech, T. R., Structural biology - The ribosome is a ribozyme. *Science* **2000,** *289* (5481), 878-879.
140. Nissen, P.; Hansen, J.; Ban, N., et al., The structural basis of ribosome activity in peptide bond synthesis. *Science* **2000,** *289* (5481), 920-930.
141. Hutchins, C. J.; Rathjen, P. D.; Forster, A. C., et al., Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic Acids Res* **1986,** *14* (9), 3627-40.
142. Prody, G. A.; Bakos, J. T.; Buzayan, J. M., et al., Autolytic processing of dimeric plant virus satellite RNA. *Science* **1986,** *231* (4745), 1577-80.

143. Ferre-D'Amare, A. R.; Zhou, K. H.; Doudna, J. A., Crystal structure of a hepatitis delta virus ribozyme. *Nature* **1998,** *395* (6702), 567-574.

144. Sharmeen, L.; Kuo, M. Y. P.; Dintergottlieb, G., et al., Antigenomic Rna of Human Hepatitis Delta-Virus Can Undergo Self-Cleavage. *J Virol* **1988,** *62* (8), 2674-2679.

145. Wu, H. N.; Lin, Y. J.; Lin, F. P., et al., Human hepatitis delta virus RNA subfragments contain an autocleavage activity. *Proc Natl Acad Sci U S A* **1989,** *86* (6), 1831-5.

146. Buzayan, J. M.; Gerlach, W. L.; Bruening, G., Nonenzymatic Cleavage and Ligation of Rnas Complementary to a Plant-Virus Satellite Rna. *Nature* **1986,** *323* (6086), 349-353.

147. Fedor, M. J., Structure and function of the hairpin ribozyme. *Journal of Molecular Biology* **2000,** *297* (2), 269-291.

148. Klein, D. J.; Ferre-D'Amare, A. R., Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science* **2006,** *313* (5794), 1752-6.

149. Winkler, W. C.; Nahvi, A.; Roth, A., et al., Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **2004,** *428* (6980), 281-6.

150. Saville, B. J.; Collins, R. A., A site-specific self-cleavage reaction performed by a novel RNA in Neurospora mitochondria. *Cell* **1990,** *61* (4), 685-96.

151. Roth, A.; Weinberg, Z.; Chen, A. G. Y., et al., A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nat Chem Biol* **2014,** *10* (1), 56-U92.

152. Weinberg, Z.; Kim, P. B.; Chen, T. H., et al., New classes of self-cleaving ribozymes revealed by comparative genomics analysis. *Nat Chem Biol* **2015,** *11* (8), 606-10.

153. Ellington, A. D.; Szostak, J. W., In vitro selection of RNA molecules that bind specific ligands. *Nature* **1990,** *346* (6287), 818-22.

154. Tuerk, C.; Gold, L., Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **1990,** *249* (4968), 505-10.

155. Lu, X. C.; Kong, K. Y. S.; Unrau, P. J., Harmonizing the growing fluorogenic RNA aptamer toolbox for RNA detection and imaging. *Chemical Society Reviews* **2023,** *52* (12), 4071-4098.

156. Ruscito, A.; DeRosa, M. C., Small-Molecule Binding Aptamers: Selection Strategies, Characterization, and Applications. *Front Chem* **2016,** *4*.

157. Shraim, A. a. S.; Abdel Majeed, B. A.; Al-Binni, M. A., et al., Therapeutic Potential of Aptamer–Protein Interactions. *ACS Pharmacology & Translational Science* **2022,** *5* (12), 1211-1227.

158. Kelly, L.; Maier, K. E.; Yan, A., et al., A comparative analysis of cell surface targeting aptamers. *Nature Communications* **2021,** *12* (1), 6275.

159. Micura, R.; Hobartner, C., Fundamental studies of functional nucleic acids: aptamers, riboswitches, ribozymes and DNAzymes. *Chem Soc Rev* **2020,** *49* (20), 7331-7353.

160. Robertson, D. L.; Joyce, G. F., Selection Invitro of an Rna Enzyme That Specifically Cleaves Single-Stranded-DNA. *Nature* **1990,** *344* (6265), 467-468.

161. Seelig, B.; Jaschke, A., A small catalytic RNA motif with Diels-Alderase activity. *Chem Biol* **1999,** *6* (3), 167-76.

162. Seelig, B.; Keiper, S.; Stuhlmann, F., et al., Enantioselective ribozyme catalysis of a bimolecular cycloaddition reaction. *Angew Chem Int Edit* **2000,** *39* (24), 4576-+.

163. Tarasow, T. M.; Tarasow, S. L.; Eaton, B. E., RNA-catalysed carbon-carbon bond formation. *Nature* **1997,** *389* (6646), 54-7.

164. Fusz, S.; Eisenfuhr, A.; Srivatsan, S. G., et al., A ribozyme for the aldol reaction. *Chem Biol* **2005,** *12* (8), 941-50.

165. Huang, F.; Yang, Z.; Yarus, M., RNA enzymes with two small-molecule substrates. *Chem Biol* **1998,** *5* (11), 669-78.

166. Lohse, P. A.; Szostak, J. W., Ribozyme-catalysed amino-acid transfer reactions. *Nature* **1996,** *381* (6581), 442-444.

167. Jenne, A.; Famulok, M., A novel ribozyme with ester transferase activity. *Chemistry & Biology* **1998,** *5* (1), 23-34.
168. Suga, H.; Lohse, P. A.; Szostak, J. W., Structural and kinetic characterization of an acyl transferase ribozyme. *J Am Chem Soc* **1998,** *120* (6), 1151-6.
169. Zhang, B.; Cech, T. R., Peptide bond formation by in vitro selected ribozymes. *Nature* **1997,** *390* (6655), 96-100.
170. Illangasekare, M.; Sanchez, G.; Nickles, T., et al., Aminoacyl-Rna Synthesis Catalyzed by an Rna. *Science* **1995,** *267* (5198), 643-647.
171. Illangasekare, M.; Yarus, M., Specific, rapid synthesis of Phe-RNA by RNA. *P Natl Acad Sci USA* **1999,** *96* (10), 5470-5475.
172. Lee, N.; Bessho, Y.; Wei, K., et al., Ribozyme-catalyzed tRNA aminoacylation. *Nat Struct Biol* **2000,** *7* (1), 28-33.
173. Saito, H.; Kourouklis, D.; Suga, H., An in vitro evolved precursor tRNA with aminoacylation activity. *Embo J* **2001,** *20* (7), 1797-1806.
174. Doudna, J. A.; Couture, S.; Szostak, J. W., A Multisubunit Ribozyme That Is a Catalyst of and Template for Complementary Strand Rna-Synthesis. *Science* **1991,** *251* (5001), 1605-1608.
175. Sczepanski, J. T.; Joyce, G. F., A cross-chiral RNA polymerase ribozyme. *Nature* **2014,** *515* (7527), 440-+.
176. Horning, D. P.; Joyce, G. F., Amplification of RNA by an RNA polymerase ribozyme. *P Natl Acad Sci USA* **2016,** *113* (35), 9786-9791.
177. Tjhung, K. F.; Shokhirev, M. N.; Horning, D. P., et al., An RNA polymerase ribozyme that synthesizes its own ancestor. *P Natl Acad Sci USA* **2020,** *117* (6), 2906-2913.
178. McDonald, R. I.; Guilinger, J. P.; Mukherji, S., et al., Electrophilic activity-based RNA probes reveal a self-alkylating RNA for RNA labeling. *Nat Chem Biol* **2014,** *10* (12), 1049-+.
179. Krochmal, D.; Shao, Y.; Li, N. S., et al., Structural basis for substrate binding and catalysis by a self-alkylating ribozyme. *Nat Chem Biol* **2022,** *18* (4), 376-384.
180. Ameta, S.; Jaschke, A., An RNA catalyst that reacts with a mechanistic inhibitor of serine proteases. *Chem Sci* **2013,** *4* (3), 957-964.
181. Wilson, C.; Szostak, J. W., In vitro evolution of a self-alkylating ribozyme. *Nature* **1995,** *374* (6525), 777-82.
182. Sharma, A. K.; Plant, J. J.; Rangel, A. E., et al., Fluorescent RNA labeling using self-alkylating ribozymes. *ACS Chem Biol* **2014,** *9* (8), 1680-4.
183. Ghaem Maghami, M.; Dey, S.; Lenz, A. K., et al., Repurposing Antiviral Drugs for Orthogonal RNA-Catalyzed Labeling of RNA. *Angew Chem Int Ed Engl* **2020,** *59* (24), 9335-9339.
184. Ghaem Maghami, M.; Scheitl, C. P. M.; Hobartner, C., Direct in Vitro Selection of Trans-Acting Ribozymes for Posttranscriptional, Site-Specific, and Covalent Fluorescent Labeling of RNA. *J Am Chem Soc* **2019,** *141* (50), 19546-19549.
185. Flemmich, L.; Heel, S.; Moreno, S., et al., A natural riboswitch scaffold with self-methylation activity. *Nat Commun* **2021,** *12* (1), 3877.
186. Jiang, H. Y.; Gao, Y. Q.; Zhang, L., et al., The identification and characterization of a selected SAM-dependent methyltransferase ribozyme that is present in natural sequences. *Nat Catal* **2021,** *4* (10), 872-+.
187. Silverman, S. K., Catalytic DNA: Scope, Applications, and Biochemistry of Deoxyribozymes. *Trends Biochem Sci* **2016,** *41* (7), 595-609.
188. Breaker, R. R.; Joyce, G. F., A DNA enzyme that cleaves RNA. *Chem Biol* **1994,** *1* (4), 223-9.
189. Santoro, S. W.; Joyce, G. F., A general purpose RNA-cleaving DNA enzyme. *Proc Natl Acad Sci U S A* **1997,** *94* (9), 4262-6.

190. Carmi, N.; Shultz, L. A.; Breaker, R. R., In vitro selection of self-cleaving DNAs. *Chem Biol* **1996,** *3* (12), 1039-46.

191. Chandra, M.; Sachdeva, A.; Silverman, S. K., DNA-catalyzed sequence-specific hydrolysis of DNA. *Nat Chem Biol* **2009,** *5* (10), 718-720.

192. Coppins, R. L.; Silverman, S. K., Rational modification of a selection strategy leads to deoxyribozymes that create native 3 '-5 ' RNA linkages. *Journal of the American Chemical Society* **2004,** *126* (50), 16426-16432.

193. Hobartner, C.; Pradeepkumar, P. I.; Silverman, S. K., Site-selective depurination by a periodate-dependent deoxyribozyme. *Chem Commun (Camb)* **2007,** (22), 2255-7.

194. Pradeepkumar, P. I.; Hobartner, C.; Baum, D. A., et al., DNA-catalyzed formation of nucleopeptide linkages. *Angew Chem Int Edit* **2008,** *47* (9), 1753-1757.

195. Ponce-Salvatierra, A.; Boccaletto, P.; Bujnicki, J. M., DNAmoreDB, a database of DNAzymes. *Nucleic Acids Research* **2020,** *49* (D1), D76-D81.

196. Silverman, S. K.; Baum, D. A., Use of Deoxyribozymes in Rna Research. *Method Enzymol* **2009,** *469*, 95-117.

197. Fokina, A. A.; Stetsenko, D. A.; Francois, J. C., DNA enzymes as potential therapeutics: towards clinical application of 10-23 DNAzymes. *Expert Opin Biol Th* **2015,** *15* (5), 689-711.

198. Zhang, N., DNAzyme as a rising gene-silencing agent in theranostic settings. *Neural Regen Res* **2022,** *17* (9), 1989-1990.

199. Zhou, Y. F.; Zhou, C. Z., NUCLEIC ACIDS Designing in vivo active DNAzymes. *Nat Chem* **2021,** *13* (4), 299-301.

200. Victor, J.; Steger, G.; Riesner, D., Inability of DNAzymes to cleave RNA in vivo is due to limited Mg concentration in cells. *Eur Biophys J Biophy* **2018,** *47* (4), 333-343.

201. Jijakli, K.; Khraiwesh, B.; Fu, W. Q., et al., The in vitro selection world. *Methods* **2016,** *106*, 3-13.

202. Joyce, G. F., Directed evolution of nucleic acid enzymes. *Annu Rev Biochem* **2004,** *73*, 791-836.

203. Lorsch, J. R.; Szostak, J. W., Chance and necessity in the selection of nucleic acid catalysts. *Accounts Chem Res* **1996,** *29* (2), 103-110.

204. Wilson, D. S.; Szostak, J. W., In vitro selection of functional nucleic acids. *Annu Rev Biochem* **1999,** *68*, 611-647.

205. and, D. S. W.; Szostak, J. W., In Vitro Selection of Functional Nucleic Acids. *Annu Rev Biochem* **1999,** *68* (1), 611-647.

206. Kore, A. R.; Vaish, N. K.; Morris, J. A., et al., In vitro evolution of the hammerhead ribozyme to a purine-specific ribozyme using mutagenic PCR with two nucleotide analogues. *J Mol Biol* **2000,** *301* (5), 1113-21.

207. Cadwell, R. C.; Joyce, G. F., Mutagenic PCR. *PCR Methods Appl* **1994,** *3* (6), S136-40.

208. Scheitl, C. P. M.; Lange, S.; Hobartner, C., New Deoxyribozymes for the Native Ligation of RNA. *Molecules* **2020,** *25* (16).

209. Flynn-Charlebois, A.; Wang, Y. M.; Prior, T. K., et al., Deoxyribozymes with 2 '-5 ' RNA ligase activity. *Journal of the American Chemical Society* **2003,** *125* (9), 2444-2454.

210. Purtha, W. E.; Coppins, R. L.; Smalley, M. K., et al., General deoxyribozyme-catalyzed synthesis of native 3 '-5 ' RNA linkages. *Journal of the American Chemical Society* **2005,** *127* (38), 13124-13125.

211. Liaqat, A.; Stiller, C.; Michel, M., et al., N(6) -Isopentenyladenosine in RNA Determines the Cleavage Site of Endonuclease Deoxyribozymes. *Angew Chem Int Ed Engl* **2020,** *59* (42), 18627-18631.

212. Huang, F. Q.; Yarus, M., 5'-RNA self-capping from guanosine diphosphate. *Biochemistry-Us* **1997,** *36* (22), 6557-6563.

213. Holmberg, A.; Blomstergren, A.; Nord, O., et al., The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. *Electrophoresis* **2005,** *26* (3), 501-510.

214. Lorsch, J. R.; Szostak, J. W., In vitro evolution of new ribozymes with polynucleotide kinase activity. *Nature* **1994,** *371* (6492), 31-6.

215. Saran, D.; Nickens, D. G.; Burke, D. H., A trans acting ribozyme that phosphorylates exogenous RNA. *Biochemistry-Us* **2005,** *44* (45), 15007-15016.

216. Bartel, D. P.; Szostak, J. W., Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science* **1993,** *261* (5127), 1411-8.

217. Griffiths, A. D.; Tawfik, D. S., Man-made enzymes - from design to in vitro compartmentalisation. *Curr Opin Biotech* **2000,** *11* (4), 338-353.

218. Agresti, J. J.; Kelly, B. T.; Jaschke, A., et al., Selection of ribozymes that catalyse multiple-turnover Diels-Alder cycloadditions by using in vitro compartmentalization. *P Natl Acad Sci USA* **2005,** *102* (45), 16170-16175.

219. Porter, E. B.; Polaski, J. T.; Morck, M. M., et al., Recurrent RNA motifs as scaffolds for genetically encodable small-molecule biosensors. *Nat Chem Biol* **2017,** *13* (3), 295-+.

220. Zaher, H. S.; Unrau, P. J., Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *Rna* **2007,** *13* (7), 1017-1026.

221. Basu, S.; Campbell, H. M.; Dittel, B. N., et al., Purification of specific cell population by fluorescence activated cell sorting (FACS). *J Vis Exp* **2010,** (41).

222. Levy, M.; Griswold, K. E.; Ellington, A. D., Direct selection of trans-acting ligase ribozymes by in vitro compartmentalization. *Rna* **2005,** *11* (10), 1555-1562.

223. Gilbert, W., Origin of Life - the Rna World. *Nature* **1986,** *319* (6055), 618-618.

224. Pressman, A.; Blanco, C.; Chen, I. A., The RNA World as a Model System to Study the Origin of Life. *Curr Biol* **2015,** *25* (19), R953-R963.

225. Nikoomanzar, A.; Chim, N.; Yik, E. J., et al., Engineering polymerases for applications in synthetic biology. *Q Rev Biophys* **2020,** *53*.

226. Wang, Y.; Ngor, A. K.; Nikoomanzar, A., et al., Evolution of a General RNA-Cleaving FANA Enzyme. *Nat Commun* **2018,** *9* (1), 5067.

227. Wang, Y. Y.; Wang, Y.; Song, D. F., et al., An RNA-cleaving threose nucleic acid enzyme capable of single point mutation discrimination. *Nat Chem* **2022,** *14* (3), 350-+.

228. Taylor, A. I.; Wan, C. J. K.; Donde, M. J., et al., A modular XNAzyme cleaves long, structured RNAs under physiological conditions and enables allele-specific gene silencing. *Nat Chem* **2022,** *14* (11), 1295-+.

229. Culbertson, M. C.; Temburnikar, K. W.; Sau, S. P., et al., Evaluating TNA stability under simulated physiological conditions. *Bioorg Med Chem Lett* **2016,** *26* (10), 2418-2421.

230. Kratschmer, C.; Levy, M., Effect of Chemical Modifications on Aptamer Stability in Serum. *Nucleic Acid Ther* **2017,** *27* (6), 335-344.

231. Anosova, I.; Kowal, E. A.; Dunn, M. R., et al., The structural diversity of artificial genetic polymers. *Nucleic Acids Res* **2016,** *44* (3), 1007-21.

232. Dunn, M. R.; McCloskey, C. M.; Buckley, P., et al., Generating Biologically Stable TNA Aptamers that Function with High Affinity and Thermal Stability. *Journal of the American Chemical Society* **2020,** *142* (17), 7721-7724.

233. Schlegel, M. K.; Xie, X. L.; Zhang, L. L., et al., Insight into the high Duplex Stability of the Simplified Nucleic Acid GNA. *Angew Chem Int Edit* **2009,** *48* (5), 960-963.

234. Vester, B.; Wengel, J., LNA (locked nucleic acid): high-affinity targeting of complementary RNA and DNA. *Biochemistry-Us* **2004,** *43* (42), 13233-41.

235. Lind, K. E.; Mohan, V.; Manoharan, M., et al., Structural characteristics of 2 '-O-(2-methoxyethyl)-modified nucleic acids from molecular dynamics simulations. *Nucleic Acids Research* **1998,** *26* (16), 3694-3699.

236. El-Khoury, R.; Damha, M. J., 2 '-Fluoro-arabinonucleic Acid (FANA): A Versatile Tool for Probing Biomolecular Interactions. *Accounts Chem Res* **2021,** *54* (9), 2287-2297.
237. Schneider, P. N.; Olthoff, J. T.; Matthews, A. J., et al., Use of fully modified 2'-O-methyl antisense oligos for loss-of-function studies in vertebrate embryos. *Genesis* **2011,** *49* (3), 117-23.
238. Nielsen, P. E.; Egholm, M.; Berg, R. H., et al., Sequence-Selective Recognition of DNA by Strand Displacement with a Thymine-Substituted Polyamide. *Science* **1991,** *254* (5037), 1497-1500.
239. Saarbach, J.; Sabale, P. M.; Winssinger, N., Peptide nucleic acid (PNA) and its applications in chemical biology, diagnostics, and therapeutics. *Curr Opin Chem Biol* **2019,** *52*, 112-124.
240. Taylor, A. I.; Pinheiro, V. B.; Smola, M. J., et al., Catalysts from synthetic genetic polymers. *Nature* **2015,** *518* (7539), 427-30.
241. Wang, Y.; Vorperian, A.; Shehabat, M., et al., Evaluating the Catalytic Potential of a General RNA-Cleaving FANA Enzyme. *Chembiochem* **2020,** *21* (7), 1001-1006.
242. Wang, Y.; Wang, Y. Y.; Song, D. F., et al., A Threose Nucleic Acid Enzyme with RNA Ligase Activity. *Journal of the American Chemical Society* **2021,** *143* (21), 8154-8163.
243. Freund, N.; Taylor, A. I.; Arangundy-Franklin, S., et al., A two-residue nascent-strand steric gate controls synthesis of 2'-O-methyl- and 2'-O-(2-methoxyethyl)-RNA. *Nat Chem* **2023,** *15* (1), 91-100.
244. Vester, B.; Lundberg, L. B.; Sorensen, M. D., et al., LNAzymes: Incorporation of LNA-type monomers into DNAzymes markedly increases RNA cleavage. *Journal of the American Chemical Society* **2002,** *124* (46), 13682-13683.
245. Wang, Y. J.; Nguyen, K.; Spitale, R. C., et al., A biologically stable DNAzyme that efficiently silences gene expression in cells. *Nat Chem* **2021,** *13* (4).
246. Nguyen, K.; Malik, T. N.; Chaput, J. C., Chemical evolution of an autonomous DNAzyme with allele-specific gene silencing activity. *Nat Commun* **2023,** *14* (1), 2413.
247. Zelin, E.; Wang, Y. M.; Silverman, S. K., Adenosine is inherently favored as the branch-site RNA nucleotide in a structural context that resembles natural RNA splicing. *Biochemistry-Us* **2006,** *45* (9), 2767-2771.
248. Baum, D. A.; Silverman, S. K., Deoxyribozyme-catalyzed labeling of RNA. *Angew Chem Int Edit* **2007,** *46* (19), 3502-3504.
249. Buttner, L.; Javadi-Zarnaghi, F.; Hobartner, C., Site-Specific Labeling of RNA at Internal Ribose Hydroxyl Groups: Terbium-Assisted Deoxyribozymes at Work. *Journal of the American Chemical Society* **2014,** *136* (22), 8131-8137.
250. Hobartner, C.; Silverman, S. K., Engineering a selective small-molecule substrate binding site into a deoxyribozyme. *Angew Chem Int Edit* **2007,** *46* (39), 7420-7424.
251. Gunther, T., Concentration, compartmentation and metabolic function of intracellular free Mg2+. *Magnes Res* **2006,** *19* (4), 225-36.
252. Felletti, M.; Stifel, J.; Wurmthaler, L. A., et al., Twister ribozymes as highly versatile expression platforms for artificial riboswitches. *Nature Communications* **2016,** *7*.
253. Park, S. V.; Yang, J. S.; Jo, H., et al., Catalytic RNA, ribozyme, and its applications in synthetic biology. *Biotechnol Adv* **2019,** *37* (8).
254. Mandal, M.; Breaker, R. R., Gene regulation by riboswitches. *Nat Rev Mol Cell Bio* **2004,** *5* (6), 451-463.
255. Loenen, W. A., S-adenosylmethionine: jack of all trades and master of everything? *Biochem Soc Trans* **2006,** *34* (Pt 2), 330-3.
256. Burke, D. H.; Gold, L., RNA aptamers to the adenosine moiety of S-adenosyl methionine: structural inferences from variations on a theme and the reproducibility of SELEX. *Nucleic Acids Res* **1997,** *25* (10), 2020-4.

257. Price, I. R.; Grigg, J. C.; Ke, A., Common themes and differences in SAM recognition among SAM riboswitches. *Biochim Biophys Acta* **2014,** *1839* (10), 931-938.

258. Batey, R. T., Recognition of S-adenosylmethionine by riboswitches. *Wiley Interdiscip Rev RNA* **2011,** *2* (2), 299-311.

259. Sun, A.; Gasser, C.; Li, F., et al., SAM-VI riboswitch structure and signature for ligand discrimination. *Nat Commun* **2019,** *10* (1), 5728.

260. Batey, R. T., Recognition of S-adenosylmethionine by riboswitches. *Wires Rna* **2011,** *2* (2), 299-311.

261. Jimenez, R. M.; Polanco, J. A.; Luptak, A., Chemistry and Biology of Self-Cleaving Ribozymes. *Trends Biochem Sci* **2015,** *40* (11), 648-661.

262. Weinberg, C. E.; Weinberg, Z.; Hammann, C., Novel ribozymes: discovery, catalytic mechanisms, and the quest to understand biological function. *Nucleic Acids Research* **2019,** *47* (18), 9480-9494.

263. Motorin, Y.; Helm, M., RNA nucleotide methylation. *Wiley Interdiscip Rev RNA* **2011,** *2* (5), 611-31.

264. Waddell, T. G.; Eilders, L. L.; Patel, B. P., et al., Prebiotic methylation and the evolution of methyl transfer reactions in living cells. *Orig Life Evol Biosph* **2000,** *30* (6), 539-48.

265. Jadhav, V. R.; Yarus, M., Coenzymes as coribozymes. *Biochimie* **2002,** *84* (9), 877-888.

266. Frye, M.; Jaffrey, S. R.; Pan, T., et al., RNA modifications: what have we learned and where are we headed? *Nat Rev Genet* **2016,** *17* (6), 365-72.

267. Traube, F. R.; Carell, T., The chemistries and consequences of DNA and RNA methylation and demethylation. *RNA Biol* **2017,** *14* (9), 1099-1107.

268. Zaccara, S.; Ries, R. J.; Jaffrey, S. R., Reading, writing and erasing mRNA methylation. *Nat Rev Mol Cell Biol* **2019,** *20* (10), 608-624.

269. Frye, M.; Harada, B. T.; Behm, M., et al., RNA modifications modulate gene expression during development. *Science* **2018,** *361* (6409), 1346-1349.

270. Bohnsack, K. E.; Höbartner, C.; Bohnsack, M. T., Eukaryotic 5-methylcytosine (m(5)C) RNA Methyltransferases: Mechanisms, Cellular Functions, and Links to Disease. *Genes (Basel)* **2019,** *10* (2).

271. Becker, S.; Schneider, C.; Crisp, A., et al., Non-canonical nucleosides and chemistry of the emergence of life. *Nat Commun* **2018,** *9* (1), 5174.

272. Schneider, C.; Becker, S.; Okamura, H., et al., Noncanonical RNA Nucleosides as Molecular Fossils of an Early Earth-Generation by Prebiotic Methylations and Carbamoylations. *Angew Chem Int Ed Engl* **2018,** *57* (20), 5943-5946.

273. Higgs, P. G.; Lehman, N., The RNA World: molecular cooperation at the origins of life. *Nat Rev Genet* **2015,** *16* (1), 7-17.

274. Doudna, J. A.; Cech, T. R., The chemical repertoire of natural ribozymes. *Nature* **2002,** *418* (6894), 222-8.

275. Pyle, A. M., Group II Intron Self-Splicing. *Annu Rev Biophys* **2016,** *45*, 183-205.

276. Ren, A.; Micura, R.; Patel, D. J., Structure-based mechanistic insights into catalysis by small self-cleaving ribozymes. *Curr Opin Chem Biol* **2017,** *41*, 71-83.

277. Attwater, J.; Raguram, A.; Morgunov, A. S., et al., Ribozyme-catalysed RNA synthesis using triplet building blocks. *Elife* **2018,** *7*.

278. Wachowius, F.; Attwater, J.; Holliger, P., Nucleic acids: function and potential for abiogenesis. *Q Rev Biophys* **2017,** *50*, e4.

279. Tjhung, K. F.; Shokhirev, M. N.; Horning, D. P., et al., An RNA polymerase ribozyme that synthesizes its own ancestor. *Proc Natl Acad Sci U S A* **2020**.

280. Ameta, S.; Jäschke, A., An RNA catalyst that reacts with a mechanistic inhibitor of serine proteases. *Chem Sci* **2013,** *4* (3), 957-964.

281. Peselis, A.; Serganov, A., Themes and variations in riboswitch structure and function. *Biochim Biophys Acta* **2014,** *1839* (10), 908-918.

282. McCown, P. J.; Corbino, K. A.; Stav, S., et al., Riboswitch diversity and distribution. *RNA* **2017,** *23* (7), 995-1011.

283. Breaker, R. R., Riboswitches and Translation Control. *Cold Spring Harb Perspect Biol* **2018,** *10* (11).

284. Lindahl, T.; Demple, B.; Robins, P., Suicide inactivation of the E. coli O6-methylguanine-DNA methyltransferase. *EMBO J* **1982,** *1* (11), 1359-63.

285. Ghaem Maghami, M.; Dey, S.; Lenz, A. K., et al., Repurposing antiviral drugs for orthogonal RNA-catalyzed labeling of RNA. *Angew Chem Int Ed Engl* **2020**.

286. Keppler, A.; Gendreizig, S.; Gronemeyer, T., et al., A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat Biotechnol* **2003,** *21* (1), 86-9.

287. Dalhoff, C.; Lukinavicius, G.; Klimasauskas, S., et al., Direct transfer of extended groups from synthetic cofactors by DNA methyltransferases. *Nat Chem Biol* **2006,** *2* (1), 31-2.

288. Oerum, S.; Degut, C.; Barraud, P., et al., m1A Post-Transcriptional Modification in tRNAs. *Biomolecules* **2017,** *7* (1).

289. Chujo, T.; Suzuki, T., Trmt61B is a methyltransferase responsible for 1-methyladenosine at position 58 of human mitochondrial tRNAs. *RNA* **2012,** *18* (12), 2269-76.

290. Reichle, V. F.; Weber, V.; Kellner, S., NAIL-MS in E. coli Determines the Source and Fate of Methylation in tRNA. *Chembiochem* **2018,** *19* (24), 2575-2583.

291. Filonov, G. S.; Kam, C. W.; Song, W., et al., In-gel imaging of RNA processing using broccoli reveals optimal aptamer expression strategies. *Chem Biol* **2015,** *22* (5), 649-60.

292. Xiong, X.; Li, X.; Yi, C., N(1)-methyladenosine methylome in messenger RNA and non-coding RNA. *Curr Opin Chem Biol* **2018,** *45*, 179-186.

293. Grozhik, A. V.; Olarerin-George, A. O.; Sindelar, M., et al., Antibody cross-reactivity accounts for widespread appearance of m(1)A in 5'UTRs. *Nat Commun* **2019,** *10* (1), 5126.

294. Safra, M.; Sas-Chen, A.; Nir, R., et al., The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* **2017,** *551* (7679), 251-255.

295. Zhou, H.; Rauch, S.; Dai, Q., et al., Evolution of a reverse transcriptase to map N(1)-methyladenosine in human messenger RNA. *Nat Methods* **2019,** *16* (12), 1281-1288.

296. Terns, M. P.; Terns, R. M., Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. *Gene Expr* **2002,** *10* (1-2), 17-39.

297. Serganov, A.; Yuan, Y. R.; Pikovskaya, O., et al., Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. Biol.* **2004,** *11* (12), 1729-41.

298. Gilbert, S. D.; Reyes, F. E.; Edwards, A. L., et al., Adaptive ligand binding by the purine riboswitch in the recognition of guanine and adenine analogs. *Structure* **2009,** *17* (6), 857-68.

299. Scheitl, C. P. M.; Mieczkowski, M.; Schindelin, H., et al., Structure and mechanism of the methyltransferase ribozyme MTR1. *Nat Chem Biol* **2022,** *18* (5), 547-555.

300. Panchapakesan, S. S. S.; Breaker, R. R., The case of the missing allosteric ribozymes. *Nat Chem Biol* **2021,** *17* (4), 375-382.

301. Kirschning, A., Coenzymes and Their Role in the Evolution of Life. *Angew Chem Int Ed Engl* **2021,** *60* (12), 6242-6269.

302. Wilson, T. J.; Lilley, D. M. J., The potential versatility of RNA catalysis. *Wiley Interdiscip Rev RNA* **2021,** *12* (5), e1651.

303. Chen, X.; Li, N.; Ellington, A. D., Ribozyme catalysis of metabolism in the RNA world. *Chem Biodivers* **2007,** *4* (4), 633-55.

304. Breaker, R. R., Imaginary Ribozymes. *ACS Chem Biol* **2020,** *15* (8), 2020-2030.

305. Jadhav, V. R.; Yarus, M., Coenzymes as coribozymes. *Biochimie* **2002,** *84* (9), 877-88.

306. Cernak, P.; Sen, D., A thiamin-utilizing ribozyme decarboxylates a pyruvate-like substrate. *Nat Chem* **2013,** *5* (11), 971-7.

307. Tsukiji, S.; Pattnaik, S. B.; Suga, H., An alcohol dehydrogenase ribozyme. *Nat Struct Biol* **2003,** *10* (9), 713-7.

308. Ishida, S.; Terasaka, N.; Katoh, T., et al., An aminoacylation ribozyme evolved from a natural tRNA-sensing T-box riboswitch. *Nat Chem Biol* **2020**.

309. Jadhav, V. R.; Yarus, M., Acyl-CoAs from coenzyme ribozymes. *Biochemistry* **2002,** *41* (3), 723-9.

310. Serganov, A.; Keiper, S.; Malinina, L., et al., Structural basis for Diels-Alder ribozyme-catalyzed carbon-carbon bond formation. *Nat Struct Mol Biol* **2005,** *12* (3), 218-24.

311. Robertson, M. P.; Scott, W. G., The structural basis of ribozyme-catalyzed RNA assembly. *Science* **2007,** *315* (5818), 1549-53.

312. Shechner, D. M.; Grant, R. A.; Bagby, S. C., et al., Crystal structure of the catalytic core of an RNA-polymerase ribozyme. *Science* **2009,** *326* (5957), 1271-5.

313. Ponce-Salvatierra, A.; Wawrzyniak-Turek, K.; Steuerwald, U., et al., Crystal structure of a DNA catalyst. *Nature* **2016,** *529* (7585), 231-4.

314. Serganov, A.; Yuan, Y.; Pikovskaya, O., et al., Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chemistry & Biology* **2004,** *11* (12), 1729-1741.

315. Wolk, S. K.; Mayfield, W. S.; Gelinas, A. D., et al., Modified nucleotides may have enhanced early RNA catalysis. *Proc Natl Acad Sci U S A* **2020,** *117* (15), 8236-8242.

316. Robertson, M. P.; Scott, W. G., A general method for phasing novel complex RNA crystal structures without heavy-atom derivatives. *Acta Crystallogr D Biol Crystallogr* **2008,** *D64* (Pt 7), 738-44.

317. Höbartner, C.; Micura, R., Chemical synthesis of selenium-modified Oligoribonucleotides and their enzymatic ligation leading to an U6SnRNA stem-loop segment. *Journal of the American Chemical Society* **2004,** *126* (4), 1141-1149.

318. Regulski, E. E.; Breaker, R. R., In-line probing analysis of riboswitches. *Methods Mol Biol* **2008,** *419*, 53-67.

319. Gaffney, B. L.; Goswami, B.; Jones, R. A., Nitrogen-15-labeled oligodeoxynucleotides. 7. Use of nitrogen-15 NMR to probe hydrogen bonding in an O6MeG.cntdot.C base pair. *Journal of the American Chemical Society* **2002,** *115* (26), 12607-12608.

320. Batey, R. T.; Gilbert, S. D.; Montange, R. K., Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* **2004,** *432* (7015), 411-5.

321. Pikovskaya, O.; Polonskaia, A.; Patel, D. J., et al., Structural principles of nucleoside selectivity in a 2'-deoxyguanosine riboswitch. *Nat Chem Biol* **2011,** *7* (10), 748-55.

322. Keller, H.; Weickhmann, A. K.; Bock, T., et al., Adenine protonation enables cyclic-di-GMP binding to cyclic-GAMP sensing riboswitches. *RNA* **2018,** *24* (10), 1390-1402.

323. Wolter, A. C.; Weickhmann, A. K.; Nasiri, A. H., et al., A Stably Protonated Adenine Nucleotide with a Highly Shifted pKa Value Stabilizes the Tertiary Structure of a GTP-Binding RNA Aptamer. *Angew Chem Int Ed* **2017,** *56* (1), 401-404.

324. Freire, F.; Cuesta, I.; Corzana, F., et al., A simple NMR analysis of the protonation equilibrium that accompanies aminoglycoside recognition: dramatic alterations in the neomycin-B protonation state upon binding to a 23-mer RNA aptamer. *Chem Commun (Camb)* **2007,** (2), 174-6.

325. Mandal, M.; Boese, B.; Barrick, J. E., et al., Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell* **2003,** *113* (5), 577-86.

326. Noeske, J.; Richter, C.; Grundl, M. A., et al., An intermolecular base triple as the basis of ligand specificity and affinity in the guanine- and adenine-sensing riboswitch RNAs. *Proc Natl Acad Sci U S A* **2005,** *102* (5), 1372-7.

327. Wilcox, J. L.; Ahluwalia, A. K.; Bevilacqua, P. C., Charged nucleobases and their potential for RNA catalysis. *Acc Chem Res* **2011,** *44* (12), 1270-9.

328. Krishnamurthy, R., Role of pK(a) of nucleobases in the origins of chemical evolution. *Acc Chem Res* **2012,** *45* (12), 2035-44.

329. Abou Assi, H.; Rangadurai, A. K.; Shi, H., et al., 2'-O-Methylation can increase the abundance and lifetime of alternative RNA conformational states. *Nucleic Acids Res* **2020,** *48* (21), 12365-12379.

330. Chatterjee, S.; Pathmasiri, W.; Plashkevych, O., et al., The chemical nature of the 2'-substituent in the pentose-sugar dictates the pseudoaromatic character of the nucleobase (pKa) in DNA/RNA. *Org Biomol Chem* **2006,** *4* (9), 1675-86.

331. Olsen, D. B.; Benseler, F.; Aurup, H., et al., Study of a hammerhead ribozyme containing 2'-modified adenosine residues. *Biochemistry* **1991,** *30* (40), 9735-41.

332. Schubert, S.; Gul, D. C.; Grunert, H. P., et al., RNA cleaving '10-23' DNAzymes with enhanced stability and activity. *Nucleic Acids Res* **2003,** *31* (20), 5982-92.

333. Wang, Y.; Nguyen, K.; Spitale, R. C., et al., A biologically stable DNAzyme that efficiently silences gene expression in cells. *Nat Chem* **2021,** *13* (4), 319-326.

334. Xu, X.; Egger, M.; Chen, H., et al., Insights into xanthine riboswitch structure and metal ion-mediated ligand recognition. *Nucleic Acids Res* **2021,** *49* (12), 7139-7153.

335. Scheitl, C. P. M.; Okuda, T.; Adelmann, J., et al., Ribozyme-Catalyzed Late-Stage Functionalization and Fluorogenic Labeling of RNA. *Angew Chem Int Ed Engl* **2023,** *62* (31), e202305463.

336. Depmeier, H.; Hoffmann, E.; Bornewasser, L., et al., Strategies for Covalent Labeling of Long RNAs. *Chembiochem* **2021,** *22* (19), 2826-2847.

337. Flamme, M.; McKenzie, L. K.; Sarac, I., et al., Chemical methods for the modification of RNA. *Methods* **2019,** *161*, 64-82.

338. Klocker, N.; Weissenboeck, F. P.; Rentmeister, A., Covalent labeling of nucleic acids. *Chem Soc Rev* **2020,** *49* (23), 8749-8773.

339. Micura, R.; Hobartner, C., Fundamental studies of functional nucleic acids: aptamers, riboswitches, ribozymes and DNAzymes. *Chemical Society Reviews* **2020,** *49* (20), 7331-7353.

340. Nguyen, T. A.; Cigler, M.; Lang, K., Expanding the Genetic Code to Study Protein-Protein Interactions. *Angew Chem Int Edit* **2018,** *57* (44), 14350-14361.

341. Spicer, C. D.; Davis, B. G., Selective chemical protein modification. *Nature Communications* **2014,** *5*.

342. Zhang, G.; Zheng, S. Q.; Liu, H. P., et al., Illuminating biological processes through site-specific protein labeling. *Chemical Society Reviews* **2015,** *44* (11), 3405-3417.

343. Kalia, J.; Raines, R. T., Advances in Bioconjugation. *Curr Org Chem* **2010,** *14* (2), 138-147.

344. Kolmel, D. K.; Kool, E. T., Oximes and Hydrazones in Bioconjugation: Mechanism and Catalysis. *Chem Rev* **2017,** *117* (15), 10358-10376.

345. Okamoto, A.; Tainaka, K.; Saito, I., A facile incorporation of the aldehyde function into DNA: 3-formylindole nucleoside as an aldehyde-containing universal nucleoside. *Tetrahedron Lett* **2002,** *43* (26), 4581-4583.

346. Dai, Q.; He, C. A., Syntheses of 5-Formyl- and 5-Carboxyl-dC Containing DNA Oligos as Potential Oxidation Products of 5-Hydroxymethylcytosine in DNA. *Org Lett* **2011,** *13* (13), 3446-3449.

347. Sato, K.; Hirose, W.; Matsuda, A., Synthesis of 5-formyl-2'-deoxyuridine and its incorporation into oligodeoxynucleotides. *Curr Protoc Nucleic Acid Chem* **2008,** *Chapter 1*, Unit 1 21.

348. Münzel, M.; Lischke, U.; Stathis, D., et al., Improved synthesis and mutagenicity of oligonucleotides containing 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine. *Chemistry* **2011,** *17* (49), 13782-8.

349. Raindlova, V.; Pohl, R.; Sanda, M., et al., Direct polymerase synthesis of reactive aldehyde-functionalized DNA and its conjugation and staining with hydrazines. *Angew Chem Int Ed* **2010,** *49* (6), 1064-6.

350. Kromer, M.; Brunderova, M.; Ivancova, I., et al., 2-Formyl-dATP as Substrate for Polymerase Synthesis of Reactive DNA Bearing an Aldehyde Group in the Minor Groove. *Chempluschem* **2020,** *85* (6), 1164-1170.

351. Wirges, C. T.; Timper, J.; Fischler, M., et al., Controlled nucleation of DNA metallization. *Angew Chem Int Ed* **2009,** *48* (1), 219-23.

352. Mukherjee, S.; Guainazzi, A.; Scharer, O. D., Synthesis of structurally diverse major groove DNA interstrand crosslinks using three different aldehyde precursors. *Nucleic Acids Res* **2014,** *42* (11), 7429-35.

353. Romanenkov, A. S.; Ustyugov, A. A.; Zatsepin, T. S., et al., Analysis of DNA-protein interactions in complexes of transcription factor NF-kappa B with DNA. *Biochemistry-Moscow+* **2005,** *70* (11), 1212-1222.

354. Tunitskaya, V. L.; Rusakova, E. E.; Memelova, L. V., et al., Mapping of T7 RNA polymerase active site with novel reagents--oligonucleotides with reactive dialdehyde groups. *FEBS Lett* **1999,** *442* (1), 20-4.

355. Kromer, M.; Bartova, K.; Raindlova, V., et al., Synthesis of Dihydroxyalkynyl and Dihydroxyalkyl Nucleotides as Building Blocks or Precursors for Introduction of Diol or Aldehyde Groups to DNA for Bioconjugations. *Chem Eur J* **2018,** *24* (46), 11890-11894.

356. Carrette, L. L.; Gyssels, E.; Loncke, J., et al., A mildly inducible and selective cross-link methodology for RNA duplexes. *Org Biomol Chem* **2014,** *12* (6), 931-5.

357. Carrette, L. L.; Morii, T.; Madder, A., Toxicity inspired cross-linking for probing DNA-peptide interactions. *Bioconjug Chem* **2013,** *24* (12), 2008-14.

358. Wilde, J. A.; Bolton, P. H.; Mazumder, A., et al., Characterization of the Equilibrating Forms of the Aldehydic Abasic Site in Duplex DNA by O-17 Nmr. *J Am Chem Soc* **1989,** *111* (5), 1894-1896.

359. Vasseur, J. J.; Peoch, D.; Rayner, B., et al., Derivatization of Oligonucleotides through Abasic Site Formation. *Nucleos Nucleot* **1991,** *10* (1-3), 107-117.

360. Pfaffeneder, T.; Hackner, B.; Truss, M., et al., The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew Chem Int Ed* **2011,** *50* (31), 7008-12.

361. Haag, S.; Sloan, K. E.; Ranjan, N., 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.

362. Carrico, I. S.; Carlson, B. L.; Bertozzi, C. R., Introducing genetically encoded aldehydes into proteins. *Nat Chem Biol* **2007,** *3* (6), 321-2.

363. Rabuka, D.; Rush, J. S.; deHart, G. W., et al., Site-specific chemical protein conjugation using genetically encoded aldehyde tags. *Nat Protoc* **2012,** *7* (6), 1052-67.

364. Deng, J.; Wilson, T. J.; Wang, J., et al., Structure and mechanism of a methyltransferase ribozyme. *Nat Chem Biol* **2022,** *18* (5), 556-564.

365. Macon, J. B.; Wolfenden, R., 1-Methyladenosine. Dimroth rearrangement and reversible reduction. *Biochemistry* **1968,** *7* (10), 3453-8.

366. Liu, H.; Zeng, T.; He, C., et al., Development of Mild Chemical Catalysis Conditions for m(1)A-to-m(6)A Rearrangement on RNA. *ACS Chem Biol* **2022,** *17* (6), 1334-1342.

367. Erlacher, M. D.; Lang, K.; Wotzel, B., et al., Efficient ribosomal peptidyl transfer critically relies on the presence of the ribose 2'-OH at A2451 of 23S rRNA. *J Am Chem Soc* **2006,** *128* (13), 4453-9.

368. Neuner, S.; Santner, T.; Kreutz, C., et al., The "Speedy" Synthesis of Atom-Specific (15)N Imino/Amido-Labeled RNA. *Chem Eur J* **2015,** *21* (33), 11634-11643.

369. Leskovskis, K.; Zakis, J. M.; Novosjolova, I., et al., Applications of Purine Ring Opening in the Synthesis of Imidazole, Pyrimidine, and New Purine Derivatives. *Eur J Org Chem* **2021,** *2021* (36), 5027-5052.

370. Budow, S.; Seela, F., 2-azapurine nucleosides: synthesis, properties, and base pairing of oligonucleotides. *Chem Biodivers* **2010,** *7* (9), 2145-90.

371. Larsen, D.; Kietrys, A. M.; Clark, S. A., et al., Exceptionally rapid oxime and hydrazone formation promoted by catalytic amine buffers with low toxicity. *Chem Sci* **2018,** *9* (23), 5252-5259.

372. Samanta, B.; Seikowski, J.; Höbartner, C., Fluorogenic Labeling of 5-Formylpyrimidine Nucleotides in DNA and RNA. *Angew Chem Int Ed* **2016,** *55* (5), 1912-6.

373. Buback, J.; Nuernberger, P.; Kullmann, M., et al., Ring-Closure and Isomerization Capabilities of Spiropyran-Derived Merocyanine Isomers. *J Phys Chem A* **2011,** *115* (16), 3924-3935.

374. Halbritter, T.; Kaiser, C.; Wachtveitl, J., et al., Pyridine-Spiropyran Derivative as a Persistent, Reversible Photoacid in Water. *J Org Chem* **2017,** *82* (15), 8040-8047.

375. Brieke, C.; Heckel, A., Spiropyran Photoswitches in the Context of DNA: Synthesis and Photochromic Properties. *Chem-Eur J* **2013,** *19* (46), 15726-15734.

376. Hertler, J.; Slama, K.; Schober, B., et al., Synthesis of point-modified mRNA. *Nucleic Acids Research* **2022,** *50* (20).

377. Hobartner, C.; Rieder, R.; Kreutz, C., et al., Syntheses of RNAs with up to 100 nucleotides containing site-specific 2'-methylseleno labels for use in X-ray crystallography. *Journal of the American Chemical Society* **2005,** *127* (34), 12035-12045.

378. Hoernes, T. P.; Faserl, K.; Juen, M. A., et al., Translation of non-standard codon nucleotides reveals minimal requirements for codon-anticodon interactions. *Nature Communications* **2018,** *9*.

379. Rahman, M. S.; Humayun, M. Z., Nebularine (9-2'-deoxy-beta-D-ribofuranosylpurine) has the template characteristics of adenine in vivo and in vitro. *Mutat Res* **1997,** *377* (2), 263-8.

380. Ryder, S. P.; Strobel, S. A., Nucleotide analog interference mapping. *Methods* **1999,** *18* (1), 38-50.

381. Nakamura, J.; Nakamura, M., DNA-protein crosslink formation by endogenous aldehydes and AP sites. *DNA Repair* **2020,** *88*, 102806.

382. Pujari, S. S.; Wu, M.; Thomforde, J., et al., Site-Specific 5-Formyl Cytosine Mediated DNA-Histone Cross-Links: Synthesis and Polymerase Bypass by Human DNA Polymerase eta. *Angew Chem Int Ed* **2021,** *60* (51), 26489-26494.

383. Raindlova, V.; Pohl, R.; Hocek, M., Synthesis of aldehyde-linked nucleotides and DNA and their bioconjugations with lysine and peptides through reductive amination. *Chem Eur J* **2012,** *18* (13), 4080-7.

384. Ghaem Maghami, M.; Dey, S.; Lenz, A. K., et al., Repurposing Antiviral Drugs for Orthogonal RNA-Catalyzed Labeling of RNA. *Angew Chem Int Ed* **2020,** *59* (24), 9335-9339.

385. McDonald, R. I.; Guilinger, J. P.; Mukherji, S., et al., Electrophilic activity-based RNA probes reveal a self-alkylating RNA for RNA labeling. *Nat Chem Biol* **2014,** *10* (12), 1049-54.

386. Wilkinson, E.; Cui, Y. H.; He, Y. Y., Roles of RNA Modifications in Diverse Cellular Functions. *Front Cell Dev Biol* **2022,** *10*, 828683.

387. Xhemalce, B., Biological functions of RNA modifications. *Brief Funct Genomics* **2021,** *20* (2), 75-76.

388. Zhou, H. Q.; Rauch, S.; Dai, Q., et al., Evolution of a reverse transcriptase to map N-1-methyladenosine in human messenger RNA. *Nature Methods* **2019,** *16* (12), 1281-+.

389. Kumar, S.; Mohapatra, T., Deciphering Epitranscriptome: Modification of mRNA Bases Provides a New Perspective for Post-transcriptional Regulation of Gene Expression. *Front Cell Dev Biol* **2021,** *9*, 628415.

390. McIntyre, A. B. R.; Gokhale, N. S.; Cerchietti, L., et al., Limits in the detection of m6A changes using MeRIP/m6A-seq. *Sci Rep-Uk* **2020,** *10* (1), 6590.

391. Gruber, A. R.; Lorenz, R.; Bernhart, S. H., et al., The Vienna RNA websuite. *Nucleic Acids Res* **2008,** *36* (Web Server issue), W70-4.

392. Flynn-Charlebois, A.; Prior, T. K.; Hoadley, K. A., et al., In vitro evolution of an RNA-cleaving DNA enzyme into an RNA ligase switches the selectivity from 3 '-5 ' to 2 '-5 '. *Journal of the American Chemical Society* **2003,** *125* (18), 5346-5350.

393. Pluckthun, O.; Siegl, J.; Bryant, L. L., et al., Dynamic changes in DNA populations revealed by split-combine selection. *Chem Sci* **2020,** *11* (35), 9577-9583.

394. Pitsch, S.; Weiss, P. A.; Jenny, L., et al., Reliable Chemical Synthesis of Oligoribonucleotides (RNA) with 2'-O-[(Triisopropylsilyl)oxy]methyl(2'-O-tom)-Protected Phosphoramidites. *Helv. Chim. Acta* **2001,** *84* (12), 3773-3795.

395. Wachowius, F.; Höbartner, C., Probing Essential Nucleobase Functional Groups in Aptamers and Deoxyribozymes by Nucleotide Analogue Interference Mapping of DNA. *J. Am. Chem. Soc.* **2011,** *133*, 14888-14891.

396. Kosutic, M.; Neuner, S.; Ren, A. M., et al., A Mini-Twister Variant and Impact of Residues/Cations on the Phosphodiester Cleavage of this Ribozyme Class. *Angew Chem Int Edit* **2015,** *54* (50), 15128-15133.

397. Höbartner, C.; Kreutz, C.; Flecker, E., et al., The Synthesis of 2′- O -[(Triisopropylsilyl)oxy] methyl ( TOM ) Phosphoramidites of Methylated Ribonucleosides ( m1G , m2G , m22G , m1 I , m3U , m4C , m6A , m62A ) for Use in Automated RNA Solid-Phase Synthesis. *Monatshefte für Chemie / Chemical Monthly* **2003,** *134* (6), 851-873.

398. Rio, D. C., Expression and purification of active recombinant T7 RNA polymerase from E. coli. *Cold Spring Harb Protoc* **2013,** *2013* (11).

399. Mortimer, S. A.; Weeks, K. M., A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* **2007,** *129* (14), 4144-5.

400. Fujii, T.; Itaya, T.; Saito, T., Purines .18. Kinetic Studies of Base-Catalyzed Conversion of 1-Alkyladenosines into N-Alkyladenosines - Effect of Substituents on Rearrangement Rate. *Chem Pharm Bull* **1975,** *23* (1), 54-61.

401. Höbartner, C.; Kreutz, C.; Flecker, E., et al., The synthesis of 2 '-O-[(triisopropylsilyl)oxy] methyl (TOM) phosphoramidites of methylated ribonucleosides (m(1)G, m(2)G, m(2)(2)G, m(1)I, m(3)U, m(4)C, m(6)A, m(2)(6)A) for use in automated RNA solid-phase synthesis. *Monatshefte für Chemie (Chemical Monthly)* **2003,** *134* (6), 851-873.

402. Moroder, H.; Kreutz, C.; Lang, K., et al., Synthesis, oxidation behavior, crystallization and structure of 2'-methylseleno guanosine containing RNAs. *J Am Chem Soc* **2006,** *128* (30), 9909-18.

403. Kabsch, W., Xds. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 2), 125-32.
404. Evans, P. R., An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr* **2011,** *67* (Pt 4), 282-92.
405. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D., et al., Phaser crystallographic software. *J Appl Crystallogr* **2007,** *40* (Pt 4), 658-674.
406. Mieczkowski, M.; Steinmetzger, C.; Bessi, I., et al., Large Stokes shift fluorescence activation in an RNA aptamer by intermolecular proton transfer to guanine. *Nat Commun* **2021,** *12* (1), 3549.
407. Terwilliger, T. C.; Grosse-Kunstleve, R. W.; Afonine, P. V., et al., Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* **2008,** *64* (Pt 1), 61-9.
408. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **2004,** *60* (Pt 12 Pt 1), 2126-32.
409. Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N., et al., Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **2012,** *68* (Pt 4), 352-67.
410. Lebedev, A. A.; Young, P.; Isupov, M. N., et al., JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr D Biol Crystallogr* **2012,** *68* (Pt 4), 431-40.
411. Moriarty, N. W.; Grosse-Kunstleve, R. W.; Adams, P. D., electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D Biol Crystallogr* **2009,** *65* (Pt 10), 1074-80.
412. Nawale, G. N.; Gore, K. R.; Hobartner, C., et al., Incorporation of 4 '-C-aminomethyl-2 '-O-methylthymidine into DNA by thermophilic DNA polymerases. *Chem Commun* **2012,** *48* (77), 9619-9621.
413. Research, G. Extinction coefficients and fluorescence data. https://www.glenresearch.com/media/folio3/productattachments/technical_bulletin/Extinctions_20210112.pdf.