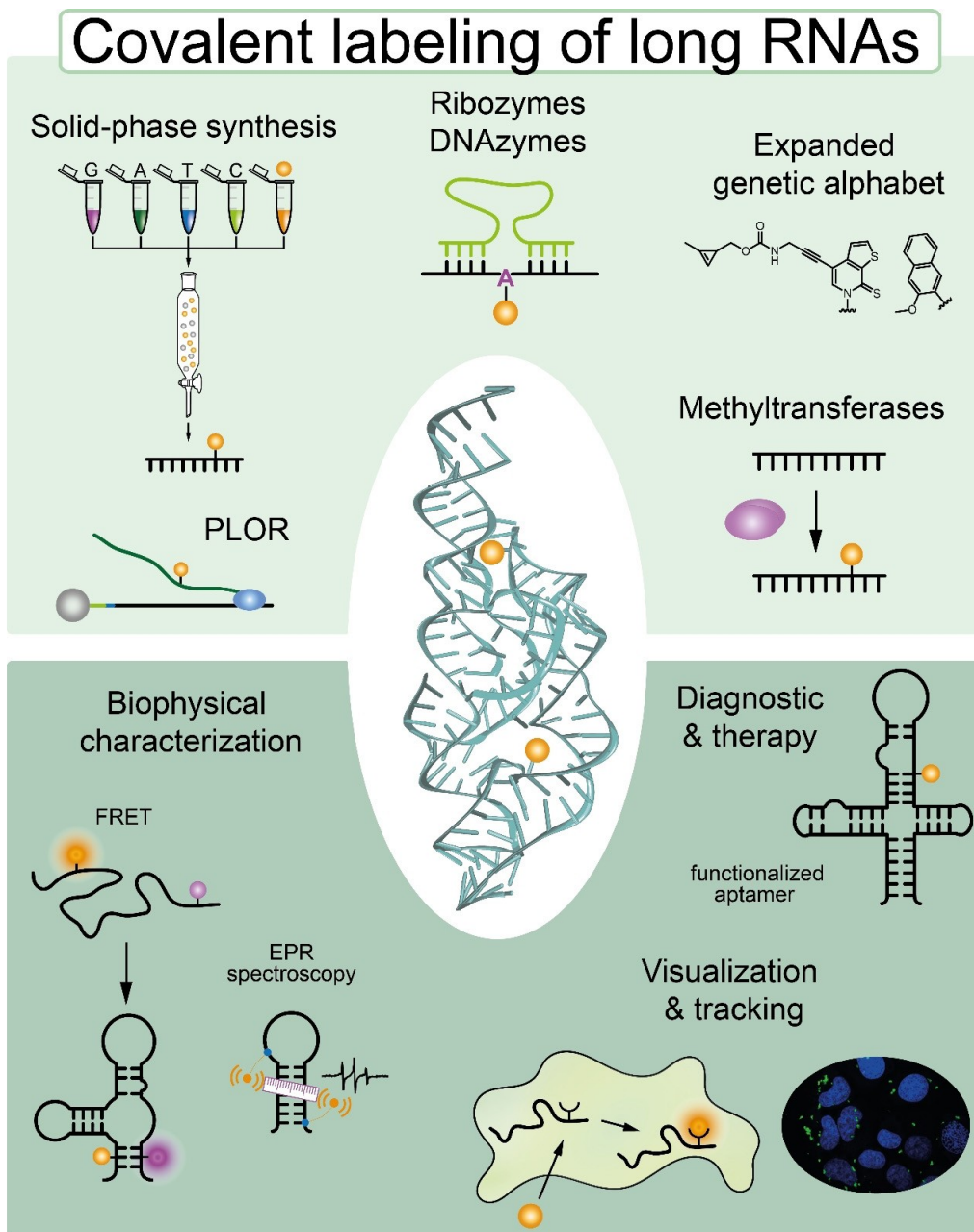


Strategies for Covalent Labeling of Long RNAs

Hannah Depmeier,^[a] Eva Hoffmann,^[a] Lisa Bornewasser,^[a] and Stephanie Kath-Schorr^{*[a]}



The introduction of chemical modifications into long RNA molecules at specific positions for visualization, biophysical investigations, diagnostic and therapeutic applications still remains challenging. In this review, we present recent approaches for covalent internal labeling of long RNAs. Topics included are the assembly of large modified RNAs via enzymatic ligation of short synthetic oligonucleotides and synthetic biology approaches preparing site-specifically modified RNAs

via *in vitro* transcription using an expanded genetic alphabet. Moreover, recent approaches to employ deoxyribozymes (DNAzymes) and ribozymes for RNA labeling and RNA methyltransferase based labeling strategies are presented. We discuss the potentials and limits of the individual methods, their applicability for RNAs with several hundred to thousands of nucleotides in length and indicate future directions in the field.

1. Introduction

A large fraction of naturally occurring non-coding ribonucleic acids (RNAs), such as long-noncoding (lnc) RNAs, possess ≥ 200 nucleotides in length.^[1] To investigate structure and function of such large RNA molecules *in vitro* and in cells, selective labeling at specific positions within their sequence is an essential prerequisite for various biophysical investigations. This review discusses numerous approaches for the preparation of long RNA molecules with internal site-specific modifications. Only methods applicable for the introduction of reporter groups within the sequence of long RNAs (several hundred nucleotides) are presented (Table 1). Widely used strategies for 5'- and 3'-end labeling of RNAs are not discussed in detail.

Different strategies exist to introduce reporter groups for internal, covalent and site-directed labeling of RNA. Reporter groups can either be directly introduced into RNA during chemical solid-phase synthesis^[2–5] or co-transcriptionally^[6–9] using the corresponding modified phosphoramidite or triphosphate building block, respectively. Stability of reporter groups is often challenged under harsh reaction conditions used in chemical RNA synthesis.^[10,11] Furthermore, the selectivity of RNA polymerases^[6,12] limits the scope of reporter groups that can be inserted co-transcriptionally into RNA. Thus, alternative approaches based on post-synthetic or post-transcriptional RNA labeling are often employed.^[10,13–15]

For the implementation of a two-step RNA labeling approach, a small reactive group is introduced first during chemical solid-phase synthesis or during *in vitro* transcription into RNA.^[14] This allows post-synthetic or post-transcriptional installation of virtually any reporter group, separating the attachment reaction from the site selection step.^[10,16,17,18] A further advantage is the small size of the molecule handle which in most cases leads to only minor disturbance of nucleic acid structures, thus allowing natural function and folding of the RNA oligonucleotide.^[19]

Reactions for post-synthetic and post-transcriptional labeling include periodate chemistry, amine chemistry, thiol chemistry and click chemistry.^[14]

Among these, especially *N*-hydroxysuccinimid (NHS) chemistry has been extensively used for nucleic acid labeling.^[20–22] For this purpose, a small amine-functionality is introduced into the oligonucleotide enabling post-synthetic or post-transcriptional reaction with an NHS-ester bearing a reporter group by forming an amido-linkage.^[20–22]

Besides this, click chemistry combines the most robust and efficient chemistries for labeling of RNA. The term click chemistry was first defined by Sharpless *et al.*^[23] referring to reactions joining together units via heteroatom links with high yields and negligible formation of byproducts under simple reaction conditions and solvents. When using click chemistry to label RNA, fast reaction kinetics, biocompatibility and bioorthogonality are key prerequisites.^[23,24]

For the installation of reporter groups, the classical copper-catalyzed azide-alkyne cycloaddition (CuAAC) forming a triazole-linkage has widely been used for DNA labeling.^[25] However, its direct use for RNA labeling was limited at first.^[13] Since copper ions can catalyze cleavage of the phosphodiester backbone in RNA, it was a challenge to ensure that RNA remains intact under CuAAC reaction conditions.^[13] Addition of Cu(I) stabilizing ligands or small amounts of acetonitrile as a co-

Table 1. Overview of methods to prepare long site-specifically labeled RNA presented in this review.

Chapter	Methods to prepare long site-specifically labeled RNA
2	Preparation of site-specifically modified long RNAs combining solid-phase synthesis and ligation
2.1.1	Ligation of RNA with T4 RNA ligase I
2.1.2	Splinted ligation of RNA with T4 DNA ligase II
2.1.3	Ligation of RNA with T4 DNA ligase
2.2.1	DNAzyme-mediated ligation of RNA
2.2.2	Ribozyme-mediated ligation of RNA
3	Synthetic biology approaches for covalent RNA labeling during <i>in vitro</i> transcription
3.1.1	Hydrogen-bonding unnatural base pairs
3.1.2	Hydrophobic unnatural base pairs
3.2.1	Transcriptional priming and enzymatic ligation
3.2.2	Position selective labeling of RNA (PLOR)
4	Post-synthetic and post-transcriptional labeling of RNAs
4.1.1	DNA-templated labeling via four-way junction formation
4.1.2	DNA-templated chemical labeling via duplex formation
4.1.3	RNA acylation at induced loops (RAIL)
4.2	Deoxyribozyme-catalyzed labeling of RNA (DECAL)
4.3	Ribozyme-catalyzed labeling of RNA
4.3.3	Methyltransferase ribozymes for RNA labeling
4.4	Methyltransferase-directed transfer of activated groups (mTAG)

[a] H. Depmeier, E. Hoffmann, L. Bornewasser, Prof. Dr. S. Kath-Schorr
 University of Cologne, Department of Chemistry
 Greinstr. 4, 50939 Cologne (Germany)
 E-mail: skathsch@uni-koeln.de

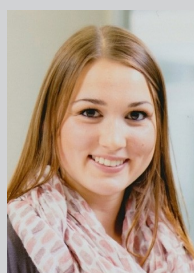
© 2021 The Authors. ChemBioChem published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

solvent during CuAAC click reactions minimized RNA degradation.^[18,26] For intracellular applications, cell toxicity of copper ions hampers the use of CuAAC chemistry for live cell RNA labeling.^[27] Therefore, copper-free variants of click reactions are indispensable for RNA labeling especially in a cellular context. Azide-alkyne cycloadditions can also be promoted by ring strain instead of copper catalysts. The so called strain promoted azide-alkyne cycloaddition (SPAAC) provides copper free conditions for the reaction between azides and alkynes.^[28] However, comparably slow reaction kinetics are detrimental for this reaction.^[29] Other bioorthogonal labeling techniques, such as inverse-electron-demand Diels Alder (IEDDA) cycloaddition reactions recently gained much attention due to fast reaction kinetics, high selectivity and high yields in aqueous media.^[30] Here, dienes such as tetrazine derivatives rapidly react with strained alkene moieties. For visualization purposes, reporter groups such as fluorophores can be coupled to tetrazines. Notably, in combination with the quenching properties of tetrazine derivatives on adjacent fluorophores emitting in the wavelength of green and red light between 400 nm and 600 nm (e.g. Tetrazine-Bodipy FL, Tetrazine-Oregon green 488, Tetrazine-BODIPY TMR-X)^[31,32], fluorescence strongly increases upon cycloaddition reactions, reducing background fluorescence and thus is a valuable tool for intracellular applications.^[31,32] Since this reaction does not require a copper catalyst, non-toxic as well as non-denaturing reaction conditions are provided, making the IEDDA reaction suitable for

investigation of complexly folded RNA and for *in cell* applications.^[19,33]

2. Preparation of Site-Specifically Modified Long RNAs Combining Solid-Phase Synthesis and Ligation

Short RNAs with internal, site-specific modifications can be prepared by chemical synthesis of RNA. By now, the most common approach for chemical synthesis of RNA is performed on a solid support and requires nucleoside 3'-phosphoramidite derivatives.^[34,35] Using the corresponding phosphoramidite building blocks, not only canonical nucleotides, but also a plethora of differently modified nucleotides can be introduced into RNA at predefined positions in a quantitative manner.^[34] Reporter groups such as fluorophores^[2] or stable paramagnetic centers, termed spin-labels^[3-5] can directly be introduced into RNA via solid-phase synthesis. Another strategy is the insertion of a small functional moiety during solid-phase synthesis allowing post-synthetic installation of virtually any desired reporter group using click chemistry.^[18] However, a major drawback of solid-phase synthesis of RNA is the limitation in length for RNA generated by routine chemical synthesis. Whereas DNA solid-phase synthesis with over 100 nucleotides (nt) is a by now well established and efficient procedure, reliable chemical synthesis of RNA oligonucleotides appeared to



Hannah Depmeier completed her training-integrated B.Sc. in Pharmaceutical Chemistry at the Cologne University of Applied Sciences in 2017, including an apprenticeship as a chemical laboratory assistant at Grünenthal GmbH (Aachen, Germany). Broadening her scope of studies, she performed an internship at the University of British Columbia (Vancouver, Canada) in Prof. Mark MacLachlan's group. In 2019, she received her M.Sc. in Biochemistry from the University of Bonn under the guidance of Prof. Stephanie Kath-Schorr and was honored with the GBM Master Award. Currently, she is pursuing her PhD in Biochemistry in Prof. Kath-Schorr's group at the University of Cologne, focusing on artificial nucleic acids with increased information storage.



Eva Hoffmann obtained her bachelor's degree in Forensic Sciences at the Bonn-Rhein-Sieg University of Applied Sciences (Germany) in 2017, including a semester abroad at Abertay University (Dundee, UK). Subsequently, she conducted her M.Sc. in Biochemistry at the University of Bonn in 2019 with Prof. Stephanie Kath-Schorr as mentor. Since then, she is working towards her PhD in the group of Prof. Kath-Schorr at the University of Cologne. In her current research, she is investigating polymerase processing of an expanded genetic alphabet.



Lisa Bornewasser graduated from the University of Bonn (Rheinische Friedrich-Wilhelms-Universität, Germany) earning her bachelor's and master's degree in chemistry. Focusing on the field of biochemistry, she joined the group of Professor Stephanie Kath-Schorr in 2017 to accomplish her master's thesis. She is now pursuing her PhD in the Kath-Schorr group where she explores new chemical labeling strategies expanding the genetic alphabet using unnatural nucleobases for RNA modification.



Stephanie Kath-Schorr obtained her PhD in the group of Prof. Dr. T. Carell in 2010. After a postdoctoral stay in the group of Prof. Dr. D. Lilley, Dundee, UK, she returned to Germany starting her independent scientific career as Liebig-Fellow (Fonds der Chemischen Industrie) at the LIMES Institute, University of Bonn in 2013. In 2018, she was awarded with the Plus 3 Perspectives Programme by the Boehringer Ingelheim Foundation. In 2020 she was appointed Professor of Organic Chemistry at the University of Cologne, Germany. Her research interests lie in the synthesis and analysis of chemically-modified ribonucleic acids. Her group develops artificial ribonucleic acids assembled from synthetic building blocks which exhibit a range of new functions ranging from detection of biologically active molecules to ribonucleic acids with catalytic activity.

be more difficult^[36,37] and is commonly restricted to a limited length of approximately 100 nucleotides.^[38] The presence of the reactive 2'-hydroxyl group of the ribose sugar requires an additional selective protecting group not necessary in solid-phase DNA synthesis.^[36,37] The 2'-hydroxyl protecting group increases steric hindrance while simultaneously lowering efficiency during solid-phase synthesis.^[36,37] Additionally, side reactions during the deprotection step as well as premature loss of the 2'-hydroxyl protecting group appear.^[37] These side reactions include hydrolytic phosphodiester backbone cleavage and 2'-3' migration resulting in 5'-2' internucleotide linkages in long RNA oligonucleotides.^[37,39] Although, ongoing advancements in RNA chemical synthesis allowed preparation of relatively long RNA with good efficiencies^[34,40] biologically relevant length of RNA is still difficult to achieve^[37] as most ribozymes,^[41,42] riboswitches^[43] or aptamers^[44] exceed the length of possible strand length that can be prepared by solid-phase RNA synthesis. Therefore, there is a need for alternatives allowing site-specific labeling of long RNAs to investigate their structure and functions.

During the last decades, strategies were developed that combine the site-specific introduction of chemically modified nucleotides during solid-phase synthesis with methods that facilitate reliable generation of longer RNA oligonucleotides

bearing modifications. These strategies mostly rely on the chemical synthesis of shorter, modified RNA strands followed by subsequent ligation to yield the desired long RNA bearing an internal site-specific functionalization. To date, these ligation strategies can be subdivided into three different approaches: enzymatic ligation,^[45–48] ligation of RNA using ribozymes^[49,50] or DNAzymes^[51–53] and chemical ligation.^[54,55]

2.1. Enzymatic ligation of synthetic oligoribonucleotides

Enzymatic ligation is one strategy to post-synthetically overcome the size limitation of a purely chemical approach to prepare site-specifically modified RNA by solid-phase synthesis.^[56] Ligation of RNA is primarily performed using bacteriophage T4 RNA ligase or T4 DNA ligase.

2.1.1. Ligation of RNA with T4 RNA ligase I

Most commonly, T4 RNA ligase I is used to join the ends of two single stranded RNAs as single stranded substrates are preferred by the enzyme (Figure 1A).^[57] Secondary structure formation (e.g. hairpin) should bring single stranded termini of acceptor

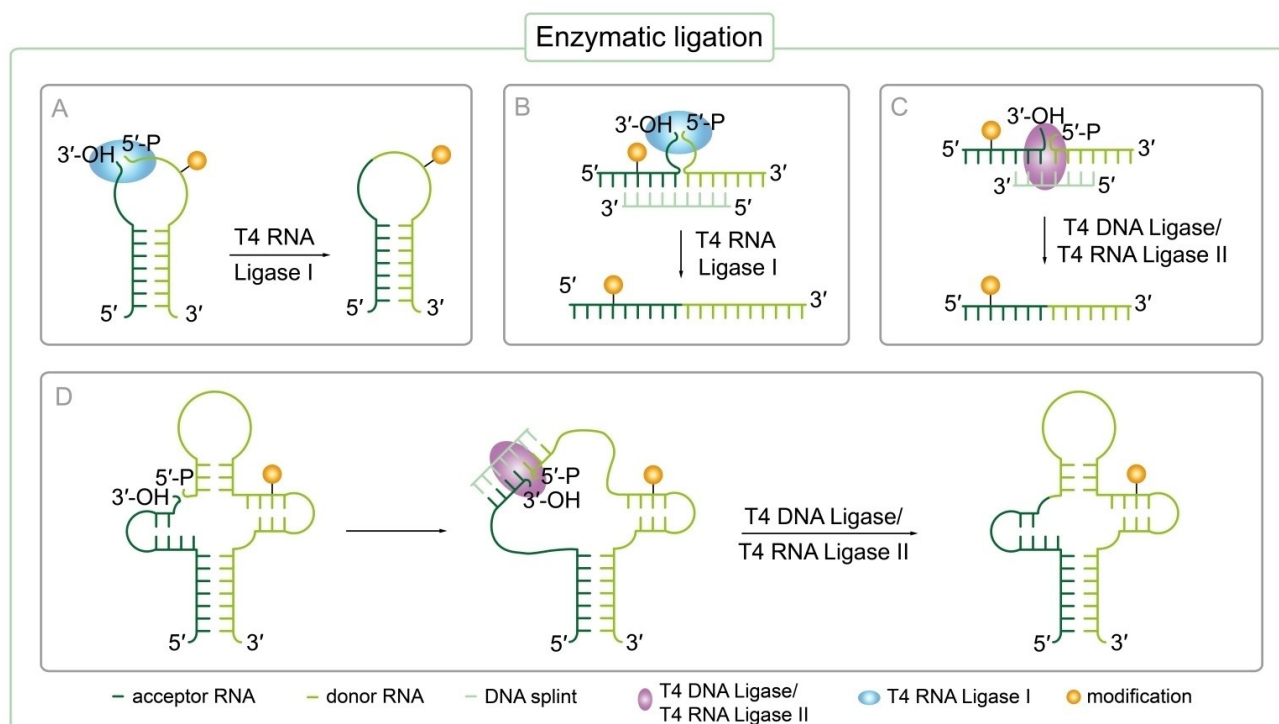


Figure 1. Enzymatic ligation strategies to prepare long modified RNA oligonucleotides beyond the size limit of chemical solid-phase synthesis. Small, site-specifically modified RNA strands prepared by RNA solid-phase synthesis can be ligated using different enzymes to generate long, site-specifically labeled RNA. The acceptor sequence is bearing a terminal 3'-OH, while the donor RNA carries a 5'-triphosphate. A) T4 RNA ligase I mediated ligation of two single stranded RNA oligonucleotides brought in close proximity due to secondary structure formation. B) T4 RNA ligase I mediated ligation of two single stranded RNAs brought in close proximity by a DNA splint. C) Templated T4 DNA ligase/T4 RNA ligase II mediated ligation of two RNA strands. A splint is used to form the necessary ternary pre-ligation complex and to bring ends to be ligated in close proximity. D) Templated T4 DNA ligase/T4 RNA ligase II mediated ligation of two RNA strands. A splint is used to break up secondary structure formation enabling the enzyme to access the ligation site and to form the necessary ternary pre-ligation complex.

and donor RNA into an entropically favored position to promote ligation (Figure 1A).^[45] Suddala *et al.* performed ligation with T4 RNA ligase I combining two shorter chemically synthesized RNA fragments.^[58] The donor RNA strand was modified with a Cyanine(Cy)5 label that was introduced during solid-phase synthesis to produce site-specifically fluorescently-labeled 75 nt long tRNA^{Gly}. Thereby, single molecule Förster resonance energy transfer (smFRET) investigation of the structure dynamics of the *glyQS* T-box riboswitch from *Bacillus subtilis* was enabled.^[58] The *glyQS* T-box riboswitch regulates expression of the gene encoding glycyl-tRNA synthase depending on the ratio of aminoacylated tRNA^{Gly}s to non-aminoacylated tRNA^{Gly}s in the cell.^[59] The *glyQS* T-box riboswitch itself was prepared by *in vitro* transcription (IVT) and visualized by hybridization of a Cy3-labeled LNA oligonucleotide to the 5' end of the RNA. Applying single smFRET they analyzed *glyQS* T-box riboswitch interaction with Cy5-labeled tRNA^{Gly} indicating a complex hierarchical sensing mechanism that reads out the tRNAs aminoacylation status by snap-lock-based trapping upon binding of the tRNA.^[58]

However, side products arising from circularization, oligomerization of RNA strands^[48,60] or repeated ligation of already formed ligation product^[61] frequently appear using T4 RNA ligase I.^[47] Therefore, it is recommended to protect the 3'-OH of the phosphor donor.^[47] Alternatively, T4 RNA ligase II can be used for RNA ligation as ligation selectivity is promoted by a splint.^[57]

2.1.2. Splinted ligation of RNA with T4 RNA ligase II

Ligation of RNA with T4 RNA ligase II makes use of a splint that serves as a bridging or guiding DNA oligonucleotide perfectly complementary to the ends of RNA strands to be joined by the enzyme increasing selectivity of this approach (Figure 1C).^[57] Using this templated ligation strategy with T4 RNA ligase II, Manz *et al.* constructed a variety of full-length 169 nt *Bacillus subtilis* *ytJ* S-adenosyl-L-methionine (SAM)-I riboswitch variants bearing FRET-labels at selected specific positions known to be affected by structural changes of the riboswitch. The SAM-I riboswitch is involved in gene expression regulation via transcription termination in *B. subtilis*. Therefore, five chemically synthesized RNA oligonucleotides bearing 5-[3-[(6-aminohexyl)amino]-3-oxo-1-propenyl]-uridine at distinct, predefined positions were post-synthetically labeled using NHS-esters of Cy3 and Cy5 prior to the ligation step. By that, the group investigated the conformational energy landscape dependent on Mg²⁺ and SAM ligand concentration of the full-length SAM-I riboswitch *in vitro* by smFRET microscopy. The authors discovered that terminator and antiterminator fold of this riboswitch coexist and are highly affected by Mg²⁺ ions. SAM ligand binding only induces a small shift towards the terminator state.^[62]

Using a similar approach, Weinrich *et al.* prepared a site-specifically spin-labeled full-length 59 mer long human immunodeficiency virus(HIV)-1 trans activation response (TAR) RNA^[63] required for trans-activation of the viral promoter and thus HI

virus replication^[64]. The site-specific introduction of spin-labels into RNA allows obtaining distance information between these two spin labels via pulsed electron paramagnetic resonance (EPR) techniques. Therefore, three short RNA sequences were synthesized via solid-phase synthesis introducing two 2-nitrobenzyloxymethyl protected 2,2,6,6-tetramethyl-piperidinyloxy (TEMPO) derivatives via the corresponding cytidine phosphoramidite followed by enzymatic ligation in one pot using T4 RNA ligase II promoted by DNA splints. The 2-nitrobenzyloxymethyl protective group is stable under harsh RNA synthesis conditions avoiding decomposition of the nitroxide spin label during RNA synthesis. Furthermore, the protecting group prevents reduction of the spin label during ligation. This is of utmost importance as ligases usually require thiols to ensure activity which are known to reduce nitroxide spin labels. Subsequent deprotection to recover the nitroxide group occurs by photolysis without damaging the RNA. Pulsed electron-electron double resonance (PELDOR) experiments in the absence and presence of arginine amide, a TAR RNA ligand, were performed to study structure dynamics and showed an increase in rigidity of TAR RNA structure upon ligand binding.^[63]

2.1.3. Ligation of RNA with T4 DNA ligase

If the ligation site is located within a highly structured region of the RNA or RNA donor and acceptor cannot form a stable pre-ligation complex themselves, ligation can alternatively be performed using T4 DNA ligase to increase ligation efficiency.^[47] As T4 RNA ligase II, T4 DNA ligase uses a ternary complex as a substrate: two RNA strands hybridized to a splint (Figure 1C).^[56] The templating DNA splint not only brings the 3'-OH of the acceptor and the 5'-phosphate of the donor RNA in close proximity^[56] but also provides good accessibility to the ligation site for the enzyme by destroying secondary structure formation of RNA strands to support ligation (Figure 1D).^[47] Using T4 DNA ligase, side product formation is reduced and efficiency is increased by templated ligation compared to T4 RNA ligation without a splint.^[56] Ligation of RNA using T4 DNA ligase in combination with a splint was first performed by Moore and Sharp in 1992 to produce RNA bearing an internal, site-specific modification.^[60] More recently, Esquiaqui *et al.* used a similar approach to prepare long RNA with position specific modifications enabling electron paramagnetic resonance line shape analysis and distance measurements. Thereby, site-specific changes in conformational dynamics of the kink-turn motif in the *Vibrio cholerae* (VC) glycine riboswitch upon K⁺, Mg²⁺ and glycine induced folding were investigated.^[38] Therefore, the group constructed a 232 nt long site-specifically spin-labeled sequence by ligating a small modified 20 nt oligoribonucleotide with a 212 mer RNA oligonucleotide prepared by IVT followed by splint-mediated ligation using T4 DNA ligase. The small 20 mer RNA was prepared by solid-phase synthesis enabling site-directed incorporation of either one 4-thiouridine or a double phosphorothioate backbone modification, followed by post-synthetic spin-labeling with either 3-(2-iodoacetamido)-PROXYL or 1-oxyl-2,2,5,5-tetramethylpyrroline, respectively. This

work demonstrates that local mobility of the spin labels located in the kink-turn motif are influenced by environmental changes such as an increase in K^+ concentration leading to a stimulated folding of the kink-turn fold of VC riboswitch that is not further stabilized by Mg^{2+} .^[38] Furthermore, it is shown that binding of the glycine ligand to the aptamer domain of the riboswitch is not influencing the local structure of the kink-turn motif although ligand binding is likely to induce structural changes in other regions of the riboswitch.^[38]

Using an analogous approach, Chen *et al.* inserted an isotopic label at a predefined position via solid-phase synthesis using an (8,1'-¹³C)-adenosine phosphoramidite building block in combination with splinted T4 DNA ligation to prepare the 96 nt long C–C chemokine receptor type 5 (CCR5) mRNA fragment for nuclear magnetic resonance (NMR) spectroscopy studies.^[65] CCR5 is proposed to take part in facilitation of viral entry of HIV-1 into human cells.^[66] In CCR5 mRNA, a pseudoknot stimulates -1 ribosomal frameshifting upon interaction with a microRNA (miRNA), so called miRNA-1224, suggesting gene expression control for HIV-1 viral replication.^[67] The labeled CCR5 mRNA fragment was used to measure structural dynamics of the functional RNA via NMR upon CCR5 RNA/miRNA-1224 complex formation, overcoming the problem of spectral overlap in RNA structure elucidation and interaction analysis using NMR.^[65]

Nonetheless, ligation using T4 RNA ligase I is often preferred due to easier purification resulting in higher yields^[47]. Furthermore, less enzyme is needed as T4 DNA ligase only slowly releases formed product and is principally slow and inefficient.^[68]

2.1.4. Splint-supported ligation of RNA with T4 RNA ligase I

In the same way as T4 DNA ligase uses a ternary RNA:RNA/DNA complex as a substrate, T4 RNA ligase I can also be used with a splint mimicking the natural ligation junction to improve sequence specificity of ligation with T4 RNA ligase I (Figure 1B). This approach was inspired by the observation, that the natural substrate of T4 RNA ligase I is the anticodon loop of tRNA^{Lys}.^[69] Hence, a DNA splint is designed in a way to hold the donor and acceptor close to each other while leaving single-stranded ends at the ligation junction.^[70] However yields of this reaction remained low most probably due to intramolecular circularization and non-optimal linker length.^[68] Nonetheless, ongoing improvements of this strategy especially regarding linker length significantly improved yield and efficiency of this reaction. Stark *et al.* created a 128 nt long RNA strand using a splint and T4 RNA ligase I.^[68] This strategy could be interesting for ligation of smaller synthetically prepared, modified RNA oligonucleotides to produce long site-specifically modified RNA to study structure, function and dynamics.

2.2. Ribozymes and DNAzymes for RNA ligation

In recent years, alternative strategies besides enzymatic ligation were developed to prepare long RNA with site-specific labels.

This includes on the one hand DNAzymes, catalytically active DNA molecules, which are able to ligate smaller RNA fragments to form long RNA oligonucleotides^[51,71] and on the other hand ribozymes which can mediate an exchange of a small, predefined patch of a long RNA with a synthetically modified small RNA via a cleavage-ligation cascade^[50].

2.2.1. DNAzyme mediated ligation of RNA

DNAzymes are artificial single stranded DNA oligonucleotides that are selected *in vitro* by directed evolution and can catalyze specific reactions, e.g. nucleic acid cleavage or ligation.^[72] The first reported DNAzyme possessing an RNA ligase activity was discovered by the working group of Scott Silverman in 2003.^[73] However, in an Mg^{2+} mediated manner, unnatural branched 2'-5' phosphodiester linkages were formed between a 2',3'-cyclic phosphate and a 5'-hydroxyl group of two RNA oligonucleotides by this DNAzyme.^[73,74] Further research in this area finally yielded DNAzymes with RNA ligase activity forming native 3'-5' linkages between two RNA strands, one bearing a 2',3'-diol and the other a 5'-triphosphate.^[51] Such DNAzymes are generally designed with two variable binding arms at its 3'- and 5'-ends complementary to the parts of the two substrate RNA strands close to the ligation site (Figure 2).^[73] Thereby, the DNAzyme is not only a catalytic unit, but at the same time represents a splint bringing the two ends of the RNA strands to be ligated in close vicinity.^[52] Büttner *et al.* used a DNAzyme catalyzed ligation to produce the site-specifically spin-labeled long SAM-III and SAM-I riboswitch domains (53 nt and 118 nt, respectively) in high yields. For this purpose, the *in vitro* selected 9DB1* deoxyribozyme was used that catalyzes the formation of native 3'-5' phosphodiester linkages. The 5'-triphosphorylated donor RNA was prepared by IVT and the acceptor RNA was prepared by solid-phase synthesis. A convertible nucleoside approach was used to generate 2,2,6,6-tetramethylpiperidin-1-oxyl (TEMPO) or 2,2,5,5-tetramethylpyrrolidin-1-oxyl (proxyl) spin-labeled RNA. Therefore, convertible *O*⁴-(4-chlorophenyl)uridine nucleosides were site-specifically introduced into RNA during solid-phase synthesis. Upon post-synthetic substitution of the *O*⁴-chlorophenyl leaving group with 4-amino-TEMPO or 3-amino-proxyl, C^{TEMPO}- and C^{proxyl}-modified RNA was obtained. Advantageously, the spin label integrity was not affected under the presented conditions for deoxyribozyme-catalyzed ligation with 9DB1* as usage of dithiothreitol (DTT) was avoided which is in turn required for enzymatic ligation. As DTT is known to reduce TEMPO spin labels, 9DB1*-mediated ligation sets a new benchmark for EPR analysis of long modified RNA. Using continuous wave (CW)-EPR, dynamics and secondary structure formation of the SAM riboswitch domains were analyzed.^[53,71]

2.2.2. Ligation of RNA with ribozymes

Apart from deoxyribozymes owning RNA ligation activity, also naturally occurring ribozymes with similar ligation activity are known^[41] and several even more efficient ribozymes were

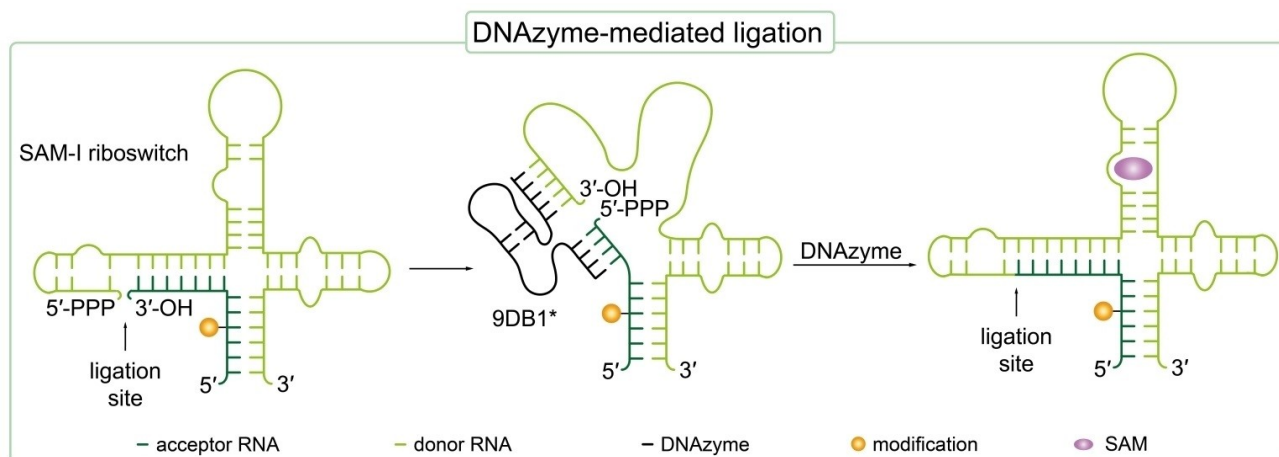


Figure 2. 9DB1* deoxyribozyme-mediated ligation to prepare a 118 nt long, site-specifically spin-labeled SAM-I riboswitch from two smaller RNA strands. The modification is inserted during chemical solid-phase synthesis of RNA. The acceptor sequence is bearing a terminal 3'-OH, while the donor RNA carries a 5'-triphosphate.^[53,71] The ligation site is indicated with an arrow.

selected *in vitro* by directed evolution.^[75] However, besides the directed evolution also rational design of such ribozymes is possible if the parental RNA structure and mechanism of action is well studied.^[49] Using the latter strategy, the group of Sabine Müller developed a ribozyme performing a cleavage/ligation cascade based on the naturally occurring hairpin ribozyme allowing site-directed RNA functionalization (Figure 3).^[50] Tandem duplication of two hairpin ribozymes yielded the so called twin ribozyme with double cleavage as well as double ligation activity which allows site-specific sequence alteration.^[50] The twin ribozyme was used to mediate an exchange of a predetermined RNA sequence in a 145 nt long oligoribonucleo-

tide with another small amino-modified RNA. The modified RNA was prepared via solid-phase synthesis introducing an amino-modified deoxythymidine building block. This allowed subsequent functionalization of the amino-modification using either isothiocyanate or *O*-succinimidyl ester derivatives of various fluorophores as well as biotin.^[50] Using the twin ribozyme, a dye-labeled 145 mer RNA could be obtained with up to 53% yield. By this method, virtually any reporter group could be attached e.g. fluorophores, isotopes and spin labels allowing FRET, NMR or EPR studies of functional RNAs beyond the limitation in length for RNA solid-phase synthesis.^[50]

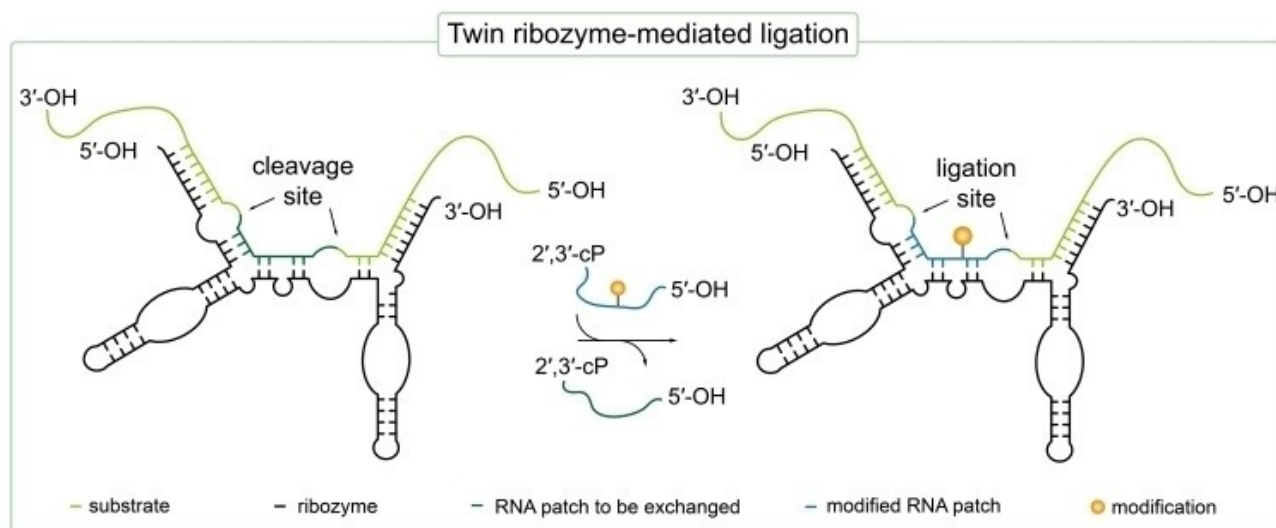


Figure 3. Cleavage/ligation cascade performed by an engineered twin ribozyme (black). A predefined patch of RNA (dark green) in the RNA substrate (light green) is exchanged with a small, synthetically prepared and modified RNA sequence (blue). At the cleavage site, a 2',3'-cyclic-phosphate (2',3'-cP) and a 5'-OH is formed which can in turn be ligated by the twin ribozyme in the same manner with a substrate bearing a 2',3'-cP and a 5'-OH group. Cleavage and ligation sites are indicated with arrows.^[50]

2.3. Chemical ligation of synthetic RNA

An alternative ligation strategy is the joining of two RNA strands using chemical ligation forming non-native linkages. Unnatural linkages can be obtained by reductive amination using periodate^[76], disulfide bond formation^[77] or click chemistry^[14]. Hence, chemical ligation should in principle be a further promising technique for the preparation of long and site-specifically modified RNA from smaller RNA fragments. However, unnatural linkages might impede the native structure and function of functional RNA making this approach relatively unsuitable for studies of biologically relevant RNAs.^[14] In 2010, El-Shagheer and Brown reported splinted CuAAC chemical ligation in combination with solid-phase synthesis to prepare a functional 100 nt hammerhead ribozyme.^[37] Shortly thereafter, ligation of two RNA strands using CuAAC chemistry was reported, reacting an 5'-azido-modified RNA and a 3'-alkyne-modified RNA strand under the presence of a copper-catalyst. It was reported that the unnatural triazole-linked backbone was not disturbing RNA function.^[18] Surprisingly, the triazole group resulting from CuAAC was found to mimic phosphodiester linkages.^[78] Additionally, Frommer *et al.* demonstrated highly efficient preparation of the flavine mononucleotide (FMN) responsive aptamer of the *ypaA* riboswitch from *B. subtilis* using solid-phase synthesis of RNA combined with CuAAC chemical ligation. Therefore, chemical ligation using CuAAC could be an interesting tool to prepare long site-specifically labeled RNA by ligation of small, modified RNA fragments to study RNA's structure and dynamics without perturbing the RNA's function.^[55]

Eventually, it should be taken into account, that previously presented strategies employing ligation of synthetically prepared, modified RNA fragments were by now primarily used to generate labeled RNA < 300 nt. Therefore, different approaches using other enzymatic strategies were established in the past few years to generate even longer labeled RNAs with better yields. These strategies are presented hereafter.

3. Synthetic Biology Approaches for Covalent RNA Labeling During *In Vitro* Transcription

In vitro transcription has widely been applied to prepare RNA up to many kb using phage-encoded polymerases, such as T7 RNA polymerase (RNAP), without encountering any length restrictions.^[79] A wide range of modified nucleotides have been used successfully in IVT reactions to label RNA molecules^[80] but transcriptions with modified nucleotides result in global RNA labeling as these nucleotides cannot be incorporated in site-specific manner.^[81] Due to the highly conserved genetic code and polymerase's substrate specificity,^[6,12] site-directed labeling of RNA by IVT remained challenging for a long time. However, different approaches have been developed to date that address these challenges to selectively label RNA during *in vitro* transcription.

3.1. RNA labeling using an expanded genetic alphabet

An approach gaining increased attention to label RNA site-specifically is the use of an expanded genetic alphabet. Expansion of the genetic alphabet with additional, unnatural base pairs (UBPs, Figure 4) enables the labeling or functionalization of nucleic acids at predefined positions via standard enzymatic reactions as they can be incorporated in sense of a third base pair.^[82] To fulfill the demands of quantitative and selective labeling, these UBPs and also functional groups or tags attached for labeling purposes must be accepted by RNA polymerases with high fidelity and efficiency. During the last decades, several groups have developed and continuously optimized novel base pair systems which can be replicated and transcribed *in vitro* with high selectivity and efficiency by standard polymerases.^[8,16,82–87] Although a few approaches of direct RNA labeling during *in vitro* transcription employing an UB functionalized with the reporter group have been presented,^[8,9] a combination of IVT and post-transcriptional labeling reactions such as CuAAC or iEDDA has mostly been applied for efficient site-specific RNA labeling. For that, an unnatural base derivative bearing a reactive handle (Figure 5, upper panel) is incorporated into RNA during IVT allowing post-transcriptional functionalization with the reporter group of interest.^[88–90] In the following, examples of RNA labeling techniques based on an expanded genetic alphabet are presented.

3.1.1. RNA labeling with hydrogen bonding UBPs

The first example of site-directed RNA labeling applying an extended genetic alphabet dates back to the year 1993.^[91] In this study, Tor and Devan employed a derivative of Benner's isoC-isoG base pair underlying hydrogen bonding which was previously shown to be incorporated on DNA and RNA level.^[92] A functionalized derivative 6-aminohexylisoG was used to *in vitro* transcribe 18 mer DNA templates containing deoxy-5-methylisoC followed by post-transcriptional labeling with NHS-biotin (Figure 4, left panel).^[91] Recently, the kappa (κ)-xanthosine (X) base pair of Benner,^[93] also forming an artificial hydrogen bonding pattern, was used to site-specifically label the 73 nt long guanine sensing riboswitch RNA from *B. subtilis* by Hegelein *et al.*^[94] For this, they synthesized the clickable X derivative 7-deaza-xanthosine bearing a terminal alkyne residue which was then site-specifically incorporated into RNA by IVT to enable post-transcriptional labeling with Cy3-azide via click chemistry (Figure 4, left panel).^[94]

In 2004, based on the 2-oxo(1H)pyridine (y) – 2-amino-6-(2-thienyl)purine (s) base pair system,^[95] Hirao and co-workers incorporated a photo-sensitive iodine-modified y derivative into an anti(Raf-1) RNA aptamer at a predefined position during transcription thereby enabling photo-crosslinking of two aptamer molecules.^[96] They also demonstrated incorporation of bulkier y derivatives, modified with biotin or fluorophores, into RNA molecules opposite s or 2-amino-6-(2-thiazolyl)purine (v) in DNA templates by T7 RNAP (Figure 4, left panel).^[97] Modification

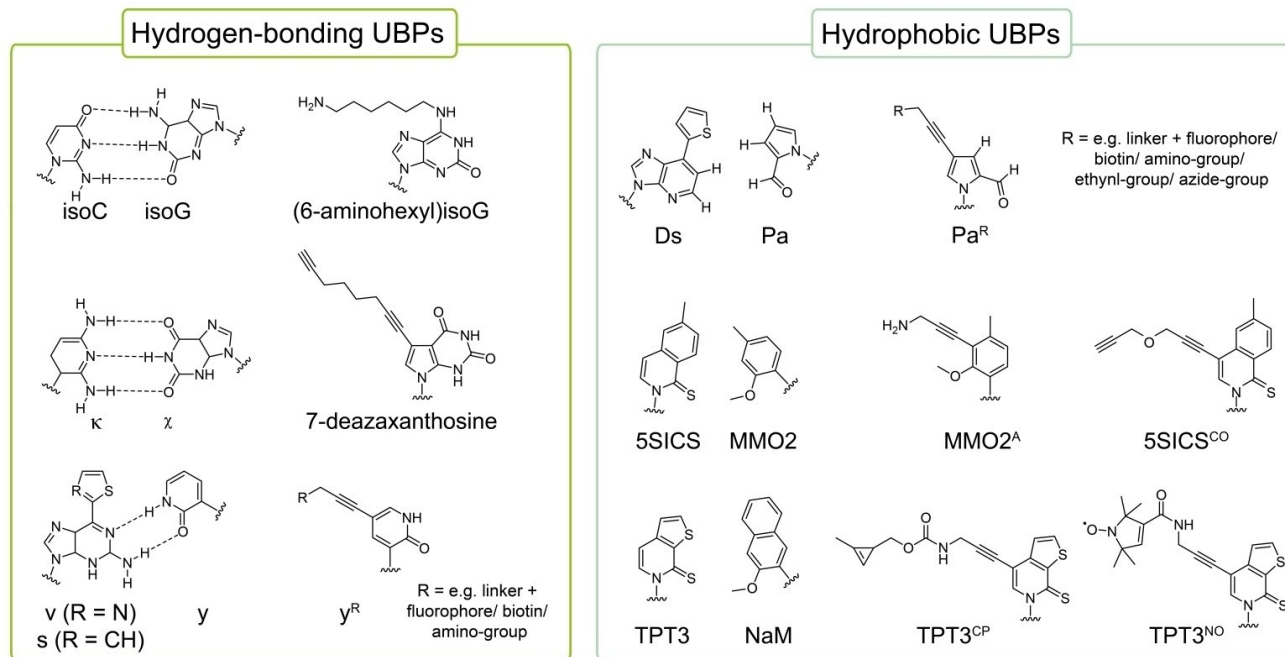


Figure 4. Unnatural base pairs and their derivatives applied for site-specific RNA labeling.

of **y** with an aminohexynyl linker additionally allowed for post transcriptional coupling with NHS-ester of 5-carboxytetramethylrhodamine or 5-carboxyfluorescein.^[98] Moreover, iodine-, biotin- and fluorophore-modified **y** bases were efficiently incorporated into RNA aptamers without disturbing its target affinity.^[96–98] However, site-directed incorporation of the **s-y** and **v-y** base pair was only achieved for **y** opposite **s** or **v**, while transcription of DNA templates containing **y** resulted in mispairing with adenine.^[99]

3.1.2. Hydrophobic UBPs for selective RNA labeling

Besides artificial hydrogen bonding base pairs, the creation of hydrophobic base pairs have evoked particular interest since Kool and co-workers proposed that selective and stable base pairing can also be controlled by hydrophobic forces and shape complementary without hydrogen bonding.^[100] To date, several hydrophobic UBPs have been developed that can be PCR amplified and transcribed selectively and have been applied for RNA labeling.^[82]

Through progressive optimization, the Hirao group has developed two hydrophobic base pairs, **Ds-Pa** and **Ds-Px**, that can be amplified and transcribed with high fidelity and efficiency (Figure 4, right panel).^[101,102] While **Ds-Px** is amplified with high specificity and fidelity,^[102] **Px** is slightly unstable under basic conditions and thus not well applicable during T7 IVT.^[103] **dDs** containing DNA molecules were therefore used to direct the site specific incorporation of **Pa** during transcription.^[9,103]

Combining these two UBPs, **dDs-dPx** for DNA template preparation via fusion PCR and **dDs-Pa** for transcription, site-specific labeling of long RNA molecules was achieved by employing **Pa** triphosphate (TP) derivatives modified with reporter groups in IVT reactions. However, bulky moieties such as fluorogenic tags or biotin directly attached to the unnatural base interfered with polymerase acceptance during enzymatic processing resulting in only moderate transcription efficiency.^[8,9] To overcome this, an ethynyl-C4 linker was attached to **Pa** allowing post-transcriptional labeling of RNA with functional groups by CuAAC (Figure 4, right panel).^[90] By this, a 75 nucleotide tRNA molecule was site-specifically labeled with azide-modified biotin or fluorophore probes. Streptavidin shift assays with biotin-functionalized RNA molecules confirmed site-specific labeling of more than 90% of transcripts.^[90] In a further approach, Hirao *et al.* site-specifically labeled a 76 mer tRNA and 260 mer RNA molecule by copper-free strain-promoted click chemistry.^[103] An azido-pentyl-modified ribonucleoside TP **N₃-Pa** TP was synthesized and efficiently transcribed opposite **dDs** by T7 RNAP. RNA molecules containing **N₃-Pa** were then functionalized with cyclooctene-modified reporter groups.^[103] In 2017, different T7-RNAP mutants were tested regarding their ability to incorporate modified **Pa** TP and 2'-fluoro-ribonucleotide triphosphates of uridine and cytosine simultaneously.^[104] A combination of both, **Pa** TP derivatives and 2'-F-modified triphosphates, in IVT reactions would allow site-specific introduction of desired functionalities and also enhance nuclease resistance of RNA molecules, respectively. The T7 RNAP variant VRS-M5 (VRS: G542V, H772R, H784S and M5: S430P, N433T, S633P, F849I, F880Y) processed Cy3-modified **Pa** TPs and 2'-F-

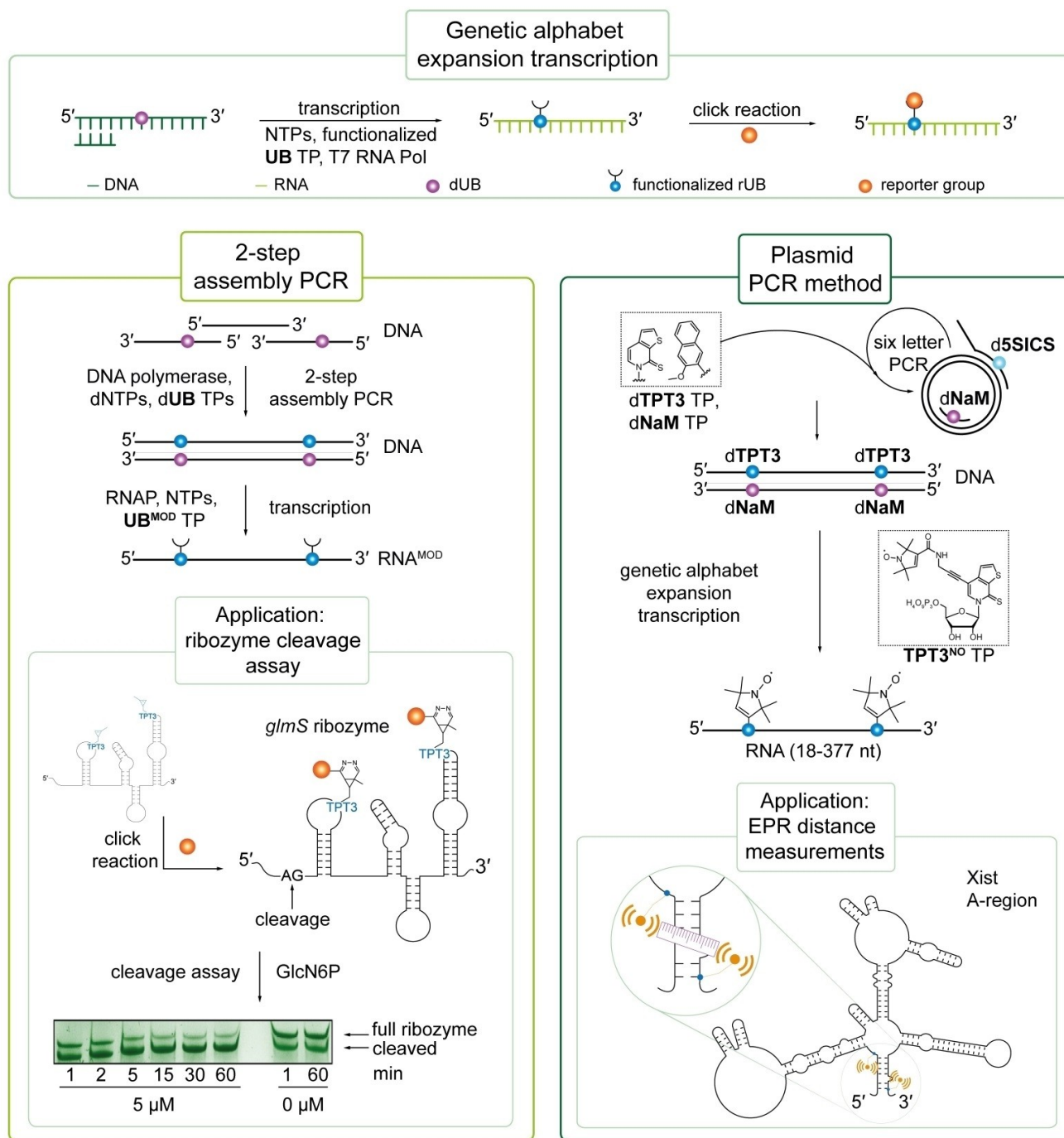


Figure 5. Site-specific RNA labeling with an UBP via genetic alphabet expansion transcription and applications thereof. Upper panel: General scheme of genetic alphabet expansion transcription: a functionalized UB triphosphate is site-specifically incorporated into RNA during T7 IVT enabling post-transcriptional labeling via click chemistry. Left panel: For site-specific labeling of lncRNA, a DNA template bearing the UBP at specific positions can be prepared by 2-step assembly PCR and is then transcribed into RNA. This technique was applied to site-specifically label the *glmS* ribozyme allowing the investigation of ribozyme self-cleavage.^[86] Right panel: Via plasmid PCR, unnatural base pairs can also be incorporated into long functional RNA > 400 nt by using modified primers. By this, nitroxyl spin labels were introduced into the A-region of Xist RNA allowing EPR distance measurements (right panel).^[5,89] Reproduced with permission from Ref. [89], Copyright 2017 Methods.

modified adenine and uracil triphosphates with high efficiency and selectivity. This might pave the way toward the generation

of highly functional RNA molecules such as aptamers and ribozymes for biotechnological or therapeutic applications.^[104]

In recent years, Romesberg's group also developed unnatural base pair candidates exhibiting stable base pairing through hydrophobic and packing forces during replication and transcription^[83–85,105] that were applied in RNA labeling approaches. By employing the **5SICS-MMO2** base pair,^[85] a long functional RNA was site-specifically labeled for the first time. As the **MMO2** scaffold tolerates free amine modifications, an amino-modified **MMO2^A** could be introduced site-specifically into the 77 nucleotide tRNA^{Tyr}_{CUA} with high fidelity during enzymatic synthesis allowing post-transcriptional labeling with NHS-Biotin of greater than 75% of the total RNA yield.^[88] In 2016, Romesberg *et al.* also reported site-directed functionalization of a 243 nucleotide fragment of the central domain of *Thermus thermophilus* 16S ribosomal RNA (rRNA) at two distinct sites using an alkyl derivative of **5SICS-NaM** and the **5SICS-MMO2^A** base pair.^[106] For transcriptions, a DNA template bearing **d5SICS** and **dNaM** was prepared by combining chemical oligonucleotide synthesis and PCR using *Thermus aquaticus* (Taq) DNA polymerase. In a first step, short fragments were chemically synthesized and joined by PCR assembly to generate two DNA oligonucleotides. Next, the two DNA fragments were fused by overlap extension PCR. After transcription, the derivatives **5SICS^{CO}** and **MMO2^A** were functionalized with azide-Cy5 and NHS-Cy3, respectively, allowing to investigate FRET properties of the rRNA in presence of different ribosomal proteins by single molecule total internal reflection fluorescence (smTIRF) microscopy.^[106] By this, Romesberg and co-workers unraveled the formation of four different conformations consisting of two three-helix junctions in open and closed states, whereby the folding states depend on protein binding events.^[106]

Based on the **NaM-TPT3** UBP, we developed an effective co-transcriptional approach for site-directed, copper-free labeling of RNA by click chemistry.^[107] For this, a **TPT3** derivative modified with a norborne linker (**TPT3^{NOR}**) was synthesized and applied in IVT reactions directed by a DNA template containing **dNaM**. **TPT3^{NOR}** was selectively incorporated opposite **dNaM** by T7 RNAP thereby allowing downstream functionalization by iEDDA reactions with tetrazine derivatives.^[107] In a next step, the approach was optimized by substituting the norbornene moiety with a methyl-cyclopropene derivative (**TPT3^{CP}**) that is not only more efficiently incorporated into RNA by T7 RNA polymerase but also more reactive in iEDDA reactions.^[16] The shorter and rigid linker system additionally enabled positioning of the reporter group in close proximity to the nucleobase. **TPT3^{CP}** was selectively incorporated during IVT and downstream functionalization with excess of a tetrazine-fluorophore conjugate in a one-pot reaction with unpurified RNA resulted in effective labeling.^[16] The approach was then applied to investigate ribozyme kinetics as it enables the introduction of labels at defined positions apart from the ribozyme's active site.^[89] For this, the well-characterized CPEB3 and *glmS* ribozymes served as models to demonstrate feasibility with co- and post-transcriptional cleavage reactions, respectively.^[89,108] To examine the 185-nucleotide fragment of the *glmS* ribozyme,^[109] a DNA template with two **dNaM** at distinct positions downstream the ribozyme's cleavage site was prepared by a 2-step PCR

assembly joining three overlapping, chemically synthesized DNA fragments (Figure 6, left panel) followed by genetic alphabet expansion transcription and post-transcriptional click labeling.^[89]

Alternatively, an UBP-modified DNA template for *in vitro* transcription can be prepared from plasmid DNA as PCR template.^[89] Here, the T7 promoter for IVT is introduced via the forward primer and the unnatural base via a modified reverse primer creating a dA (plasmid)-**dNaM** (primer) mismatch (Figure 6, right panel). By this, **TPT3^{CP}** was selectively incorporated into a 401 nt long fragment of the A-region of the lncRNA Xist from *M. musculus* (X inactive specific transcript).^[89]

Introduction of **TPT3^{CP}** during transcription reactions further allows site-directed spin labeling (SDSL) of RNA. We synthesized a tetrazine nitroxyl spin label **Tet^{NO}** and reported SDSL of **TPT3^{CP}** modified RNA by iEDDA cycloaddition reactions for the first time.^[110] A short self-complementary 18 mer RNA, forming a duplex with two **TPT3^{CP}**-A mismatches, was labeled with **Tet^{NO}**. Despite the long flexible linker, the intermolecular distance distribution between both spin labels in the RNA duplex was determined by pulsed EPR techniques (PELDOR).^[110] Recently, we showed direct incorporation of small pyrroline nitroxide spin labels via the **TPT3-dNaM** IVT approach which provides a rapid access to large spin labeled RNA molecules.^[5] The *glmS* ribozyme was successfully labelled with nitroxyl spin label **TPT3^{NO}** (Figure 5, right panel) and inter-nitroxide distance distributions in the *glmS* construct were determined by PELDOR.^[5] The approach also enabled to obtain structural information on the folding of lncRNA that remains still challenging by crystallography: the 377 nt long fragment of the A-region of the long non-coding RNA Xist was structurally explored by PELDOR for the first time (Figure 6, right panel).^[5] PELDOR studies reinforced an existing folding model proposed by Fang *et al.*,^[111] which suggests stable duplex formation between 5'- and 3'-end.^[5] This SDSL approach enables not only direct introduction of spin labels into long RNA during IVT but can also facilitate structural insights of largely unexplored functional RNA *in vitro* and potentially in cells.^[5] Later in 2020, an analogous approach applying an alkyne-modified **TPT3** TP for transcriptions and post-transcriptional CuAAC for SDSL of RNA was presented.^[112]

The discussed examples clearly demonstrate suitability and applicability of UBPs for site-specific labeling of RNA without disturbing the global fold of complex RNA molecules. Currently, limited availability of unnatural bases, especially modified with reactive handles, constitutes the major drawback of this technique making it difficult to access for non-chemical laboratories. Apart from that, expansion of the genetic alphabet enables straightforward introduction of labels at any desired position of the RNA of interest arising from almost excellent acceptance of the established UBPs by standard polymerases.^[8,16,82–87] Compatibility with diverse click reactions, moreover, allows post-transcriptional functionalization with virtually any modification of choice.^[16,90,94,103,106,107,110]

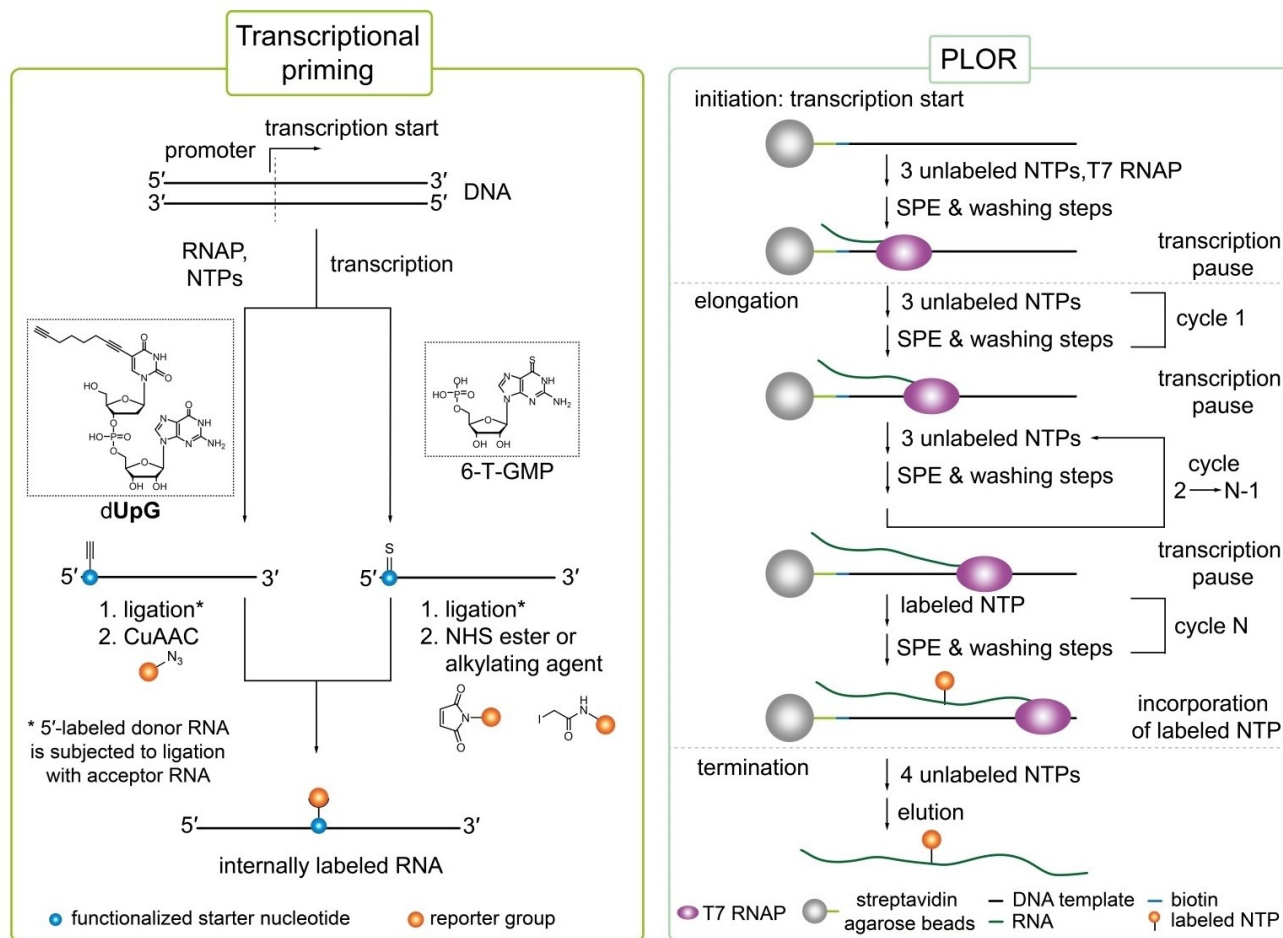


Figure 6. Selective RNA labeling via transcriptional priming (left) and position selective labeling of RNA (PLOR, right panel). Left panel: In transcriptional priming, a functionalized starter nucleotide is selectively incorporated at the 5'-end of RNA. Enzymatic ligation of 5'-labeled donor RNA to an acceptor RNA results in an internal functionalized nucleobase that can be further modified via a compatible click or crosslinker chemistry.^[113,115] Right panel: PLOR is divided into initiation, elongation and termination. Transcription proceeds on a DNA template immobilized to streptavidin agarose beads. Transcriptional pausing is caused by the lack of a canonical triphosphate. Elongation consists of many cycles each lacking one canonical NTP to cause transcriptional pausing. This allows for labeling with a modified NTP at a specific position. To avoid cross-contamination between individual cycles, thorough washing after pausing is essential (SPE = solid-phase extraction).^[117,119]

3.2. Alternative RNA modification strategies via *in vitro* transcription

Modification of one of the canonical nucleoside triphosphates for *in vitro* transcription usually results in random labeling throughout the entire RNA sequence.^[81] However, a few IVT approaches have been established which enable internal introduction of canonical base analogs at specific positions.

3.2.1. Site-specific RNA labeling by combining transcriptional priming and enzymatic ligation

A segmental approach for internal introduction of labels into RNA at a desired position combines transcriptional priming with an initiator (or starter) nucleotide and splinted ligation as developed by the group of Jäschke (Figure 6, left panel).^[113] 5'-

Labeling of RNA by transcriptional priming has been widely studied over the last decades^[114] and allows to incorporate various guanosine or adenosine monophosphate (GMP; AMP) and dinucleotide monophosphate analogs selectively at the 5'-end due to its lack of triphosphate. For internal labeling, the transcript (donor RNA) is coupled to another RNA fragment (acceptor RNA) by splinted ligation.^[113] Nevertheless, the starter (di)nucleotide must retain its 5'-OH to enable phosphorylation and subsequent ligation. The RNA can either be modified directly by transcriptional priming or by post-transcriptional functionalization depending on the size of the desired label. A rather small reactive handle or label, however, is more advisable as the modification has to be accepted by three different enzymes – a polymerase, kinase and ligase.^[113] Thus, Jäschke and co-workers synthesized a “clickable” octadiynyl UpG dinucleotide (OdUpG) which was selectively incorporated at the 5'-end by T3, T7 and SP6 polymerases (72%, 80% and 82%

labeling efficiency, respectively).^[113] After transcription, the primed donor RNA was phosphorylated by T4 polynucleotide kinase and ligated to acceptor RNA using a mixture of T4 RNA ligase 1, T4 RNA ligase 2 and T4 DNA ligase. Ligated RNA molecules can be internally labeled by coupling azide cargos to the octadynyl moiety via CuAAC.^[113]

In a similar approach, Lebars *et al.* spin-labeled a 55-nt RNA fragment covering K-turn and specifier-loop domains from T-box leader RNA of *Bacillus subtilis* site-specifically.^[115] For this, RNA was selectively functionalized at its 5'-end during T7 IVT by employing commercially available 6-thioguanosine-5'-O-monophosphate (6-T-GMP) that allows labeling with virtually any probe by easily reacting with iodacetamido- or NHS-compounds due to its nucleophilic nature (Figure 6, left panel). After splinted ligation of 5'-modified donor RNA to acceptor RNA, the thioguanosine-modified base was subjected to coupling reaction with the nitroxide reagent 3-(2-iodoacetamido)-proxyl. To guarantee stability of the spin label, RNA was purified prior to coupling reaction as the nitroxide moiety is prone to reduction under ligation conditions by DTT.^[71] Investigation by EPR spectroscopy indicated almost quantitative spin-labeling efficiency of $\geq 95\%$.^[115] Additionally, the proxyl-spin label did not perturb global RNA folding of the 55 nt T-box fragment as demonstrated by NMR spectroscopy. By combining 5'-priming and isotopic labeling with ¹³C-modified NTPs of the acceptor RNA fragment, the authors also demonstrated compatibility of selective spin-labeling with segmental isotopic labeling enabling combined structural and dynamical investigation by EPR and NMR spectroscopy.^[115]

In principle, this labeling strategy can be adapted to characterize any large RNA molecule of interest by adjusting number and length of fragments.^[115] For example, two spin labels could be introduced by 5'-priming of two defined fragments of the RNA molecule of interest that can then be assembled with the remaining fragments by ligation.^[115] To date, several dinucleotide or GMP analogs with various reporter or functional groups are commercially available making the approach accessible for biological laboratories. The combination of two or more different functionalized starter nucleotides, such as alkyne or thio-modified nucleotide monophosphates, would also allow for labeling with diverse reporter groups or tags after RNA ligation. However, this segmental labeling approach is laborious due to many reaction and purification steps required to obtain fully intact labeled RNA. Moderate ligation efficiencies (as discussed in chapter 2.1) might also frequently result in relatively low yields of labeled RNA.

3.2.2. Position selective labeling of RNA

A completely different methodology for the internal labeling of RNA at specific positions is PLOR – position selective labeling of RNA – as developed by Wang and co-workers. (Figure 6, right panel).^[116] The method takes advantage of the elongation complex (EC) formed between T7 RNAP, DNA template and nascent RNA during transcription. This complex can persist biochemical manipulations or purification procedures allowing

to force pausing and restarting of the transcription process (also known as stepwise walking).^[117] A 5'-biotinylated DNA template containing the T7 promoter is coupled to streptavidin agarose beads that serve as solid support.^[116] Steric hindrance of the beads, possibly affecting transcription efficiency, is obviated through a 15–30 bp long linker between the biotin moiety and T7 promoter sequence.^[118] In contrast to standard transcription, PLOR is initiated with three or fewer NTP types resulting in stalling at the first position where the absent NTP is required.^[116] To ensure stable transition from the less stable initiation complex to EC, the first pausing event is proposed to occur at least nine but optimally eleven to thirteen nucleotides downstream of initiation start.^[118] Then, NTPs are completely removed from the polymerase/DNA/RNA complex by repetitive cycles of solid-phase extraction (SPE) and washing of the beads.^[116] A fresh NTP mixture is added containing the modified nucleotide of interest to allow elongation until the next pause point caused by a missing NTP is reached and the washing step is repeated. This cycle of elongation can be replicated multiple times until all modifications of choice are incorporated into the growing RNA sequence. However, selective labeling requires careful design and portioning of the RNA sequence and elongation steps in such fragments that the desired modified nucleotide can be incorporated at only one specific position.^[116] To terminate transcription, all four canonical NTPs are added to complete synthesis of the transcript. The transcript falls off from the complex after termination and can be collected with the liquid phase. Re-initiation of transcription can be prevented by addition of heparin, a well-known transcription inhibitor, or exclusion of guanine triphosphates (GTP) during termination if permitted by the RNA sequence. After thorough washing, the immobilized DNA template and RNAP can be recycled for the next round of PLOR to achieve higher RNA yields as only one transcript is produced per template and cycle. PLOR has been established with a fully robotic platform due to extensive washing of the probes.^[116,119] Thus, this new technology is restricted to laboratories with access to robotic platforms or requires otherwise time-consuming manual operation.

PLOR was originally applied to selectively label two RNA molecules, a 71 nucleotide aptamer domain of the adenine riboswitch (riboA71) from *Vibrio vulnificus* and a 104 nt fragment of the turnip crinkle virus (TCV) RNA, with isotopes or fluorophore to investigate structure and dynamics.^[116,120] Heavy-atom labeled as well as various fluorophore-modified NTPs (Cy3, Cy5, AlexaFluor 488, AlexaFluor 555) have been incorporated at desired positions in RNA applying PLOR.^[116,119,120] Different riboA71 derivatives were labeled at predefined, structurally important positions with isotope-labeled nucleotides allowing investigation of RNA folding by NMR spectroscopy. Labeling of a single nucleotide of RNA is, rather than uniform labeling, often more purposeful for unambiguous structural investigation of a local position by NMR spectroscopy.^[116,120] For example, RiboA71 was site-specifically labeled with ¹³C/¹⁵N-uridine-5'-TP at residue U39, which participates to adenine binding, to study coexistence of conformers as previously proposed by a four-state model. Obtained NMR data revealed four distinct peaks supporting the coexistence of

four conformations in presence of adenine as stated by the model. Recently, PLOR was also used to selectively introduce 5'-iodouridine in riboA71 for crystallization.^[121] In principle, PLOR can be applied to incorporate various types of modifications or labels as long as it is compatible with polymerase acceptance. To overcome limitations in the variety of labeling tags, 5'-aminoallyl-modified NTPs were incorporated into RNA during PLOR enabling downstream functionalization with basically any functional group derivatized with an NHS-ester.^[116]

4. Post-synthetic and Post-transcriptional Labeling of RNAs

Besides introduction of modified nucleotides during solid phase or enzymatic synthesis, naturally occurring RNA can also be labeled post-synthetically or post-transcriptionally. To date, a variety of different approaches has been reported employing chemical modifications^[122–128] or catalytically active deoxyribozymes,^[129–132] ribozymes^[133–135] and enzymes^[136–140] to accomplish modification reactions on unmodified RNA. Starting from pre-existing RNA sequences, that can either be prepared *in vitro*, by synthetic or enzymatic strategies or emerge from endogenous origin, the RNA oligonucleotides undergo labeling reactions introducing functional reporter groups such as fluorophores, affinity tags or small molecule handles for subsequent modification with reporter groups.^[14] In principle, post-synthetic and post-transcriptional RNA labeling approaches can be applied to short and long sequences equally and do not face limitations concerning long RNAs. However, special attention needs to be paid addressing site-specificity and universality of such approaches regarding applicability for all types of RNA and not only for those with intrinsic distinctive features.^[14] As previously mentioned, a wide range of labeling techniques for unmodified RNA were developed by now. Admittedly, some techniques include methyltransferases for the introduction of modifications at the 5'-cap structure^[141] or at the 3'-terminus^[142] or poly(A) tail.^[143] Other approaches utilizing self-alkylating ribozymes,^[144] tRNA-guanine transglycosylases (RNA-TAG)^[145] or tRNA^{lle2}-agmatidine synthetase (Tias)^[146] are based on larger structural elements that need to be fused to the RNA of interest. As this review focuses on internal covalent and site-specific labeling of long RNA with minimal impact on RNA's native folding and function, the aforementioned approaches will not be described in the following section.

4.1. Chemical strategies for site-specific labeling of unmodified RNA

In unmodified RNA oligonucleotides, a plethora of functionalization sites are available as the 2'-OH residue of the ribose sugar possesses nucleophilic character and therefore provides multiple options for chemical modifications.^[147,148] Special care has to be taken to address site-specificity and chemoselectivity to select only a certain target site of the RNA oligonucleotide.^[148]

This can be achieved through the use of DNA helper strands assigned to guide a functional group to the desired site of modification and to protect the remaining RNA from chemical reagents.^[123,124,127,128]

4.1.1. DNA-templated chemical labeling via four-way junction formation

Applying a DNA-templated chemical labeling strategy, Jahn *et al.* demonstrated the site-specific introduction of a thiol group for subsequent attachment of labels for both synthetic and endogenous RNA.^[124] The thiol functionalization can be attached to predetermined internal nucleotides of the acceptor RNA by formation of a four-way junction with a donor DNA linked to an activated carboxylic acid group and two site-specific DNA guiding strands. The DNA donor strand will then be coupled to a specific internal 2'-OH group of the acceptor RNA via the activated carboxylic acid group. Subsequently, the adjacent disulfide bond in the donor's linker is cleaved eventually yielding the thiol-modified acceptor RNA. Afterwards, site-specifically thiol-modified target RNA can further be labeled with a reporter group which was demonstrated by coupling of a maleimide-functionalized fluorophore.^[124]

4.1.2. DNA-templated chemical labeling via duplex formation

An alternative chemical labeling strategy was developed by Freisinger and co-workers using a DNA-templated approach for sequence-specific generation of etheno-adducts in single-stranded DNA.^[122] Thereby, a guiding DNA bearing an alkylating agent is hybridized to the target DNA to bring the target base and alkylating agent in close proximity. Using this approach, the group of Freisinger obtained only moderate yields up to 30%.^[122] Improving yields up to 65%, Sigel, Freisinger and co-workers demonstrated DNA-templated formation of a 12-propargyl-etheno-adenine-modified single-stranded 16 mer DNA and its RNA equivalent and subsequent fluorophore labeling via CuAAC.^[123] This method in principle allowed the incorporation of bioorthogonal groups into single stranded regions of both DNA and RNA of unrestricted length. Fluorescent labeling of a large RNA oligonucleotide was established for the 633 nt long D135-L14 group II intron ribozyme construct derived from the *Saccharomyces cerevisiae* Sc.ai5γ.^[123] In 2018, they also presented site-specific dual-color labeling of the regulatory 275 nt long *btuB* riboswitch from *Escherichia coli* (*E. coli*) by combining internal functionalization with oxidative opening of the 3'-terminal ribose and subsequent conjugation to two different fluorophores.^[128] Native folding and function of the *btuB* RNA riboswitch were studied using smFRET to characterize the conformational equilibrium of the *btuB* riboswitch upon binding of its cofactor adenosylcobalamin. The experiment proved that not only nucleotides within single-stranded regions but also within RNAs with complex secondary structures can be targeted.^[128]

4.1.3. RNA acylation at induced loops

More recently, Kool and co-workers developed the technique RNA acylation at induced loops (RAIL) for *in vitro* functionalization of RNA at specific 2'-OH groups within the sequence.^[127] RAIL employs complementary helper DNA strands to expose loops or gaps at desired predefined sites within the RNA of interest to enable site-selective 2'-OH acylation with an acylimidazole reagent, e.g. nicotinyl acylimidazole azide (NAI-N₃, Figure 7). The remaining RNA segments and their reactive 2'-OH groups are protected by the formed DNA-RNA duplex structures. After acylation reaction, the helper DNA can be degraded by DNase digest.^[126] Utilizing this strategy, the desired mono-acylated adduct was formed predominantly and only a minor fraction of side product acylated at the adjacent nucleotide 5' to the induced bulge was observed.^[127] Addressing acylation sites in close proximity to both, 5'- and 3'-end, of the RNA revealed a lack of selectivity for the gap induction approach as reliable hybridization of a short helper DNA strand complementary to the 5'- or 3'-end of the RNA cannot be ensured.^[127] RAIL was employed to selectively control the catalytic activity of a 81 nt tandem hammerhead ribozyme (TR) with dual catalytic cores (3TR and 5TR). Specific acylation at the 3TR core guided to the target site via an induced gap, subsequently suppressed the cleavage rate by 5-fold and retained 85% of the initial rate of untreated TR with the 5TR substrate.^[127] In comparison, site-specific acylation of the 5TR core facilitated by an induced loop strategy suppressed the cleavage rate by 4-fold and retained 93% of the initial rate of untreated TR with the 3TR substrate. Thus, this method provides local control of RNA acylation in multifunctional RNA and high yields of acylation-based suppression of local RNA function with little off-target acylation.^[127] Furthermore, Kool's group tested dual labeling of 65 nt small nucleolar SNORD78 RNA for FRET experiments via serial RAIL labeling.^[127] Successive labeling was performed by first loop induced acylation with NAI-N₃ at G14 and subsequent SPAAC reaction with Alexa488-azide, followed by a second acylation step at A49 and subsequent click reaction with tetramethylrhodamine(TAMRA)-azide.^[127] Taken together, the RAIL method is well suited for internal, site-directed acylation and further functionalization particularly for longer RNAs. RNAs with several hundred nucleotides in length could in

principle be selectively acylated, if helper DNAs are prepared enzymatically or several synthetic helper DNAs are arranged along the RNA of interest.^[127] The major disadvantage of this approach is the formation of 20–30% secondary acylation products adjacent to the targeted position.^[127] Interestingly, the acylating reagent NAI-N₃ that has so far been used for click-based conjugation as well as for blocking and caging^[125] forms a reversible adduct on RNA which can be removed by bioorthogonal deacylation with phosphines.^[127]

4.2. Deoxyribozymes for labeling of unmodified RNA oligonucleotides

DNAzymes can not only be used to ligate RNA as described in section 2.2 but can also be employed for post-transcriptional or post-synthetic labeling of RNA.^[129–132] For that purpose, special deoxyribozymes exist that can either link a modified RNA strand or a single nucleotide to a pre-existing RNA.^[129,149] A major advantage of this approach is the high labeling specificity facilitated by specially designed base pairing interactions between target RNA and DNAzyme.^[129,149] In comparison to RNA, facile and low-priced preparation of the deoxyribozyme by standard solid-phase DNA synthesis and DNA's high stability in principle allows a broad application in biochemical laboratories.

4.2.1. Deoxyribozyme-catalyzed labeling introducing phosphodiester branches

Introducing the approach of deoxyribozyme-catalyzed labeling (DECAL), Silverman and co-workers effectively utilized the 10-DM24 DNAzyme for modifying RNA at specific internal positions (Figure 8).^[129] Therefore, a tagging RNA is created by standard *in vitro* transcription introducing a 5-aminoallyl-modified cytidine at its second position. The primary amino group on the aminoallyl-RNA is then functionalized by reaction with an NHS-ester derivative to obtain labeled tagging RNA. Using the 10-DM24 DNAzyme which hybridizes to both, target RNA and tagging RNA, the labeled tagging RNA is attached to an internal 2'-OH group of the target RNA to result in a 2',5'-phosphodi-

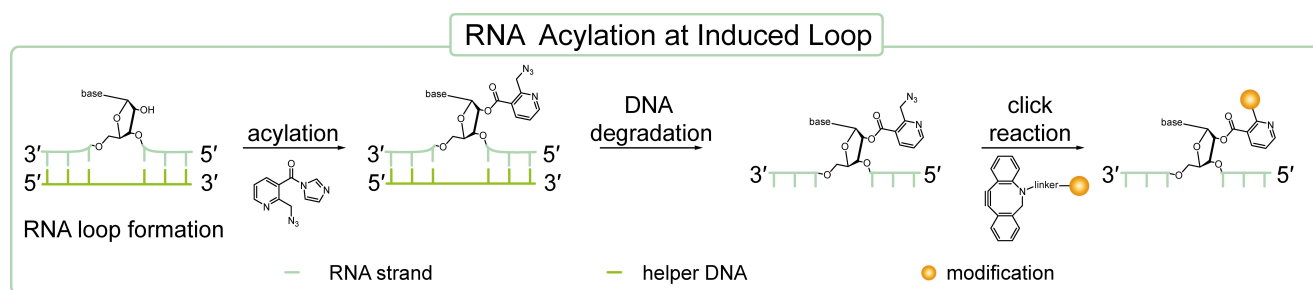


Figure 7. Site-selective RNA labeling via RNA acylation at induced loop (RAIL) structures.^[127] A complementary helper DNA strand is hybridized exposing a loop at a predefined site within the endogenous RNA of interest to enable site-selective 2'-OH acylation with an acylimidazole reagent (here: nicotinyl acylimidazole azide). Remaining RNA segments and their 2'-OH groups are shielded by the helper DNA-RNA duplex. After acylation, the helper DNA strand is degraded by DNase treatment. Site-selectively acylated RNA can be further modified by CuAAC reactions with functionalized alkyne-conjugates.

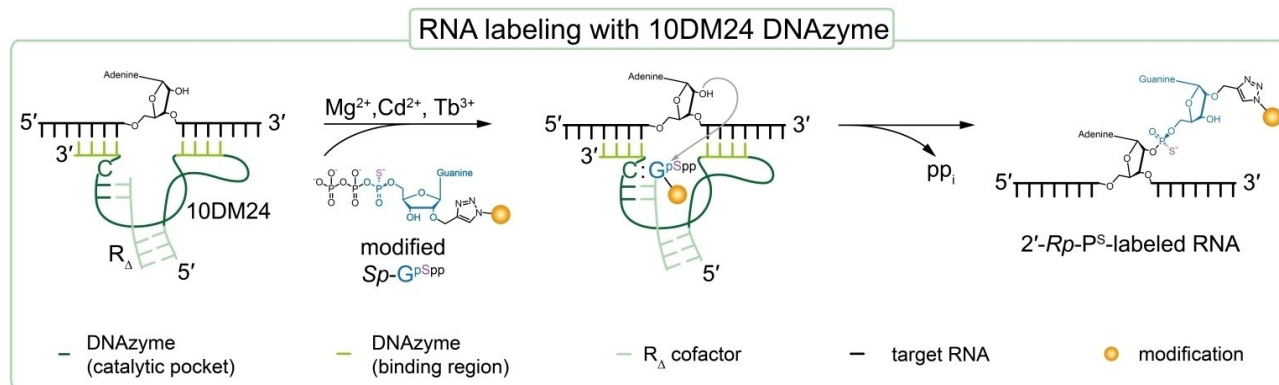


Figure 8. Refined deoxyribozyme-catalyzed labeling (DECAL) of pre-existing RNA with the 10-DM24 DNAzyme applying assisting heavy metal ions and the phosphorothioate-modified GTP derivative *Sp*-GTP^S.^[130–131] The further engineered 10-DM24 deoxyribozyme catalyzes the attachment of modified guanine mononucleotides to a 2'-OH group within the RNA strand of interest. Addition of the oligonucleotide cofactor R_A compensates for a formerly longer oligonucleotide substrate and is necessary for sufficient conversion. The deoxyribozyme's binding arms hybridize to the target RNA and can be adapted for individual target sites to enable site-specific RNA modification. Phosphorothioate-modified *Sp*-GTP^S is used as ligation substrate to prevent subsequent phosphodiesterase hydrolysis of the 10-DM24 catalyzed 2',5'-branch modification due to the particular steric configuration of the newly formed linkage in 2'-*Rp*-P^S-labeled RNA. Tb³⁺ acts as an accelerating cofactor that enables efficient conversions at pH 7.5 and reduces the required concentration of modified GTP triphosphate. Thiophilic Cd²⁺ ions are used to prevent phosphorothioate interfering effects on the metal ion coordination by the 10-DM24/RNA/GTP^S ternary complex.

er bond formation yielding a branched oligonucleotide. The label can be attached to different sites within the target RNA by respective modification of the DNAzyme's binding arms.^[129] By this, Silverman reported the successful application of DECAL to install two fluorophores at specific internal sites of the 160 nt long P4–P6 *Tetrahymena* group I intron P4–P6 RNA enabling to study Mg²⁺ dependent RNA folding by FRET.^[129]

Based on the aforementioned approach, Höbartner and Silverman reported further engineering of the 10-DM24 deoxyribozyme to accept mononucleotides as ligation substrates instead of 5'-triphosphorylated oligonucleotides.^[149] Catalyzed ligation of free GTP mononucleotides by the engineered 10-DM24 DNAzyme requires the auxiliary addition of an oligonucleotide cofactor to stabilize the deoxyribozyme's secondary and tertiary structure.^[149]

Additionally, Höbartner and co-workers showed that lanthanide cofactors, in particular Terbium (Tb³⁺), are able to accelerate DNA-catalyzed synthesis of branched RNA.^[132] They presented a general post-transcriptional labeling approach for endogenous RNA with different ribose-modified guanine triphosphates and Tb³⁺ as an accelerating cofactor to install functional groups such as azides and primary amines, affinity reagents, fluorophores, spin labels and cross-linkers at 2'-OH groups of internal adenines within *in vitro* transcribed RNA (Figure 8).^[130] This method provides several advantages: first of all, the use of commercially available modified mononucleotides ensures a small modification size compared to the earlier used tagging oligonucleotides and at the same time reduces preparatory work.^[130] Secondly, utilizing Tb³⁺ as accelerating cofactor for the DNAzyme enables efficient conversions at physiological pH and reduces the concentration of GTP derivatives required for labeling. Overall yields > 80% could be obtained for most labeled GTP analogs providing at the same

time fast reaction rates of $k_{\text{obs}} \approx 1 \text{ min}^{-1}$.^[130] The group of Höbartner used this method to introduce two fluorophores as FRET pair on the SAM II riboswitch RNA by two successive DNAzyme catalyzed labeling reactions to study Mg²⁺-induced folding of its 52 nt long aptamer domain in the presence and absence of SAM ligand.^[130] The observed results were consistent with previously reported data,^[150] proving full functionality of the labeled riboswitch.^[130] Furthermore, they proved applicability of their approach to label larger RNA transcripts *in vitro*. For the 120 nt long spliceosomal U6 snRNA, ten different labeling positions for methylanthraniloyl-G (MANT-G) were tested in total revealing influence of the secondary structure on efficient deoxyribozyme binding and catalysis.^[130] To improve DNA-catalyzed labeling of complexly folded RNAs, the use of specially designed disruptor oligonucleotides breaking up secondary structure formation might facilitate the DNAzyme's hybridization to the target RNA increasing labeling efficiency. Using this approach including a disruptor oligonucleotide, DNA-catalyzed labeling of the 156 nt long *ydaO* riboswitch RNA with Cy3-G and Cy5-G was successively performed yielding double labeled long RNA allowing further FRET studies.^[130]

4.2.2. Deoxyribozyme-catalyzed labeling introducing phosphorothioate branches

In 2017, Höbartner and co-workers reported almost quick and complete hydrolysis of the 10-DM24-catalyzed 2',5'-phosphodiester branch modification in a 209 nt long *UBC4* pre-mRNA transcript after incubation with purified *Saccharomyces cerevisiae* debranchase (Dbr1).^[131] Hence, the use of this labeling technique is restricted to *in vitro* applications only. However, as it is known that Dbr1 is unable to hydrolyze phosphorothioate

linkages with *Rp*-configuration,^[151] this characteristic can be exploited to prevent phosphodiesterase hydrolysis of the label. Therefore, phosphorothioate-modified, fluorescent GTP derivatives of *Sp*-GTP⁵ were synthesized as 10-DM24-catalyzed labeling of RNA with *Sp*-GTP⁵ results in an inversion of the phosphorothioate configuration yielding 2'-*Rp*-GMP⁵-labeled RNA. Moreover, the labeling conditions for 10-DM24 were optimized to efficiently incorporate *Sp*-GTP⁵ by addition of the thiophilic metal ion Cd²⁺ to prevent phosphorothioate interfering effects on the metal ion coordination by the 10-DM24/RNA/GTP⁵ ternary complex. Although reactions with *Sp*-GTP⁵ derivatives generally showed a slower reaction rate compared to GTP derivatives, the resulting 2'-*Rp*-GMP⁵-labeled RNA was found to be mostly resistant to debranching by yeast Dbr1.^[131] Yet, applications for sensitive RNA substrates in a cellular context are limited by the need to form a three-helix junction^[129] and the necessity for assisting heavy metal ions for improved deoxyribozyme activity.^[130–132]

4.3. Ribozymes for labeling of unmodified RNA oligonucleotides

A promising approach to label endogenous RNA *in vivo* can be found in utilizing genetically encoded ribozymes.

4.3.1. Ribozyme-catalyzed labeling introducing phosphodiester branches

Maghami *et al.* reported direct *in vitro* selection of trans acting 2'-5' adenylyl transferase ribozymes for covalent and site-specific RNA labeling with *N*⁶-(6-aminohexyl)-ATP and *N*⁶-fluorophore labeled ATP analogs at specific internal 2'-OH groups resulting in 2'-5'-phosphodiester bond branched RNA.^[133] In the presence of total cellular RNA, the most efficient and specific ribozyme, FH14, was employed to fluorescently label the 120 nt long *E. coli* 5 s rRNA at three different positions using a fluorophore-modified ATP analog. Therefore, three variants of the FH14 ribozyme were designed with individual binding arms of 8–10 nt flanking the different labeling sites. Combining two or three ribozyme variants, multiple labeling could be achieved.^[133] Admittedly, it is known that the branched 2'-5'-phosphodiester bonds were readily cleaved by natural debranching enzymes resulting in loss of the installed label^[131] therefore hampering the scope of this labeling approach for cellular applications. In addition, not only modified ATP analogs but also cellular, ubiquitous ATP is used as a substrate by the ribozyme and competes for incorporation.^[152] In return, the *N*⁶-modified ATP can also be incorporated by cellular ATP dependent enzymatic reactions which will cause unspecific ribozyme-independent background labeling.^[153]

4.3.2. Ribozyme-catalyzed labeling introducing phosphonate ester branches

Recently, Höbartner and co-workers evolved advanced ribozymes for bioorthogonal RNA labeling attaching derivatives of the acyclic nucleoside phosphonate Tenofovir via branched phosphonate ester junctions to the 2'-OH of an adenosine within the target RNA (Figure 9).^[134] Tenofovir is an antiviral lipophilic and cell-permeable prodrug used against HIV and hepatitis B virus (HBV) infections.^[154] Applying the Tenofovir-transferase ribozyme FJC9 and fluorescently-modified Tenofovir, the 120 nt long *E. coli* 5S rRNA was efficiently labeled in the presence of cellular total RNA. Additionally, using Cy5-modified Tenofovir and Tenofovir-transferase ribozyme FJ1, six different target sites within *E. coli* 16S and 23S rRNA, 1500 nt and 2900 nt long, respectively were modified.^[134] Notably, the phosphonate ester bonds formed by Tenofovir-transferase ribozymes were more stable towards Dbr1 compared to natural phosphodiester linkages formed during previously described labeling approaches using DNAzymes and ribozymes. This aspect taken together with the benefit of avoiding cross-interactions with non-specific cellular substrates, as described for previous nucleotide labeling analogues,^[130,132,133,149,152,153] renders this method in principle possible for future applications in a cellular context. Moreover, simultaneous orthogonal labeling installing both a phosphonate ester and phosphodiester branching modification by application of two orthogonal ribozymes could be achieved on either a single common target RNA (Figure 9) or on two individual target RNAs. More in detail, simultaneous labeling of 23S rRNA by the FJ1 ribozyme using Cy5-Tenofovir in combination with labeling of 16S rRNA by FH14 with FAM-ATP in the presence of cellular total RNA was demonstrated.^[133]

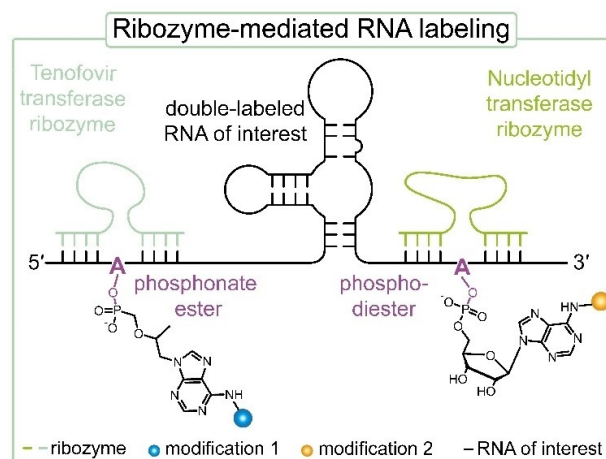


Figure 9. Site-specific orthogonal double labeling of an endogenous RNA of interest using nucleotidyl and Tenofovir transferase ribozymes.^[134] The *in vitro* selected 2'-5' adenylyl transferase ribozyme FH14 can be utilized for internal labeling of 2'-OH groups with *N*⁶-modified ATP analogs yielding 2'-5'-phosphodiester bond branched RNA. Additionally, the *in vitro* selected Tenofovir transferase ribozyme FJ1 can be applied for internal 2'-OH group labeling with *N*⁶-modified acyclic nucleoside phosphonate Tenofovir derivatives yielding 2'-5'-phosphonate ester bond branched RNA.

4.3.3. Ribozyme-catalyzed labeling by methyltransferase ribozymes

More recently, Scheitl *et al.* reported *in vitro* selection of a methyltransferase ribozyme for site-specific RNA methylation.^[135] The methyltransferase ribozyme MTR1 catalyzes the transfer of a methyl group from the cofactor O⁶-methylguanine (m⁶G) to a specific adenine yielding modified 1-methyladenosine (m¹A) in its target RNA sequence. Site-specific methylation of adenosines was shown for a target tRNA in the presence of total *E. coli* tRNA *in vitro*.^[135] However, the methyltransferase ribozyme was not yet applied for site-specific RNA labeling other than methylation but represents a promising tool for future applications in RNA labeling using e.g. fluorescently labeled benzylguanine derivatives.^[135] Nevertheless, when developing such an application, circumvention of undesired cross-interactions with non-specific cellular substrates needs to be addressed.

4.4. Methyltransferases for enzymatic labeling of RNA oligonucleotides

RNA methyltransferases catalyze the transfer of methyl groups to target nucleotides in RNA.^[147] Many methyltransferases employ the cofactor *S*-adenosyl-L-methionine (AdoMet or SAM) as methyl-group donor.^[147] RNA methylation can be applied for post-transcriptional modification of coding and non-coding RNA.^[147,155] Exploiting the substrate promiscuity of certain methyltransferases, these enzymes can be used for RNA labeling introducing a variety of functionalities. By that, the meth-

yltransferase-directed transfer of activated groups (mTAG) can be used for covalent and sequence-specific labeling of pre-existing RNA.^[136,139,140]

Liu and co-workers first found that human RNA N⁶-methyladenosine (m⁶A) methyltransferase METTL3-METTL14 (METTL3-14) to a certain extent exhibits co-substrate promiscuity and is able to site-specifically transfer an allyl group to the N⁶-position of adenosine.^[138] Therefore, a distinctive consensus motif needs to be present in the RNA of interest that serves as a recognition sequence for the methyltransferase and defines the methylation site. For METTL3-14, the consensus motif comprises five nucleobases in the DRACH motif (D = A, G or U; R = G or A; H = A, C or U; underlined A represents the modification site), in which the central A will be modified.^[156] METTL3-14 is able to generate N⁶-allyl labeled RNA utilizing the synthetic cofactor allyl-SAM.^[138]

Recently, Rentmeister and co-workers presented sequence-specific chemo-enzymatic labeling and photocaging of RNA applying the mRNA methyltransferases METTL3-14 and METTL16 (Figure 10).^[137] Utilizing the co-substrate promiscuity of the two methyltransferases, they tested different analogs of the methyltransferases' natural co-substrate AdoMet to transfer functional groups to the N⁶-position of target adenosines within the RNA of interest.^[137]

Therefore, RNA sequences with the respective individual consensus motifs for each methyltransferase were prepared including the previously described DRACH motif for METTL3-14. The consensus motif for METTL 16 comprises the nine nucleobases UACAGAGAA (underlined A represents the modification site) which need to be posed in a loop or hairpin structure within the target RNA.^[157] Testing the transfer of propargyl

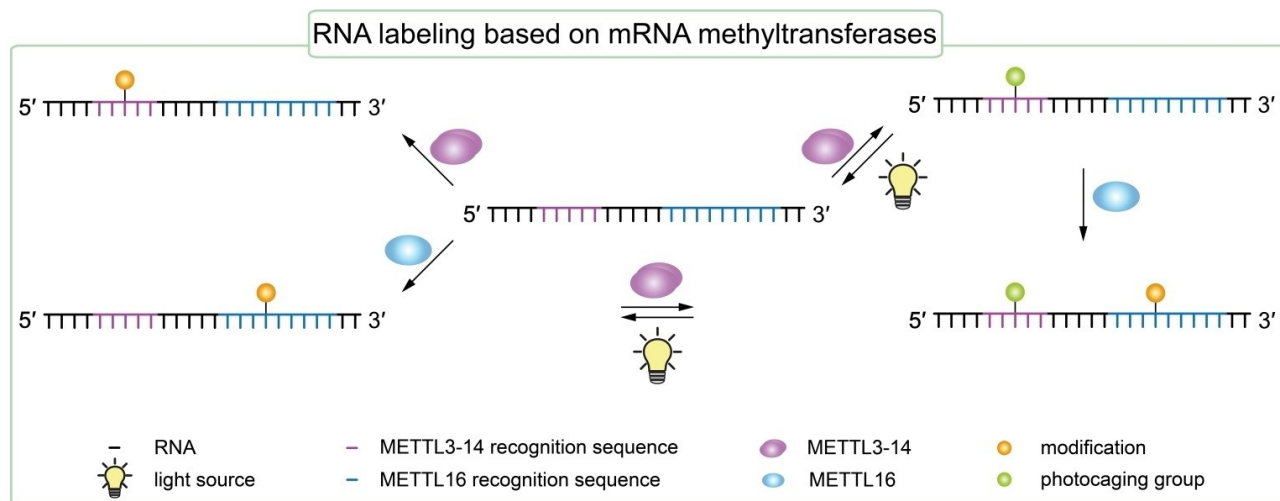


Figure 10. Tag-free internal RNA labeling based on mRNA methyltransferases METTL3-14 and METTL16. The approach of methyltransferase-directed transfer of activated groups (mTAG) can be used for covalent and sequence-specific modification of endogenous RNA.^[144] Methyltransferases are applied to transfer various functional groups from *S*-adenosyl-L-methionine (SAM or AdoMet) analogs to the N⁶-position of target adenosines within a RNA of interest. Therefore, a distinct recognition sequence defining the modification site for the methyltransferase needs to be included in the RNA sequence. Additionally, METTL3-14 is able to transfer photoactive benzylic AdoMet analogs like *ortho*-nitrobenzyl (ONB) and 6-nitropiperonyl (NP) groups. Upon irradiation with light, the previously attached photoactive modification is removed from the RNA of interest. Both methyltransferases, METTL3-14 and METTL16 can be used in a combined approach for orthogonal labeling of different sites within the same RNA strand. After sequential modification with a propargyl group and photoactive ONB group, the latter can be removed reversibly by irradiation with light.^[137]

groups, the selenium-based AdoMet analog SeAdoYn was used which was previously found to be more stable and to improve the transfer efficiency.^[158] Both, METTL3-14 and METTL16, were shown to transfer propargyl groups with efficiencies corresponding to either 82% or up to 69% relative yield compared to unmodified AdoMet, respectively. METTL3-14 additionally accepted photoactive benzylic AdoMet analogs like *ortho*-nitrobenzyl (ONB) and 6-nitropiperonyl (NP) groups for transfer with more than 50% relative yield compared to AdoMet. Moreover, METTL16 was able to transfer an azidobutenyl group, but only yielded low amounts of labeled RNA.^[137] Subsequent functionalization of propargylated and azidobutenyl-modified RNA with different reporter groups using either CuAAC or SPAAC, respectively, was successfully demonstrated.^[137] Eventually, 1900 nt long firefly luciferase (FLuc) mRNA which naturally comprises the DRACH motif was post-synthetically propargylated and then fluorescently click-labeled with Cy5-azide employing CuAAC.^[137] Additionally, innovative RNA photocaging experiments proved that the installation of ONB and NP groups by METTL3-14 on short, modified RNAs is reversible as both modifications can be removed by irradiation with UV light without damaging the RNA.^[137] It was further demonstrated, that both methyltransferases, METTL3-14 and METTL-16, can be used in combination for orthogonal labeling of a 26 nt long RNA at different sites. Sequential modification was performed first applying METTL16 to transfer a propargyl group followed by the transfer of an ONB group by METTL3-14.^[137] The presented labeling approach is promising for future application such as dual color labeling for FRET studies on endogenous RNA as well as activation of previously photocaged RNA. Site-specificity of the presented approach is limited as the required consensus motifs are spanning only a few nucleobases. Potential interference with consensus motifs naturally present in long, functional RNA sequences^[159] is imaginable and will cause unspecific labeling by methyltransferases at undesired sites. Referring to earlier *in vivo* metabolic labeling experiments with a propargyl-modified amino acid precursor as methyltransferase substrate eventually yielding propargylated RNA,^[160] an interesting adaptation for site-specific *in vivo* RNA labeling may be developed in the future. For this, care has to be taken to address undesired cross-interactions with cellular AdoMet or SAM substrates.

5. Summary and Outlook

In recent years, rapid progress has been made in the development of new methods to prepare large RNA molecules site-specifically decorated with reporter groups such as fluorophores, biotin or spin labels in the laboratory. Biophysical investigations, for example folding studies via FRET or EPR spectroscopy of large RNAs are feasible by now. Nevertheless, facile, homogenous and high-yielding labeling strategies for diagnostic and therapeutic applications still remain challenging.

The assembly of large RNAs via enzymatic ligation from short, synthetic oligonucleotides is still a widely used technique. Synthetic biology approaches to prepare site-specifically modi-

fied RNAs via *in vitro* transcription using an expanded genetic alphabet are now easier to implement and huge progress has been made in the development of novel catalytic nucleic acids for RNA labeling and RNA methyltransferase based labeling strategies. In the future, the developed approaches need to be extended allowing efficient intracellular labeling to study large naturally occurring RNA molecules in their native environment. Expanded genetic alphabet based labeling techniques and methods labeling endogenous RNAs will certainly pave the way for future *in cell* investigations to shine light on the complex folding pathways and the diverse regulatory functions of large non-coding RNA molecules in cells.

Acknowledgements

This work has received financial support from the Boehringer Ingelheim Stiftung. Open access funding enabled and organized by Projekt DEAL. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: click chemistry · bioorthogonal labeling · lncRNA · ribozymes · RNA labeling

- [1] a) J. S. Mattick, I. V. Makunin, *Hum. Mol. Genet.* **2006**, *15 Spec No 1*, R17–29; b) P. Kapranov, J. Cheng, S. Dike, D. A. Nix, R. Duttagupta, A. T. Willingham, P. F. Stadler, J. Hertel, J. Hackermüller, I. L. Hofacker, I. Bell, E. Cheung, J. Drenkow, E. Dumais, S. Patel, G. Helt, M. Ganesh, S. Ghosh, A. Piccolboni, V. Sementchenko, H. Tammanna, T. R. Gingeras, *Science* **2007**, *316*, 1484–1488.
- [2] S. Kirk, *Bioorg. Med. Chem.* **2001**, *9*, 2295–2301.
- [3] C. Höbartner, G. Sicoli, F. Wachowius, D. B. Gophane, S. T. Sigurdsson, *J. Org. Chem.* **2012**, *77*, 7749–7754.
- [4] T. Weinrich, M. Gränz, C. Grünewald, T. F. Prisner, M. W. Göbel, *Eur. J. Org. Chem.* **2017**, 491–496.
- [5] C. Domnick, F. Eggert, C. Wuebben, L. Bornewasser, G. Hagelueken, O. Schiemann, S. Kath-Schorr, *Angew. Chem. Int. Ed.* **2020**, *59*, 7891–7896; *Angew. Chem.* **2020**, *132*, 7965–7970.
- [6] N. Milisavljevič, P. Perliková, R. Pohl, M. Hocek, *Org. Biomol. Chem.* **2018**, *16*, 5800–5807.
- [7] S. G. Srivatsan, Y. Tor, *J. Am. Chem. Soc.* **2007**, *129*, 2044–2053.
- [8] M. Morohashi, M. Kimoto, A. Sato, R. Kawai, I. Hirao, *Molecules* **2012**, *17*, 2855–2876.
- [9] M. Kimoto, R. Yamashige, S. Yokoyama, I. Hirao, *J. Nucleic Acids* **2012**, *2012*, 230943.
- [10] K. Krell, D. Harijan, D. Ganz, L. Doll, H.-A. Wagenknecht, *Bioconjugate Chem.* **2020**, *31*, 990–1011.
- [11] R. K. Hartmann, A. Bindereif, A. Schön, E. Westhof, *Handbook of RNA Biochemistry*, Wiley, Hoboken, **2015**.
- [12] S. Wu, J. Wang, X. Pu, L. Li, Q. Li, *Biophys. J.* **2018**, *114*, 1755–1761.
- [13] J. T. George, S. G. Srivatsan, *Methods* **2017**, *120*, 28–38.
- [14] E. Paredes, M. Evans, S. R. Das, *Methods* **2011**, *54*, 251–259.
- [15] S. H. Weisbrod, A. Marx, *Chem. Commun.* **2008**, 5675–5685.
- [16] F. Eggert, S. Kath-Schorr, *Chem. Commun.* **2016**, *52*, 7284–7287.
- [17] a) K. Fauster, M. Hartl, T. Santner, M. Aigner, C. Kreutz, K. Bister, E. Ennifar, R. Micura, *ACS Chem. Biol.* **2012**, *7*, 581–589; b) M. B. Walunj, A. A. Tanpure, S. G. Srivatsan, *Nucleic Acids Res.* **2018**, *46*, e65.
- [18] E. Paredes, S. R. Das, *ChemBioChem* **2011**, *12*, 125–131.

- [19] A. M. Pyka, C. Domnick, F. Braun, S. Kath-Schorr, *Bioconjugate Chem.* **2014**, *25*, 1438–1443.
- [20] J. Schoch, S. Ameta, A. Jäschke, *Chem. Commun.* **2011**, *47*, 12536–12537.
- [21] S. Solomatin, D. Herschlag in *Methods in Enzymology*, v. 468, 469 (Ed.: D. Herschlag), Academic Press/Elsevier, San Diego, **2009**.
- [22] T. Ledoan, R. Auger, A. Benjahad, J. P. Tenu, *Nucleosides Nucleotides* **1999**, *18*, 277–289.
- [23] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021; *Angew. Chem.* **2001**, *113*, 2056–2075.
- [24] E. M. Sletten, C. R. Bertozzi, *Angew. Chem. Int. Ed.* **2009**, *48*, 6974–6998; *Angew. Chem.* **2009**, *121*, 7108–7133.
- [25] a) A. H. El-Sagheer, T. Brown, *Chem. Soc. Rev.* **2010**, *39*, 1388–1405; b) P. M. E. Gramlich, C. T. Wirges, A. Manetto, T. Carell, *Angew. Chem. Int. Ed.* **2008**, *47*, 8350–8358; *Angew. Chem.* **2008**, *120*, 8478–8487.
- [26] E. Paredes, S. R. Das, *Bioorg. Med. Chem. Lett.* **2012**, *22*, 5313–5316.
- [27] a) N. J. Agard, J. M. Baskin, J. A. Prescher, A. Lo, C. R. Bertozzi, *ACS Chem. Biol.* **2006**, *1*, 644–648; b) P. N. Asare-Okai, E. Agustin, D. Fabris, M. Royzen, *Chem. Commun.* **2014**, *50*, 7844–7847; c) D. C. Kennedy, C. S. McKay, M. C. B. Legault, D. C. Danielson, J. A. Blake, A. F. Pegoraro, A. Stolor, Z. Mester, J. P. Pezacki, *J. Am. Chem. Soc.* **2011**, *133*, 17993–18001.
- [28] N. J. Agard, J. A. Prescher, C. R. Bertozzi, *J. Am. Chem. Soc.* **2004**, *126*, 15046–15047.
- [29] C. P. Ramil, Q. Lin, *Chem. Commun.* **2013**, *49*, 11007–11022.
- [30] a) E. Kozma, O. Demeter, P. Kele, *ChemBioChem* **2017**, *18*, 486–501; b) A.-C. Knall, C. Slugovc, *Chem. Soc. Rev.* **2013**, *42*, 5131–5142.
- [31] B. L. Oliveira, Z. Guo, G. J. L. Bernardes, *Chem. Soc. Rev.* **2017**, *46*, 4895–4950.
- [32] N. K. Devaraj, S. Hilderbrand, R. Upadhyay, R. Mazitschek, R. Weissleder, *Angew. Chem. Int. Ed.* **2010**, *122*, 2931–2934.
- [33] M. L. Blackman, M. Royzen, J. M. Fox, *J. Am. Chem. Soc.* **2008**, *130*, 13518–13519.
- [34] C. S. Chow, S. K. Mahto, T. N. Lamichhane, *ACS Chem. Biol.* **2008**, *3*, 30–37.
- [35] M. Flamme, L. K. McKenzie, I. Sarac, M. Hollenstein, *Methods* **2019**, *161*, 64–82.
- [36] A. Somoza, *Chem. Soc. Rev.* **2008**, *37*, 2668–2675.
- [37] A. H. El-Sagheer, T. Brown, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 15329–15334.
- [38] J. M. Esquiaqui, E. M. Sherman, J.-D. Ye, G. E. Fanucci, *Biochemistry* **2016**, *55*, 4295–4305.
- [39] C. B. Reese, *Org. Biomol. Chem.* **2005**, *3*, 3851–3868.
- [40] a) Y. Shiba, H. Masuda, N. Watanabe, T. Ego, K. Takagaki, K. Ishiyama, T. Ohgi, J. Yano, *Nucleic Acids Res.* **2007**, *35*, 3287–3296; b) W. S. Marshall, R. J. Kaiser, *Curr. Opin. Chem. Biol.* **2004**, *8*, 222–229.
- [41] D. M. J. Lilley, *Curr. Opin. Struct. Biol.* **2005**, *15*, 313–323.
- [42] S. V. Park, J.-S. Yang, H. Jo, B. Kang, S. S. Oh, G. Y. Jung, *Biotechnol. Adv.* **2019**, *37*, 107452.
- [43] A. Serganov, D. J. Patel, *Nat. Rev. Genet.* **2007**, *8*, 776–790.
- [44] G. Mayer, *Angew. Chem. Int. Ed.* **2009**, *48*, 2672–2689; *Angew. Chem.* **2009**, *121*, 2710–2727.
- [45] R. Rieder, C. Höbartner, R. Micura, *Methods Mol. Biol.* **2009**, *540*, 15–24.
- [46] B. M. Akiyama, M. D. Stone in *Methods in Enzymology*, v. 468, 469 (Ed.: D. Herschlag), Academic Press/Elsevier, San Diego, Calif, **2009**.
- [47] K. Lang, R. Micura, *Nat. Protoc.* **2008**, *3*, 1457–1466.
- [48] M. J. Moore, C. C. Query, *Methods Enzymol.* **2000**, *317*, 109–123.
- [49] R. Welz, K. Bossmann, C. Klug, C. Schmidt, H.-J. Fritz, S. Müller, *Angew. Chem. Int. Ed.* **2003**, *42*, 2424–2427; *Angew. Chem.* **2003**, *115*, 2526–2530.
- [50] S. Vauléon, S. A. Ivanov, S. Gwiazda, S. Müller, *ChemBioChem* **2005**, *6*, 2158–2162.
- [51] W. E. Purtha, R. L. Coppins, M. K. Smalley, S. K. Silverman, *J. Am. Chem. Soc.* **2005**, *127*, 13124–13125.
- [52] S. K. Silverman, D. A. Baum in *Methods in Enzymology*, v. 468, 469 (Ed.: D. Herschlag), Academic Press/Elsevier, San Diego, **2009**.
- [53] K. Wawrzyniak-Turek, C. Höbartner, *Methods Enzymol.* **2014**, *549*, 85–104.
- [54] A. H. El-Sagheer, T. Brown in *Chemoselective and Bioorthogonal Ligation Reactions: Concepts and Applications* (Eds.: I. L. Medintz, P. Dawson, W. R. Algar), Wiley-VCH, Weinheim, **2017**.
- [55] J. Frommer, R. Hieronymus, T. Selvi Arunachalam, S. Heeren, M. Jenckel, A. Strahl, B. Appel, S. Müller, *RNA Biol.* **2014**, *11*, 609–623.
- [56] C. J. Kershaw, R. T. O’Keefe, *Methods Mol. Biol.* **2012**, *941*, 257–269.
- [57] N. M. Nichols, S. Tabor, L. A. McReynolds, *Curr. Protoc. Mol. Biol.* **2008**, Chapter 3, Unit3.15.
- [58] K. C. Suddala, J. Cabello-Villegas, M. Michnicka, C. Marshall, E. P. Nikonowicz, N. G. Walter, *Nat. Commun.* **2018**, *9*, 1896.
- [59] T. M. Henkin, *Biochim. Biophys. Acta.* **2014**, *1839*, 959–963.
- [60] M. J. Moore, P. A. Sharp, *Science* **1992**, *256*, 992–997.
- [61] C. Höbartner, R. Rieder, C. Kreutz, B. Puffer, K. Lang, A. Polonskaia, A. Serganov, R. Micura, *J. Am. Chem. Soc.* **2005**, *127*, 12035–12045.
- [62] C. Manz, A. Y. Kobitski, A. Samanta, B. G. Keller, A. Jäschke, G. U. Nienhaus, *Nat. Chem. Biol.* **2017**, *13*, 1172–1178.
- [63] T. Weinrich, E. A. Jaumann, U. Scheffer, T. F. Prisner, M. W. Göbel, *Chem. Eur. J.* **2018**, *24*, 6202–6207.
- [64] T. Kulinski, M. Olejniczak, H. Huthoff, L. Bielecki, K. Pachulska-Wieczorek, A. T. Das, B. Berkhout, R. W. Adamiak, *J. Biol. Chem.* **2003**, *278*, 38892–38901.
- [65] B. Chen, A. P. Longhini, F. Nußbaumer, C. Kreutz, J. D. Dinman, T. K. Dayie, *Chem. Eur. J.* **2018**, *24*, 5462–5468.
- [66] a) D. C. Chan, P. S. Kim, *Cell* **1998**, *93*, 681–684; b) S. A. Gallo, C. M. Finnegan, M. Viard, Y. Raviv, A. Dimitrov, S. S. Rawat, A. Puri, S. Durell, R. Blumenthal, *Biochim. Biophys. Acta Biomembr.* **2003**, *1614*, 36–50.
- [67] A. T. Belew, A. Meskauskas, S. Musalgaonkar, V. M. Advani, S. O. Sulima, W. K. Kasprzak, B. A. Shapiro, J. D. Dinman, *Nature* **2014**, *512*, 265–269.
- [68] M. R. Stark, J. A. Pleiss, M. Deras, S. A. Scaringe, S. D. Rader, *RNA* **2006**, *12*, 2014–2019.
- [69] M. Amitsur, R. Levitz, G. Kaufmann, *EMBO J.* **1987**, *6*, 2499–2503.
- [70] a) J. D. Bain, C. Switzer, *Nucleic Acids Res.* **1992**, *20*, 4372; b) M. R. Stark, S. D. Rader, *Methods Mol. Biol.* **2014**, *1126*, 137–149.
- [71] L. Büttner, J. Seikowski, K. Wawrzyniak, A. Ochmann, C. Höbartner, *Bioorg. Med. Chem.* **2013**, *21*, 6171–6180.
- [72] G. F. Joyce, *Annu. Rev. Biochem.* **2004**, *73*, 791–836.
- [73] A. Flynn-Charlebois, Y. Wang, T. K. Prior, I. Rashid, K. A. Hoadley, R. L. Coppins, A. C. Wolf, S. K. Silverman, *J. Am. Chem. Soc.* **2003**, *125*, 2444–2454.
- [74] Y. Wang, S. K. Silverman, *J. Am. Chem. Soc.* **2003**, *125*, 6880–6881.
- [75] a) D. P. Bartel, J. W. Szostak, *Science (New York, N. Y.)* **1993**, *261*, 1411–1418; b) L. Jaeger, M. C. Wright, G. F. Joyce, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 14712–14717; c) M. P. Robertson, J. R. Hesselberth, A. D. Ellington, *RNA* **2001**, *7*, 513–523; d) E. H. Eklund, J. W. Szostak, D. P. Bartel, *Science* **1995**, *269*, 364–370.
- [76] L. Bellon, C. Workman, J. Scherrer, N. Usman, F. Wincott, *J. Am. Chem. Soc.* **1996**, *118*, 3771–3772.
- [77] D. Jaikaran, M. D. Smith, R. Mehdizadeh, J. Olive, R. A. Collins, *RNA* **2008**, *14*, 938–949.
- [78] A. H. El-Sagheer, T. Brown, *Q. Rev. Biophys.* **2015**, *48*, 429–436.
- [79] a) I. D. Pokrovskaya, V. V. Gurevich, *Anal. Biochem.* **1994**, *220*, 420–423; b) F. T. Edelman, A. Niedner, D. Niessing, *Methods* **2014**, *65*, 333–341.
- [80] a) A. D. Keefe, S. T. Cload, *Curr. Opin. Chem. Biol.* **2008**, *12*, 448–456; b) C. Y. Jao, A. Salic, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 15779–15784.
- [81] S. R. Das, R. Fong, J. A. Piccirilli, *Curr. Opin. Chem. Biol.* **2005**, *9*, 585–593.
- [82] D. A. Malyshev, F. E. Romesberg, *Angew. Chem. Int. Ed.* **2015**, *54*, 11930–11944; *Angew. Chem.* **2015**, *127*, 12098–12113.
- [83] Y. J. Seo, G. T. Hwang, P. Ordoukhanian, F. E. Romesberg, *J. Am. Chem. Soc.* **2009**, *131*, 3246–3252.
- [84] L. Li, M. Degardin, T. Lavergne, D. A. Malyshev, K. Dhami, P. Ordoukhanian, F. E. Romesberg, *J. Am. Chem. Soc.* **2014**, *136*, 826–829.
- [85] Y. J. Seo, S. Matsuda, F. E. Romesberg, *J. Am. Chem. Soc.* **2009**, *131*, 5046–5047.
- [86] Z. Yang, A. M. Sismour, P. Sheng, N. L. Puskar, S. A. Benner, *Nucleic Acids Res.* **2007**, *35*, 4238–4249.
- [87] S. Hoshika, N. A. Leal, M.-J. Kim, M.-S. Kim, N. B. Karalkar, H.-J. Kim, A. M. Bates, N. E. Watkins, H. A. Santa Lucia, A. J. Meyer, S. Das Gupta, J. A. Piccirilli, A. D. Ellington, J. Santa Lucia, Jr., M. M. Georgiadis, S. A. Benner, *Science* **2019**, *363*, 884–887.
- [88] Y. J. Seo, D. A. Malyshev, T. Lavergne, P. Ordoukhanian, F. E. Romesberg, *J. Am. Chem. Soc.* **2011**, *133*, 19878–19888.
- [89] F. Eggert, K. Kulikov, C. Domnick, P. Leifels, S. Kath-Schorr, *Methods* **2017**, *120*, 17–27.
- [90] T. Ishizuka, M. Kimoto, A. Sato, I. Hirao, *Chem. Commun.* **2012**, *48*, 10835–10837.
- [91] Y. Tor, P. B. Dervan, *J. Am. Chem. Soc.* **1993**, *115*, 4461–4467.
- [92] C. Switzer, S. E. Moroney, S. A. Benner, *J. Am. Chem. Soc.* **1989**, *111*, 8322–8323.
- [93] J. A. Piccirilli, T. Krauch, S. E. Moroney, S. A. Benner, *Nature* **1990**, *343*, 33–37.

- [94] A. Hegelein, D. Müller, S. Größl, M. Göbel, M. Hengesbach, H. Schwalbe, *Chem. Eur. J.* **2020**, *26*, 1800–1810.
- [95] T. Fujiwara, M. Kimoto, H. Sugiyama, I. Hirao, S. Yokoyama, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2221–2223.
- [96] M. Kimoto, M. Endo, T. Mitsui, T. Okuni, I. Hirao, S. Yokoyama, *Chem. Biol.* **2004**, *11*, 47–55.
- [97] K. Moriyama, M. Kimoto, T. Mitsui, S. Yokoyama, I. Hirao, *Nucleic Acids Res.* **2005**, *33*, e129.
- [98] R. Kawai, M. Kimoto, S. Ikeda, T. Mitsui, M. Endo, S. Yokoyama, I. Hirao, *J. Am. Chem. Soc.* **2005**, *127*, 17286–17295.
- [99] I. Hirao, Y. Harada, M. Kimoto, T. Mitsui, T. Fujiwara, S. Yokoyama, *J. Am. Chem. Soc.* **2004**, *126*, 13298–13305.
- [100] a) S. Moran, R. X.-F. Ren, S. Rumney, E. T. Kool, *J. Am. Chem. Soc.* **1997**, *119*, 2056–2057; b) J. C. Morales, E. T. Kool, *Nat. Struct. Biol.* **1998**, *5*, 950–954; c) T. J. Matray, E. T. Kool, *J. Am. Chem. Soc.* **1998**, *120*, 6191–6192.
- [101] I. Hirao, M. Kimoto, T. Mitsui, T. Fujiwara, R. Kawai, A. Sato, Y. Harada, S. Yokoyama, *Nat. Methods* **2006**, *3*, 729–735.
- [102] R. Yamashige, M. Kimoto, Y. Takezawa, A. Sato, T. Mitsui, S. Yokoyama, I. Hirao, *Nucleic Acids Res.* **2012**, *40*, 2793–2806.
- [103] T. Someya, A. Ando, M. Kimoto, I. Hirao, *Nucleic Acids Res.* **2015**, *43*, 6665–6676.
- [104] M. Kimoto, A. J. Meyer, I. Hirao, A. D. Ellington, *Chem. Commun.* **2017**, *53*, 12309–12312.
- [105] A. W. Feldman, F. E. Romesberg, *Acc. Chem. Res.* **2018**, *51*, 394–403.
- [106] T. Lavergne, R. Lamichhane, D. A. Malyshev, Z. Li, L. Li, E. Sperling, J. R. Williamson, D. P. Millar, F. E. Romesberg, *ACS Chem. Biol.* **2016**, *11*, 1347–1353.
- [107] C. Domnick, F. Eggert, S. Kath-Schorr, *Chem. Commun.* **2015**, *51*, 8253–8256.
- [108] D. M. Chadalavada, E. A. Gratton, P. C. Bevilacqua, *Biochemistry* **2010**, *49*, 5321–5330.
- [109] W. C. Winkler, A. Nahvi, A. Roth, J. A. Collins, R. R. Breaker, *Nature* **2004**, *428*, 281–286.
- [110] C. Domnick, G. Hagelueken, F. Eggert, O. Schiemann, S. Kath-Schorr, *Org. Biomol. Chem.* **2019**, *17*, 1805–1808.
- [111] R. Fang, W. N. Moss, M. Rutenberg-Schoenberg, M. D. Simon, *PLoS Genet.* **2015**, *11*, e1005668.
- [112] Y. Wang, V. Kathiresan, Y. Chen, Y. Hu, W. Jiang, G. Bai, G. Liu, P. Z. Qin, X. Fang, *Chem. Sci.* **2020**, *11*, 9655–9664.
- [113] A. Samanta, A. Krause, A. Jäschke, *Chem. Commun.* **2014**, *50*, 1313–1316.
- [114] a) F. Huang, J. He, Y. Zhang, Y. Guo, *Nat. Protoc.* **2008**, *3*, 1848–1861; b) M. Hafner, E. Vianini, B. Albertoni, L. Marchetti, I. Grüne, C. Gloeckner, M. Famulok, *Nat. Protoc.* **2008**, *3*, 579–587; c) J. Wolf, V. Dombos, B. Appel, S. Müller, *Org. Biomol. Chem.* **2008**, *6*, 899–907; d) N. Li, C. Yu, F. Huang, *Nucleic Acids Res.* **2005**, *33*, e37; e) J. C. Macosko, M. S. Pio, I. Tinoco, Y. K. Shin, *RNA* **1999**, *5*, 1158–1166.
- [115] I. Lebars, B. Vilen, S. Bourbigot, P. Turek, P. Wolff, B. Kieffer, *Nucleic Acids Res.* **2014**, *42*, e117.
- [116] Y. Liu, E. Holmstrom, J. Zhang, P. Yu, J. Wang, M. A. Dyba, de Chen, J. Ying, S. Lockett, D. J. Nesbitt, A. R. Ferré-D'Amaré, R. Sousa, J. R. Stagno, Y.-X. Wang, *Nature* **2015**, *522*, 368–372.
- [117] a) T. H. Tahir, D. Temiakov, M. Anikin, V. Patlan, W. T. McAllister, D. G. Vassilyev, S. Yokoyama, *Nature* **2002**, *420*, 43–50; b) Y. Sohn, H. Shen, C. Kang in *Methods Enzymol.* Elsevier, **2003**.
- [118] X. Zhang, M. Li, Y. Liu, *RNA Biol.* **2020**, *17*, 1009–1017.
- [119] Y. Liu, E. Holmstrom, P. Yu, K. Tan, X. Zuo, D. J. Nesbitt, R. Sousa, J. R. Stagno, Y.-X. Wang, *Nat. Protoc.* **2018**, *13*, 987–1005.
- [120] Y. Liu, P. Yu, M. Dyba, R. Sousa, J. R. Stagno, Y.-X. Wang, *Methods* **2016**, *103*, 4–10.
- [121] J. R. Stagno, P. Yu, M. A. Dyba, Y.-X. Wang, Y. Liu, *PLoS One* **2019**, *14*, e0215555.
- [122] D. Eglhoff, I. A. Oleinich, E. Freisinger, *ACS Chem. Biol.* **2015**, *10*, 547–553.
- [123] D. Eglhoff, I. A. Oleinich, M. Zhao, S. L. B. König, R. K. O. Sigel, E. Freisinger, *ACS Chem. Biol.* **2016**, *11*, 2558–2567.
- [124] K. Jahn, E. M. Olsen, M. M. Nielsen, T. Tørring, R. Mohammad Zadegan, E. S. Andersen, K. V. Gothelf, J. Kjems, *Bioconjugate Chem.* **2011**, *22*, 95–100.
- [125] A. Kadina, A. M. Kietrys, E. T. Kool, *Angew. Chem. Int. Ed.* **2018**, *57*, 3059–3063; *Angew. Chem.* **2018**, *130*, 3113–3117.
- [126] R. C. Spitale, R. A. Flynn, Q. C. Zhang, P. Crisalli, B. Lee, J.-W. Jung, H. Y. Kuchelmeister, P. J. Batista, E. A. Torre, E. T. Kool, H. Y. Chang, *Nature* **2015**, *519*, 486–490.
- [127] L. Xiao, M. Habibian, E. T. Kool, *J. Am. Chem. Soc.* **2020**, *142*, 16357–16363.
- [128] M. Zhao, F. D. Steffen, R. Börner, M. F. Schaffer, R. K. O. Sigel, E. Freisinger, *Nucleic Acids Res.* **2018**, *46*, e13.
- [129] D. A. Baum, S. K. Silverman, *Angew. Chem. Int. Ed.* **2007**, *46*, 3502–3504; *Angew. Chem.* **2007**, *119*, 3572–3574.
- [130] L. Büttner, F. Javadi-Zarnaghi, C. Höbartner, *J. Am. Chem. Soc.* **2014**, *136*, 8131–8137.
- [131] T. J. Carrocci, L. Lohe, M. J. Ashton, C. Höbartner, A. A. Hoskins, *Chem. Commun.* **2017**, *53*, 11992–11995.
- [132] F. Javadi-Zarnaghi, C. Höbartner, *J. Am. Chem. Soc.* **2013**, *135*, 12839–12848.
- [133] M. Ghaem Maghami, C. P. M. Scheitl, C. Höbartner, *J. Am. Chem. Soc.* **2019**, *141*, 19546–19549.
- [134] M. Ghaem Maghami, S. Dey, A.-K. Lenz, C. Höbartner, *Angew. Chem.* **2020**, *132*, 9421–9425; *Angew. Chem. Int. Ed.* **2020**, *59*, 9335–9339.
- [135] C. P. M. Scheitl, M. G. Maghami, A.-K. Lenz, C. Höbartner, *Nature* **2020**, *587*, 663–667.
- [136] Y. Motorin, J. Burhenne, R. Teimer, K. Koynov, S. Willnow, E. Weinhold, M. Helm, *Nucleic Acids Res.* **2011**, *39*, 1943–1952.
- [137] A. Ovcharenko, F. P. Weissenboeck, A. Rentmeister, *Angew. Chem. Int. Ed.* **2021**, *60*, 4098–4103; *Angew. Chem.* **2021**, *133*, 4144–4149.
- [138] X. Shu, Q. Dai, T. Wu, I. R. Bothwell, Y. Yue, Z. Zhang, J. Cao, Q. Fei, M. Luo, C. He, J. Liu, *J. Am. Chem. Soc.* **2017**, *139*, 17213–17216.
- [139] G. Lukinavičius, V. Lapiene, Z. Stasevskij, C. Dalhoff, E. Weinhold, S. Klimasauskas, *J. Am. Chem. Soc.* **2007**, *129*, 2758–2759.
- [140] M. Tomkuvienė, B. Clouet-d'Orval, I. Cerniauskas, E. Weinhold, S. Klimasauskas, *Nucleic Acids Res.* **2012**, *40*, 6765–6773.
- [141] a) J. M. Holstein, L. Anhäuser, A. Rentmeister, *Angew. Chem. Int. Ed.* **2016**, *55*, 10899–10903; *Angew. Chem.* **2016**, *128*, 11059–11063; b) J. M. Holstein, D. Schulz, A. Rentmeister, *Chem. Commun.* **2014**, *50*, 4478–4481; c) J. M. Holstein, D. Stummer, A. Rentmeister, *Protein Eng. Des. Sel.* **2015**, *28*, 179–186; d) J. M. Holstein, D. Stummer, A. Rentmeister, *Chem. Sci.* **2015**, *6*, 1362–1369; e) F. Muttach, N. Muthmann, D. Reichert, L. Anhäuser, A. Rentmeister, *Chem. Sci.* **2017**, *8*, 7947–7953; f) F. Muttach, A. Rentmeister, *Angew. Chem. Int. Ed.* **2016**, *55*, 1917–1920; *Angew. Chem.* **2016**, *128*, 1951–1954; g) D. Schulz, J. M. Holstein, A. Rentmeister, *Angew. Chem. Int. Ed.* **2013**, *52*, 7874–7878; *Angew. Chem.* **2013**, *125*, 8028–8032.
- [142] a) M. Mickute, M. Nainyte, L. Vasiliauskaitė, A. Plotnikova, V. Masevičius, S. Klimašauskas, G. Vilkaitis, *Nucleic Acids Res.* **2018**, *46*, e104; b) A. Osipenko, A. Plotnikova, M. Nainyte, V. Masevičius, S. Klimašauskas, G. Vilkaitis, *Angew. Chem. Int. Ed.* **2017**, *56*, 6507–6510; *Angew. Chem.* **2017**, *129*, 6607–6610; c) A. Plotnikova, A. Osipenko, V. Masevičius, G. Vilkaitis, S. Klimašauskas, *J. Am. Chem. Soc.* **2014**, *136*, 13550–13553.
- [143] L. Anhäuser, S. Hüwel, T. Zobel, A. Rentmeister, *Nucleic Acids Res.* **2019**, *47*, e42.
- [144] a) R. I. McDonald, J. P. Guiling, S. Mukherji, E. A. Curtis, W. I. Lee, D. R. Liu, *Nat. Chem. Biol.* **2014**, *10*, 1049–1054; b) A. K. Sharma, J. J. Plant, A. E. Rangel, K. N. Meek, A. J. Anamisis, J. Hollien, J. M. Heemstra, *ACS Chem. Biol.* **2014**, *9*, 1680–1684.
- [145] a) S. C. Alexander, K. N. Busby, C. M. Cole, C. Y. Zhou, N. K. Devaraj, *J. Am. Chem. Soc.* **2015**, *137*, 12756–12759; b) F. Ehret, C. Y. Zhou, S. C. Alexander, D. Zhang, N. K. Devaraj, *Mol. Pharm.* **2018**, *15*, 737–742; c) D. Zhang, C. Y. Zhou, K. N. Busby, S. C. Alexander, N. K. Devaraj, *Angew. Chem. Int. Ed.* **2018**, *57*, 2822–2826; *Angew. Chem.* **2018**, *130*, 2872–2876; d) C. Y. Zhou, S. C. Alexander, N. K. Devaraj, *Chem. Sci.* **2017**, *8*, 7169–7173.
- [146] F. Li, J. Dong, X. Hu, W. Gong, J. Li, J. Shen, H. Tian, J. Wang, *Angew. Chem. Int. Ed.* **2015**, *54*, 4597–4602; *Angew. Chem.* **2015**, *127*, 4680–4685.
- [147] Y. Motorin, M. Helm, *Wiley Interdiscip. Rev.: RNA* **2011**, *2*, 611–631.
- [148] W. A. Velema, E. T. Kool, *Nat. Chem. Rev.* **2020**, *4*, 22–37.
- [149] C. Höbartner, S. K. Silverman, *Angew. Chem. Int. Ed.* **2007**, *46*, 7420–7424; *Angew. Chem.* **2007**, *119*, 7564–7568.
- [150] B. Chen, X. Zuo, Y.-X. Wang, T. K. Dayie, *Nucleic Acids Res.* **2012**, *40*, 3117–3130.
- [151] a) R. Mourani, M. J. Damha, *Nucleosides Nucleotides Nucleic Acids* **2006**, *25*, 203–229; b) N. E. Clark, A. Katolik, K. M. Roberts, A. B. Taylor, S. P. Holloway, J. P. Schuermann, E. J. Montemayor, S. W. Stevens, P. F. Fitzpatrick, M. J. Damha, P. J. Hart, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 14727–14732.
- [152] G. Martin, W. Keller, *RNA* **1998**, *4*, 226–230.
- [153] M. Grammel, H. Hang, N. K. Conrad, *ChemBioChem* **2012**, *13*, 1112–1115.

- [154] a) E. de Clercq, A. Holý, *Nat. Rev. Drug Discovery* **2005**, *4*, 928–940; b) U. Pradere, E. C. Garnier-Amblard, S. J. Coats, F. Amblard, R. F. Schinazi, *Chem. Rev.* **2014**, *114*, 9154–9218.
- [155] T. G. Waddell, L. L. Eilders, B. P. Patel, M. Sims, *Origins Life Evol. Biospheres* **2000**, *30*, 539–548.
- [156] P. Wang, K. A. Doxtader, Y. Nam, *Mol. Cell* **2016**, *63*, 306–317.
- [157] a) K. E. Pendleton, B. Chen, K. Liu, O. V. Hunter, Y. Xie, B. P. Tu, N. K. Conrad, *Cell* **2017**, *169*, 824–835.e14; b) H. Shima, M. Matsumoto, Y. Ishigami, M. Ebina, A. Muto, Y. Sato, S. Kumagai, K. Ochiai, T. Suzuki, K. Igarashi, *Cell Rep.* **2017**, *21*, 3354–3363; c) A. Ruskowska, M. Ruskowski, Z. Dauter, J. A. Brown, *Sci. Rep.* **2018**, *8*, 5311; d) K. A. Doxtader, P. Wang, A. M. Scarborough, D. Seo, N. K. Conrad, Y. Nam, *Mol. Cell* **2018**, *71*, 1001–1011.e4; e) A. S. Warda, J. Kretschmer, P. Hackert, C. Lenz, H. Urlaub, C. Höbartner, K. E. Sloan, M. T. Bohnsack, *EMBO Rep.* **2017**, *18*, 2004–2014.
- [158] a) J. M. Holstein, F. Muttach, S. H. H. Schiefelbein, A. Rentmeister, *Chem. Eur. J.* **2017**, *23*, 6165–6173; b) S. Willnow, M. Martin, B. Lüscher, E. Weinhold, *ChemBioChem* **2012**, *13*, 1167–1173.
- [159] a) D. Dominissini, S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, S. Osenberg, K. Cesarkas, J. Jacob-Hirsch, N. Amariglio, M. Kupiec, R. Sorek, G. Rechavi, *Nature* **2012**, *485*, 201–206; b) K. D. Meyer, Y. Saletore, P. Zumbo, O. Elemento, C. E. Mason, S. R. Jaffrey, *Cell* **2012**, *149*, 1635–1646.
- [160] K. Hartstock, B. S. Nilges, A. Ovcharenko, N. V. Cornelissen, N. Püllen, A.-M. Lawrence-Dörner, S. A. Leidel, A. Rentmeister, *Angew. Chem. Int. Ed.* **2018**, *57*, 6342–6346; *Angew. Chem.* **2018**, *130*, 6451–6455.

Manuscript received: April 4, 2021

Revised manuscript received: May 26, 2021

Accepted manuscript online: May 27, 2021

Version of record online: June 17, 2021