**Univerzita Karlova**

**Přírodovědecká fakulta**

Studijní program: Organická chemie

Studijní obor: Organická chemie

*Study programme: Organic chemistry*

*Branch of study: Organic chemistry*


**Mgr. Mária Brunderová**

# Reaktivní modifikace RNA pro biokonjugace s proteiny a nové enzymové metody pro syntézu RNA s modifikovanými bázemi

*Reactive modifications of RNA for bioconjugations with proteins and new enzymatic methods for the synthesis of base-modified RNA*

Disertační práce

*Doctoral thesis*

Vedoucí práce/*Supervisor:* prof. Ing. Michal Hocek, CSc., DSc.:

Praha, 2023

Disertační práce byla vypracována na Ústavu organické chemie a biochemie Akademie věd České republiky v Praze v období září 2019 – září 2023.

## Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce, ani její podstatná část, nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, dne……

……

Mgr. Mária Brunderová

*Failure should be our teacher, not our undertaker.*
*Failure is delay, not defeat. It is a temporary detour, not a dead end.*
*Failure is something we can avoid only by saying nothing, doing nothing, and being nothing.*

*Denis Waitley*

## Acknowledgement

Most importantly, I would like to express my deepest gratitude to my supervisor, prof. Michal Hocek, not only for the opportunity to work in his group on several challenging projects, that contributed to the development of my scientific skills, but also for his help, support, patience, and useful tips throughout my entire doctoral study.

I would also like to extend my thanks to all my collaborators, whose work was an integral part of my projects. My thanks go to Vojtěch Havlíček and Ján Matyašovský for synthesis of several compounds included in this study. Additionally, I appreciate the contribution of members of the MS department – Marta Vlková, Martin Hubálek, Kvetoslava Kertisová, Kateřina Nováková, and Edita Kofroňová, and the NMR department – Lenka Poštová Slavětínská and Radek Pohl of the IOCB. They assisted me with proteomic analysis, HR-ESI-MS measurements, MS-MALDI spectra acquisition, and NMR characterisation of prepared compounds.

My gratitude extends to all members of the Hocek group for creating an inspiring working environment, with a special acknowledgment to Nemanja Milisavljević, for his advice, patience, and for sharing his knowledge and experience, and help particularly during my first RNA transcription experiments.

My very special thanks go to Matouš Krömer. Not only for his always useful advice or tips, unforgettable happy moments, and never-ending inspiring scientific talks, but especially for his help, and support during my low and frustrating moments in my career, when nothing seemed to work and for his patience when teaching and explaining me new things.

Last but not least, my deepest gratitude goes to my family, especially to my mom, dad, and brother, for always believing in me and my abilities. I am thankful for their support, love, patience, and understanding throughout my entire Ph.D. study.

This work was supported by the Czech Science Foundation; EXPRO (20-00885X) and by the European Regional Development Fund; OP RDE (CZ.02.1.01/0.0/0.0/16_019/0000729).

## Abstract

This doctoral thesis focuses on the enzymatic synthesis of base-modified RNA probes with diverse functional groups, including reactive cross-linking, hydrophobic and fluorescent moieties, or affinity tags. The construction of nucleobase-modified oligonucleotides is accomplished either through conventional *in vitro* transcription with T7 RNA polymerase or by an innovative approach leveraging engineered mutant DNA polymerases and primer extension reaction (PEX).

In the first section of the thesis, a novel ribonucleoside triphosphate building block with reactive chloroacetamide functionality was synthesised using an aqueous Pd-catalysed Sonogashira cross-coupling reaction, directly applied on iodinated nucleotide. The chloroacetamide modified triphosphate was then tested as a putative substrate for T7 RNA polymerase in *in vitro* transcription reaction, aiming to construct RNA probes with one or multiple reactive groups. The selectivity of chloroacetamide-modified RNA for thiol-, or cysteine-, and histidine-containing (bio)molecules was demonstrated by model bioconjugation reactions and cross-linking experiments with three RNA-binding proteins of diverse structures and functions. The efficient formation of RNA-protein covalent adducts was confirmed by western blot or gel, and mass spectrometry analyses conducted under denaturing conditions. Identification of the RNA-binding sites and targeted amino acids residues within the model proteins was performed through proteomic analysis of cleavage products. The effectiveness and potential applicability of the novel modified RNA probe was further demonstrated by cross-linking with extracted cellular proteins, where sequence specific protein recognised and cross-linked only to its cognate modified RNA probe.

In the second part of the thesis, a novel methodology was developed for constructing base-modified RNA oligonucleotides with varying lengths and number of unnatural nucleotides utilising engineered thermostable DNA polymerases and PEX. Modified nucleotides, featuring diverse functional groups attached to position 5- of pyrimidines or position 7- of 7-deazapurines, were investigated as suitable substrates for two engineered polymerases – TGK and SFM4-3. Both mutant polymerases demonstrated efficiency in incorporating either one or four modified nucleotides into a single RNA strand. The increased activity of TGK polymerase was showcased by synthesising fully modified RNA probes, wherein all four canonical nucleotides were replaced with modified counterparts bearing distinct functional groups. This surpassed the performance of the conventional T7 RNA polymerase *in vitro* transcription. Conversely, SFM4-3 polymerase failed in achieving this level of activity. Efficient elongation was also demonstrated with a DNA primer, facilitating the construction of intriguing DNA-RNA hybrids. Moreover, a straightforward methodology utilising selective 2'-deoxyuridine cleavage was developed to enable the removal of DNA primer from the synthesised part of RNA, facilitating the generation of hypermodified RNA probes where each

nucleobase is modified. Additionally, a method for more challenging the selective labelling of RNA at specific positions, utilising combination of single nucleotide incorporation (SNI) and PEX, was successfully developed. RNA probes decorated with two distinct fluorophores at internal positions were employed for fluorescence structural studies. Importantly, this approach was also applied to the synthesis of region-modified or point-modified messenger RNAs, revealing for the first time, that a single 5-methylcytidine modification within protein-coding region significantly enhanced protein production in both *in vivo* and *in cellulo* experiments when compared to its natural or fully modified mRNA counterparts.

# Abstrakt

Tato disertační práce je zaměřena na enzymovou syntézu na bázi modifikovaných RNA s různými funkčními skupinami, včetně reaktivních pro síťování proteinů, hydrofobních a fluorescenčních skupin nebo afinitních značek. Konstrukce oligonukleotidů modifikovaných na bázi je zabezpečena buď konvenční *in vitro* transkripcí s T7 RNA polymerázou, nebo inovativním přístupem využívajícím upravené mutantní DNA polymerázy a reakci prodlužování primerů (PEX).

V první části práce byl nasyntetizován nový ribonukleosid trifosfátový stavební blok s reaktivní chloracetamidovou funkční skupinou pomocí Sonogashirovy kaplingové reakce ve vodní fázi katalyzované Pd, přímo aplikované na jodovaný nukleotid. Chloracetamidem modifikovaný trifosfát byl poté testován jako vhodný substrát pro *in vitro* transkripci s T7 RNA polymerázou s cílem zkonstruovat RNA sondy s jednou nebo více reaktivními skupinami. Selektivita chloracetamidem modifikovaných RNA pro thiol, nebo cystein a histidin obsahující (bio)molekuly byla ukázána modelovými biokonjugačními reakcemi a experimenty pro síťování se třemi RNA-vazebnými proteiny s různými strukturami a funkcemi. Účinná tvorba kovalentních RNA-protein aduktů byla potvrzena western blotem, gelovou elektroforézou a hmotnostní analýzou, které byly prováděny za denaturačních podmínek. Identifikace RNA-vazebných míst a cílových aminokyselinových zbytků v modelových proteinech byla provedena pomocí proteomické analýzy produktů štěpení. Účinnost a potenciální použitelnost nové modifikované RNA sondy byla dále předvedena na příkladu síťování s extrahovanými buněčnými proteiny, kde vybraný sekvenčně-specifický protein rozpoznal a kovalentně zachytil pouze RNA sondu obsahující jeho cílovou sekvenci.

Ve druhé části práce byla vyvinuta nová metodika pro konstrukci na bázi modifikovaných RNA oligonukleotidů s různou délkou a počtem nepřirozených nukleotidů s využitím uměle upravených termostabilních DNA polymeráz a PEX reakce. Modifikované nukleotidy obsahující různé funkční skupiny připojené k pozici 5 u pyrimidinů nebo k pozici 7 u 7-deazapurinů byly zkoumány jako vhodné substráty pro dvě upravené polymerázy – TGK a SFM4-3. Obě mutantní polymerázy prokázaly účinnost při začleňování jak jednoho, tak čtyř modifikovaných nukleotidů do jednoho řetězce RNA. Zvýšená aktivita TGK polymerázy byla ukázána na syntéze plně modifikovaných RNA sond, kde byly všechny čtyři kanonické nukleotidy nahrazeny modifikovanými protějšky nesoucími vzájemně odlišné funkční skupiny. Tento výsledek překonal konvenční *in vitro* transkripci s T7 RNA polymerázou. Naopak polymeráza SFM4-3 nedosáhla této úrovně aktivity. Účinné prodloužení bylo provedeno také v případě DNA primeru, pro konstrukci neobvyklých DNA-RNA hybridů. Navíc byla vyvinuta přímočará metodika využívající selektivního vyštěpení 2'-deoxyuridinu, která umožňuje odstranění DNA primeru ze syntetizované části RNA, což

usnadňuje tvorbu hypermodifikovaných RNA sond, kde je každá nukleobáze modifikována. Dále byla úspěšně vyvinuta metoda pro poměrně náročné selektivní značení RNA na specifických pozicích využívající kombinaci inkorporace jednoho nukleotidu (SNI) a PEX. Pro strukturní fluorescenční studie byly použity interně značené RNA sondy se dvěma odlišnými fluorofory. Je třeba zdůraznit, že tento přístup byl také aplikován pro syntézu messengerových RNA modifikovaných v určité oblasti nebo v bodě, přičemž bylo poprvé odhaleno, že jediná modifikace v podobě 5-methylcytidinu v oblasti kódující pro protein významně zvýšila produkci proteinu jak v *in vivo*, tak v buněčných experimentech ve srovnání s jeho přirozenými nebo plně modifikovanými mRNA protějšky.

## Publications of the author incorporated in this thesis

Brunderová, M.; Krömer, M.; Vlková, M. and Hocek, M.: "Chloroacetamide-Modified Nucleotide and RNA for Bioconjugations and Cross-Linking with RNA-Binding Proteins" *Angew. Chem. Int. Ed.* **2023**, *62*, e202213764.

## Publications of the author not incorporated in this thesis

Krömer, M.♦; Brunderová, M.♦; Ivancová, I.; Poštová Slavetínská, L.; Hocek, M.: "2-Formyl-dATP as Substrate for Polymerase Synthesis of Reactive DNA bearing an Aldehyde Group in the Minor Groove" *ChemPlusChem* **2020**, *85*, 1164.

♦These authors contributed equally.

## List of abbreviations

| | |
|---|---|
| Arg | arginine |
| ASO | antisense oligonucleotides |
| bp | base pair |
| BSA | bovine serum albumin |
| CA | chloroacetamide |
| cDNA | complementary DNA |
| circRNA | circular RNA |
| CLIP | cross-linking immunoprecipitation |
| CRISPR | clustered regularly interspaced short palindromic repeats |
| crRNA | CRISPR RNA |
| CuAAC | copper(I)-catalysed azide-alkyne cycloaddition |
| Cy3 | cyanine-3 |
| Cy5 | cyanine-5 |
| Cys | cysteine |
| DBPs | DNA-binding proteins |
| DIPEA | *N,N'*-diisopropylethylamine |
| DMEDA | *N,N'*-dimethylethylenediamine |
| DMSO | dimethyl sulfoxide |
| DMTr | dimethoxytrityl |
| dNTPs | natural 2'-deoxyribonucleotides |
| **dN$^{X}$TP**s | modified 2'-deoxyribonucleotides |
| dPAGE | denaturing polyacrylamide gel electrophoresis |
| dsDNA | double-stranded DNA |
| dsRNA | double-stranded RNA |
| DTT | dithiothreitol |
| dU | 2'-deoxyuridine |
| eCLIP | enhanced cross-linking immunoprecipitation |
| EDC | 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide |
| EMSA | electrophoretic mobility shift assay |
| ESI-LC-MS | electrospray ionisation liquid chromatography-mass spectrometry |
| FAM | fluorescein amidite |
| FL | fluorescein |
| FRET | Förster (fluorescence) resonance energy transfer |

| | |
|---|---|
| Gal1 | galectin 1 |
| GalNAc | *N*-acetylgalactosamine |
| GSH | reduced glutathione |
| H2A | histone H2A |
| hAgo2 | human argonaute 2 |
| HF | hydrofluoric acid |
| HITS-CLIP | high throughput sequencing cross-linking immunoprecipitation |
| HIV-RT | HIV reverse transcriptase |
| HPLC | high-performance liquid chromatography |
| HuR | human antigen R |
| IEDDA | inverse electron-demand Diels-Alder |
| kb | kilobase |
| LNA | locked nucleic acid |
| lncRNA | long non-coding RNA |
| Lys | lysine |
| lysoz. | lysozyme |
| m1Ψ | $N^1$-methylpseudouridine |
| $m^5C$ | 5-methylcytosine |
| $m^6A$ | $N^6$-methyladenosine |
| $m^7G$ | $N^7$-methylguanosine |
| miRNA | microRNA |
| mRNA | messenger RNA |
| MS-MALDI-TOF | matrix-assisted laser desorption/ionization time-of-flight |
| nPAGE | native polyacrylamide gel electrophoresis |
| nt | nucleotide |
| PAR-CLIP | photoactivatable ribonucleoside-enhanced CLIP |
| PCR | polymerase chain reaction |
| PCT | polymerase chain transcription |
| PEX | primer extension |
| PLOR | position-selective RNA labelling |
| **$rA^{CA}TP$** | CA-modified ribonucleotide triphosphate |
| **$rA^{I}TP$** | 7-iodo-7-deazaadenosine-5′-*O*-triphosphate |
| RBPs | RNA-binding proteins |
| RNAP | RNA polymerase |

| | |
|---|---|
| **rN$^{I}$TP**s | iodinated ribonucleotides |
| rNTPs | natural ribonucleotides |
| **rN$^{X}$TP**s | modified ribonucleotides |
| rRNA | ribosomal RNA |
| rSAP | shrimp alkaline phosphatase |
| s2U | 2-thiouridine |
| SELEX | systematic evolution of ligands by exponential enrichment |
| sgRNA | single guide RNA |
| siRNA | small interfering RNA |
| SNI | single nucleotide incorporation |
| snoRNA | small nucleolar RNA |
| snRNA | small nuclear RNA |
| SPAAC | strain-promoted azide-alkyne cycloaddition |
| SSB | single-strand binding protein |
| ssDNA | single-stranded DNA |
| ssRNA | single-stranded RNA |
| T7 RNAP | T7 RNA polymerase |
| TEAA | triethylammonium acetate buffer |
| TPPTS | 3,3′,3″-phosphanetriyltris(benzenesulfonic acid) trisodium salt |
| tracrRNA | trans-activating RNA |
| tRNA | transfer RNA |
| Trp | tryptophan |
| Tyr | tyrosine |
| UDG | uracil-DNA glycosylase |
| UTRs | untranslated regions |
| WB | western blot analysis |
| XNA | xenonucleic acids |
| Ψ | pseudouridine |

# Table of contents

# 1 Introduction

## 1.1 History of ribonucleic acid (RNA)

The discovery of nucleic acids is dated back to 1869, when Swiss biochemist Friedrich Miescher isolated a novel substance from nuclei of white blood cells and named it "nuclein"[1]. This initial finding set the stage for further exploration. The first evidence of RNA, playing a pivotal role in flow of genetic information, is in Crick's description of his "Central Dogma of Molecular Biology" during year 1953, which outlines that genetic information in DNA is transcribed into RNA, and subsequently, RNA is translated into proteins[2] (Figure 1). This was later confirmed by the discovery of mRNA by S. Brenner and F. Gros[3,4]. In 1956, Rich and Davies made another significant contribution by publishing pioneering work on nucleic acid hybridisation reactions. Their groundbreaking findings revealed that two RNA strands, much like DNA, could adopt a comparable configuration through complementarity in base-pairing[5]. This discovery highlighted the structural similarities between RNA and DNA, exploring new possibilities for understanding the complexity of nucleic acids and paving the way for subsequent investigations into their structure and function.

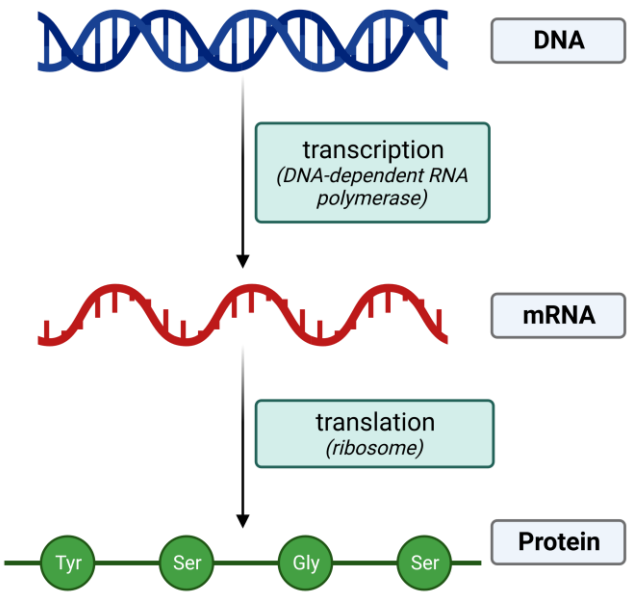



**Figure 1. Predicted flow of genetic information in a biological system by Francis Crick.**

## 1.2 Structure and functions of RNA

From the chemical point of view, RNA closely resembles its DNA counterpart. This polynucleotide biomacromolecule is composed of four canonical building blocks known as

ribonucleoside triphosphates (ribonucleotides). The heterocyclic nucleobases can be divided into two groups, purines and pyrimidines. Adenine (A), uracil (U), cytosine (C), and guanine (G) nucleobases are attached to 1'-position of a five-carbon ring (ribose sugar) through a C-N glycosidic bond. The 5'-carbon bears a triphosphate functional group. The presence of 2'-hydroxyl group in RNA imparts nucleophilic character to the molecule, contributing to its reactivity and introducing level of instability to RNA. In eukaryotic cells, RNA is synthesised in nucleus by a DNA-dependent RNA polymerase in a process called transcription. During this process, the DNA template is transcribed into a newly generated RNA strand. The α-position of phosphate group of the incoming nucleotide reacts with the 3'-hydroxyl group of the attached nucleotide through an $S_N2$ mechanism, resulting in the formation of stable bond and pyrophosphate as a leaving group[6] (Figure 2). In addition, single-stranded RNA has the capability to fold into hairpin structures[7] and when two complementary RNA strands interact, they can form A- or Z-RNA helix[8,9]. Furthermore, interaction with additional RNA strands under certain conditions can lead to the formation of triplex[10] or even G-quadruplex[11] structures. These three-dimensional structures serve not only as binding sites for proteins[12], but also play functional roles, such as exhibiting catalytic activity. RNA enzymes, termed as ribozymes, are a class of highly structured RNAs, exemplify molecules with nucleic acid cleavage activity[13].

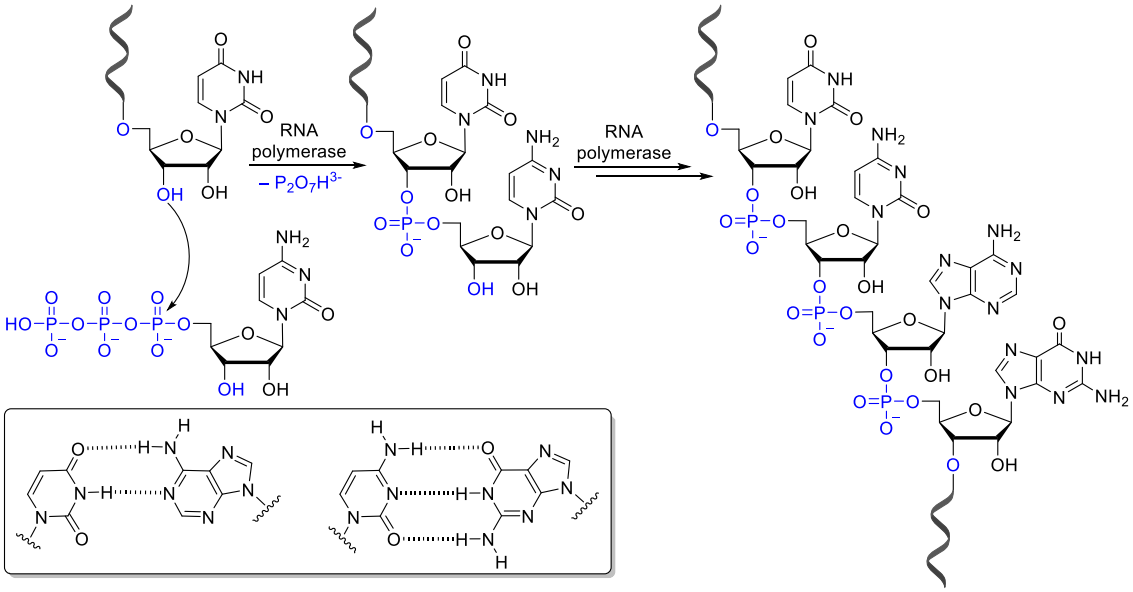



**Figure 2. Synthesis of RNA strand and canonical Watson-Crick base pairing in RNA molecules.**

There are several classes of RNA molecules, each playing distinct roles within the cellular machinery. The major representatives include messenger RNA (mRNA), which carries genetic information for protein synthesis[14]; transfer RNA (tRNA), essential for transferring amino acids to

the growing protein chain during translation[15]; ribosomal RNA (rRNA), a crucial component of ribosome involved in protein synthesis[16]; long non-coding RNA (lncRNA), participating in various cellular processes without encoding for proteins[17]; microRNA (miRNA), involved in post-transcriptional gene regulation[18]; small nuclear RNA (snRNA), essential for splicing processes[19]; small nucleolar RNA (snoRNA), guiding modifications of other RNAs[20]; ribozymes, catalytic RNA molecules[21]; and the recently identified circular RNA (circRNA), which although less abundant, exhibit stability due to their covalent loop structure[22] (Figure 3). The exploration of this field is ongoing, with continuous discoveries of new RNA classes and a growing understanding of their roles.

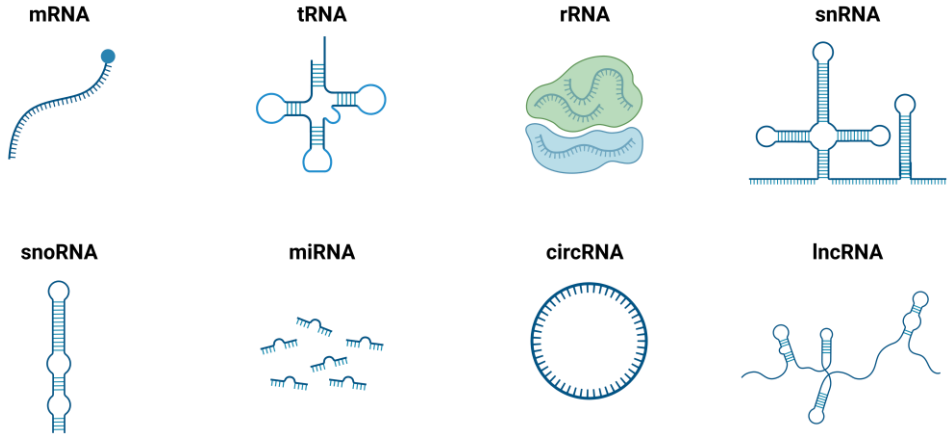



**Figure 3. Overview of selected RNA molecules and their structures.**

In addition to its canonical A, U, C, G nucleotide components, RNA undergoes chemical modifications that significantly contribute to its structural diversity and functional versatility[23]. Various modifications have been identified on all four nucleobases, including $N^6$-methyladenosine ($m^6A$), pseudouridine (Ψ), 5-methylcytosine ($m^5C$), and $N^7$-methylguanosine ($m^7G$). Furthermore, there are reports of sugar modifications such as 2'-*O*-methyl functional group (2'-*O*-Me) (Figure 4). These modifications do not influence only proper folding and stability of RNA, but also play crucial roles in diverse molecular processes such as transcription, splicing, and translation of RNA[24]. To date, over 170 chemical modifications have been identified in both coding and non-coding RNAs, and this number continues to grow as more advanced detection methods are developed[25]. These modifications are introduced into RNA by "writers", removed by "erasers", and further modified by enzymes referred to as "modifiers"[26]. The presence of modifications is pervasive throughout nearly all cellular RNAs, with a notable abundance in functional RNAs, such as tRNA, rRNA, and mRNA[27]. Pseudouridine, recognised as the "fifth nucleobase", was the first identified nucleoside modification, isolated from bulk yeast RNA in 1957[28]. A significant advancement of understanding

these modifications emerged from recent pioneering investigations by Karikó and colleagues, demonstrating that numerous naturally occurring modifications in human RNA, e.g., Ψ, 2-thiouridine (s2U), and $m^5C$, reduce immunogenicity[29].

The multifaceted nature of RNA extends its significance beyond a transmitter of genetic information, contributing to various cellular processes and expanding the scope of its biological impact. Therefore, RNA has attracted considerable attention and is now extensively studied.

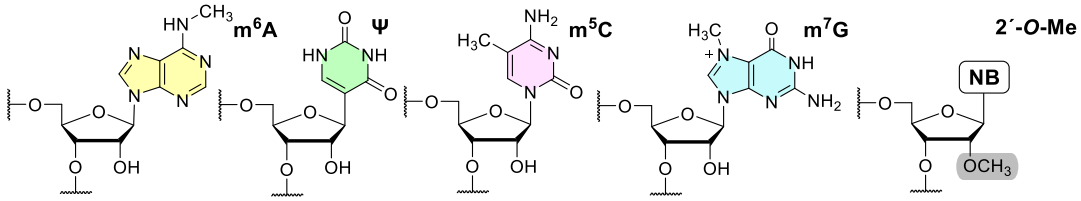



**Figure 4. Selected examples of naturally occurring modifications on nucleobase and ribose within RNA.**

## 1.3 Chemical (solid support-based) synthesis of RNA

Besides isolation from biological samples, generation of RNA can be accomplished through two primary methods, solid support-based chemical synthesis and/or enzymatic polymerisation using rNTPs building blocks[30]. The predominant approach for incorporating various chemical functionalities into RNA oligonucleotides is solid phase synthesis, recognised as the gold standard procedure[31]. Since its development in the 1980s by Caruthers and Beaucage originally for short DNA oligonucleotide synthesis[32–34], it has found application even in generation of modified RNA probes, including those used as small interfering RNAs (siRNA)[35], short RNA hairpins[36] or as antisense oligonucleotides (ASO)[37]. In contrast to DNA, presence of the 2'-hydroxyl sugar group renders RNA not only more labile than DNA, but also introduces complications in the steps of chemical synthesis, making it approximately twice more expensive than DNA synthesis. Existence of the 2'-OH group necessitates additional protection and deprotection steps during the reaction cascade and forms sterical hindrance during condensation reaction (Figure 5)[38,39]. Employing a four-step reaction cycle for step-by-step integration of phosphoramidite building blocks enables selective introduction of modifications and labels at the sugar, base, or phosphate backbone of nucleic acid. Oligonucleotide synthesis occurs in the 3'→5' direction and upon cleavage from the solid support, the generated RNA molecules are usually purified through high-performance liquid chromatography (HPLC)[38].

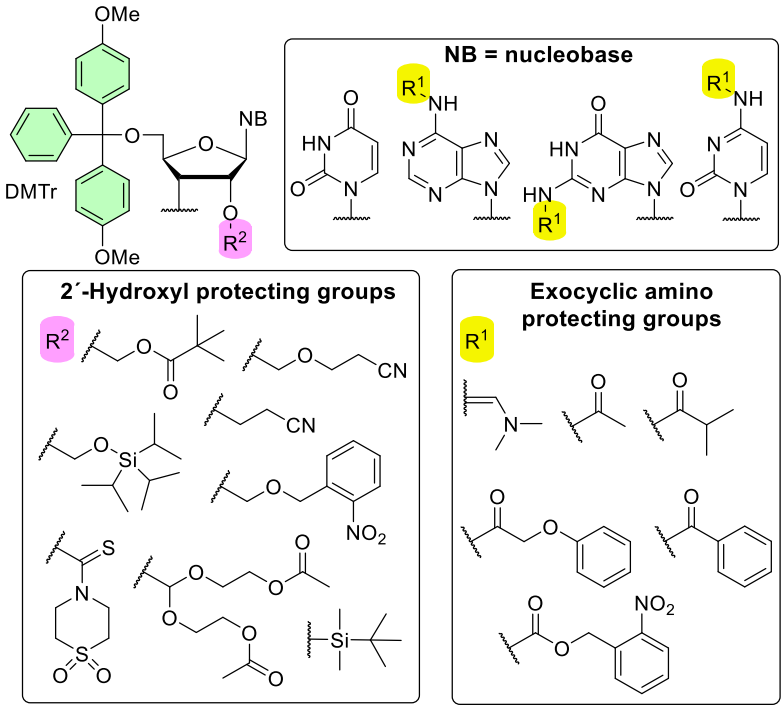
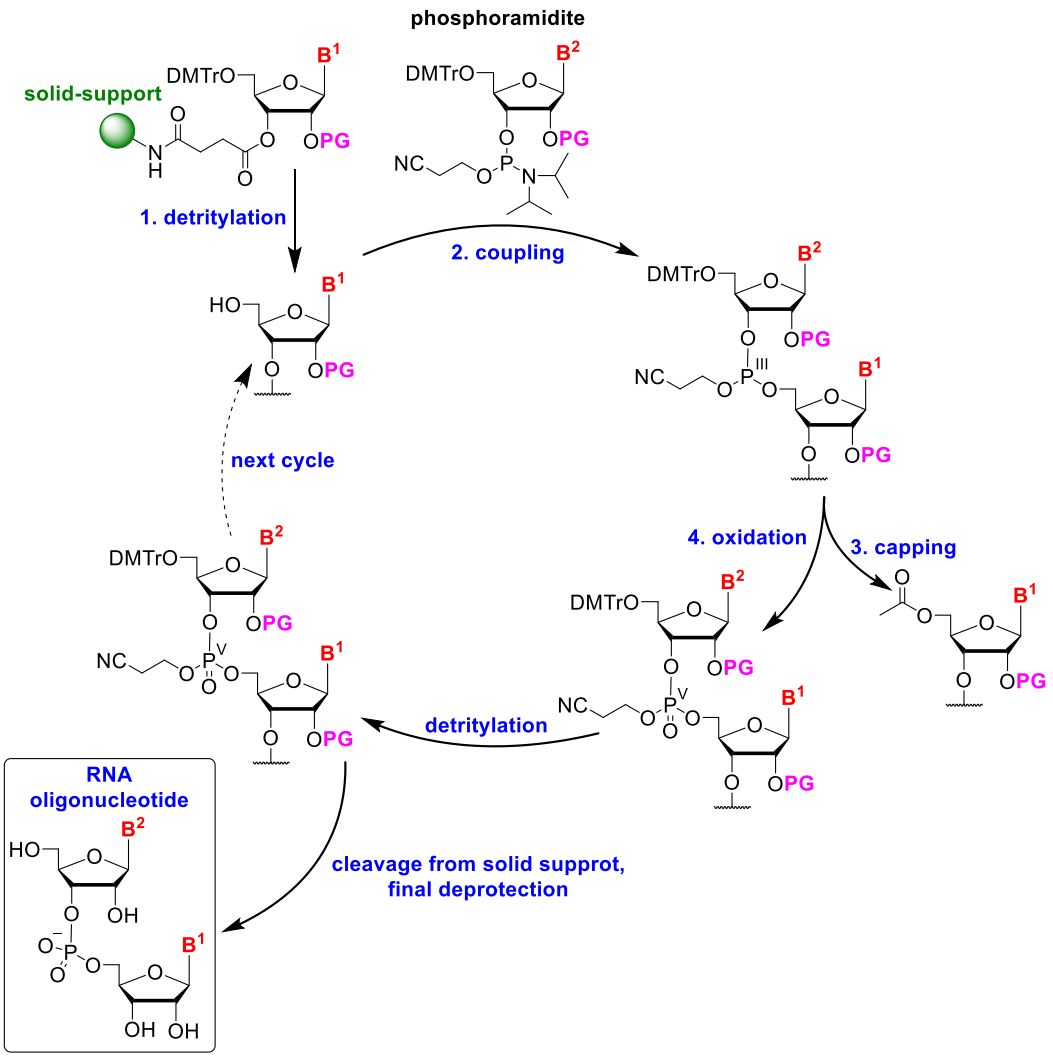



**Figure 5. Selected nucleobase and sugar protecting groups used in chemical synthesis of RNA oligonucleotides.**

The initiation step of RNA chemical synthesis involves removing the dimethoxytrityl (DMTr) protecting group from the 5'-hydroxyl functionality of the first ribonucleoside attached to solid support (controlled-pore glass or polystyrene) through the 3'-hydroxyl. Following the detritylation step, typically performed by trichloroacetic or dichloroacetic acid and removal of protecting group in washing step, the coupling reaction takes place[39]. The incoming nucleoside, activated by addition of an azole catalyst (e.g., 1*H*-tetrazole, 5-ethylthio-1*H*-tetrazole, 5-benzylthio-1*H*-tetrazole)[40,41], undergoes nucleophilic substitution with free 5'-hydroxyl group, along with elimination of tetrazole complex as a leaving group and forming a new phosphite triester bond[39]. As the coupling step never achieves full conversions, a capping step must be implemented to protect the unreacted 5'-hydroxyl moiety, that could form unwanted deletion by-products in next reaction steps. This is usually performed by a mixture of DMAP with acetic anhydride and lutidine[38]. The newly formed phosphite triester bond [P(III)] of the coupling product is oxidised to a more stable phosphotriester bond [P(V)] upon treatment, usually with iodine or *tert*-butyl hydroperoxide[39]. After complete synthesis of desired oligonucleotide sequence, the RNA probe is released from solid support, followed by deprotection steps to remove phosphate and nucleobase protecting groups, typically performed in concentrated aqueous ammonia. Final steps involve deprotection of sugar moiety to generate the desired RNA oligonucleotide product (Scheme 1). Chemical solid phase method offers potential for complete automation and scale-up in production of substantial quantities of modified RNA probes[38].




**Scheme 1. Process for solid support-based chemical synthesis of RNA oligonucleotides.**

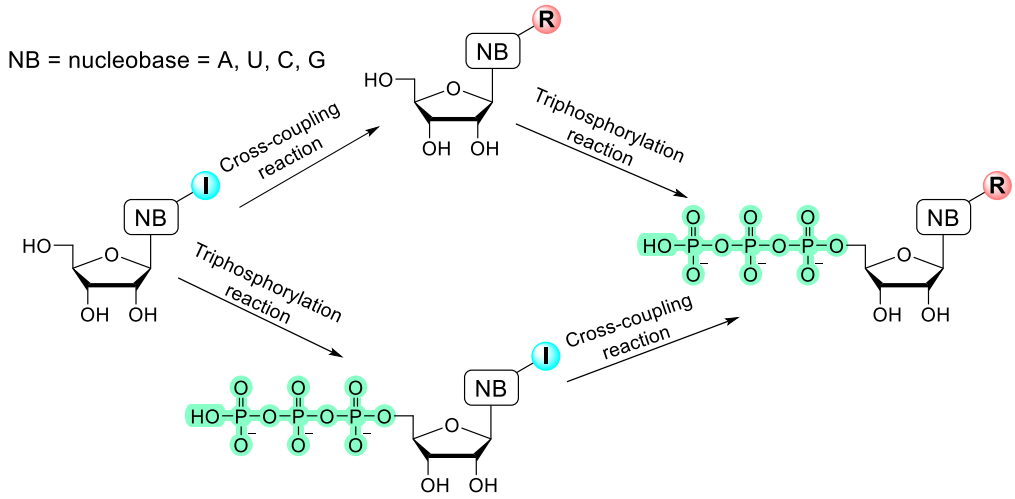
Despite demonstrated success even in synthesising a 110 nt long RNA probe,[42] it is essential to acknowledge certain limitations associated with this method. While this approach mainly excels in generating shorter modified RNA probes, typically below 60 nt, attempting the synthesis of longer oligonucleotides in reasonable yields remains a challenge. Although a single elongation cycle efficiency can reach 99%, overall yield drops exponentially with each additional cycle. For instance, the 50 nt long RNA probe yield may decrease to 37%, making this method impractical for generation of long RNA oligonucleotides[30]. Moreover, application of harsh reaction conditions in the process may not always align with requirements of sensitive or reactive functional groups[38], therefore alternative, more environmentally friendly, and milder options are required.

## 1.4 Enzymatic synthesis of RNA

Enzymatic synthesis emerges as an environmentally sustainable and gentle alternative for RNA production, particularly valuable when dealing with very reactive modifications incompatible with traditional chemical synthesis or aiming at generation of long RNA strands. The key player in enzymatic RNA synthesis is the RNA polymerase, an enzyme that catalyses the formation of RNA strand from a DNA template. This process involves incorporation of either natural or modified ribonucleoside triphosphate building blocks in a template-dependent fashion. Mild reaction conditions provide compatibility with a diverse range of modifications making it suitable for tailoring RNA molecules to specific needs and offering advantages in situations where classical chemical synthesis falls short[43].

### 1.4.1 Synthesis of ribonucleoside triphosphate building blocks

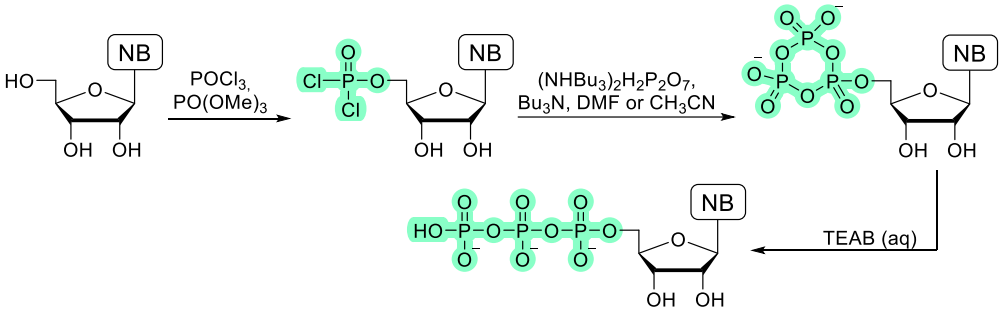
While a universally applicable and high-yielding method for synthesis of modified nucleotides is still a big challenge, recent advancements have contributed to overall ease of obtaining these building blocks. There are two established pathways leading to base-modified nucleotides. One method involves the phosphorylation of modified nucleosides, whereas the other is based on functionalisation of the triphosphate building block itself, under mild conditions (Scheme 2).




**Scheme 2. Synthetic pathways leading to modified ribonucleoside triphosphates.**

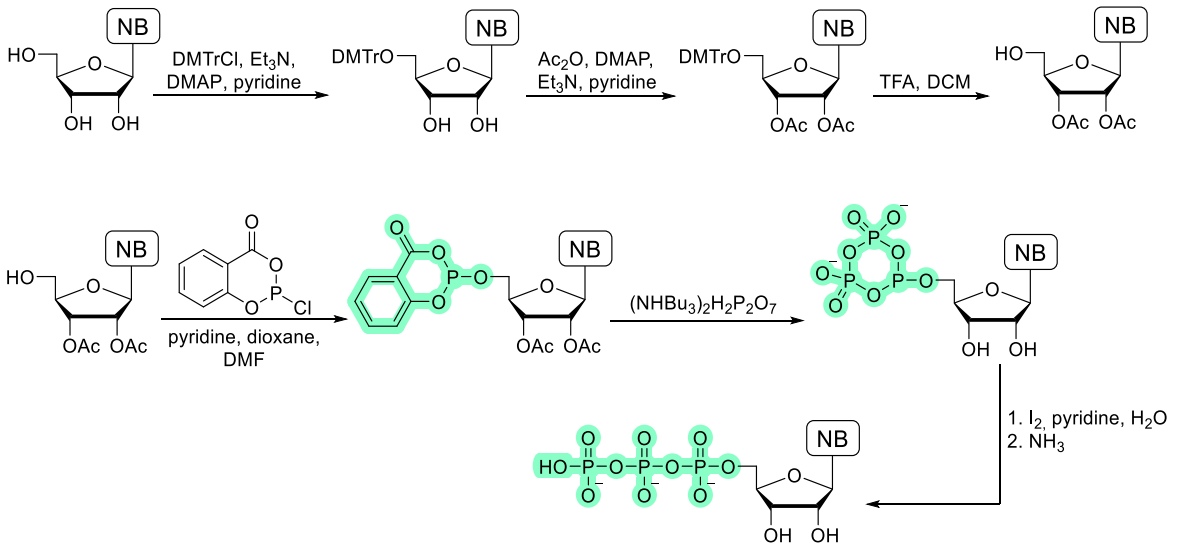
Until now, various procedures have been developed for phosphorylation of modified nucleosides. One of the pioneering and most common approaches for preparation of nucleoside triphosphates was reported by Yoshikawa in 1976[44,45]. This one-pot reaction starts with regioselective 5'-monophosphorylation of an unprotected nucleoside using electrophilic phosphorous oxychloride

($POCl_3$) in trimethylphosphate [$PO(OMe_3)$]. The generated highly reactive phosphorodichloridate intermediate subsequently undergoes an *in situ* reaction with bis(tri-*n*-butylammonium) pyrophosphate leading to cyclic triphosphate. Finally, hydrolysis in a triethylammonium bicarbonate buffer yields the desired triphosphate (Scheme 3). However, it is important to note that this triphosphorylation reaction is extremely sensitive to water, usage of highly electrophilic phosphoric reagent is not universally compatible with all functionalised nucleosides, and since selectivity to primary 5'-hydroxyl group is not complete, HPLC separation of regioisomeric by-products is a necessary step. Over time, the procedure has undergone minor improvements to enhance its effectiveness. Replacement of DMF by acetonitrile as a solvent for pyrophosphate notably reduced presence of monophosphate by-products[46]. Additionally, temperature screening experiments have revealed, that lowering the temperature enhances reaction regioselectivity[47].



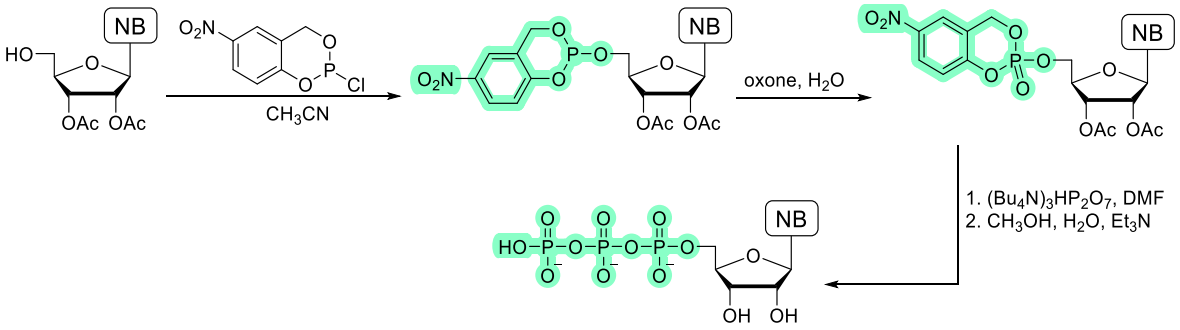

**Scheme 3. Yoshikawa triphosphorylation protocol.**

An alternative method, known as the "one-pot, three-steps synthesis", was developed by Ludwig and Eckstein in the late 1980s[48]. This approach is based on usage of [P(III)] reagents, e.g., salicyl chlorophosphite (2-chloro-4*H*-1,3,2-benzodioxaphosphorin-4-one), in the first phosphorylation step. Because of the highly reactive and non-regioselective character of [P(III)] reagent, the appropriate 2', 3'-*O*-protection of nucleoside is needed, prior to phosphorylation step. A commonly employed protection cascade involves tritylation of 5'-position hydroxyl group, followed by acetylation at 2'- and 3'-*O*-positions. Subsequent deprotection of DMTr group results in appropriately 2',3'-*O*-protected precursor with free primary 5'-hydroxyl group. Afterwards, reaction with salicyl phosphorochlorite leads to formation of the activated 5'-phosphite intermediate. Displacement of salicylic acid, facilitated by tris(tetra-*n*-butylammonium) hydrogen pyrophosphate forms cyclic phosphorous intermediate. Finally, iodine-mediated oxidation and deprotection of acetyl groups by ammonia generates the desired triphosphate (Scheme 4). Although tedious strategy, the protection of hydroxyl groups diminishes generation of side-products e.g., regioisomers, mono-, di-, or oligo-phosphates, facilitating the subsequent HPLC purification[49].



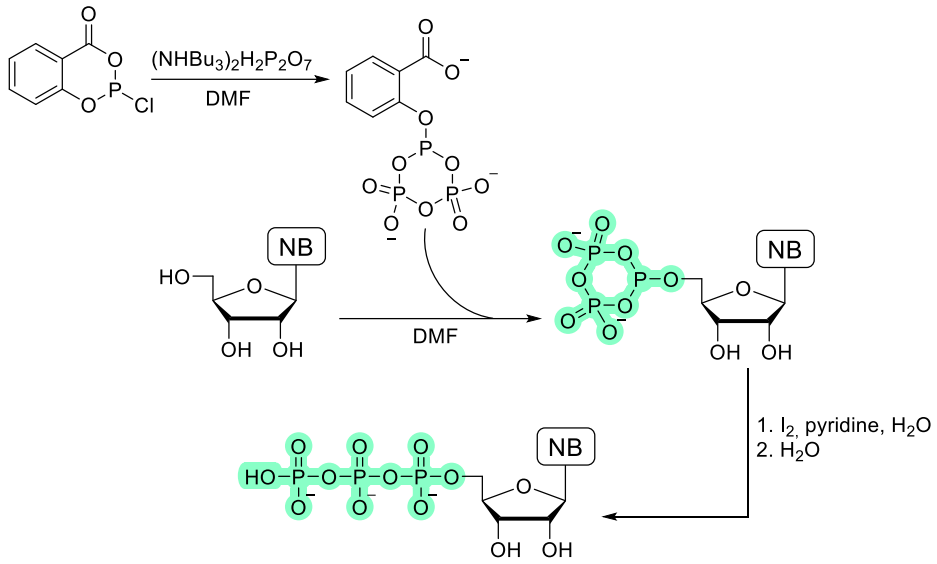

**Scheme 4. Common protecting method of nucleoside used for the Ludwig-Eckstein triphosphorylation strategy.**

Afterwards, an improved Ludwig-Eckstein strategy employing 5-nitro-cycloSal-phosphochloridite was published by Warnecke and Meier[50]. The strategical introduction of electron-withdrawing nitro group increased the electrophilic character and reactivity of phosphorus atom (Scheme 5).



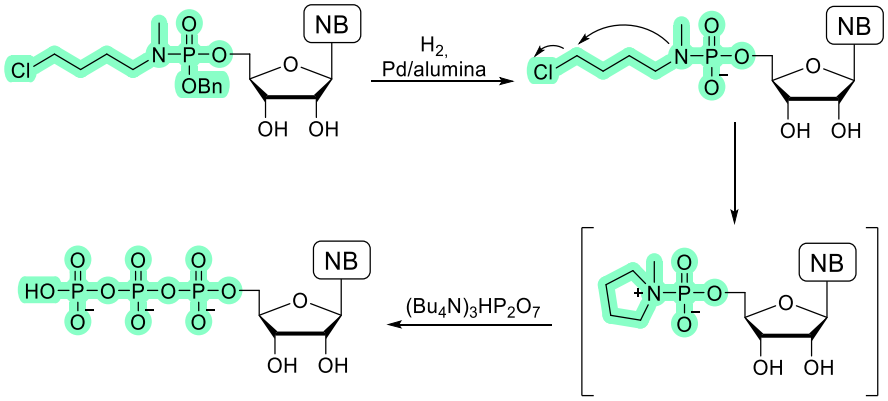

**Scheme 5. Triphosphorylation reaction *via* cycloSal activated nucleosides.**

Few years later in 2011, an alternative "hybrid" method combining the advantages of both Yoshikawa and Ludwig-Eckstein strategies was published by Caton-Williams[51]. The *in situ* generated bulky phosphitylating reagent, prepared by reaction of salicyl chlorophoshite with pyrophosphate, selectively reacts with primary 5'-OH group of unprotected nucleosides (Scheme 6).

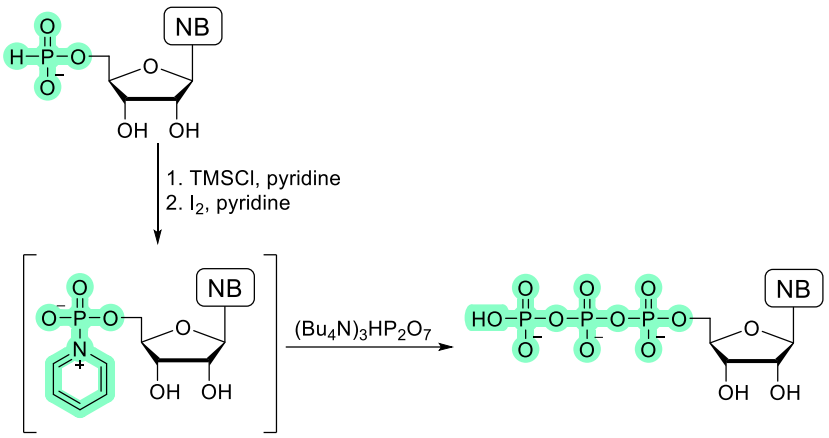



**Scheme 6. "Hybrid" one-pot triphosphorylation reaction.**

An completely innovative method utilising reactive zwitterionic intermediates was developed by Borch[52]. The process begins with catalytic hydrogenation of a non-protected nucleoside with an *O*-benzyl-phosphoramidate ester. The activated intermediate undergoes spontaneous rearrangement resulting in formation of pyrrolidinium phosphoramidate zwitterion, which undergoes *in situ* reaction with pyrophosphate, yielding the corresponding triphosphate (Scheme 7). Later on, Peterson and colleagues proposed an alternative approach utilising more readily available *H*-phosphonate nucleoside as a starting compound[53]. The process involves silylation and oxidation, forming a pyridinium phosphoramidate, which reacts with pyrophosphate in following step (Scheme 8).




**Scheme 7. Synthesis of ribonucleoside triphosphates by original Borch method.**

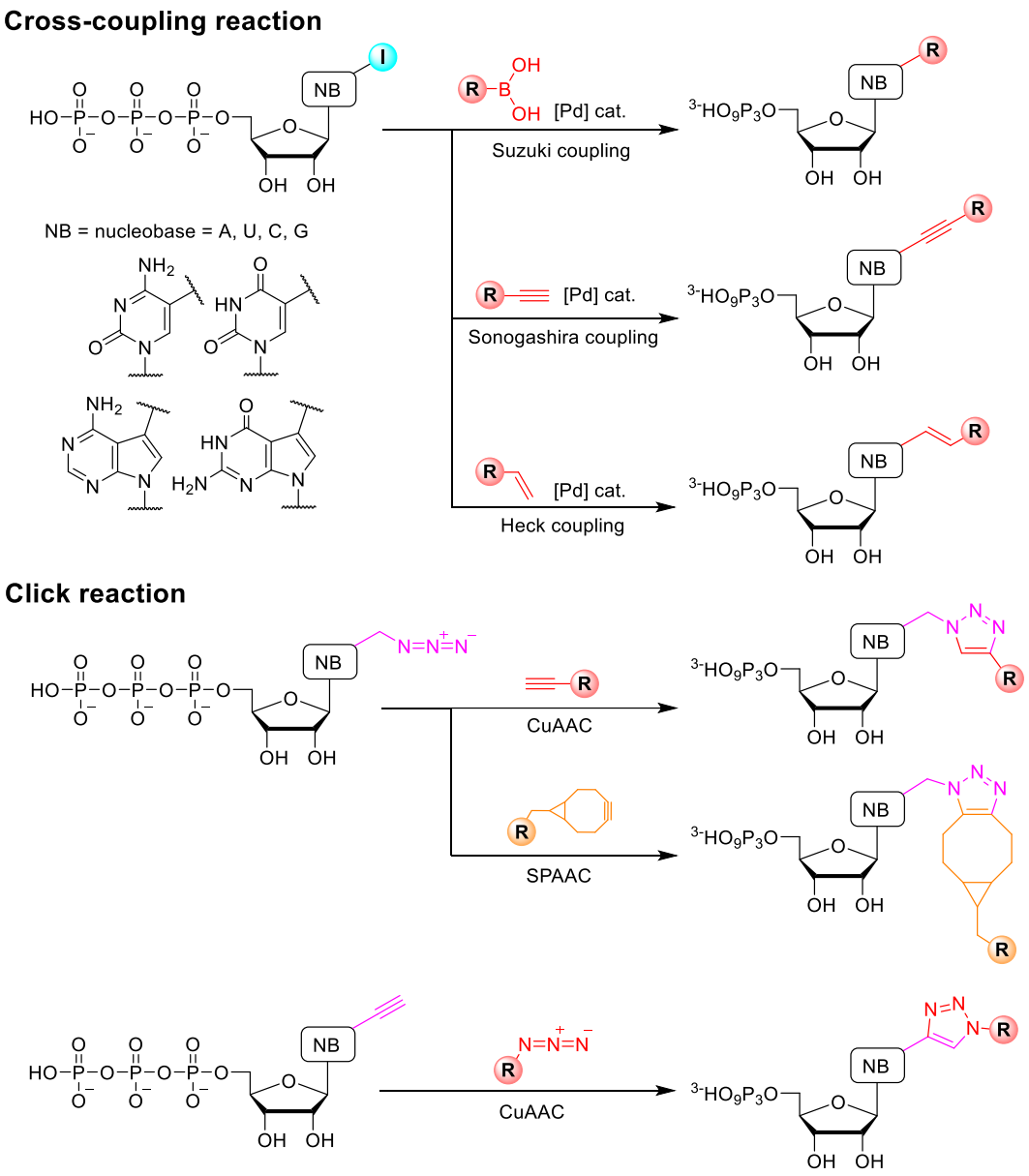



**Scheme 8. Improved Borch methodology with *H*-phosphonate nucleosides.**

To address disadvantages associated with triphosphorylation reactions, alternative pathways have been developed for construction of modified nucleotides. Metal catalysed cross-coupling reactions emerge as suitable post-phosphorylation derivatisation approach for formation of stable carbon-carbon (C–C) bonds[54]. Aqueous-phase cross-coupling reactions, facilitated by water-soluble ligands often employing hydrophilic 3,3′,3″-phosphanetriyltris(benzenesulfonic acid) trisodium salt (TPPTS)[55], offer a direct means of functionalisation of unprotected nucleotides[56,57]. Among them, the most common is Sonogashira cross-coupling of iodinated nucleotide building blocks with terminal alkynes. The first evidence of Sonogashira reaction was reported in 1990 for synthesis of a fluorophore-labelled nucleotide[58]. Rapid development of this synthetic method has been supported by straightforward alkyne accessibility and broad tolerance to functional groups. Reactions are usually carried out in a mixture of water and acetonitrile, in presence of TPPTS ligand, triethylamine or DIPEA as a base and CuI as a copper source. Common challenges include side-reactions of 5-alkynyl-modified uridine triphosphates, leading to furanopyrimidine-2-ones, upon long reaction times or high temperatures in presence of copper(I) co-catalyst[59]. This can be bypassed by copper-free versions of the conventional approach[60]. Suzuki coupling involves usage of boronic acids, esters, or trifluoroboronates and a vinyl or arylhalogenide nucleotide, as a second partner. Milisavljević et al. recently published an extensive study on feasibility and efficiency of introducing diverse sterically demanding modifications through the application of Pd-catalysed Sonogashira and Suzuki coupling reactions[43]. Unlike other conventional Pd-catalysed C-C bond-forming reactions, the Heck coupling has limited examples available[61] (Scheme 9).

For introducing reactive groups that may not withstand higher temperatures needed for cross-coupling reactions and require neutral pH or other milder conditions, click chemistry and related reactions provide a viable approach for implementation of a moiety of interest to a nucleotide.

Copper(I)-catalysed azide-alkyne cycloaddition (CuAAC) or metal-free strain-promoted azide-alkyne cycloaddition (SPAAC) play pivotal roles in derivatisation of nucleotides bearing azide groups with either terminal (in the case of CuAAC) or cyclic (in SPAAC) alkynes. This process leads to formation of a stable 1,2,3-triazole linkage, as extensively described in literature[62–66]. Alternatively, azido-modified nucleotides can undergo further modification through Staudinger reaction. Another convenient option involves employing click chemistry on nucleotides bearing terminal alkynes with molecules bearing azide modifications[67]. Ribonucleoside triphosphates with terminal alkyne modifications were fluorescently labelled with four distinct dyes (fluorescein, BODIPY, rhodamine, coumarin) with azido modifications[68] (Scheme 9). Functionalities are commonly tethered to C-5 position of pyrimidines and/or C-7 position of 7-deazapurines, that directs them towards the major groove upon incorporation. This orientation helps minimise structural perturbations and ensures good acceptance by polymerases[69–73].




**Scheme 9. Reactions for derivatisation of nucleoside triphosphates.**

### 1.4.2 Enzymes for RNA synthesis

Key strategy for enzymatic synthesis of RNA currently relies on usage of T7 bacteriophage RNA polymerase. This is due to its high efficiency, low error rate, independence on cofactors, simple promoter, and single subunit structure. The discovery of bacteriophage T7 RNA polymerase (T7 RNAP) is dated to 1970 when it was first isolated from T7 bacteriophage-infected *Escherichia coli* (*E. coli*) cells. This DNA-dependent polymerase is a simple single-subunit 98 kDa enzyme. T7 RNAP comprises of N-terminal and polymerase domain; the latter one can be further divided into three sub-domains – thumb, palm, and finger[74]. Although its structure is relatively similar to Klenow fragment

of *E. coli* DNA polymerase, unlike DNA polymerases, it does not require any short oligonucleotide fragment (primer) to start RNA synthesis. The enzyme displays exceptional specificity for T7 promoter, demonstrating a complete lack of affinity for any other closely related promoter sequences. This includes also T3 promoter, which is the most similar, and differs only in one triplet[75]. It effectively transcribes dsDNA templates to RNA without the need for additional transcription factor. Although it does not exhibit proofreading activity, errors occur at relatively low average frequency ~ ($10^{-4}$)[76]. Conversely, the enzyme exhibits limited tolerance to bulky or 7-deazaguanine modifications, necessitates an optimal GGG trinucleotide initiation sequence[77,78], lacks thermal stability and terminates when transcribed structured DNA templates[43,79].

Naturally occurring rNTPs are in cells in concentrations significantly higher (between 10 to 100-fold) than deoxyribonucleotide (dNTPs) congeners. To prevent undesired incorporation of rNTPs into the genome during DNA synthesis, all wild-type DNA polymerases feature a distinctive "steric gate" residue. This residue acts as steric barrier for triphosphates carrying any functional group at 2′-sugar position, thereby excluding them from the polymerase active site[80]. Historical efforts in DNA and RNA polymerase engineering[81,82] aimed to unravel this natural protection mechanism, as well as facilitating primer-dependent RNA synthesis. Initial attempts in evolution of DNA polymerases for RNA synthesis encountered challenges as the elongation process aborted shortly after incorporation of first few ribonucleotides[83–85]. Nevertheless, significant progress has been made over the last few decades, leading to identification of new mutant variants of DNA polymerases that show promise in addressing these challenges. In the Marx group, a strategic starting point was the modified *Thermococcus species* 9°N-7 DNA polymerase, known as Therminator polymerase (9°N: A485L), which has a residual RNA polymerase activity[83]. Through generation of random mutants of the parental polymerases using error-prone PCR[86], and subsequent screening of their activities, they identified one candidate capable to perform primer-dependent RNA synthesis. The selected mutant polymerase (Therminator: L408Q) efficiently incorporated up to 58 ribonucleotides in 3 hours starting from an RNA primer and additionally demonstrated an enhanced ability to accept C5-base modified **dN$^{X}$TP**s, compared to the parental polymerase[87]. In 2012, Holliger and his colleagues made a significant breakthrough involving engineering replicative DNA polymerase of *Thermococcus gorgonarius* (Tgo) (Figure 6). Their investigation underscored the critical role of a specific mutation within the polymerase thumb subdomain (E664K) when combined with a conventional steric gate relieving mutation (Y409G). This dual engineering approach markedly boosted polymerase's RNA synthesis capabilities, showcasing its proficiency in primer-dependent synthesis of RNA strands, reaching lengths of 0.7 to 2 kb. Additionally, it was found that TGK mutant polymerase is capable to initiate primer dependent RNA synthesis from a wide range of chemistries,

including DNA, RNA, locked nucleic acids (LNA) or 2'-*O*-Me, modified primers and incorporate not only 2'-OH, but also 2'-F and 2'-azido modifications[88]. Shortly after, Chaput et al. conducted a comprehensive investigation into the effects of mutations (Y409G, A485L, and E664K) previously recognised for enhancing rNTPs incorporation efficiency by Tgo DNA polymerase[88]. They transferred these mutations to homologous B-family DNA polymerases, specifically 9°N, KOD, Tgo, and Deep Vent, aiming to study and elucidate their impact[89]. Surprisingly, only Tgo and DeepVent polymerases were proficient in synthesis along difficult random templates, while KOD and 9°N were remarkably less active.

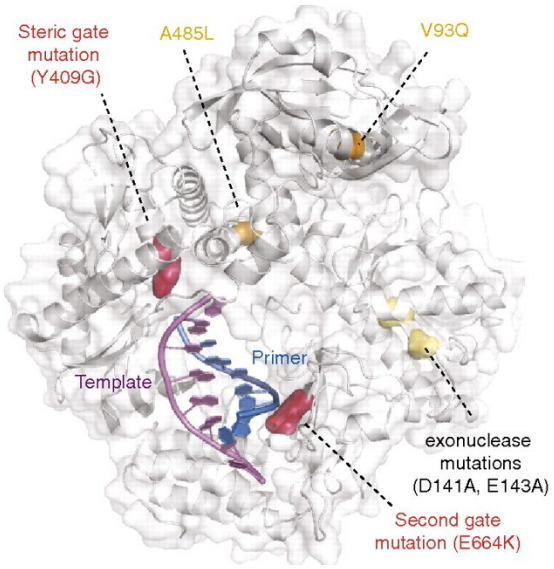



**Figure 6. RNA synthesis-enabling mutations in TGK polymerase are mapped on the structure of a closely related Pfu polymerase[88].**

In parallel, the Romesberg group undertook a groundbreaking study that involved introducing strategic mutations into the Stoffel fragment (SF) of *Thermococcus aquaticus* (Taq) DNA polymerase I, specifically targeting SFR1, SFR2, and SFR3, by phage display[90,91]. Among the pool of generated polymerases, only Taq-SFM4-3 mutant retained thermostability characteristics of parental DNA enzyme and proved to be useful in a RNA analogue of the polymerase chain reaction (PCR), known as polymerase chain transcription (PCT) (Figure 7). SFM4-3 represents the first mutant derivative from a heat-stable (A-family) DNA polymerase, with ability to synthesise RNA. Application of PCT proved to be a useful technique, facilitating amplification of RNA fragments from a single DNA template, primed with two distinct DNA oligonucleotides. Impressively, when PCT was employed, the amplification levels relative to DNA template achieved substantial ranges, reaching $10^3$ to $10^5$-fold higher compared to conventional T7 RNAP transcription methods. This pivotal development significantly expands the accessibility of RNA oligonucleotides from limited amounts of DNA

templates, pushing the scale to levels previously attainable only through solid phase chemical synthesis[92,93]. Notably, PCT was efficient in incorporation of purine ribonucleotides containing 2'-F modification, however, synthesis with 2'-F modified pyrimidines was inefficient. The PCT found application even in post-transcriptional labelling performed by click chemistry, after incorporation of 2'-azido modified nucleotides as reported by Shao et al[94].

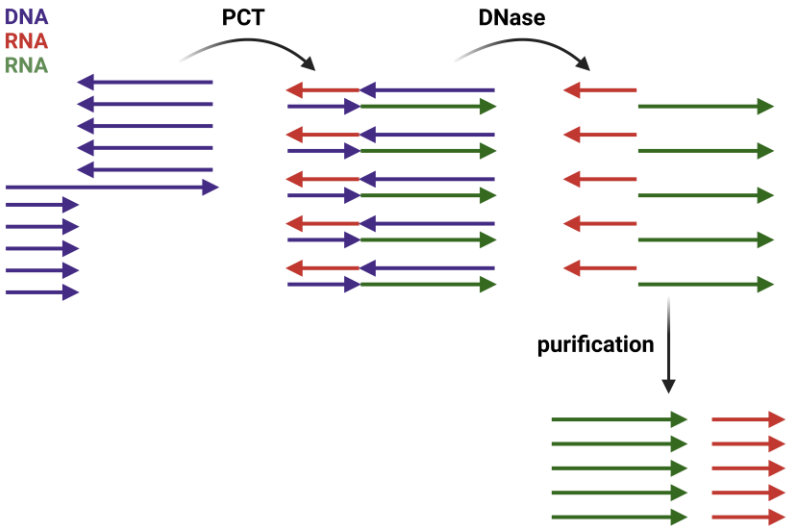



**Figure 7. Mechanism of polymerase chain transcription (PCT) reaction.**

### 1.4.3 Diverse enzymatic approaches for construction of RNA oligonucleotides

Up to date, T7 RNAP IVT reaction is one of the most used methods for generation of RNA probes. T7 RNAP uses rNTPs as building blocks for synthesis of new RNA strand complementary to dsDNA template of choice, composed of two together annealed complementary single-stranded oligonucleotides (sense and antisense strand)[95,96]. The IVT reactions are typically conducted in Tris-HCl buffer with neutral pH at 37 °C. The inclusion of $Mg^{2+}$ cations is essential as they serve as a critical cofactor role in facilitating the proper functioning of the enzyme[97]. Additionally, dithiothreitol (DTT) is incorporated into reaction as a reducing agent maintaining the activity of enzyme. Other optional additives that improve IVT process include Triton X-100, DMSO, PEG8000, BSA, etc., although concentrations of these agents needs to be optimised on a case-by-case basis[98]. The initiation of transcription by T7 RNAP starts with recognition and binding to a specific T7 promoter sequence in dsDNA. The consensus promoter region strictly needs to have a double-stranded structure, while the template itself can be either single- or double-stranded, with no apparent difference observed in transcription process[99]. The initiation step of T7 RNAP transcription is relatively unstable and involves an abortive cycling phase. This phase is characterised by generation

of short transcripts consisting of 8 to 10 bases[100]. After overcoming this stage and successful incorporation of the first ~10 nucleotides, the enzyme-template-transcript forms highly stable complex ("transcription bubble"), that moves along the anti-sense template in the 3'→5' direction[101] (Figure 8). The effectiveness of IVT is dependent upon various factors, with one of the pivotal determinants being the nature of the initial nucleotide integrated into nascent RNA. Initiating the process with a minimum of one guanosine nucleotide is essential, but substantial augmentation of IVT process efficiency occurs when up to three consecutive guanosine nucleotides are utilised[77,78]. Upon completing the polymerisation process, the enzyme disengages from DNA template (run-off transcription), releasing synthesised RNA transcript with 5'-end triphosphate residue originating from first incorporated rNTP. This cycle is then repeated multiple times leading up to milligram quantities of RNA[95].

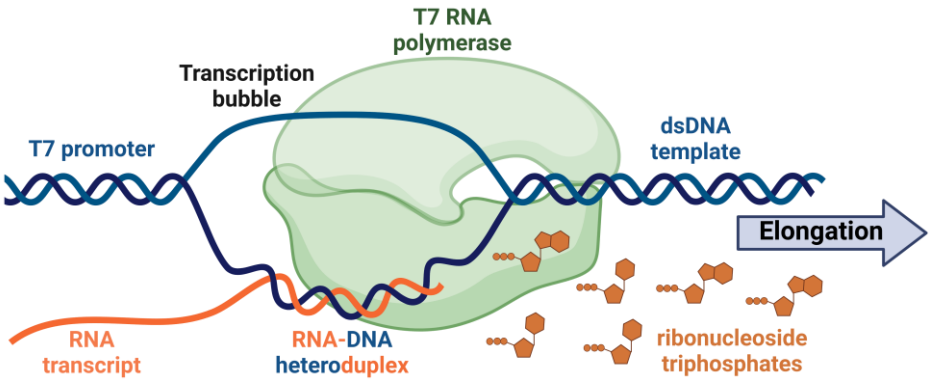



**Figure 8. T7 RNA polymerase transcription process.**

On the other hand, due to T7 RNAP's strong specificity for the promoter sequence, incorporation of any user-defined 5'- sequence is constrained. Furthermore, IVT results in non-templated addition of one or several nucleotides, most commonly rATP, at the 3'-end because of inconsistent polymerase run-off. These species are challenging to separate from the desired full-length product. A feasible solution to address 3'-end inhomogeneity involves introducing 2'-*O*-methyl modified nucleotides in the last two positions of dsDNA template (at 5'-end of anti-sense strand)[102]. In addition, recent findings indicate that addition of DMSO can further contribute to reducing the generation of unwanted by-products during IVT process and enhancing production of RNA with uniform 3'-ends[103,104]. An alternative approach to achieving RNA with homogeneous 3'-ends involves utilising ribozymes, which are self-cleaving RNA molecules[105,106]. These ribozymes, characterised by their relatively compact size, are directly inserted in the dsDNA template, and serve as valuable tools for precisely cleaving RNA strands[107,108]. The cleavage mechanism is catalysed in presence of divalent

metal ions, such as $Mg^{2+}$ or $Mn^{2+}$, resulting in formation of two RNA strands, the catalytic domain (ribozyme) and targeted RNA sequence[107]. Monitoring of IVT process and detection of *de novo* synthesised RNA transcript is dependent on utilising radioactive nucleotides, for example [α-$^{32}$P]-rGTP or [α-$^{32}$P]-rATP. While this approach is highly sensitive and capable of detecting even small quantities of RNA, it comes with drawbacks of being labour-intensive, requiring specialised equipment, environmentally unfriendly, and having a limited duration due to short half-life of radioisotopes.

Despite some noted limitations, T7 RNAP IVT method remains the gold standard for enzymatic synthesis of modified RNA. It has been demonstrated that the wild-type T7 RNAP exhibits remarkable versatility by accepting a broad range of modified ribonucleotide building blocks[43]. The use of a biotin affinity tag-modified rUTP was the first successful report for modified IVT, dating back to 1981[109]. Since then, the flexibility of T7 RNAP was showcased in synthesis of modified RNA with diverse functionalities, including incorporation of reactive groups suitable for post-transcriptional labelling[110,111] (such as iodine[112], ethynyl[43,113] or alkyne[114], azido[115], vinyl[111] and *trans*-cyclooctene[116] modifications), cross-linking moieties like diazirines[117] for capturing RNA-protein interactions, amino acid-like side chains[101], azobenzene[118], amino or thiol modifications[119] for aptamer selections, as well as introduction of fluorescent nucleotides[120–122]. Remarkably, this method even enables incorporation of nucleotides with unnatural bases[123]. T7 RNAP IVT is therefore powerful and adaptable tool for decorating RNA with a variety of structural and functional modifications (Figure 9).

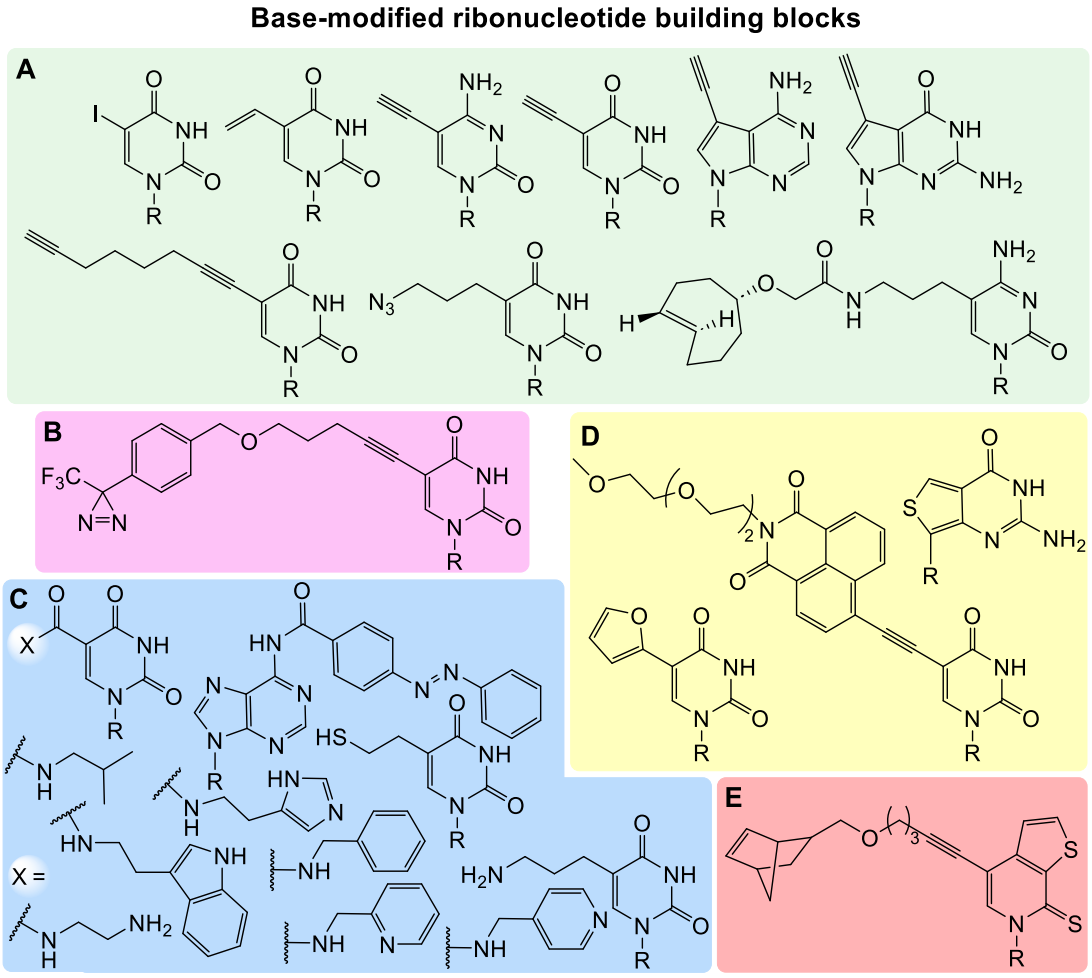



**Figure 9. Structures of selected base-modified ribonucleoside triphosphates accepted by T7 RNAP. A – nucleotides with reactive groups for post-transcriptional labelling, B – diazirine modified nucleotide for cross-linking with proteins, C – modified nucleotides for SELEX, D – fluorescent and E – unnatural nucleotides, R – ribose.**

A novel approach involving combination of T7 RNAP IVT and solid phase synthesis for position-selective RNA labelling (PLOR), with modified **rN$^{X}$TP**s, was recently developed by Wang and colleagues[124]. This method takes advantage of the highly stable enzyme-template-transcript elongation complex, which can withstand manipulation procedures and purification steps, allowing for pausing and restarting the IVT reaction[125,126]. In this method, a 5'-biotinylated dsDNA template with a T7 promoter is attached to streptavidin agarose beads, that serve as a solid support (Figure 10). In contrast to conventional transcription, PLOR begins with incorporation of a maximum of three different types of rNTPs, causing a pause at position where the absent rNTP is to be incorporated. Subsequently, unincorporated rNTPs are removed from enzyme-template-transcript complex through solid phase extraction. A new mixture of rNTPs, including the modified one, allows elongation until the complex reaches a next pause point. After thorough washing, next cycle of elongation is initiated.

These cycles are repeated until all modified nucleotides of interest are incorporated into growing RNA chain. Transcription is completed upon addition of all four canonical rNTPs. The full-length transcript falls off from the complex and can be collected from liquid phase. The DNA template attached to agarose support can be recycled and used for generation of new RNA transcript. In theory, PLOR can be employed to introduce a variety of modified **rN$^{X}$TP**s, if they are suitable substrates for T7 RNAP. By applying this procedure, RNA probes underwent successful labelling at precise positions using isotopically or fluorescently modified nucleotides for structural studies. However, specific labelling requires a complex experimental design and considering that total yield of transcript rapidly drops with each additional pause-restart cycle and washing step.

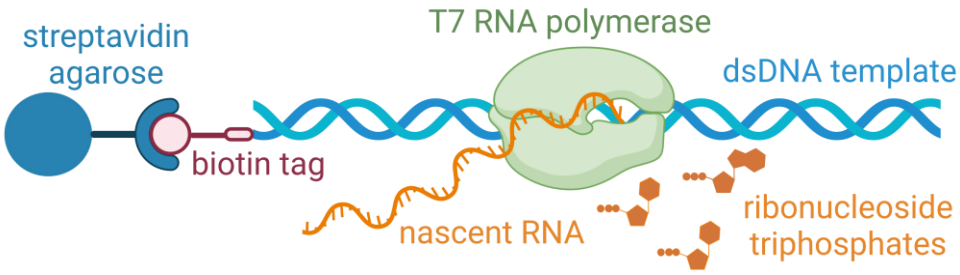



**Figure 10. Synthesis of RNA from dsDNA template attached to solid support by PLOR.**

The primer extension reaction (PEX) serves as a commonly employed biochemical technique for nucleic acid synthesis. Although, this method was initially used for constructing natural or highly functionalised DNA oligonucleotides[127–129], it is also finding its place in RNA synthesis, mainly in construction of xenonucleic acids (XNA) with a plethora of sugar analogues [88,92]. PEX reaction involves extending a short oligonucleotide sequence, referred to as a primer, which is complementary to a longer sequence known as a template. This extension occurs through the incorporation of nucleotide building blocks. During PEX reaction, the incoming nucleotide with its 5'-triphosphate residue, reacts with free 3'-OH group of the RNA primer, releasing diphosphate. The process continues until the starting primer is fully prolonged along DNA template. Primer can originate from either DNA or RNA (Figure 11) and can be labelled at 5'-terminus with fluorophores such as fluorescein amidite (FAM) or cyanine dyes (Cy5, Cy3) for visualisation through fluorescent scanning of the synthesised product using denaturing polyacrylamide gel electrophoresis (dPAGE). Beyond fluorophores various options exist for 5'-end labelling, such as affinity tags (biotin, digoxigenin), lipophilic modifications (cholesterol, palmitate), electrochemically active groups (ferrocene, viologen), or click chemistry handles (alkynes, azides).

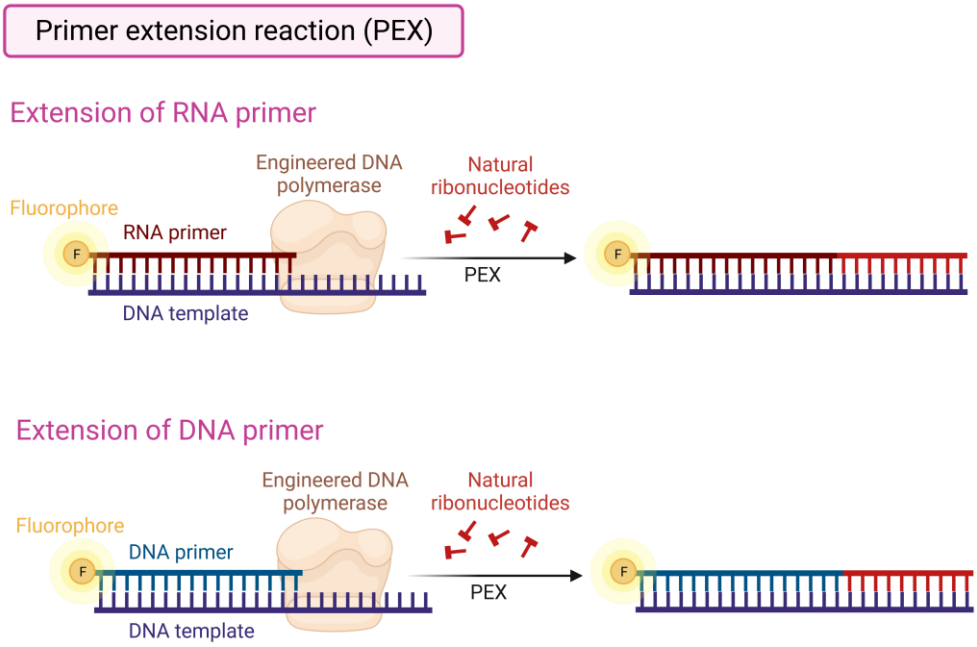



**Figure 11. Incorporation of natural rNTPs by PEX and with usage of engineered DNA polymerases, starting from either RNA or DNA primer.**

Both, primer and template can be tagged, with biotin labelling of template being particularly useful. A biotin-labelled template can be removed after PEX reaction with streptavidin magnetic beads. Denaturation through basic pH or elevated temperature results in recovery of the ssRNA product[130,131]. Stability of streptavidin-template interaction can be further enhanced by usage of dual-biotin labelled template[132]. This method is commonly referred to as "magnetoseparation" or "strand separation". An alternative method for DNA template removal involves digestion by DNase, which non-specifically cleaves DNA template into small fragments, leaving the RNA strand intact. In this case, it is not necessary to use a costly labelled DNA template. The ssRNA product is usually isolated by simple silica spin column purification (Figure 12). However, this approach is not suitable for PEX from a DNA primer, as the nuclease would destroy both template and primer. Another template removal technique includes digestion by lambda exonuclease for removing 5'-phosphorylated templates[133,134] or purification and separation by preparative dPAGE followed by isolation through gel extraction[135].

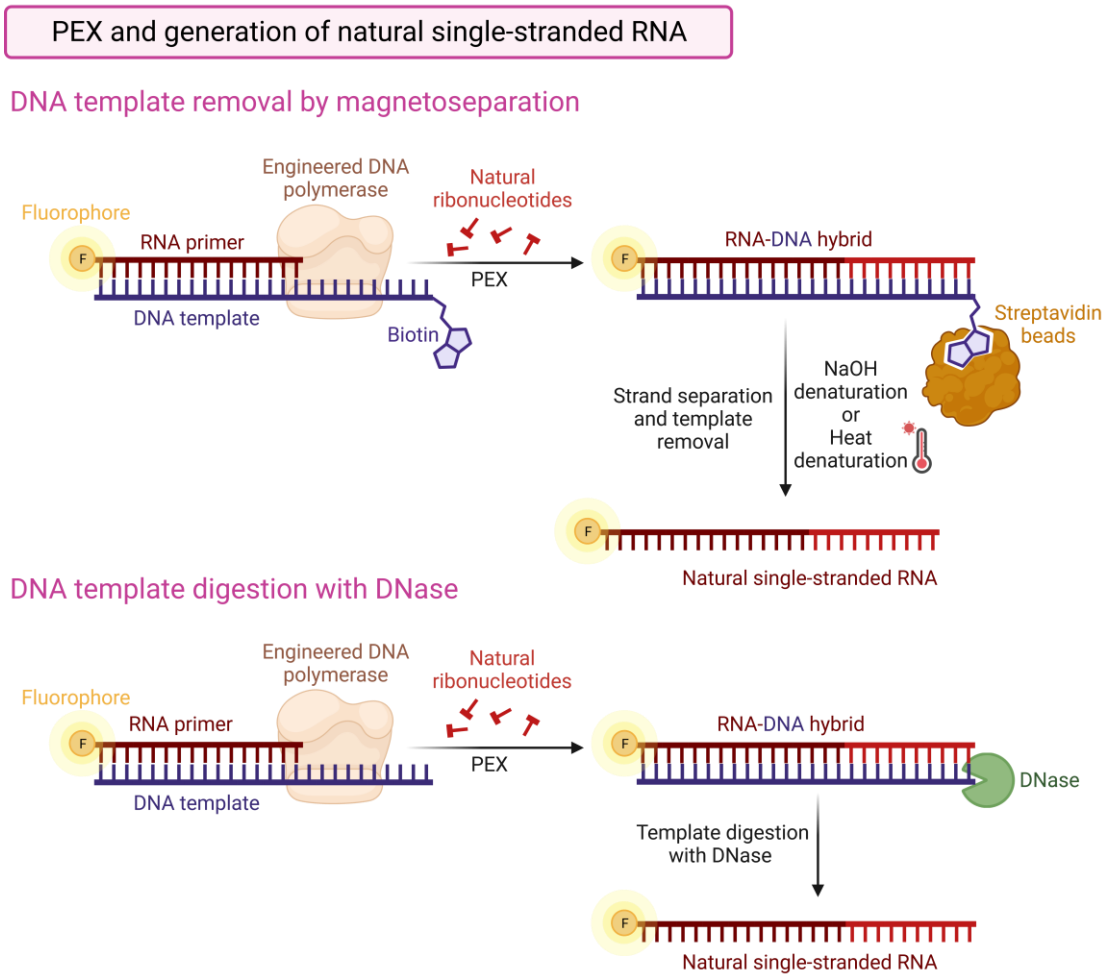



**Figure 12. PEX reaction followed by generation of single-stranded RNA either by template removal with magnetoseparation or digestion with DNase.**

Similarly to T7 RNAP, polymerases in PEX, particularly those lacking 3'-5' exonuclease proofreading activity, may add extra nucleotides at 3'-end of the synthesised strand (non-templated addition). Complete cessation or at least reduction of non-templated nucleotide addition can be achieved when employing 5'-*ortho*-twisted intercalating nucleic acid template (Figure 13)[136].

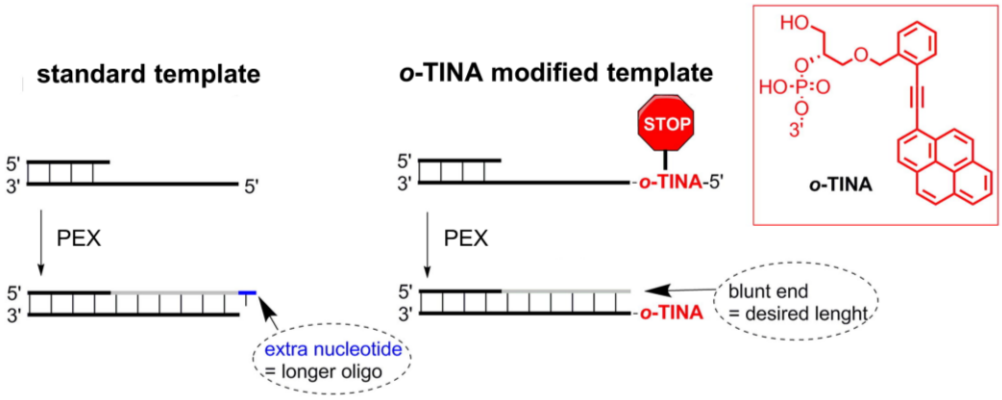



**Figure 13. Primer extension reaction with non-modified and/or 5'-end *o*-TINA labelled DNA template. Modified figure from literature[136].**

Although PEX is a robust method for enzymatic synthesis of nucleic acid, it results in a uniformly modified oligonucleotide at each position. A recent development by Ménová et al. established single nucleotide incorporation method (SNI), enabling site-specific internal labelling of DNA oligonucleotides[137]. In this method, DNA primer is initially extended by a single modified deoxynucleotide (**dN$^{X}$TP**), followed by addition of an abundant amount of all four canonical nucleotides (dNTPs), resulting in full extension along the template. This leads to insertion of the modified nucleotide immediately after primer, followed by natural ones (Figure 14).

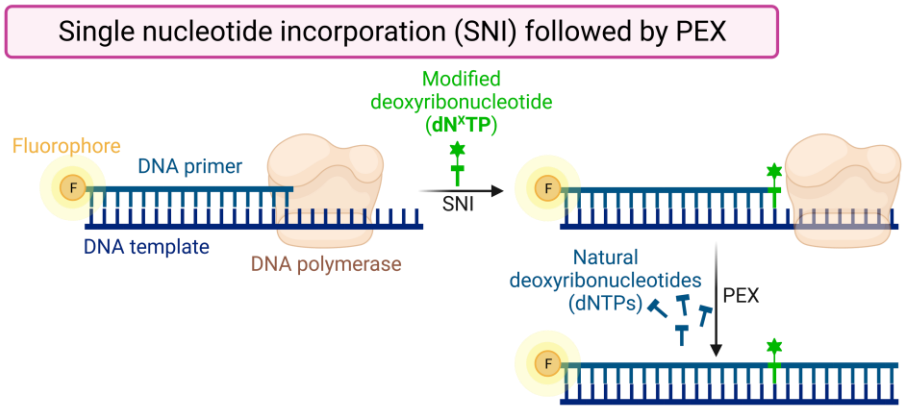



**Figure 14. Strategy for internal DNA labelling employing SNI and PEX reaction.**

For shorter RNA oligonucleotides (< 100 nt), the analysis of extension products typically involves dPAGE and mass spectrometry techniques such as matrix-assisted laser desorption/ionization time-of-flight (MS-MALDI-TOF) or electrospray ionisation liquid

chromatography-mass spectrometry (ESI-LC-MS). In case of longer RNA sequences (e.g., mRNA), characterisation is achieved by low-density dPAGE, native agarose or Sanger sequencing.

Applications of PEX are constrained by length of ssDNA template, typically restricted to around 200-300 nt with poor yields in solid phase synthesis. To overcome this limitation and produce longer ssDNA templates exceeding 300 nt, an alternative approach involves combining PCR with strand separation techniques. PCR is a robust method that can generate dsDNA fragments spanning several kb in length, starting from small quantities of template. The PCR process is facilitated by thermostable polymerases capable of withstanding harsh conditions, including high temperatures exceeding 90°C during denaturation step[138]. In this process, two primers, designed to be complementary to dsDNA template, are utilised in excess. PCR procedure involves cycles of denaturation (strand separation), annealing and partial extension of primers along template, and final full extension. This leads to exponential amplification, producing millions to billions of copies of a specific DNA sequence (Figure 15). To generate ssDNA, one of the two primers is typically labelled at 5'-end either with biotin or phosphate. Upon PCR, template removal is ensured by digestion with lambda exonuclease[134] (applied for phosphate-labelled primer) or strand separation using streptavidin magnetic particles[131,132] (in case of biotin-labelled primer).

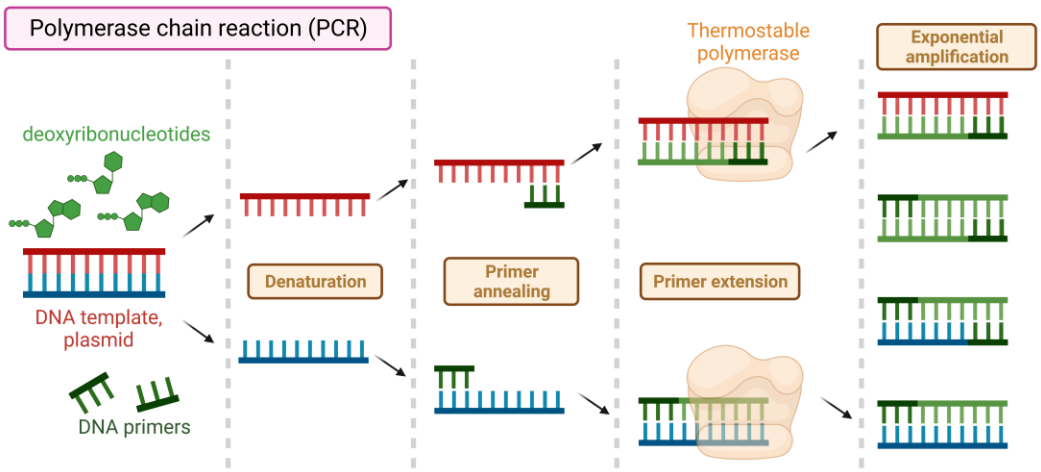



**Figure 15. Principle of polymerase chain reaction (PCR).**

### 1.4.4 Post-synthetic derivatisation pathways leading to modified RNA probes

Although the enzymatic synthesis offers the capability to introduce various chemical functionalities into RNA, there is a noticeable decrease in polymerase efficiency when incorporating bulky functional groups[43]. Consequently, incorporating a less perturbing reactive functionality into RNA opens possibility for post-synthetic methods to functionalise RNA. Reactive labels such as

alkynes, azides, iodine, etc., are small and are commonly used in post-synthetic or post-transcriptional reactions. They induce minor structural changes and are well tolerated by RNA polymerases[110,139,140]. Chemoselective Suzuki-Miyaura cross-coupling reaction of iodo-modified RNA with various florescent of affinity-tagged boronic acids was ensured by water soluble catalytic system[112]. Besides Suzuki coupling, another Pd-mediated cross-coupling reaction was well studied for derivatisation of RNA. Stille coupling was performed with vinyl and heterocyclic tributylstannanes of increasing bulkiness[141]. Recently, the Srivatsan group reported vinyl-labelled RNA with a dual purpose. They reported for the first time oxidative Heck reaction with boronic acids or esters and inverse electron-demand Diels-Alder (IEDDA) with tetrazines bearing fluorescent or affinity tags[111]. The first successful RNA labelling using IEDDA reaction was presented by the Jäschke group in 2011[142]. Norbornene-modified RNAs were labelled with electron-deficient dienes, dansyl- or biotin-modified tetrazine counterparts. Biorthogonal IEDDA was applied also for fluorescent RNA labelling. The significantly more reactive *trans*-cyclooctene-modified RNA was post-transcriptionally labelled with fluoresceine-bearing tetrazine[116]. Alkyne-decorated RNA transcripts were efficiently derivatised *via* copper-catalysed CuAAC reaction with azide-bearing fluorescent tags, affinity labels, sugars or groups resembling amino acid side-chains[114]. On the other hand, azide-modified RNA can be used for copper- free SPAAC with activated cyclooctynes, beyond the CuAAC, or additionally employed as a substrate for Staudinger reaction[143]. Efficient insertion of azide-modified nucleotides, for post-transcriptional click labelling, has been demonstrated not only by T7 RNAP IVT, but also by poly(A) polymerase[144].

In addition, an effective method for internal RNA labelling using methyltransferases and synthetic analogues of *S*-adenosyl-L-methionine, as substrates, was demonstrated by Rentmeister et al[145]. A small vinyl group was successfully used in photo-click reaction with diphenyltetrazol[146] along with transfer of alkyne or azido-modification for SPAAC[147,148], what was additionally validated also by other researchers[149]. Another well-established technique for incorporating functional groups into RNA is splinted ligation. This process involves joining two short RNA oligonucleotide fragments – typically, one non-modified prepared by IVT, and the other modified, usually obtained through solid phase synthesis. In this method, a ssDNA template, known as a splint, acts as a bridge. These two short RNA fragments hybridise to complementary DNA template, bringing them into proximity for ligation of 3'-hydroxyl group (acceptor) of one RNA with 5'-end monophosphate (donor) of second RNA strand. This results in production of a long RNA oligonucleotide with internal modifications placed at specific positions[150–154].

Finally, native structure of RNA, featuring 3'-terminal 2',3'-*cis*-diol (vicinal) moiety, was utilised for direct chemical labelling. Usage of sodium periodate ($NaIO_4$) facilitated ring opening

(oxidation), resulting in a 1,5-dialdehyde functionality. This reactive intermediate was subsequently labelled with primary amines carrying fluorophores (such as FAM, Cy5, Cy3) or biotin, in presence of a reducing agent ($NaBH_3CN$), forming a stable morpholine derivative[155].

## 1.5 Revealing the importance of modified RNA

### 1.5.1 Application of cross-linking for studying RNA-protein interactions

RNA-protein interactions are fundamental processes in molecular biology, playing a crucial role in various cellular functions. RNA-binding proteins (RBPs) are key players in these interactions, recognising specific RNA sequences or structures and forming dynamic complexes[156,157]. Therefore, a comprehensive understanding of RNA-protein interactions is essential for deciphering cellular pathways, understanding disease mechanisms, and advancing the development of innovative therapeutic approaches and biotechnological applications. Consequently, numerous methods have been developed for uncovering RNA-protein interactions or sequencing of RNA molecules interacting with a protein of interest, categorising them into two groups[158–160].

Protein-centric methods aim to discover unknown RNA molecules that interact with known RBPs. Among them, cross-linking immunoprecipitation (CLIP) has found widespread applications for identification of novel RNA targets of specific proteins. They rely on UV-triggered cross-linking between uracil nucleobases and neighbouring amino acid residue of proteins, that stabilise the whole complex. Affinity purification with antibodies against the protein is employed to enrich covalently attached RNA molecules[161,162]. Great leap forward has been made by integration of high throughput sequencing in the workflow (HITS-CLIP). Reverse transcription of the RNA-protein complex produces mutation, deletion, or truncation at the cross-linked nucleotide, which enables to footprint exact binding sequence. After PCR amplification of cDNA and preparation of library, deep sequencing reveals exact sequences of interacting RNA molecules[163]. Several improved protocols for sequencing library preparation, such as enhanced CLIP (eCLIP)[164] or individual-nucleotide resolution CLIP (iCLIP)[165], have been established. Notably, inclusion of photoactivatable ribonucleotides in the protocol (PAR-CLIP), enhances cross-linking yield, contributing to improved coverage and sensitivity in detecting RNA-protein interactions[166].

While protein-centric methods provide insights into the sequences of bound RNA molecules, the discovery of unknown RBPs poses a significantly greater challenge. Unlike sequencing, current mass spectrometry methods lack the capacity for amplification, and the sequence diversity of proteome exceeds that of four nucleotides present in nucleic acids. Nevertheless, advancements in the field of liquid chromatography and mass spectrometry, characterised by enhanced separation power and increased detection sensitivity, have paved the way for the development of several RNA-centric

approaches aimed at identifying novel RNA-associated proteins[167]. The most straightforward method harnesses the polyadenylation of mRNA transcripts. UV-cross-linked mRNA is subsequently isolated using poly(T) magnetic beads with successive washes to eliminate non-RBPs[168,169]. However, this method is limited to polyadenylated molecules, which are not universally present in all classes of RNA. Broader range of RBP-RNA cross-links can be enriched based on the physico-chemical properties of these complexes. Methods such as guanidinium thiocyanate-phenol-chloroform[170] or acidic phenol-toluol[171] extraction result in accumulation of RNA-protein complexes at interphases, thereby enabling effective separation from non-cross-linked RNA or proteins. Moreover, RNA strongly adsorbs to silica gel beads in presence of chaotropic reagents. Following the tryptic digestion of RNA-protein complexes, the resulting peptide-RNA conjugates are captured by silica beads[172]. Hence, combination of organic solvent extraction and silica-bead capture effectively removes non-crosslinked peptides[173].

### 1.5.2 Methods for capturing RNA interacting partners by cross-linking

Numerous methods exist for identification and characterisation of RNA-RNA and RNA-protein interactions, offering valuable insights into functional roles of RNA and its involvement in diverse cellular processes. RNA pull-down assay is a valuable technique designed to investigate specific RNA-protein interactions by capturing RNA-protein complexes through affinity-based methods without formation of a stable covalent bond between interacting partners. In this assay, the RNA sequence of interest is immobilised onto a solid support, such as magnetic beads or a chip, followed by incubation with a cellular lysate containing a diverse mixture of proteins. One of the simplest methods for RNA immobilisation involves end-labelling with biotin for subsequent fishing out with streptavidin magnetic beads[174]. Additionally, specific RNA aptamer sequences can be employed as an alternative approach for implementing tags[175,176]. The underlying principle of this method is that only proteins with specific affinities for immobilised RNA sequence will interact, forming relatively stable complexes that can withstand pull-down isolation. After the incubation period, non-specifically bound proteins are typically washed away, retaining only those proteins that have specific interactions with the RNA probe. These RNA-protein complexes are then eluted from solid support for further analysis[160] (Figure 16). While the RNA pull-down assay provides some insights into RNA-protein interactions, it is crucial to consider certain setbacks. A critical weakness is inefficiency in capturing weak or transient RNA-protein interactions, which can be solved by employing covalent bonds (cross-linking) between the interacting partners.

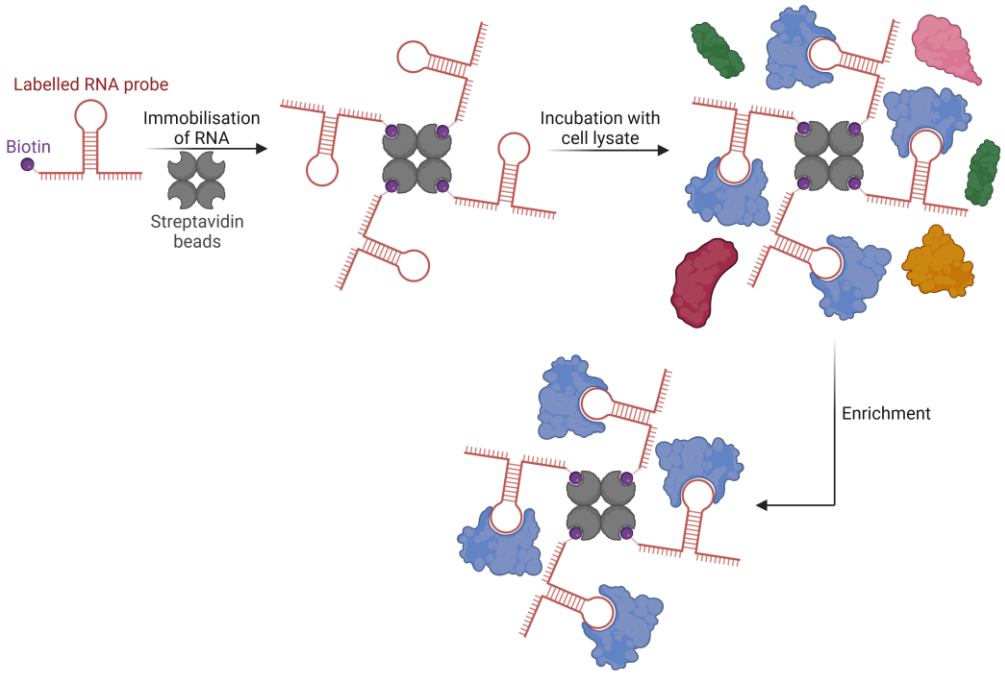



**Figure 16. Non-covalent pull-down methodology for identification of RBPs.**

External cross-linking agents play a crucial role in study of RNA structures or interacting partners, offering a notable advantage by eliminating the need for chemical or enzymatic synthesis of modified RNA probes. These agents are readily accessible and are employed to generate both intra- and intermolecular cross-links[177] (Figure 17). Photoreactive compounds, such as psoralen and its derivatives, represent a widely used class of external RNA cross-linking reagents, that have been leveraged in investigating RNA-RNA interactions. Psoralen's mechanism involves a [2+2]-photocycloaddition, where covalent bond formation occurs between two opposing pyrimidine nucleobases[178,179]. Another noteworthy option is utilisation of dithiothreitol (DTT) as an RNA-protein cross-linking agent. In this approach, upon exposure to UV light, DTT selectively forms covalent bonds between cysteine (Cys) residues within RBPs and the uracil nucleobase of RNA probe in proximity[180]. Another class of cross-linking agents includes formaldehyde or 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC). In contrast to methods relying on UV light activation, these agents facilitate formation of covalent bonds between RNA molecules and their interacting partners without the need of external activator. Formaldehyde cross-linking involves the reversible formation of methylene bridges between amino groups of proteins, preferentially with lysine (Lys) residues[181], and nucleophilic sites on RNA, or between two RNA strands, providing access to fixation of RNA-protein or RNA-RNA interactions under mild conditions[182]. On the other hand, EDC activates the carboxyl groups of proteins, facilitating amide bond formation with neighboring

amino groups of nucleic acid bases[183,184]. Despite the potential advantages offered by external cross-linking agents for studying RNA-protein interactions, their utilisation is relatively limited. This restraint is attributed to several factors, including inefficient cross-linking reactions, the cytotoxic nature of these agents for living cells, and the propensity for non-specific protein-protein conjugations and excessive cross-linking of indirect interaction partners. These challenges pose significant hurdles in achieving reliable and interpretable results.

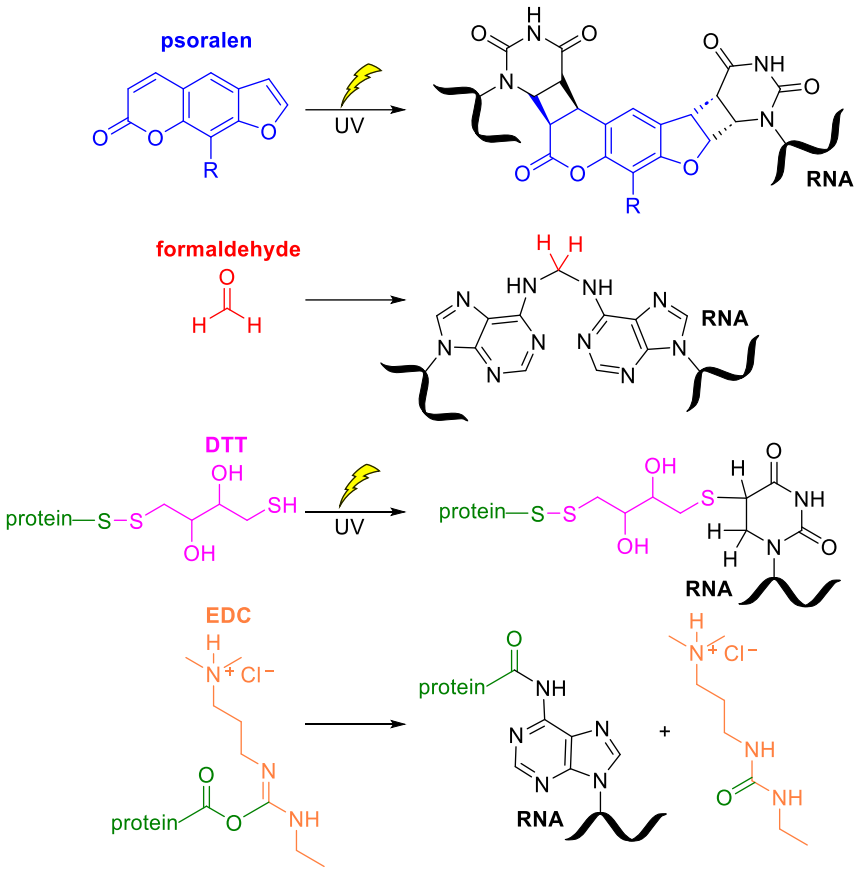



**Figure 17. Overview of selected external cross-linking agents and their mechanism of action.**

RNA photocross-linking has become a valuable tool for studying the dynamics of RNA-protein interactions, providing insights into structural details and functional roles of these complexes. UV light-induced photocross-linking presents a distinct advantage in maintaining the integrity of natural RNA-protein interactions, that might be compromised when external cross-linking agents are employed. The irreversible covalent bond formation induced by short-wavelength UVC (254 nm) primarily involves the uracil nucleobases (Figure 18) of non-modified RNA probes, with other bases like adenine, guanine, and cytosine exhibiting rare instances of cross-linking. The non-modified uridines are cross-linked at position C-5, leaving the Watson-Crick pairing intact[185]. It is noteworthy,

however, that usage of UVC for RNA-protein cross-linking has limitations due to its potential damage to nucleic acids causing phosphodiester bond cleavage or unwanted inter- and intra-strand cross-linking. This method is also of low efficiency and is consequently restricted in its *in vivo* applications.

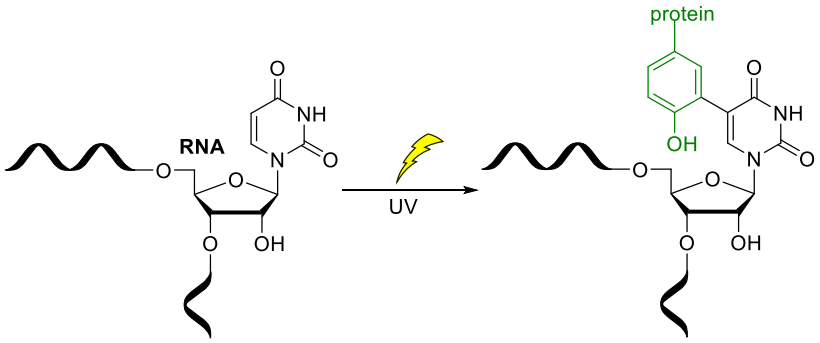



**Figure 18. UVC induced RNA cross-linking with aromatic amino acid residues through uracil nucleobase.**

An alternative strategy to improve RNA-protein cross-linking efficiency and overcoming upper-mentioned drawbacks of conventional photocross-linking involves utilising modified nucleotide building blocks with photoactivatable functional groups. Notably, 5-halogenated pyrimidines[186–189] (such as 5-iodouridine, 5-iodocytidine or 5-bromouridine) and thiol-modified ribonucleosides (including 4-thiouridine and 6-thioguanosine)[190] were extensively studied and widely employed for this purpose (Figure 19). The thiol-modified building blocks are preferred due to high stability and simple structure tolerated by RNA polymerases[191]. These photosensitive moieties become selectively activated upon exposure to long-wavelength UV light (>310 nm), a range where natural nucleotides do not undergo cross-linking. Consequently, this activation leads to the formation of photo-adducts with diverse groups, influenced by the reaction mechanism involving either short-lived radical intermediates or direct photoaddition. Contrary to its natural counterpart, 4-thiouridine nucleobase undergoes a reaction at C4 position with the amino acid residue, leading to changes in its base-pairing properties[185]. Reactivity towards these nucleoside analogues varies among amino acid side chains. Aromatic amino acids such as phenylalanine (Phe), tyrosine (Tyr), and tryptophan (Trp) exhibit predominant reactivity, followed by Lys and Cys, contributing to overall diversity[185,188].

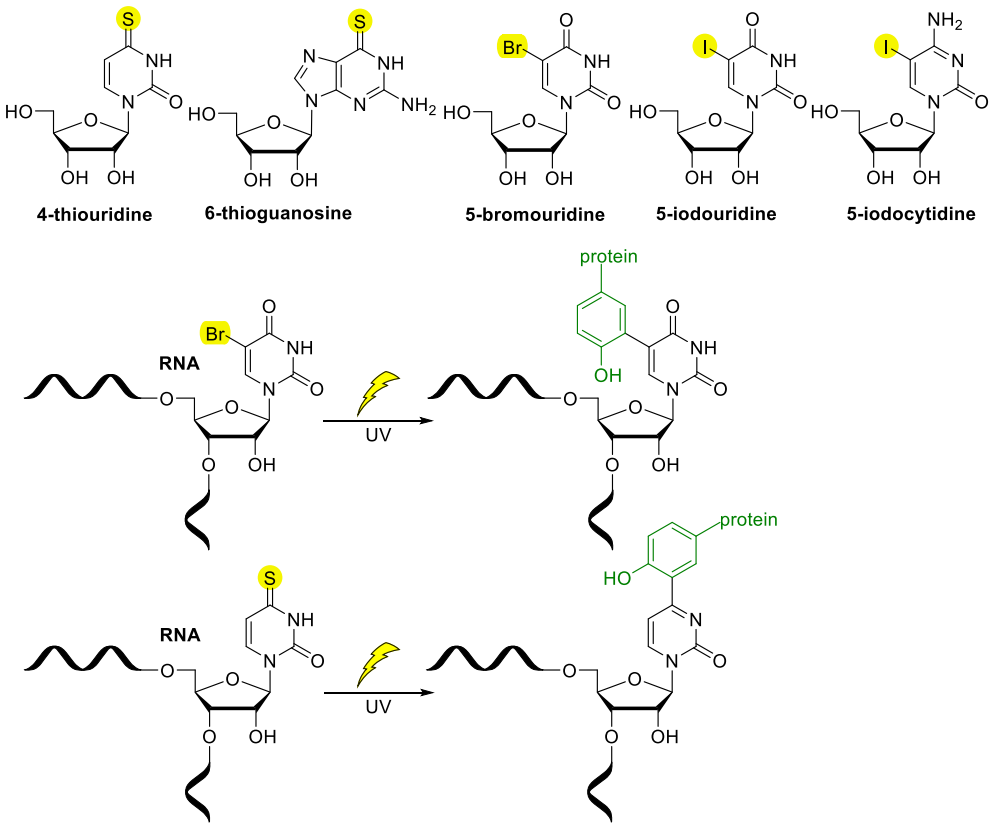



**Figure 19. The structures of modified ribonucleotide analogues with photoactivatable functionalities and their UV induced photocross-linking with aromatic amino acids.**

Other, more UV-sensitive functional groups, such as aromatic azides, benzophenones, or diazirines, have been utilised for RNA-protein cross-linking. These groups, upon irradiation, generate highly reactive intermediates, including nitrenes, diradicals, or carbenes[192–194] (Figure 20). Arylazide modifications are simple to prepare and were used for interstrand cross-linking of RNA duplex[195] or RNA-protein cross-linking[196]. Upon exposure to UV light with a wavelength of 250 nm, the functional groups initiate generation of nitrene intermediates, facilitating cross-linking reactions. Despite the efficacy of this approach, there are limitations associated with its usage. The damaging effects of short-wavelength UV light may limit the scope of experiments or applications. On the other hand, benzophenone and diazirine functionalities are activated with biologically more relevant long wavelength UV light at 365 nm. Diazirine functional groups stand out as particularly efficient for RNA-protein cross-linking due to their capability to react not only with typically unreactive C-H but additionally with N-H, O-H and S-H bonds[197], showcasing superior cross-linking efficiency, especially when compared to sulphur- or halogen-modified nucleosides[198]. Incorporation of diazirines into RNA is predominantly achieved through chemical synthesis[117,198] or by post-synthetic methods, such as cross-coupling reactions[199]. There are only sporadic examples of T7 RNAP-

mediated incorporation into nascent RNA strand[117]. Diazirine-decorated RNA probes have been leveraged in photocross-linking reactions for identification of RBPs[198–200]. Notable advantage of diazirine modification lies in its relatively small size, stability, and high reactivity. Upon exposure to UV light irradiation, diazirine is activated, leading to formation of a highly reactive carbene species by eliminating nitrogen. Diazirine functional groups, despite their utility in cross-linking reactions with proteins, are associated with relatively low yields, often falling below 15%. One of the primary drawbacks contributing to this limitation is quenching of the reactive species generated during the diazirine activation process. Small molecules, such as water, can readily quench these reactive intermediates and decrease the efficiency of cross-linking reactions[117].

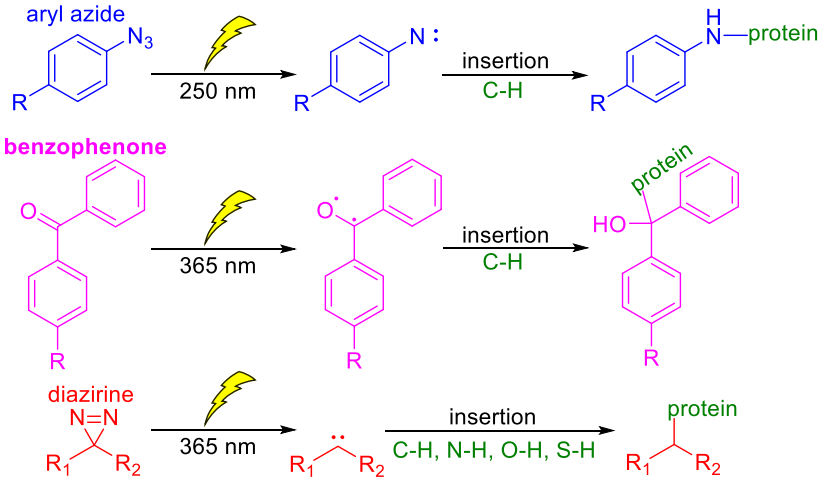



**Figure 20. Reactivity of aryl azide, benzophenone and diazirine modified species in cross-linking with RBPs.**

Addressing the upper mentioned challenges associated with UV-induced nucleic acid damage, non-specific amino acid targeting, inadequate efficiency, and low yields in conventional cross-linking methodologies necessitates a strategic shift towards the integration of reactive functional groups designed for amino acid-specific targeting. This innovative approach holds significant promise in overcoming the limitations inherent in traditional UV-based cross-linking methods, providing a more precise and controlled mechanism for investigating nucleic acid-protein interactions without the need of external cross-linking reagents or the activation of reactive functionalities. The complexity of achieving reactions with specific amino acid residues within a myriad of reactive functional groups, including carboxylic acids, amides, amines, hydroxyls, and thiols, poses a considerable challenge in terms of chemo- and regioselectivity. Moreover, these reactions must adhere to biologically ambient conditions characterised by temperatures at 37 °C, neutral pH levels ranging from 6 to 8, and the use of aqueous solvents to ensure preservation of

protein structure and function[201]. Given that many canonical amino acids function as nucleophiles, introducing selective electrophiles into nucleic acids becomes a viable strategy for their targeted capturing. The installation of these reactive functionalities on the phosphate or sugar backbone is an option, but nucleobase emerges as the optimal position, eliminating structural changes to oligonucleotide, preventing disruption of natural RNA-protein interactions, and enabling enzymatic incorporation. In case of DNA, pioneering works have been predominantly published by Hocek and colleagues. They utilised novel **dN$^{X}$TP**s bearing reactive functionalities for preparation of modified DNA probes, enabling identification of DNA-binding proteins (DBPs), fluorescent DNA labelling, and bioconjugation with peptides in both major and minor grooves of double-stranded helix[202–210]. Vinylsulfonamide and acrylamide reactive groups have been employed to selectively target Cys residues through Michael addition[207,208]. Later on, a highly reactive chloroacetamide functionality has been designed to target both Cys and histidine (His) amino acids in peptides and proteins[206]. Despite the broad spectrum of potential targets, Lys are particularly attractive due to their abundant occurrence. Aldehyde functional groups have been strategically placed in the major[211] and/or minor grooves[204] of dsDNA to form reversible Schiff-base or stable covalent bonds with Lys residues upon reductive amination with sodium cyanoborohydride ($NaBH_3CN$). In pursuit for even more robust methods, a squaramate modification has been introduced to react specifically with Lys in peptides and proteins. Notably, this approach eliminates the need for harmful external reagents such as $NaBH_3CN$ to stabilise DNA-protein interaction, as is the case with aldehyde moieties[205]. Recently developed 1,3-diketone[203] and glyoxal[202] functional groups provide additional options for selectively targeting arginine- (Arg-) containing peptides and proteins, showcasing the continual improvement and expansion of cross-linking methodologies for precise and diverse amino acid interactions (Figure 21).

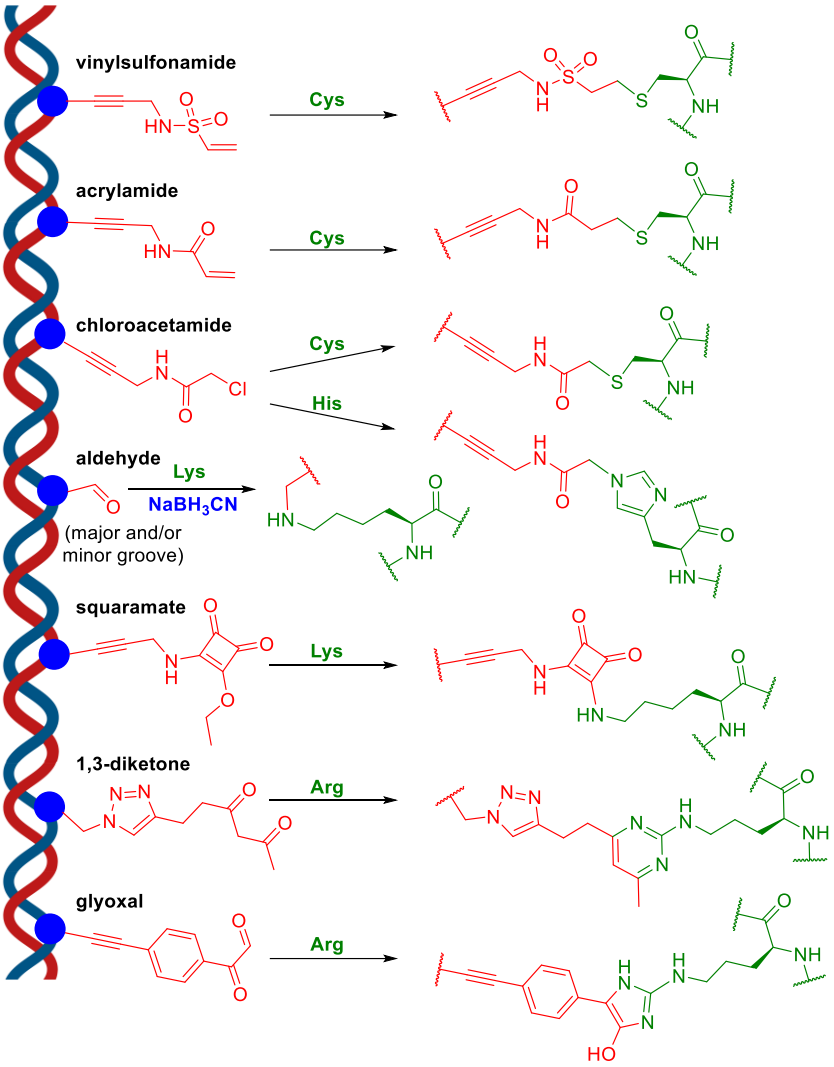



**Figure 21. Overview of modified nucleotides with reactive functionalities incorporated into DNA for selective cross-linking with specific amino acid residues of DBPs.**

On the other hand, in case of RNA there are only a handful of reports about reactive functional groups for selective targeting of specific amino acid residues. A notable advancement in this area involves attachment of the squaramate functional group at 2'-position of the RNA sugar backbone, as reported in recent study. The 2'-azido functional group of chemically synthesised RNA probe was in the first step reduced to amino group followed by post-synthetic functionalisation with diethyl squarate. The prepared squaramate-modified RNA oligonucleotide readily reacted with Lys residues of peptide and aminoacyl-transferase $FemX_{Wv}$[212](Figure 22). In addition to this method, alternative strategies for selective amino acid cross-linking involve the utilisation of 5-halo and 5-azapyrimidines, such as 5-fluorocytidine[213] and 5-azacytidine[214], in metabolic labelling. These compounds are incorporated into nascent RNA by cellular RNAP. Due to poor stability of 5-azacytidine and its conjugates, 5-

fluorocytidine is often preferred. Interestingly, the deamination of 5-fluorocytidine results in formation of 5-fluorouridine, which can capture uridine-methyltransferases, thereby broadening the scope of investigated enzymes. These modified nucleotides selectively target catalytic Cys residues of RNA methyltransferases, forming irreversible covalent bonds (Figure 22). This interaction is identified through an anti-RBP antibody pulldown strategy, facilitating selective capture and study of RNA-protein interactions[213,214]. Both methods exhibit limitations in their applicability for identification of a broad range of RBPs. These limitations stem from either intricate multi-step synthesis required for generating of the reactive RNA probe limited only to targeting Lys amino acids[212], or restriction to metabolic labelling followed by targeting only specific groups of enzymes acting on RNA[213,214].

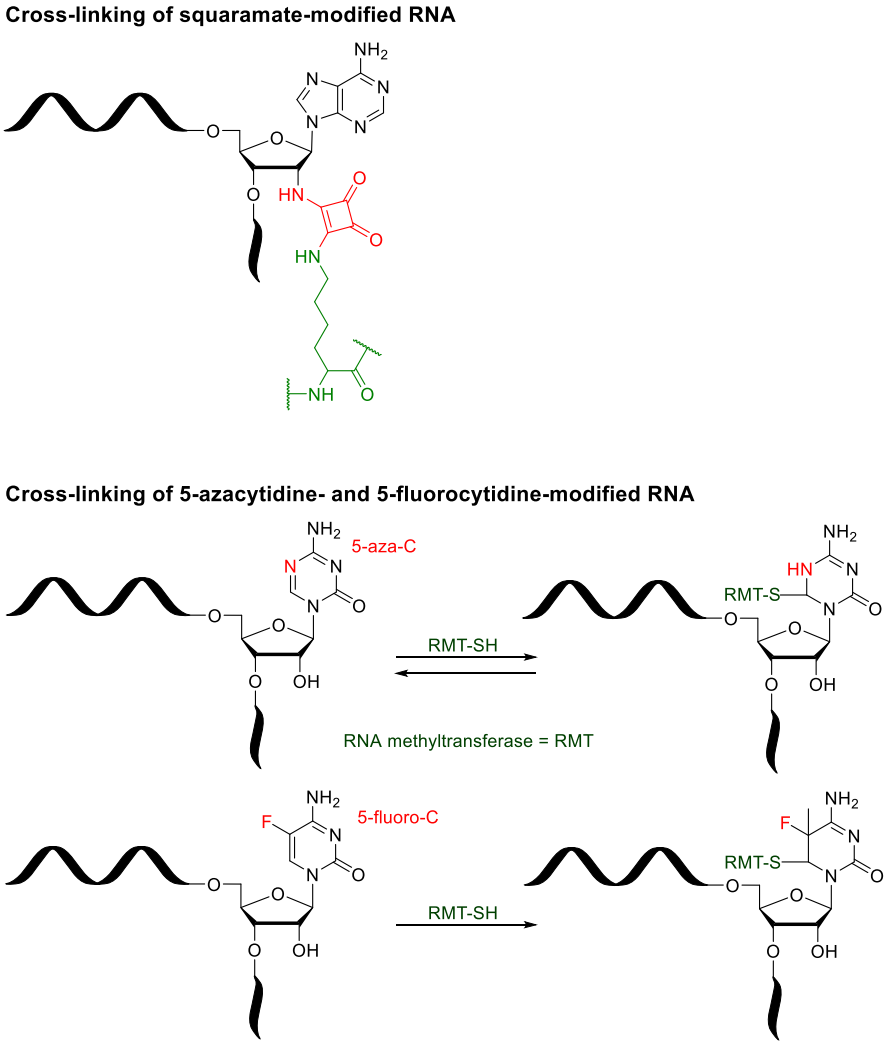



**Figure 22. Reported examples of selective RNA cross-linking with specific amino acid residues.**

### 1.5.3 RNA-based therapeutic agents

Modified RNA has found applications in development of next generation therapeutics for conditions difficult-to-target by conventional small molecule drugs. Oligonucleotides, unlike antibodies, can be prepared both enzymatically and chemically, which enables manufacturing scale-up, flexible derivatisation and modification to enhance stability. However, thanks to their highly charged backbone, intracellular delivery needs to be mediated by complexation with cationic lipids or hydrophobic group attachment[215]. Additionally, unmodified RNA is highly prone to nuclease degradation and as such needs to be stabilised by shielding from outer environment (e.g., within liposomes). Another option is introducing unnatural backbone or phosphate modifications. Most widely used in FDA-approved oligonucleotide drugs are ribose modifications 2'-*O*-methoxyethyl (2'-MOE), 2'-*O*-methyl and 2'-fluoro groups together with phosphate modifications – phosphorothioates[216]. RNA therapeutics can be divided into oligonucleotides (antisense oligonucleotides, siRNA, aptamers) and long mRNA. RNA-guided nucleases have not reach clinical applications yet, however many studies are currently ongoing (Figure 23).

ASO are usually short 12-25 nt long ssRNA-derived synthetic oligonucleotides. Upon delivery to cells, they can affect RNA fate depending on base-pairing complementarity. They can act through several distinct mechanisms of action. RNAse H involving routes lead to degradation of target RNA, which decreases pathological protein expression. Binding to mRNA prevents ribosome from protein translation. First approved drug – fomivirsen mediated degradation of UL123 protein from cytomegalovirus[217]. ASO can be also used to eliminate proteins truncated due to premature aberrant stop codon. Contrary, protein levels can be increased by reshaping untranslated regions (UTRs), which enhances translation initiation. Chemically, most ASO contain a mix of differently modified nucleotides[218]. Nucleobase modifications are recently emerging as a strategy to modulate side-effects and toxicity and warrant further exploration[219].

Unlike ASO, siRNAs are formed from two annealed complementary 21-25 nt long RNA oligonucleotides. They act through RNA interference (RNAi) pathway. Target mRNA is cleaved and degraded with Argonaute-2 (Ago-2) upon loading into RNA-induced silencing complex (RISC)[220]. siRNA approved drugs are currently used in therapy of diseases affecting liver, where siRNA can be readily accumulated due to *N*-acetylgalactosamine (GalNAc) derivatisation. Chemical modifications are generally analogous to the ASO. Several siRNA containing modifications to nucleobases are in the stage of research and development[221]. Notably, triazole-based nucleobase enhances binding to Ago-2 pockets and leads to lower off-target cleavage[222].

Aptamers are oligonucleotide entities that can act as specific affinity probes mimicking antibody-antigen interactions[223] by adopting diverse secondary structures[224]. They offer numerous

advantages over antibodies, such as enhanced pH and thermal stability, reduced toxicity and immunogenicity[225], and crucially, more straightforward and cost-effective synthesis[226–228]. Aptamers are selected from a large pool of random sequences by a process known as SELEX (Systematic Evolution of Ligands by Exponential enrichment). The term was initially introduced in 1990 by Tuerk and Gold who documented the selection of an RNA oligonucleotide binding to bacteriophage T4 DNA polymerase[229]. Simultaneously, Ellington and Szostak independently reported an RNA aptamer binding to a small organic dye[230]. Since then, a considerable number of aptamers have been acquired by SELEX. Despite their ready accessibility, RNA aptamers in their unmodified form are highly susceptible to nuclease degradation, rendering them generally unsuitable for therapeutic applications. Efforts to address these issues typically involve incorporation of specific modifications at 3'-ends of nucleic acid chains with inverted thymidine or biotin, replacement of sugar 2'-OH with F-, $NH_2$-, or MeO-groups and concerning the phosphodiester linkage, common approach involves substitution with phosphorothioate or methylphosphonate analogs to enhance nuclease stability[231–234]. Despite these efforts, success rate in obtaining aptamers with both high affinity and specificity remains disappointingly low. To date, only one modified RNA aptamer has received approval for clinical use. In December 2004, the US FDA granted approval for pegaptanib sodium (Macugen), an RNA aptamer against vascular endothelial growth factor (anti-VEGF), for the treatment of all forms of neovascular age-related macular degeneration (AMD)[235]. Beside this, several modified RNA aptamers treatment of coronary artery disease[236], myeloma and non-Hodgkin lymphoma[237], chronic inflammatory diseases[238], age-related macular degeneration[239], anemia[240], hemophilia[241] are currently undergoing clinical trials or are in the developmental pipeline[231,242–244].

mRNA therapy currently attracted attention during SARS-Cov-2 pandemic outbreak, where development of efficient vaccines demonstrated highly modular and flexible nature of mRNA molecules, and thereby rapid access to functional and safe disease prevention[245]. However, history of mRNA medicines started long time ago, when mRNA importance for therapy has been recognised. Key milestones in development were discovery of nucleobase modifications by Karikó and Weissman, that led to enhanced translation, and more importantly, to shielding from innate immunity receptors[29,246]. Another important step was design and implementation of novel lipidic nanoparticles for efficient intracellular delivery[247]. Development of mRNA constructs comprises thorough optimization of 5'- and 3'-UTRs, coding sequence, poly(A) tail, cap, and carrier. Crucial point is inclusion of nucleobase modifications. Apart from cap structures, best explored and widely used for coding region are $N^1$-methylpseudouridine (m1Ψ), $m^5C$, s2U and $m^6A$[248]. mRNA is commonly manufactured by T7 RNAP IVT from linearised plasmids[249]. Here an important factor is efficient purification, as dsRNA contaminants from IVT can impair translation and elicit immune response.

Although costly and tedious, HPLC purification has been found to be most effective in removal of these by-products[250].

The discovery of clustered regularly interspaced short palindromic repeats, CRISPR-based immune defence system in bacteria[251], along with elucidation of its mode of action[252], has rekindled interest in gene engineering and therapy. The most widely used CRISPR-Cas9 system comprises the Cas9 nuclease protein, which binds to trans-activating RNA (tracrRNA) and CRISPR RNA (crRNA), guiding the nuclease to complementary sequence. Upon binding, the nuclease generates a blunt-ended dsDNA break. This break is then reconnected by nonhomologous end joining, leading to indels, and ultimately resulting in gene knockout. Alternatively, it can be repaired by homologous recombination with a sister chromosome or a provided template. This versatility allows for insertion of desired sequences or changes to chromosomal loci of interest[253]. It has been discovered that tracrRNA and crRNA can be fused to single guide RNA (sgRNA) which is about 100 nt long. This simplifies the whole system and enhances efficiency and specificity[252]. The sgRNA tightly associates with Cas9 protein in multiple regions, and introduction of modifications has been proposed to enhance its stability and specificity towards DNA recognition[254]. Chemical synthesis of sgRNA is preferred, as most backbone modifications are incompatible with T7 RNAP. The introduction of 2'-*O*-methyl-3'-thiophosphonoacetate has been particularly effective in significantly enhancing genome editing[255]. Another study demonstrated the advantages of incorporating 2'-F groups[256]. Although nucleobase modifications have not been explored to date, they certainly merit further investigation and interest.

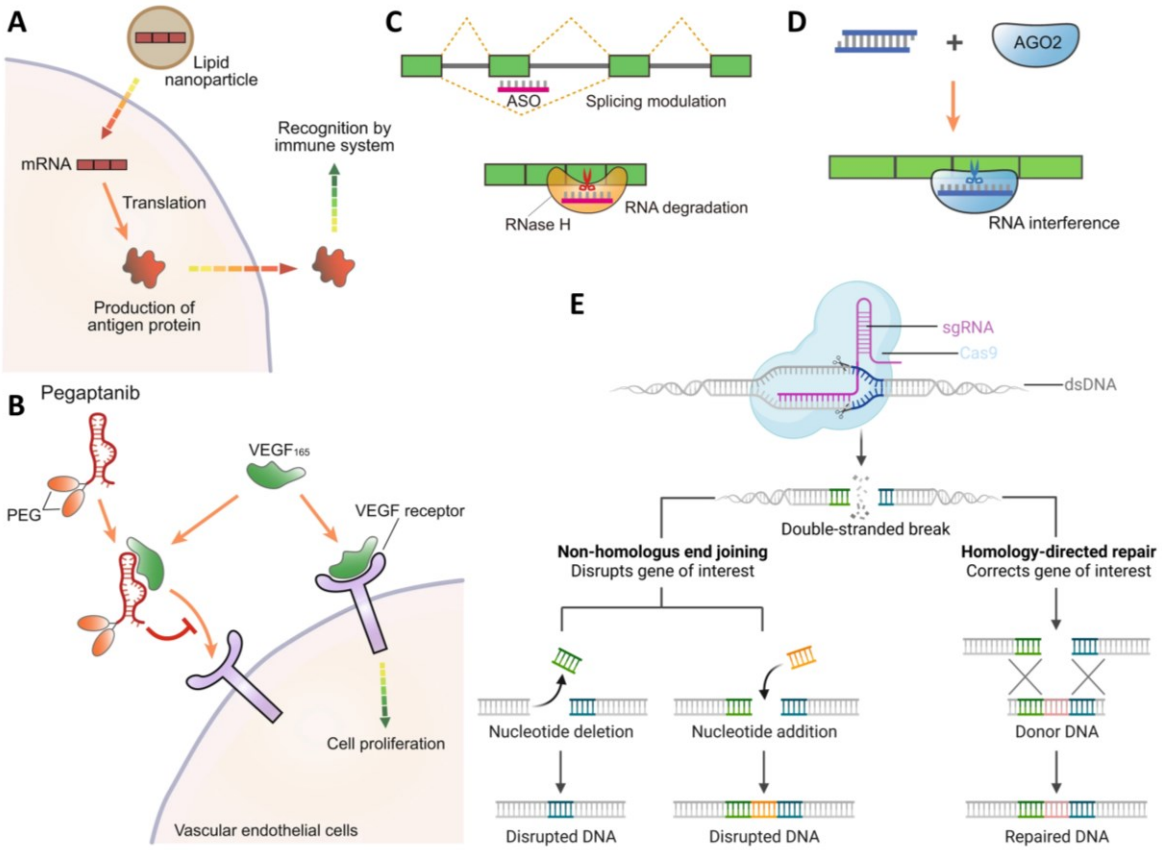



**Figure 23. Mechanism of action of selected RNA therapeutics. A – Application of mRNA vaccines. mRNA produces proteins recognised by immune system. B – Aptamer binds to VEGF and hinders its interaction with the VEGF receptor. C – ASOs modulating splicing or inducing RNase H-mediated target cleavage. D – siRNA interacting with AGO protein and inducing cleavage of its mRNA target. E – CRISPR-Cas9-based gene editing mechanism. Modified figure from literature[257].**

# 2 Aims of the thesis

1. Design and synthesise the chloroacetamide-bearing **rA$^{CA}$TP** building block and optimise conditions for IVT of modified RNA probes with varying number of reactive functional groups.
2. Investigate bioconjugations of CA-modified RNA with thiol- or Cys- and His-residues of (bio)molecules and peptides. Study cross-linking reactions with various RNA-binding proteins and perform identification of targeted amino acid residues.
3. Develop and establish an alternative method for enzymatic synthesis of base-modified RNA by PEX reaction leveraging engineered DNA polymerases. Explore a library of variously modified **rN$^{X}$TP**s to delineate the scope of accepted modifications by mutant polymerases.
4. Develop a novel approach for site-specific RNA labelling and construction of hypermodified RNA polymers through DNA primer removal.
5. Design a strategy for synthesis of nucleobase-modified mRNAs at specific positions and investigate the impact of the modified position in translation studies.

## 2.1 Rationale of the thesis aims

Recent advancements in cross-linking techniques have significantly improved our understanding of RNA-protein interactions, contributing to the broader insight into RNA biology. However, certain limitations hinder further advancements in this area. Notably, the Hocek group has pioneered methods for spontaneous proximity-triggered cross-linking, by selective targeting amino acid residues of DNA-interacting proteins. Among these, incorporation of the chloroacetamide moiety into DNA major groove has demonstrated exceptional efficacy in capturing Cys and His within DBPs. In this context, our goal was to apply this methodology to the field of RNA and expand the portfolio of cross-linking methods. To achieve this, I was tasked to design and develop a simple pathway that would lead to creation of the chloroacetamide-modified ribonucleotide building block (**rA$^{CA}$TP**). My objective was to directly tether the reactive moiety to the iodinated nucleotide *via* straightforward Sonogashira cross-coupling reaction, as has been successfully demonstrated previously. The construction of modified RNA typically employs the established T7 RNAP *in vitro* transcription. I planned to identify conditions under which T7 RNAP would accept **rA$^{CA}$TP** and leave it intact upon insertion into RNA strand. This was expected not to be entirely straightforward, considering the strongly electrophilic nature of chloroacetamide, presence of Cys within T7 RNAP, and the need to maintain a reducing environment during transcription reaction. Following the establishment of a feasible procedure for constructing modified RNA, the subsequent step would involve exploring potential applications for bioconjugation and cross-linking. Attaching thiol-bearing

small molecules, orthogonal to known click reactions, would enable the post-transcriptional derivatisation of modified RNA. The ultimate goal was to determine whether RBPs could be selectively captured over non-RBPs containing Cys and/or His. To achieve this, I planned to evaluate a number of known RBPs, either commercially available or recombinantly expressed in our laboratory. Assuming the reaction proceeded selectively, the next step would be identification of RNA binding sites within the proteins by mass spectrometry. Successful completion of this part would pave the way for next generation of RNA cross-linking probes, significantly expanding the coverage of identified RNA-RBP interactions and offering deep insights into the structures of RNA-protein complexes.

In the second section of my thesis, I was expected to be focused on the utilisation of alternative enzymatic methods for constructing modified RNA. Recent advancements in directed protein evolution have facilitated the creation of thermostable DNA-dependent DNA polymerases exhibiting significantly enhanced RNA-polymerising activity. Application of these enzymes for incorporation of base-modified nucleotides has not been investigated, yet it promises achieving so far elusive tasks, where T7 RNAP IVT fails to give any reasonable output due to known limitations. To tackle this objective, my plan involved utilising a library of chemically diverse base-modified **rN$^{X}$TP**s to functionalise RNA probes with hydrophobic, reactive cross-linking, fluorescent functional groups, or affinity tags. The aim was to compare two engineered thermostable polymerases with this collection of **rN$^{X}$TP**s in PEX reactions, characterise the synthesised RNA probes by gel electrophoresis and mass spectrometry analysis and examine enzyme kinetics and performance under challenging conditions using difficult templates or mixtures of up to four modified nucleotides. After delineating the scope of accepted modifications, my strategy included developing a method to eliminate undesirable primer regions, resulting in a synthesised RNA segment composed entirely of modified nucleotides. This would involve usage of a DNA oligonucleotide to prime the PEX reaction and enzymatic digestion of both template and primer regions. Such heavily modified RNA polymers could hold the potential for diverse applications in biotechnology and medicine. Crucially, currently available methods for introducing labels into RNA at specific positions lack robustness and often require expensive instrumentation or involve inconvenient protocols. By adapting methods developed for SNI in DNA, my goal was to achieve selective RNA labelling at multiple sites or in challenging homopolymeric sequences. This approach would provide access to valuable probes bearing fluorescent, spin, or isotopic labels, facilitating the biophysical investigation of RNA molecules. Ultimately, extending this methodology to long mRNA molecules was part of the plan, aiming to obtain region- or single-site-modified mRNAs with epitranscriptomic groups. Subsequent *in vitro*

and *in cellulo* translation assays of these modified mRNA molecules would offer insight into dependence of translation efficiency as a function of site of modification.

# 3 Results and discussion

## 3.1 Chloroacetamide-modified nucleotide and RNA for bioconjugations and cross-linking with cysteine- or histidine-containing peptides and RNA-binding proteins

*This study was performed in collaboration with M. Krömer, M. Vlková and L. Poštová Slavětínská. All experiments in this chapter were performed by me, unless otherwise stated. M. Krömer performed LC-MS separation and analysis of* ***20RNA_1A$^{CA\text{-}pept\text{-}(+)\text{-}C}$*** *and* ***20RNA_1A$^{CA\text{-}HIV\text{-}RT}$*** *conjugates. M. Vlková performed ESI-MS analysis of* ***20RNA_1A$^{CA\text{-}HuR}$*** *conjugate and nano-LC-MS/MS analysis of RNA-protein digests. L. Poštová Slavětínská performed NMR spectra acquisition and interpretation of* ***rA$^{CA}$TP****.*

### 3.1.1 Introduction

Proteins that recognise and interact with RNA in various ways are collectively referred to as RBPs. RBPs play a pivotal role in different biological processes, underscoring the importance of understanding RNA-protein interactions[157,158,160,167,258]. Consequently, several methods have been employed to unravel these complex relationships. Early pioneering studies relying on non-covalent pulldown techniques faced challenges due to weak and ineffective RNA-protein interactions, resulting in the elusiveness of many RBPs. In response, alternative methods were implemented to address these limitations.

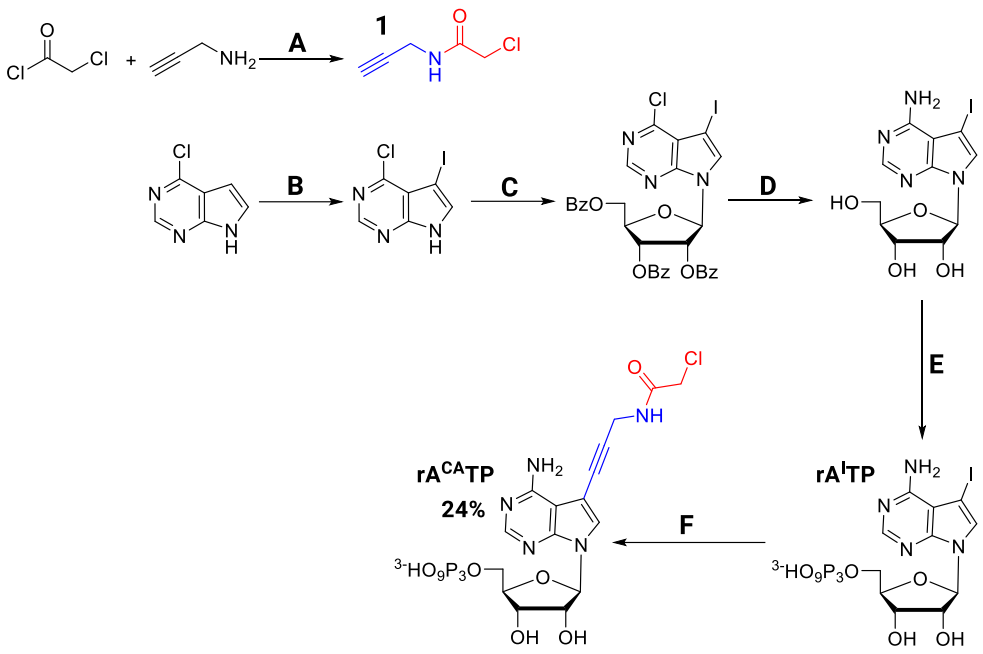
Cross-linking, i.e., the formation of a stable covalent bond between RNA and the protein of interest, remains the most effective tool for capturing and identifying RBPs, particularly those with low binding affinities to RNA. Most of these methods rely on UVC light-induced photocross-linking of natural RNA occurring mainly through uracil nucleobase[185,259]. Alternatively, more sophisticated approaches involve the use of base-modified nucleosides with halogen[186–188] or sulphur[166,190,191] atoms for metabolic RNA labelling or diazirine-modified RNA probes prepared by chemical[117,199] or enzymatic[117] synthesis. Upon UV light irradiation, these modifications generate highly reactive species (radicals, carbenes) that non-specifically react with neighbouring amino acids *via* C-H activations. Another option involves addition of external chemical cross-linking reagents[180,182,184]. While these methods are widely adopted today, they are not without deep-rooted limitations across a broad spectrum of applications. Common drawbacks include low efficacy, poor cross-linking yields, need for zero-distance contact, non-specific targeting of amino acids, and the use of potentially damaging UV light or toxic chemical agents. These challenges have spurred the emergence of

alternative methods. In the case of DNA, some of these drawbacks have been mitigated through even more challenging amino acid-specific cross-linking with DBPs. This field has recently undergone rapid development, introducing a variety of modifications for selective targeting of Cys[206,207], His[206], Arg[202,203], and predominantly Lys[204,205,211].

Conversely, only a handful of examples is reported for selective RNA cross-linking, with most of them limited to RNA methyltransferases only. Notably, the identified examples largely revolve around utilisation of 5-fluoro- or 5-azapyrimidine-modified nucleosides[213,214], demonstrating a specific targeting mechanism primarily directed at Cys residues. To extent of my current knowledge, beside these documented cases, there is an absence of reports detailing the application of other reactive nucleobase-modifications designed for the selective targeting of specific amino acids within RBPs. This insuffucuency underscores the need for development of RNA cross-linking strategies, especially those designed to elucidate interactions involving specific amino acids.

### 3.1.2 Synthetic pathway to chloroacetamide-modified ribonucleoside triphosphate

To expand the toolbox of available modifications, I synthesised a novel ribonucleoside triphosphate bearing the reactive chloroacetamide (CA) functional group in analogy to previously published work. As documented in previous studies, this gentle electrophile demonstrates selectivity in bioconjugations with Cys and His amino acids within proteins[206]. In this study, the CA functionality was tethered to position C-7 of 7-deazaadenine base through a rigid aminopropargyl linker. Based on earlier experiments involving base-modified DNA, it is known that functional groups positioned at this specific site on 7-deazaadenine nucleobase are effectively accommodated within the major groove, minimise structural perturbations and are well-tolerated by polymerases[69,71,73]. Therefore, I assumed that a similar approach might be applicable also for RNA. Numerous researchers have showcased the elegance and simplicity of single-step Sonogashira cross-coupling reactions involving iodinated ribonucleoside triphosphates (**rN$^{I}$TP**s) and functionalised terminal acetylenes[43]. Thus, the known iodinated 7-deazaadenosine-5'-*O*-triphosphate (**rA$^{I}$TP**) prepared according to standard procedure[260] was reacted with *N*-(propargyl)-chloroacetamide (**1**)[261] in analogy to previously published work[206]. The cross-coupling reaction was performed in a mixture of water and acetonitrile (2:1) in presence of $Pd(OAc)_2$, CuI, TPPTS and DIPEA at 60°C for 4 h. The desired 7-{3-[*N*-(2-chloroacetamido)]-prop-2-yn-1-yl}-7-deazaadenosine-5'-*O*-triphosphate (**rA$^{CA}$TP**) was after reversed-phase and ion exchange HPLC purification followed by ion-exchange chromatography with Dowex 50WX8 in $Na^+$ cycle isolated in satisfactory 24% yield (Scheme 10). Identity and purity of the prepared compound was proved by NMR and HR-ESI-MS analysis.




**Scheme 10. Synthetic pathway leading to modified rA$^{CA}$TP. Reaction conditions for synthesis of compound 1[261] – A: $Et_3N$, DCM, 0 °C → room temperature, 3 h. Reaction conditions for iodination – B: *N*-iodo-succinimide, dry DMF, rt, overnight. Reaction conditions for glycosylation – C: i) *N*,*O*-bis(trimethylsilyl)acetamide, dry $CH_3CN$, rt, 15 min, ii) 1-*O*-acetyl-2,3,5-tri-*O*-benzoyl-D-ribofuranose, trimethylsilyl trifluoromethanesulfonate, 80 °C, 3h. Reaction conditions for deprotection and substitution – D: 25% $NH_3$ (aq), dioxane, 120 °C, overnight. Reaction conditions for triphosphorylation – E: i) $POCl_3$, $PO(OMe)_3$, 0°C, 4h, ii) $Bu_3N$, bis(tri-*n*-butylammonium)pyrophosphate, dry $CH_3CN$, 0 °C, 6 h, iii) 2 M TEAB. Reaction conditions for Sonogashira cross-coupling – F: 10 mol% CuI, 5 mol% $Pd(OAc)_2$, 10 mol% TPPTS, 1, DIPEA, $H_2O/CH_3CN$ (2:1), 60 °C, 4 h.**

### 3.1.3 Enzymatic synthesis of chloroacetamide-modified RNA with T7 RNA polymerase

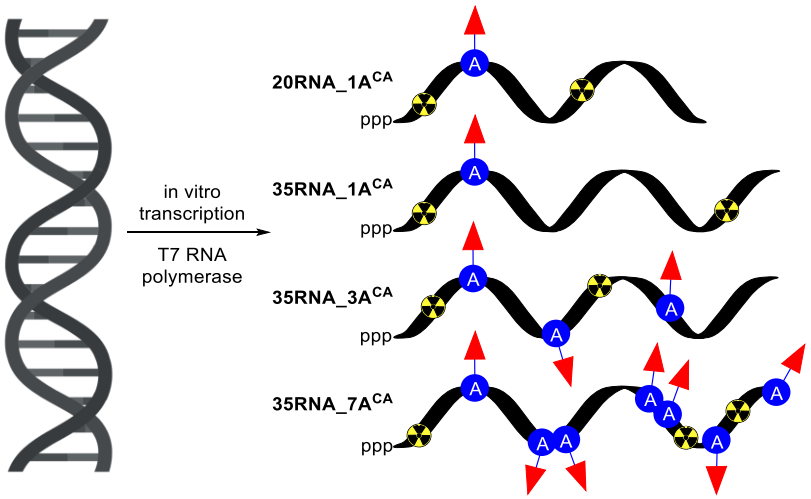



**Figure 24. T7 RNAP IVT of dsDNA templates. Synthesis of short RNA with one CA-modification or longer ones with up to 7 modifications.**

The key nucleotide building block, **rA$^{CA}$TP** was subsequently tested as a putative substrate in IVT reactions with various dsDNA templates. For enzymatic synthesis of RNA, I selected the widely used T7 RNAP since it was previously shown to smoothly incorporate a pallet of diverse base-modified nucleotides[43]. First, I tested IVT reactions with 37 bp long dsDNA template (**20DNA_1A-ds**, Table 1) encoding for a short 20 nt RNA probe with one CA modification (**20RNA_1A$^{CA}$**, Table 2, Figure 24).

**Table 1. List of DNA oligonucleotides fore generation of dsDNA template used in this study.**

| DNA oligonucleotide name | Sequence (5'→ 3') | 5'-end modification | Length [nt] |
|---|---|---|---|
| **20DNA_1A-s** | *TAATACGACTCACTATA*GGGCCCCTATTGTCTCTCTC | – | 37 |
| **20DNA_1A-as** | [mG][mA]GAGAGACAATAGGGGCCC*TATAGTGAGTCGTATTA* | 2'-*O*-Me-G, 2'-*O*-Me-A | 37 |
| **21DNA_3A-bind-s** | *TAATACGACTCACTATA*GGGTGATTTTATTTTATTCTC | – | 38 |
| **21DNA_3A-bind-as** | [mG][mA]GAATAAAATAAAATCACCC*TATAGTGAGTCGTATTA* | 2'-*O*-Me-G, 2'-*O*-Me-A | 38 |
| **21DNA_3A-non-bind-s** | *TAATACGACTCACTATA*GGGTCACGTGACGCCAGTCCC | – | 38 |
| **21DNA_3A-non-bind-as** | [mG][mG]GACTGGCGTCACGTGACCC*TATAGTGAGTCGTATTA* | 2'-*O*-Me-G, 2'-*O*-Me-G | 38 |
| **35DNA_1A-s** | TAATACGACTCACTATAGGGCCCCTATTGTCTCTCTCTTCTCTGCTGTTTCC | – | 52 |
| **35DNA_1A-as** | [mG][mG]AAACAGCAGAGAAGAGAGAGACAATAGGGGCCC*TATAGTGAGTCGTATTA* | 2'-*O*-Me-G, 2'-*O*-Me-G | 52 |
| **35DNA_3A-s** | *TAATACGACTCACTATA*GGGCCCGTATGTTACTTGCTCTTATCGTCTCTCGC | – | 52 |
| **35DNA_3A-as** | [mG][mC]GAGAGACGATAAGAGCAAGTAACATACGGGCCC*TATAGTGAGTCGTATTA* | 2'-*O*-Me-G, 2'-*O*-Me-C | 52 |
| **35DNA_7A-s** | *TAATACGACTCACTATA*GGGCTTGCACGTGAATCGCTCTTAATGGATCGCGA | – | 52 |
| **35DNA_7A-as** | [mU][mC]GCGATCCATTAAGAGCGATTCACGTGCAAGCCC*TATAGTGAGTCGTATTA* | 2'-*O*-Me-U, 2'-*O*-Me-C | 52 |

s = Sense strand used for generation of double stranded DNA template. as = Anti-sense strand used for generation of double stranded DNA template. T7 RNAP promoter region is highlighted in italics.

The transcription reactions were performed with a mixture of either all four natural rNTPs (rATP, rUTP, rCTP, rGTP) for synthesis of natural RNA (**20RNA_1A**), a mixture of **rA$^{CA}$TP**, rUTP, rCTP, rGTP for generation of CA-modified RNA (**20RNA_1A$^{CA}$**) or with omitting rATP for negative control reaction. In each case, a small amount of radioactively labelled [α-$^{32}$P]-GTP was spiked in for visualisation of the newly synthesised RNA probe and monitoring of the whole IVT process. The

transcriptions were conducted at an optimal reaction temperature of 37 °C for 2 hours in a T7 RNAP reaction buffer supplemented with additives (triton X-100, $MgCl_2$) and RNase inhibitor. Reaction products were analysed on dPAGE followed by visualisation through phosphor imaging. dPAGE analysis proved formation of full-length modified **20RNA_1A$^{CA}$** product represented by a slower moving band at about the same level where its natural congener (**20RNA_1A**) migrated (Figure 25). Additionally, by scaling up the reaction and by leaving out the radioactive label, I prepared **20RNA_1A$^{CA}$** probe, that was after purification confirmed by MS-MALDI analysis (Figure 26). Calculated mass was in good agreement with the measured one (Table 3).

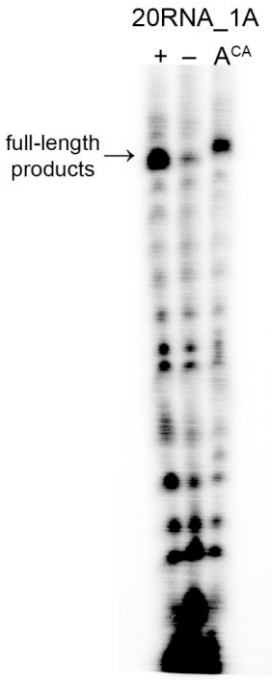



**Figure 25. 20% dPAGE analysis of T7 RNAP IVT with 20DNA_1A-ds template. Gel legend: (+) positive control (20RNA_1A), all natural rNTPs (rATP, rCTP, UTP, rGTP/[α-$^{32}$P]-rGTP); (–) negative control, mixture of rCTP, rUTP, rGTP/[α-$^{32}$P]-rGTP and $H_2O$; (A$^{CA}$) modification (20RNA_1A$^{CA}$), mixture of rA$^{CA}$TP, rUTP, rCTP, rGTP/[α-$^{32}$P]-rGTP. Visualisation by phosphor imager.**

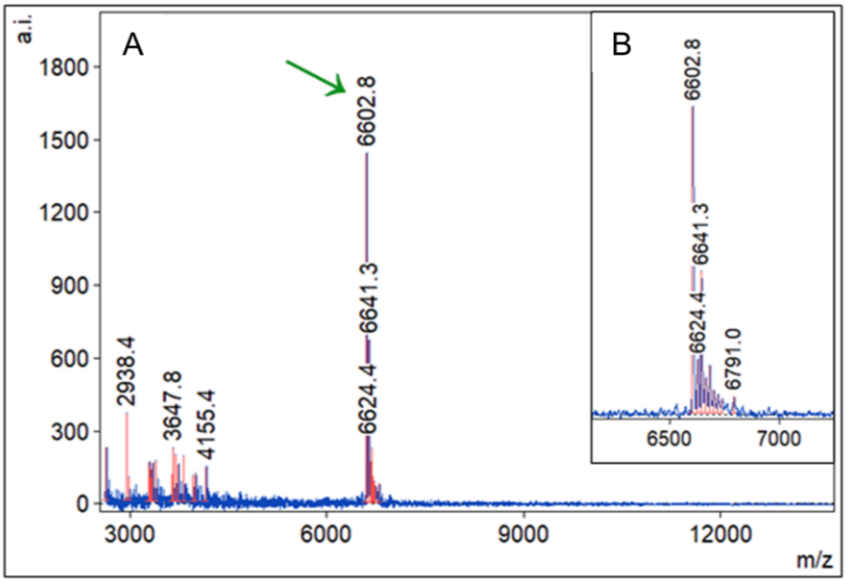



**Figure 26. MS-MALDI-TOF spectrum of 20RNA_1A$^{CA}$. A – full spectrum; B – magnified area of interest; calculated mass: 6601.2 Da; found mass: 6602.8 Da; Δ = 1.6 Da. The peak at m/z = 6641.3 Da can be assigned to the adduct [20RNA_ 1A$^{CA}$ + K$^+$]. The peak at m/z = 6624.4 Da can be assigned to the adduct [20RNA_ 1A$^{CA}$ + Na$^+$].**

**Table 2. List of RNA oligonucleotides prepared in this study.**

| RNA oligonucleotide name | Sequence (5'→ 3') | 3'-End modification | Length [nt] |
|---|---|---|---|
| **20RNA_1A$^{CA}$** | pppGGGCCCCU**A$^{CA}$**UUGUCUCUCUC | – | 20 |
| **21RNA_1A$^{CA}$-Cy5** | pppGGGCCCCU**A$^{CA}$**UUGUCUCUCUCC-Cy5 | Cy5 | 21 |
| **21RNA_1A$^{CA}$-Bio** | pppGGGCCCCU**A$^{CA}$**UUGUCUCUCUCC-Bio | Bio | 32 |
| **21RNA_3A$^{CA}$-bind** | pppGGGUG**A$^{CA}$**UUUU**A$^{CA}$**UUUU**A$^{CA}$**UUCUC | – | 21 |
| **21RNA_3A$^{CA}$-non-bind** | pppGGGUC**A$^{CA}$**CGUG**A$^{CA}$**CGCC**A$^{CA}$**GUCCC | – | 21 |
| **35RNA_1A$^{CA}$** | pppGGGCCCCU**A$^{CA}$**UUGUCUCUCUCUUCUCUGCUGU UUCC | – | 35 |
| **35RNA_3A$^{CA}$** | pppGGGCCCGU**A$^{CA}$**UGUU**A$^{CA}$**CUUGCUCUU**A$^{CA}$**UCGUC UCUCGC | – | 35 |
| **35RNA_7A$^{CA}$** | pppGGGCUUGC**A$^{CA}$**CGUG**A$^{CA}$A$^{CA}$**UCGCUCUU**A$^{CA}$A$^{CA}$**U GG**A$^{CA}$**UCGCG**A$^{CA}$** | – | 35 |
| **36RNA_1A$^{CA}$-Cy5** | pppGGGCCCCU**A$^{CA}$**UUGUCUCUCUCUUCUCUGCUGU UUCCC-Cy5 | Cy5 | 36 |
| **36RNA_3A$^{CA}$-Cy5** | pppGGGCCCGU**A$^{CA}$**UGUU**A$^{CA}$**CUUGCUCUU**A$^{CA}$**UCGUC UCUCGCC-Cy5 | Cy5 | 36 |
| **36RNA_7A$^{CA}$-Cy5** | pppGGGCUUGC**A$^{CA}$**CGUG**A$^{CA}$A$^{CA}$**UCGCUCUU**A$^{CA}$A$^{CA}$**U GG**A$^{CA}$**UCGCG**A$^{CA}$**C-Cy5 | Cy5 | 36 |
| **20RNA_1A** | pppGGGCCCCUAUUGUCUCUCUC | – | 20 |
| **21RNA_1A-Cy5** | pppGGGCCCCUAUUGUCUCUCUCC-Cy5 | Cy5 | 21 |
| **21RNA_1A-Bio** | pppGGGCCCCUAUUGUCUCUCUCC-Bio | Bio | 21 |
| **21RNA_3A-bind** | pppGGGUGAUUUUAUUUUAUUCUC | – | 21 |
| **21RNA_3A-non-bind** | pppGGGUCACGUGACGCCAGUCC | – | 21 |

| | | | |
|---|---|---|---|
| **35RNA_1A** | pppGGGCCCCUAUUGUCUCUCUCUUCUCUGCUGUUUCC | – | 35 |
| **35RNA_3A** | pppGGGCCCGUAUGUUACUUGCUCUUAUCGUCUCUCGC | – | 35 |
| **35RNA_7A** | pppGGGCUUGCACGUGAAUCGCUCUUAAUGGAUCGCGA | – | 35 |
| **36RNA_1A-Cy5** | pppGGGCCCCUAUUGUCUCUCUCUUCUCUGCUGUUUCCC-Cy5 | Cy5 | 36 |

Modified nucleobases with CA modification are marked in bold. ppp = Triphosphate residue at the 5'-terminus of the synthesised RNA transcripts. Cy5 = cyanine-5 modification, Bio = biotin modification.

Even after providing evidence by dPAGE and MS, that T7 RNAP is proficient in incorporation at least one CA-modified ribonucleotide triphosphate (**rA$^{CA}$TP**) into RNA, a crucial step remained. It was important to thoroughly investigate and exclude the formation of any potential side products stemming from transcription reaction. First, I investigated whether the CA functionality of RNA probe does not cross-link to T7 RNAP, since it contains twelve reduced Cys residues[262], that could serve as possible cross-linking targets. For this purpose, IVT reactions performed either with natural rNTPs or with a mixture of **rA$^{CA}$TP**, rUTP, rCTP, rGTP were directly analysed on SDS-PAGE. This experiment showcased the absence of any by-product formation in case of modified RNA, since no band with lower mobility corresponding to cross-linking product was observed (Figure 27). Additionally, I explored the possibility of T7 RNAP inhibition by cross-linking to free **rA$^{CA}$TP** during IVT. To unravel this, I performed IVT with either a mixture of all four natural rNTPs or with four natural rNTPs containing extra addition of **rA$^{CA}$TP**. To my satisfaction, both IVT experiments yielded about the same amount of desired RNA product (**20RNA_1A**), ruling out any significant inhibitions of T7 RNAP by the modified nucleotide (Figure 28).

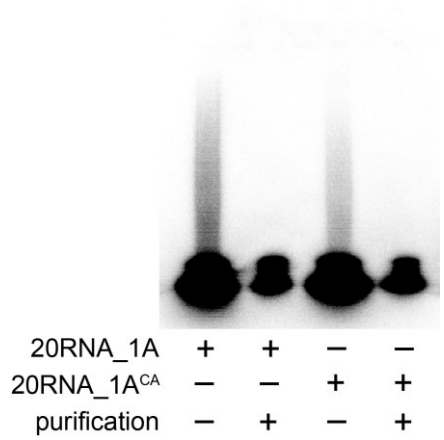



**Figure 27. 10% SDS-PAGE analysis of either crude or purified T7 RNAP IVT with 20DNA_1A-ds template. Comparison of IVT with a mixture of all 4 natural rNTPs (rATP, rUTP, rCTP, rGTP/[α-$^{32}$P]-rGTP), natural 20RNA_1A or a mixture of rA$^{CA}$TP, rUTP, rCTP, rGTP/[α-$^{32}$P]-rGTP, modified 20RNA_1A$^{CA}$. Visualisation by phosphor imager.**

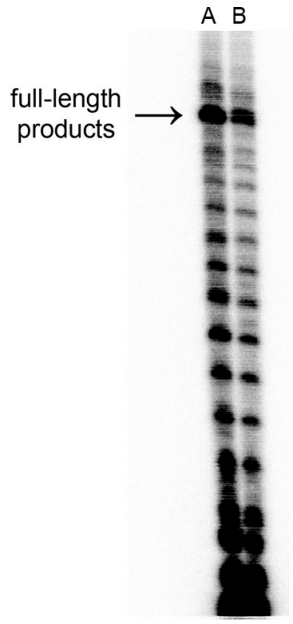



**Figure 28. 20% dPAGE analysis of T7 RNAP IVT with 20DNA_1A-ds template. Comparison of IVT reaction with: A – all four natural rNTPs (rATP, rUTP, rCTP, rGTP/[α-$^{32}$P]-rGTP) and IVT with: B – all four natural rNTPs containing extra addition of modified rA$^{CA}$TP in same concentration as natural rATP (rA$^{CA}$TP, rATP, rUTP, rCTP, rGTP/[α-$^{32}$P]-rGTP). Visualisation by phosphor imager.**

Next, I tested the modified **rA$^{CA}$TP** in IVT with more challenging 52 bp long dsDNA templates (**35DNA_1A-ds**, **35DNA_3A-ds**, **35DNA_7A-ds**, Table 1) encoding for 35 nt long RNA with one (**35RNA_1A$^{CA}$**), three (**35RNA_3A$^{CA}$**) or even seven (**35RNA_7A$^{CA}$**) CA-modified nucleotides (Figure 24, Table 2). Also, in this case the monitoring of IVT reactions was ensured by radioactive labelling and visualisation by phosphor imaging. Reactions proceeded smoothly with all aforementioned DNA templates. Full-length RNA products were formed in good conversions, even in case of RNA with seven CA modifications (**35RNA_7A$^{CA}$**) as confirmed by dPAGE (Figure 29). Furthermore, upon preparation in larger quantities and subsequent purification through silica-based columns, the desired products (**35RNA_1A$^{CA}$**, **35RNA_3A$^{CA}$**, **35RNA_7A$^{CA}$**) were validated through MS-MALDI (Table 3).

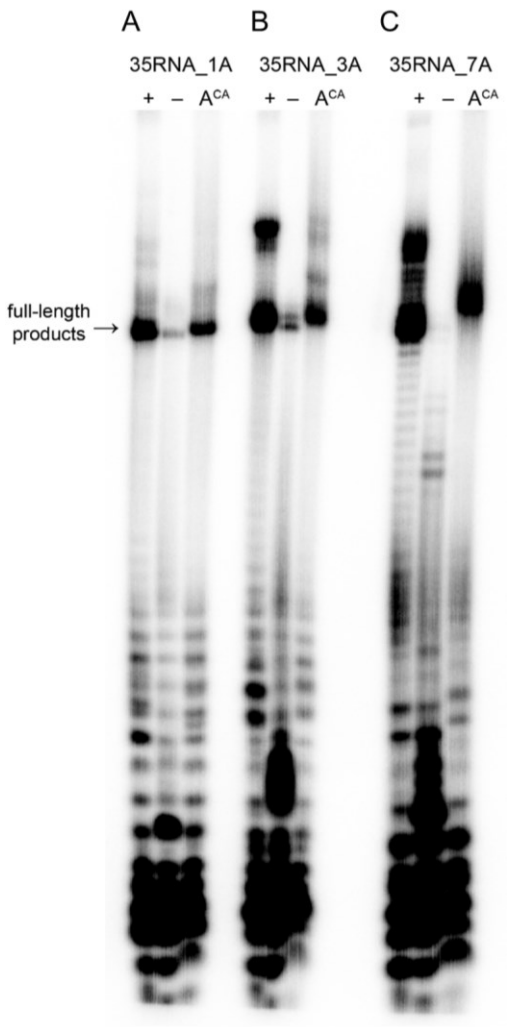


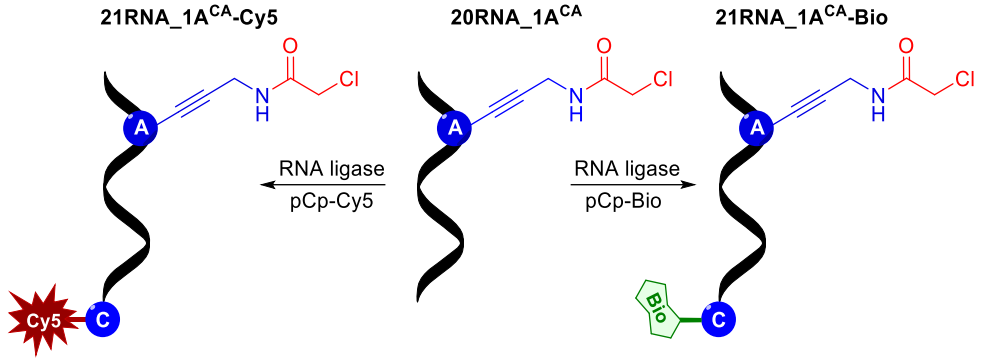

**Figure 29. 12.5% dPAGE analysis of T7 RNAP IVT with: A – 35DNA_1A, B – 35DNA_3A or C – 35DNA_7A templates. Gel legend: (+) positive control (35RNA_1A, 35RNA_3A, 35RNA_7A), all natural rNTPs (rATP, rUTP, rCTP, rGTP/[α-$^{32}$P]-rGTP); (–) negative control, mixture of rUTP, rCTP, rGTP/[α-$^{32}$P]-rGTP and $H_2O$; ($A^{CA}$) modification (35RNA_1A$^{CA}$, 35RNA_3A$^{CA}$, 35RNA_7A$^{CA}$), mixture of rA$^{CA}$TP, rUTP, rCTP, rGTP/[α-$^{32}$P]-rGTP. Visualisation by phosphor imager.**

**Table 3. MS data of prepared RNA oligonucleotides.**

| Analysed RNA | Calculated mass | Measured mass |
|---|---|---|
| **20RNA_1A** | 6472.2 Da | 6476.3 Da |
| **21RNA_1A-Cy5** | 7595.8 Da | 7599.3 Da |
| **21RNA_1A-Bio** | 7183.3 Da | 7186.8 Da |
| **21RNA_3A-bind**. | 6830.1 Da | 6830.0 Da |
| **21RNA_3A-non-bind**. | 6941.9 Da | 6941.0 Da |
| **35RNA_1A** | 11138.3 Da | 11138.5 Da |
| **20RNA_1A$^{CA}$** | 6601.2 Da | 6602.8 Da |
| **21RNA_1A$^{CA}$-Cy5** | 7724.3 Da | 7729.7 Da |
| **21RNA_1A$^{CA}$-Bio** | 7312.2 Da | 7316.2 Da |

| | | |
|---|---|---|
| **21RNA_3A$^{CA}$-bind**. | 7216.4 Da | 7216.0 Da |
| **21RNA_3A$^{CA}$-non-bind**. | 7327.6 Da | 7328.0 Da |
| **35RNA_1A$^{CA}$** | 11266.9 Da | 11267.5 Da |
| **35RNA_3A$^{CA}$** | 11650.1 Da | 11651.3 Da |
| **35RNA_7A$^{CA}$** | 12376.6 Da | 12378.0 Da |

### 3.1.4 3'-end RNA labelling *via* ligation




**Scheme 11. RNA ligation with pCp-Cy5 and pCp-Bio using T4 RNA ligase. For structure of pCp-Cy5 and pCp-Bio see Figure 103.**

To avoid the use of potentially hazardous and environmentally unfriendly radioactive labelling in subsequent experiments, and with the aim of achieving more flexible visualisation of RNA, an alternative approach involving non-radioactive, stable, fluorescent label was employed. Therefore, I tagged the prepared RNA oligonucleotides (**20RNA_1A$^{CA}$**, **35RNA_1A$^{CA}$**, **35RNA_3A$^{CA}$**, **35RNA_7A$^{CA}$**, Table 2) with nucleoside-monophosphate bearing Cy5 fluorophore (Scheme 11). Since, T7 RNAP IVT can yield numerous short or aborted transcripts[100], I opted to purify the full-length RNA products by dPAGE with subsequent isolation from gel by extraction. The ligation reaction proceeded between the 5'-phosphate group of donor bearing the fluorescent tag, cytidine-5'-phosphate-3'-(Cy5-amino-hexyl)phosphate (**pCp-Cy5**, Figure 103), and free 3'-OH of acceptor represented by RNA probe. Reaction catalysed by T4 RNA ligase in presence of DMSO, to unfold RNA secondary structures, reached full conversion after overnight incubation at 16 °C. Identity of the desired fluorescent Cy5-labelled products (**21RNA_1A$^{CA}$-Cy5**, **36RNA_1A$^{CA}$-Cy5**, **36RNA_3A$^{CA}$-Cy5**, **36RNA_7A$^{CA}$-Cy5**, Table 2) was confirmed by dPAGE analysis, where slower-moving band was visible in Cy5 channel, when compared to unlabelled RNA standards visualised by non-specific RNA staining with SybrGold (Figure 30). In case of **21RNA_1A$^{CA}$-Cy5**, the identity was also proved by MS-MALDI (Table 3). Following same reaction conditions, I prepared biotin-labelled RNA probe (**21RNA_1A$^{CA}$-Bio**) for capturing RNA through its strong and specific interaction with streptavidin[174], a protein that binds tightly to biotin. In this case cytidine-5'-

phosphate-3'-(biotin-amino-hexyl)phosphate (**pCp-Bio**, Figure 103) was used. Formation of the biotin-labelled RNA product (**21RNA_1A$^{CA}$-Bio**) was proved by MS analysis (Figure 31), where only mass of biotinylated oligonucleotide was observed (Table 3), proving full conversion of ligation reaction.

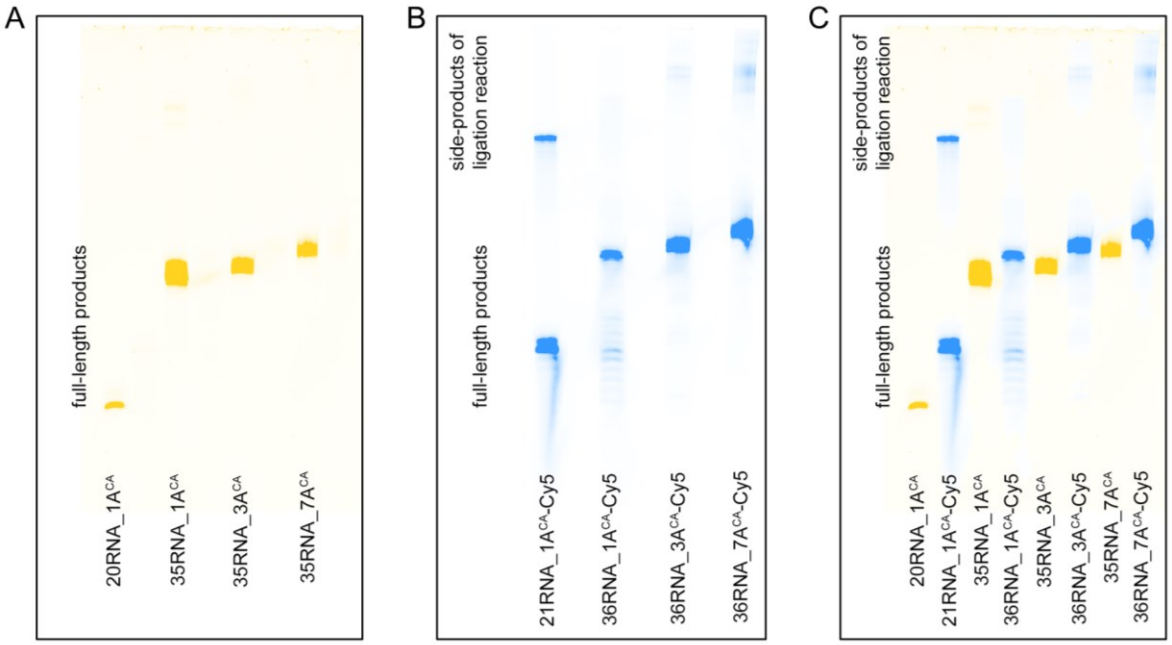



**Figure 30. 15% dPAGE analysis of ligation reaction with pCp-Cy5. Gel legend: A – SybrGold channel scan. Post-staining of dPAGE with SybrGold for visualisation of non-fluorescently labelled RNA prepared by T7 RNAP IVT (20RNA_1A$^{CA}$, 35RNA_1A$^{CA}$, 35RNA_3A$^{CA}$, 35RNA_7A$^{CA}$). B – Cy5 channel scan. Visualisation of fluorescent Cy5-labelled RNA prepared by ligation with pCp-Cy5 (21RNA_1A$^{CA}$-Cy5, 36RNA_1A$^{CA}$-Cy5, 36RNA_3A$^{CA}$-Cy5, 36RNA_7A$^{CA}$-Cy5). C – SybrGold and Cy5 channel scans merged. Visualisation of non-fluorescent and Cy5-labelled RNA probes.**

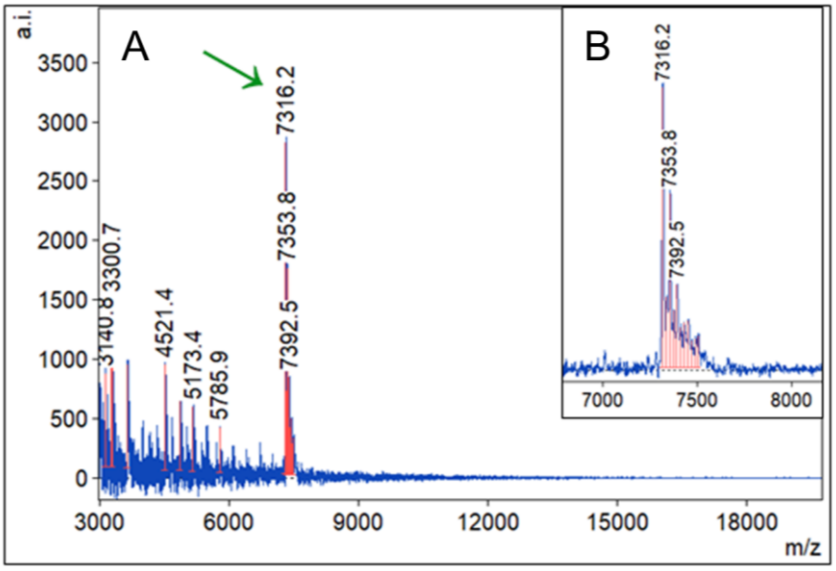



**Figure 31. MS-MALDI-TOF spectrum of 21RNA_1A$^{CA}$-Bio. A – full spectrum; B – magnified area of interest; calculated mass: 7312.2 Da; found mass: 7316.2 Da; Δ = 3.97 Da. The peak at *m*/*z* = 7353.8 Da can be assigned to the adduct [21RNA_1A$^{CA}$-Bio + K$^+$]. The peak at *m*/*z* = 7392.5 Da can be assigned to the adduct [21RNA_1A$^{CA}$-Bio + 2K$^+$].**

## 3.1.5 Model bioconjugation reactions

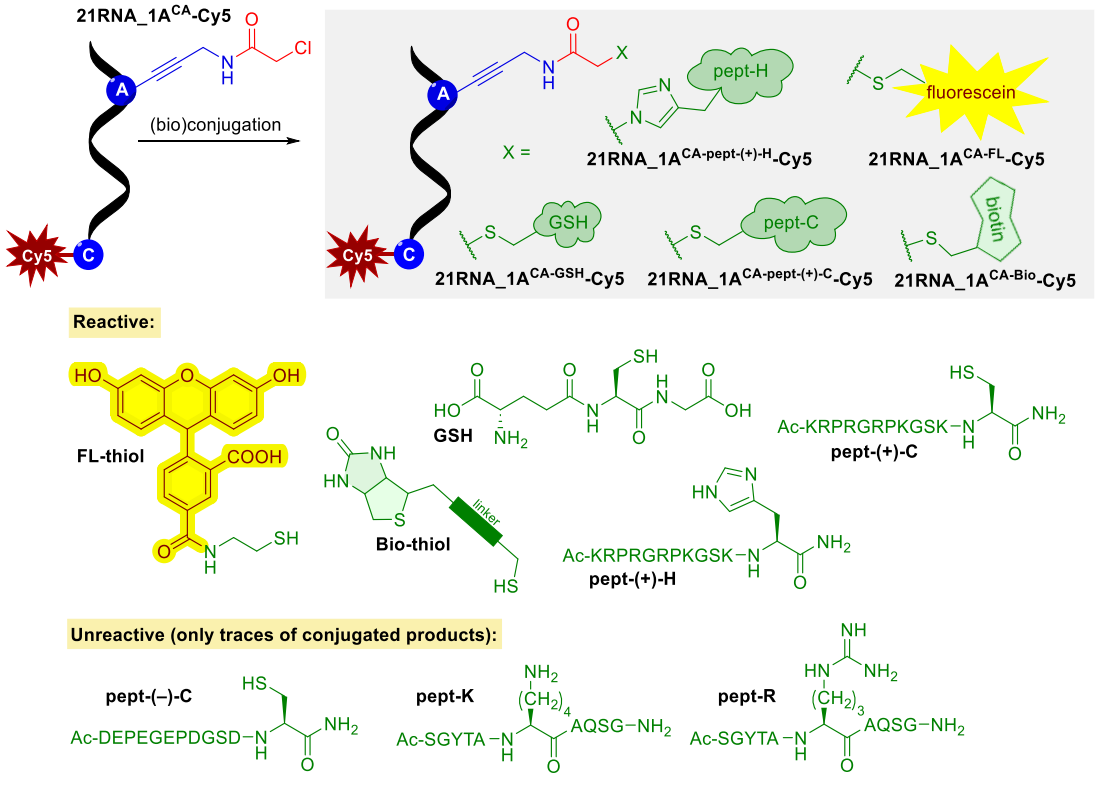



**Figure 32. Overview of bioconjugations of modified RNA probe with various (bio)molecules and peptides. For full chemical structures see Figure 104.**

Bioconjugation emerges as a versatile tool, enhancing the functionality and utility of RNA probes. With successful enzymatic synthesis procedure for modified RNA in my hands, I moved forward to assess the reactivity of these species. The fluorescently labelled modified RNA probe (**21RNA_1A$^{CA}$-Cy5**) was tested in bioconjugation reaction with Cys- and His-containing peptides and small (bio)molecules with free thiol functionality (Figure 32, Figure 104). I selected this RNA sequence due to its shorter length, offering improved resolution on dPAGE and facilitating a more straightforward analysis of bioconjugation products through MS. Reactions were performed in presence of non-nucleophilic triethylammonium acetate buffer (TEAA) at pH 8 to enhance thiols reactivity with great excess of peptides or (bio)molecules since no proximity effect was expected. Bioconjugations with biotin- (Bio) and/or fluorescein- (FL) thiol were performed with 10000 equiv. excess. Reactions proceeded in moderate either 21% (Bio-thiol) or 24% yield (FL-thiol) according to ImageJ quantification of dPAGE (Figure 33). Identity of desired conjugation products (**20RNA_1A$^{CA-Bio}$**, **20RNA_1A$^{CA-FL}$**) was confirmed by MS analysis (Table 6). Moreover, formation of the FL-labelled conjugation product (**20RNA_1A$^{CA-FL}$**) was confirmed by fluorescence measurements (Figure 34). After removal of unreacted thiol, enhanced fluorescence was observed in case of modified RNA (**20RNA_1A$^{CA}$**) conjugated with FL-thiol, after excitation at 490 nm. This phenomenon was not observed for negative control reaction of FL-thiol with natural RNA (Figure 34).

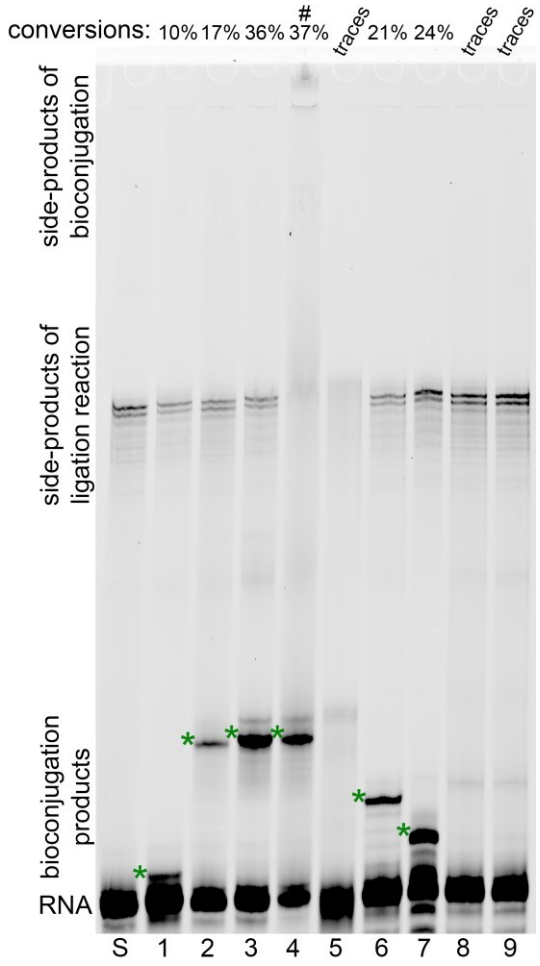



**Figure 33. 22.5% dPAGE analysis of bioconjugation reaction of modified 21RNA_1A$^{CA}$-Cy5 with (bio)molecules and various peptides. Gel legend: (S) modified RNA (21RNA_1A$^{CA}$-Cy5), standard; modified 21RNA_1A$^{CA}$-Cy5 in reaction with: (1) 10000 equiv. of reduced glutathione (GSH); (2) 10000 equiv. of pept-(+)-H; (3) 100 equiv. of pept-(+)-C; (4) 2500 equiv. of pept-(+)-C; (#) significant formation of side-product affects the calculated conversion, side-product formation in 18% yield, total conversion of all products is 55%; (5) 2500 equiv. of pept-(–)-C; (6) 10000 equiv. of Bio-thiol; (7) 10000 equiv. of FL-thiol; (8) 10000 equiv. of pept-K; (9) 10000 equiv. of pept-R. (*) Desired products of bioconjugation reaction. Cy5 channel scan.**

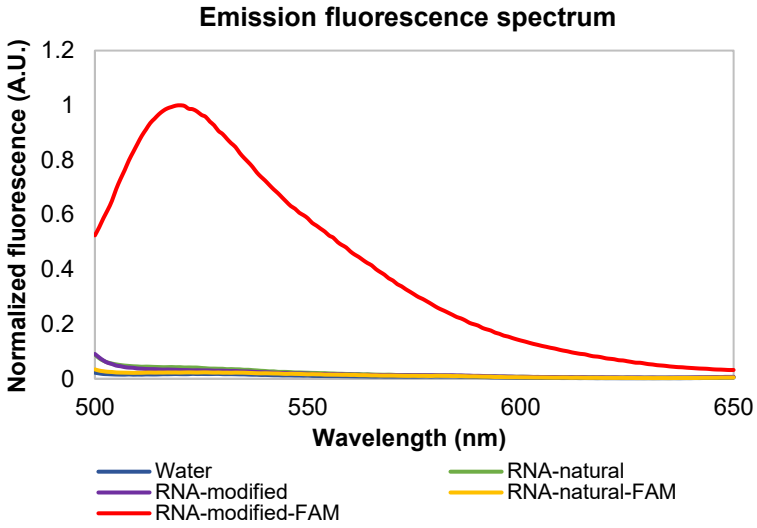



**Figure 34. Normalised emission fluorescence spectra of either modified 20RNA_1A$^{CA}$ or natural 20RNA_1A after bioconjugation with FL-thiol. Enhanced fluorescence observed in case of modified 20RNA_1A$^{CA}$ after reaction with FL-thiol, while no fluorescence observed in case of natural 20RNA_1A after reaction with FL-thiol.**

For bioconjugations with peptides, I tested a small tripeptide represented by reduced glutathione (**GSH**), a set of either positively (**pept-(+)-C**) or negatively (**pept-(–)-C**) charged peptides with one Cys residue, positively charged peptide containing one His (**pept-(+)-H**) and two peptides lacking any Cys or His, but rich in Lys (**pept-K**) or Arg (**pept-R**), as negative control reactions (Figure 32). As it was expected, positively charged Cys-containing peptide (**pept-(+)-C**) was the most reactive one, forming the desired RNA-peptide conjugate (**20RNA_1A$^{CA\text{-}pept\text{-}(+)\text{-}C}$**) in remarkable 36% yield with only 100 equiv. of **pept-(+)-C**, according to dPAGE quantification (Figure 33) and subsequent MS analysis (Table 6). Surprisingly, when further increasing amount of the peptide (up to 2500 equiv.), formation of a less mobile side-product was observed on dPAGE (Figure 35). This phenomenon was observed only in case of **pept-(+)-C**, but not with the negatively charged peptide (**pept-(–)-C**), although it was used in same excess for bioconjugation reactions (Figure 33). The formed side-product most probably corresponds to a non-covalent aggregate of peptide with RNA and/or ligation by-products. This hypothesis was additionally supported by LC-MS analysis of the RNA-peptide product (**20RNA_1A$^{CA\text{-}pept\text{-}(+)\text{-}C}$**), where only peaks corresponding either to unreacted RNA (**20RNA_1A$^{CA}$**) or to the desired **20RNA_1A$^{CA\text{-}pept\text{-}(+)\text{-}C}$** conjugate were detected (Figure 36, Table 4).

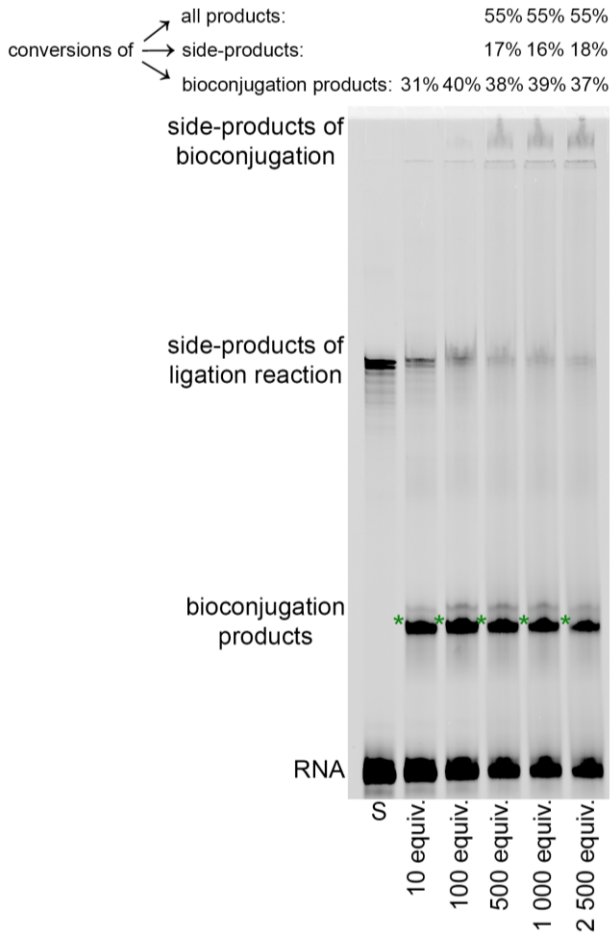



**Figure 35. 22.5% dPAGE analysis of bioconjugation reaction of modified 21RNA_1A$^{CA}$-Cy5 with various concentrations of positively charged pept-(+)-C. Gel legend: (*) desired products of bioconjugation reaction; significant formation of side-product affects the calculated conversions in reaction with 500 equiv., 1000 equiv. and 2500 equiv. of pept-(+)-C; (S) modified RNA (21RNA_1A$^{CA}$-Cy5), standard.**

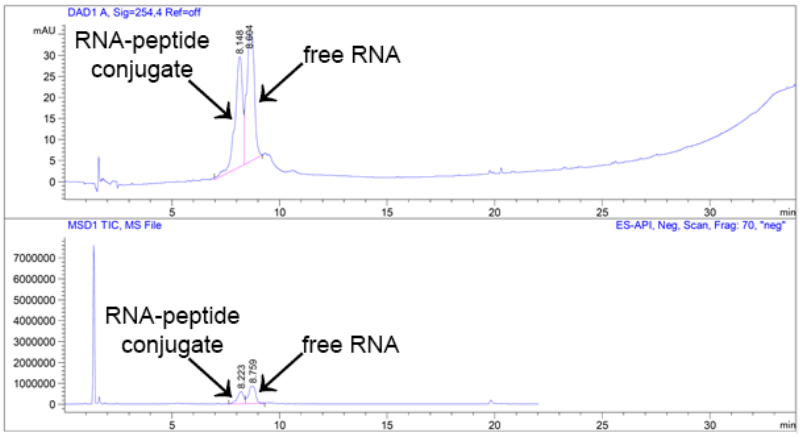



**Figure 36. LC-MS separation chromatogram of 20RNA_1A$^{CA\text{-}pept\text{-}(+)\text{-}C}$ conjugate.**

**Table 4. MS data of LC-MS separation of 20RNA_1A$^{CA}$ and 20RNA_1A$^{CA\text{-}pept\text{-}(+)\text{-}C}$ conjugate.**

| Analysed component | Calculated mass | Measured mass |
|---|---|---|
| **free RNA, 20RNA_1A$^{CA}$** | 6601.2 Da | 6600.0 Da |
| **conjugate, 20RNA_1A$^{CA\text{-}pept\text{-}(+)\text{-}C}$** | 7974.6 Da | 7974.0 Da |

On the other hand, significantly lower reactivity of the targets was observed either in case with slightly negatively charged **GSH** tripeptide or cationic, but less nucleophilic His-containing peptide (**pept-(+)-H**). Although, I used 10000 equiv. excess in both cases the desired products were formed in low either 10% (**GSH**) or 17% yield (**pept-(+)-H**) according to dPAGE (Figure 33). All successfully formed RNA-peptide conjugates (**20RNA_1A$^{CA\text{-}GSH}$**, **20RNA_1A$^{CA\text{-}pept\text{-}(+)\text{-}C}$**, **20RNA_1A$^{CA\text{-}pept\text{-}(+)\text{-}H}$**) were further characterised by MS (Table 6). In case of negatively charged Cys-containing peptide (**pept-(–)-C**) or peptides without any Cys or His residue (**pept-K** or **pept-R**), only traces of bioconjugation products were detected by dPAGE (Figure 33). These results further support selectivity of CA-modified RNA (**21RNA_1A$^{CA}$-Cy5**) towards Cys and His amino acids (Table 5). To solidify these findings, I performed negative control reactions under same conditions with natural RNA (**21RNA_1A-Cy5**), lacking any CA modification. As it was expected, no bioconjugation products were observed in this case (Figure 37).

**Table 5. List of (bio)molecules and peptides used in bioconjugations and reaction conversion analysis.**

| Reagent | Equivalents | Conversion |
|---|---|---|
| **FL-thiol** | 10000 | 24% |
| **Bio-thiol** | 10000 | 21% |
| **reduced GSH** | 10000 | 10% |
| **pept-(+)-H** | 10000 | 17% |
| **pept-(+)-C** | 100 | 36% |
| **pept-(–)-C** | 2500 | traces |
| **pept-K** | 10000 | traces |
| **pept-R** | 10000 | traces |

**Table 6. MS data of bioconjugations products.**

| Analysed conjugate | Calculated mass | Measured mass |
|---|---|---|
| **20RNA_1A$^{CA\text{-}FL}$** | 7002.2 Da | 7003.8 Da |
| **20RNA_1A$^{CA\text{-}biotin}$** | 7352.8 Da | 7355.4 Da |
| **20RNA_1A$^{CA\text{-}GSH}$** | 6872.1 Da | 6873.5 Da |
| **20RNA_1A$^{CA\text{-}pept\text{-}(+)\text{-}C}$** | 7974.6 Da | 7977.4 Da |
| **20RNA_1A$^{CA\text{-}pept\text{-}(+)\text{-}H}$** | 8008.6 Da | 8010.1 Da |

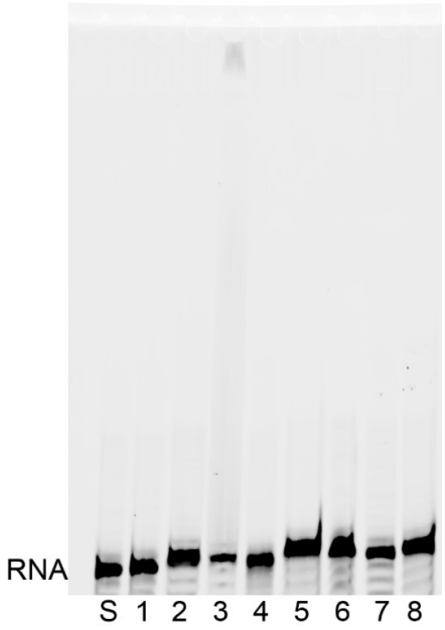



**Figure 37. 22.5% dPAGE analysis of bioconjugation reaction of natural 21RNA_1A-Cy5 with (bio)molecules and various peptides. Gel legend: (S, RNA) natural RNA (21RNA_1A-Cy5), standard; natural 21RNA_1A-Cy5 in reaction with: (1) 10000 equiv. of reduced glutathione (GSH); (2) 10000 equiv. of pept-(+)-H; (3) 2500 equiv. of pept-(+)-C; (4) 2500 equiv. of pept-(–)-C; (5) 10000 equiv. of Bio-thiol; (6) 10000 equiv. of FL-thiol; (7) 10000 equiv. of pept-K; (8) 10000 equiv. of pept-R. Cy5 channel scan.**

### 3.1.6 Cross-linking reactions with RNA-binding proteins

In order to prove utility of our approach for investigating RNA-protein interactions, I turned my focus to the ultimate goal of this study, cross-linking with RBPs. I selected three distinct RBPs with diverse biological functions, each of them containing several either Cys and/or His amino acids, serving as possible cross-linking targets for our CA-modified RNA probe. I tested human argonaute 2 protein (hAgo2)[263], human antigen R (HuR/ELAVL1)[264] and HIV reverse transcriptase (HIV-RT)[265]. Selection of RBPs with varied functions and distinct RNA structure and/or sequence preferences aimed to represent a wide spectrum of putative RNA-protein interactions. First, I needed to examine whether the CA functionality does not prevent binding of proteins and if the modified RNA probe (**21RNA_1A$^{CA}$-Cy5**) is recognised by RBPs. Hence, RNA-protein binding studies with each of the studied RBPs and CA-modified **21RNA_1A$^{CA}$-Cy5** or natural **21RNA_1A-Cy5** (as a positive control) were performed at physiologically relevant neutral pH in HEPES buffer. Reactions were performed by overnight incubation of RNA with either equimolar ratio (HuR) or with slight two equiv. excess (hAgo2, HIV-RT) of RBPs, contrary to peptide or small molecule bioconjugations, since strong proximity effect was expected in this case. Electrophoretic mobility shift assay (EMSA) was performed for analysis of RNA-protein complexes. 5% Native-PAGE (nPAGE) proved, for both natural and modified RNA with each of the studied proteins, formation of RNA-protein complex

represented by a slower moving band when comparing to protein-free nucleic acid standard (Figure 38).

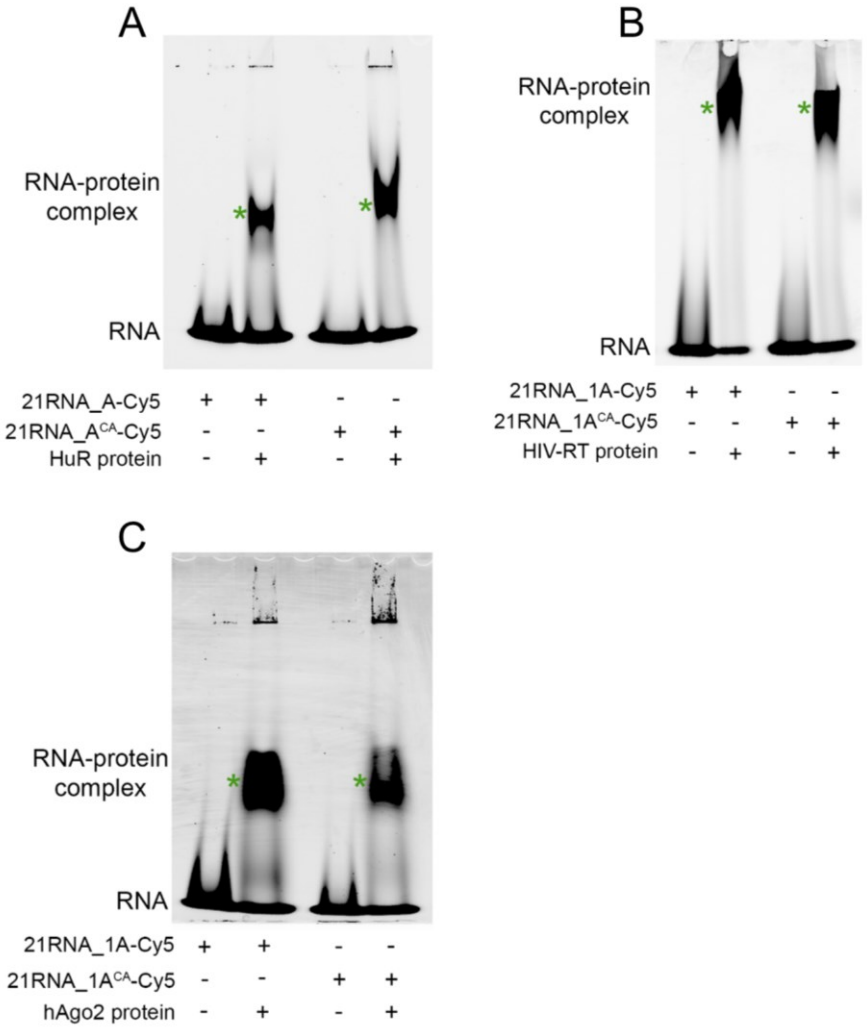



**Figure 38. EMSA analysis of natural 21RNA_1A-Cy5 and modified 21RNA_1A$^{CA}$-Cy5 with HuR protein (A), HIV-RT (B) and hAgo2 protein (C). Gel legend: (*) RNA-protein complex. Cy5 channel scan.**

To explore formation of a stable covalent bond between modified RNA (**21RNA_1A$^{CA}$-Cy5**) and RBPs, I performed analysis of these reactions on SDS-PAGE under denaturing conditions. In order to elucidate the temporal progression of cross-linking reaction, I conducted kinetic study involving **21RNA_1A$^{CA}$-Cy5** and model HIV-RT. SDS-PAGE analysis revealed formation of stable covalently linked RNA-protein complex (**21RNA_1A$^{CA\text{-}HIV\text{-}RT}$-Cy5**) already after 30 min in 3% conversion (quantified by ImageJ). By further increasing incubation time (up to 6 h), the cross-linked product was formed in a fairly high amount (24% yield), that is already exceeding conversions

achieved by classical photocross-linking methods. Remarkable 49% conversion was reached by incubation for 24 h, as revealed by SDS-PAGE (Figure 39). This extended duration of incubation resulted in enhanced cross-linking yield, underscoring the effectiveness of the developed approach in comparison to traditional photocross-linking techniques.

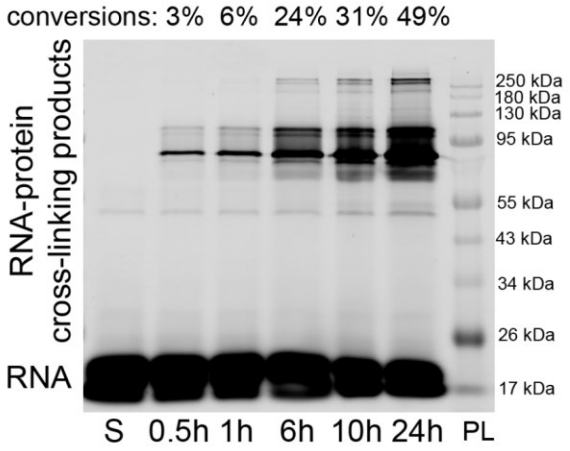



**Figure 39. 10% denaturing SDS-PAGE analysis of cross-linking reaction of modified 21RNA_1A$^{CA}$-Cy5 with HIV-RT in different time periods (0.5 h, 1 h, 6 h, 10 h, 24 h). Gel legend: (S; RNA) modified RNA (21RNA_1A$^{CA}$-Cy5), standard; (PL) pre-stained protein ladder. Cy5 channel scan.**

Therefore, all following cross-linking reactions with model proteins were performed by overnight incubations to achieve maximum attainable conversions. In all cases, cross-linking adducts were formed in reasonable yields (22-30%, Table 7) with only two equiv. of proteins. The formation of covalent conjugates represented by lower-mobile bands was tracked by SDS-PAGE (Figure 40 – A). Additionally, negative control reactions were carried with natural RNA congener (**21RNA_1A-Cy5**) and each of the studied RBPs. No formation of RNA-protein cross-linking product was observed on SDS-PAGE (Figure 40 – B).

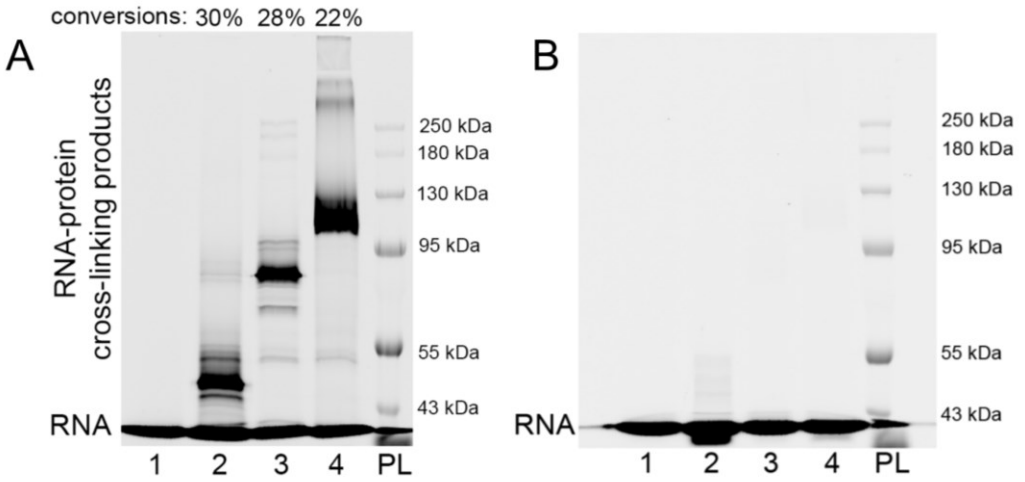



**Figure 40. 7% denaturing SDS-PAGE analysis of cross-linking reactions of modified 21RNA_1A$^{CA}$-Cy5 (Figure – A) or natural 21RNA_1A-Cy5 (Figure – B) with various RBPs. Gel legend (Figure – A): (RNA, 1) modified RNA (21RNA_1A$^{CA}$-Cy5), standard; modified 21RNA_1A$^{CA}$-Cy5 in reaction with (2) HuR; (3) HIV-RT (4) hAgo2; (PL) pre-stained protein ladder. Gel legend (Figure – B): (RNA, 1) natural RNA (21RNA_1A-Cy5), standard; natural 21RNA_1A-Cy5 in reaction with (2) HuR; (3) HIV-RT (4) hAgo2; (PL) pre-stained protein ladder. Cy5 channel scan.**

In order to validate the specificity of the prepared modified RNA probe (**21RNA_1A$^{CA}$-Cy5**), a series of control reactions was performed with various weakly- or non-RBPs. Each of these proteins was designedly chosen for its content of Cys and/or His amino acids. Among the non-RBPs, bovine serum albumin (BSA), galectin 1 (Gal1), lysozyme (lysoz.) were tested, while representative weakly-RBPs included single-strand binding protein (SSB) and human recombinant histone H2A (H2A). The outcomes of these control reactions were evaluated by SDS-PAGE, revealing either absence (0%) or minimal amounts (2-3%) of cross-linking products for weakly- and non-RBPs. Notably, an exception was observed with the highly positively charged histone, which formed a covalent conjugate with 9% yield (Figure 41, Table 7). This observation shows that a strong proximity effect and the presence of targeted amino acids within the protein recognition site are essential for the formation of a stable covalent bond with the protein of interest.

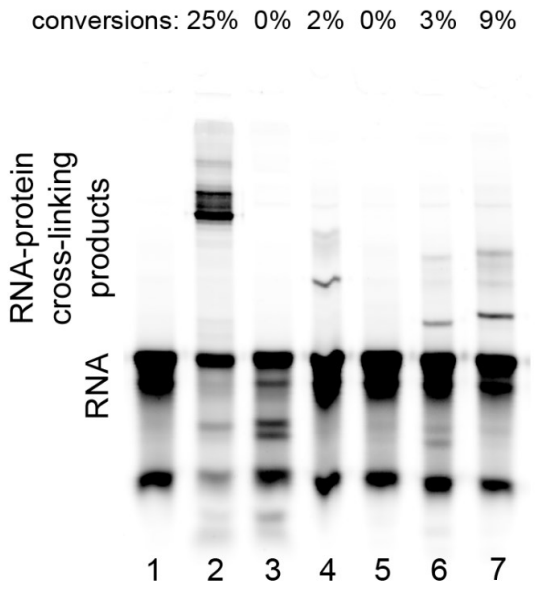



**Figure 41. 15% denaturing SDS-PAGE analysis of cross-linking reaction of modified 21RNA_1A$^{CA}$-Cy5 with various proteins. (RNA, 1) modified RNA (21RNA_1A$^{CA}$-Cy5), standard; modified 21RNA_1A$^{CA}$-Cy5 in reaction with: (2) HuR, control reaction; (3) BSA; (4) SSB; (5) Gal1; (6) lysoz.; (7) histone H2A. Cy5 channel scan.**

**Table 7. Analysis of cross-linking efficacy and selectivity with various proteins.**

| Protein | | Equivalents | Conversion |
|---|---|---|---|
| **RNA binding proteins** | HuR | 2 | 30% |
| | HIV-RT | 2 | 28% |
| | hAgo2 | 2 | 22% |
| **Weak- or non-RNA binding proteins** | BSA | 2 | 0% |
| | SSB | 2 | 2% |
| | Gal1 | 2 | 0% |
| | lysoz. | 2 | 3% |
| | histone H2A | 2 | 9% |

To investigate the complex relationship between the quantity of reactive modifications, cross-linking efficiency, and selectivity for target proteins, I prepared three distinct modified RNA probes with an increasing number of CA functionalities within a single strand. These modified RNAs were bearing either one (**36RNA_1A$^{CA}$-Cy5**), three (**36RNA_3A$^{CA}$-Cy5**), or seven (**36RNA_7A$^{CA}$-Cy5**) CA modifications. The aim was to compare their performance in cross-linking reactions with the RBP represented by the previously studied HuR. As a control for off-target interactions, the non-RBP BSA was included in the investigation. All reactions were conducted under identical conditions, employing 2 equiv. of proteins. Upon analysis of SDS-PAGE results, an interesting pattern emerged. On one hand, an increase in the number of CA modifications positively correlated with enhanced cross-

linking efficiency with the RBP, as depicted in Figure 42. However, a simultaneous reduction in selectivity was observed, particularly evident in the case of the heavily modified **36RNA_7A$^{CA}$-Cy5**, demonstrated by its cross-linking with BSA (Figure 43, Table 8). Therefore, an optimal balance between cross-linking specificity and efficiency must be considered case by case.

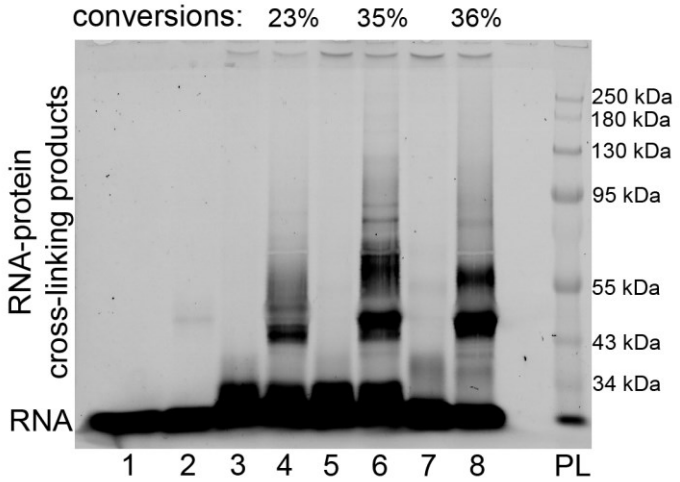



**Figure 42. 7% denaturing SDS-PAGE analysis of cross-linking reaction of either natural 36RNA_1A-Cy5 or various modified RNA probes (36RNA_1A$^{CA}$-Cy5, 36RNA_3A$^{CA}$-Cy5, 36RNA_7A$^{CA}$-Cy5) with HuR protein. Gel legend: (1) natural 36RNA_1A-Cy5, standard; (2) natural 36RNA_1A-Cy5 in reaction with HuR; (3) modified 36RNA_1A$^{CA}$-Cy5, standard; (4) modified 36RNA_1A$^{CA}$-Cy5 in reaction with HuR; (5) modified 36RNA_3A$^{CA}$-Cy5, standard; (6) modified 36RNA_3A$^{CA}$-Cy5 in reaction with HuR; (7) modified 36RNA_7A$^{CA}$-Cy5, standard; (8) modified 36RNA_7A$^{CA}$-Cy5 in reaction with HuR; (PL) pre-stained protein ladder. Cy5 channel scan.**

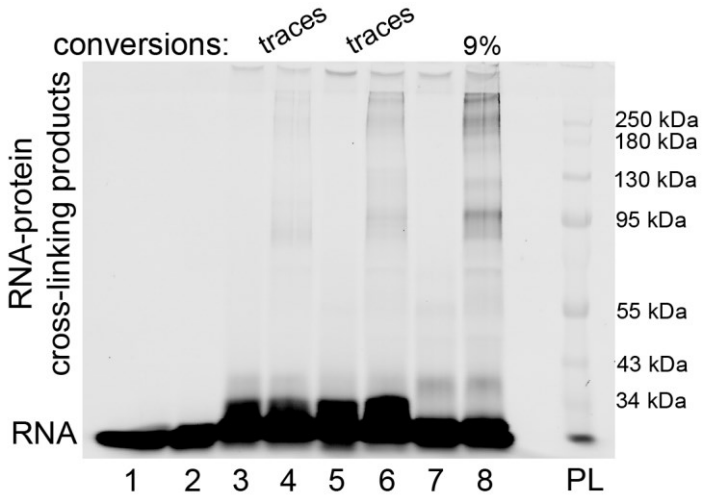



**Figure 43. 7% denaturing SDS-PAGE analysis of cross-linking reaction of either natural 36RNA_1A-Cy5 or various modified RNA probes (36RNA_1A$^{CA}$-Cy5, 36RNA_3A$^{CA}$-Cy5, 36RNA_7A$^{CA}$-Cy5) with BSA protein. Gel legend: (1) natural 36RNA_1A-Cy5, standard; (2) natural 36RNA_1A-Cy5 in reaction with BSA; (3) modified 36RNA_1A$^{CA}$-Cy5, standard; (4) modified 36RNA_1A$^{CA}$-Cy5 in reaction with BSA; (5) modified 36RNA_3A$^{CA}$-Cy5, standard; (6) modified 36RNA_3A$^{CA}$-Cy5 in reaction with BSA; (7) modified 36RNA_7A$^{CA}$-Cy5, standard; (8) modified 36RNA_7A$^{CA}$-Cy5 in reaction with BSA; (PL) pre-stained protein ladder. Cy5 channel scan.**

**Table 8. Analysis of cross-linking efficiency and selectivity depending on number of CA modifications in RNA.**

| RNA | Protein | Equivalents | Conversion |
|---|---|---|---|
| **36RNA_1A$^{CA}$-Cy5** | HuR | 2 | 23% |
| **36RNA_1A$^{CA}$-Cy5** | BSA | 2 | traces |
| **36RNA_3A$^{CA}$-Cy5** | HuR | 2 | 35% |
| **36RNA_3A$^{CA}$-Cy5** | BSA | 2 | traces |
| **36RNA_7A$^{CA}$-Cy5** | HuR | 2 | 36% |
| **36RNA_7A$^{CA}$-Cy5** | BSA | 2 | 9% |

Furthermore, the efficacy of cross-linking between the modified RNA and the protein of interest was proved by western blot (WB) analysis. In this case, non-fluorescently labelled modified **20RNA_1A$^{CA}$** or the natural **20RNA_1A** (serving as a negative control) were incubated with HuR or HIV-RT, followed by SDS-PAGE and subsequent WB analysis. After proteins were transferred to a membrane, incubation with the corresponding primary antibodies (anti-HuR or anti-HIV-RT) was carried out, followed by the application of fluorescent secondary antibodies. This procedure revealed the formation of stable RNA-protein conjugates. Notably, the cross-linked RNA-protein complexes, represented by bands exhibiting lower mobility in comparison to unreacted free proteins, were only formed in reactions involving the modified **20RNA_1A$^{CA}$** (Figure 44 – A, B). Contrary to small HuR or HIV-RT, for the large hAgo2 protein, no apparent shift between unreacted protein and the RNA-protein complex was observed when using the small, non-labelled **20RNA_1A$^{CA}$** (Figure 44 – C). To

overcome this challenge and generate a "super-shift" enabling the differentiation between free protein and the cross-linked complex, biotinylated RNA (**21RNA_1A$^{CA}$-Bio**) was employed. Considering the high affinity of streptavidin to biotin, forming one of the strongest non-covalent interactions in nature[266], the cross-linked RNA-protein complex was treated with fluorescent APC streptavidin, followed by SDS-PAGE and WB analysis with primary (anti-hAgo2) and secondary fluorescent antibodies. This strategic use of biotinylated RNA facilitated better resolution between free and conjugated hAgo2 protein with modified RNA (**21RNA_1A$^{CA}$-Bio**), as illustrated in Figure 44 – D.

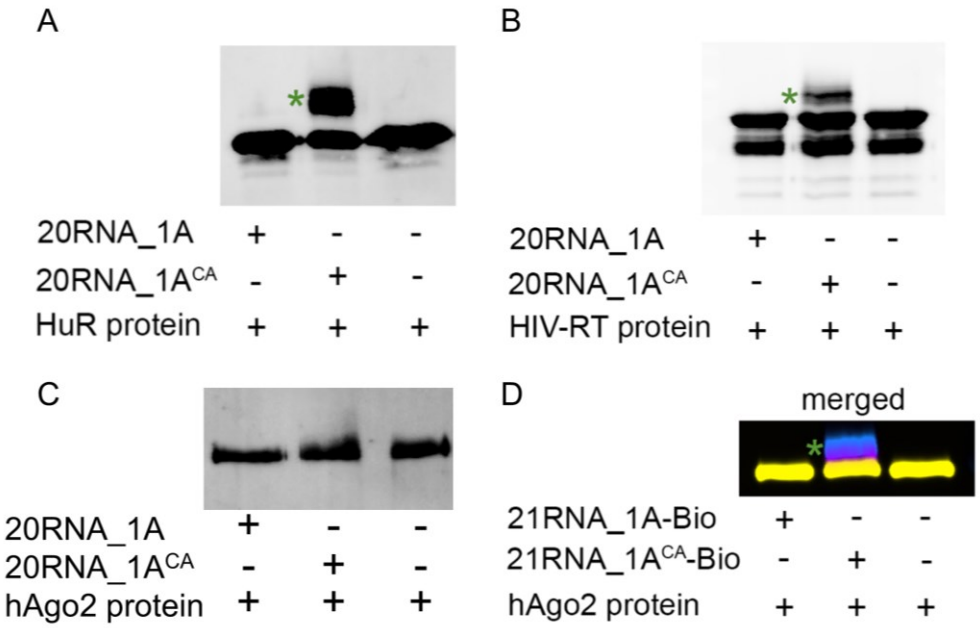



**Figure 44. Analysis of cross-linking reaction of either modified 20RNA_1A$^{CA}$ or natural 20RNA_1A with: A – HuR, B – HIV-RT, C – hAgo2 and analysis of cross-linking reaction of either modified 20RNA_1A$^{CA}$-BIO or natural 20RNA_1A-BIO with hAgo2 (D) by western blot. Gel legend: (*) RNA-protein cross-links. Alexa Fluor 488 channel scan of the membrane for visualisation of protein and RNA-protein complexes. APC channel scan for visualisation of RNA-streptavidin complex (D).**

### 3.1.7 ESI-MS identification of RNA-protein conjugates

The identification process involved subjecting the intact RNA-protein conjugates to LC-ESI-MS, where they were desalted, thermally denatured, and converted into gas-phase ions. The resulting mass spectrum provided valuable information about the mass-to-charge ratios of the ions, allowing for the determination of the molecular masses of the RNA-protein complexes. This step was crucial for confirming the successful formation of covalent bonds between the modified RNA and the target proteins, further validating the SDS-PAGE outcomes of the cross-linking reactions. For analysis of whole RNA-protein conjugates by intact ESI-MS, I performed cross-linking reactions of **20RNA_1A$^{CA}$** with all previously studied RBPs. In case of HuR protein and HIV-RT we were able

to acquire mass of RNA-protein complexes (Figure 105 for **20RNA_1A$^{CA-HuR}$**, **20RNA_1A$^{CA-HIV-RT}$**), however in case of hAgo2 because of detergents present in protein storage buffer, we were not able to obtain any relevant results. Moreover, we performed LC-MS analysis of **20RNA_1A$^{CA-HIV-RT}$** conjugate. Although separation of these species is extremely challenging and established procedures are missing, we succeeded to separate all components present in the reaction mixture (Figure 45). Efficient separation and ionisation of all components was ensured by usage of basic TEAB containing mobile phases with high concentration of TEA. We acquired mass spectra of unreacted free RNA (**20RNA_1A$^{CA}$**), unreacted free protein (HIV-RT) and desired **20RNA_1A$^{CA-HIV-RT}$** conjugate (Figure 46, Table 9).

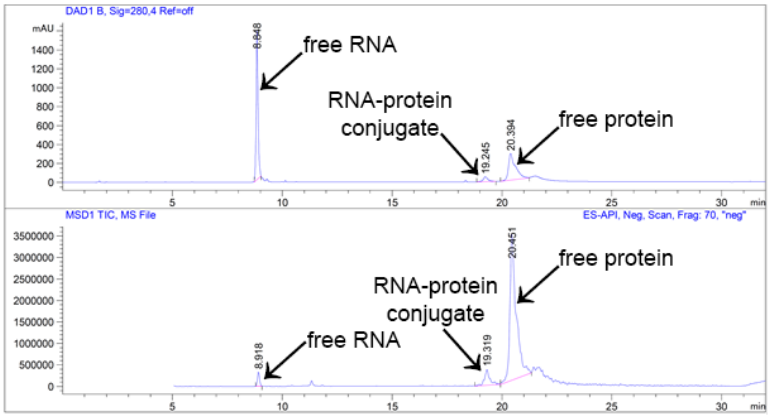



**Figure 45. LC-MS separation chromatogram of cross-linking reaction of 20RNA_1A$^{CA}$ with HIV-RT.**

**Table 9. MS data of 20RNA_1A$^{CA-HuR}$ conjugate and from LC-MS separation of 20RNA_1A$^{CA}$, HIV-RT and 20RNA_1A$^{CA-HIV-RT}$.**

| Analysed component | Calculated mass | Measured mass |
|---|---|---|
| **conjugate, 20RNA_1A$^{CA-HuR}$** | 44937.9 Da* | 44936.8 Da* |
| **free RNA, 20RNA_1A$^{CA}$** | 6601.2 Da | 6598.0 Da |
| **free protein, HIV-RT** | 65529.3 Da | 65528.0 Da |
| **conjugate, 20RNA_1A$^{CA-HIV-RT}$** | 72094.0 Da | 72092.0 Da |

* Calculated and measured mass of protein with loss of N-terminal methionine.

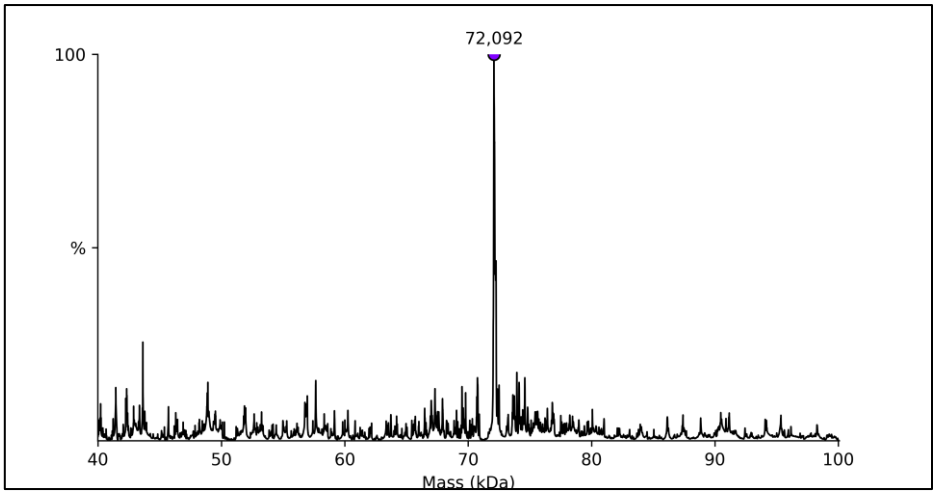



**Figure 46. Deconvoluted mass spectrum of 20RNA_1A$^{CA\text{-}HIV\text{-}RT}$ conjugate from LC-MS analysis. Calculated mass: 72094.0 Da, measured mass: 72092.0 Da.**

### 3.1.8 Identification of cross-linked amino acids by proteomics

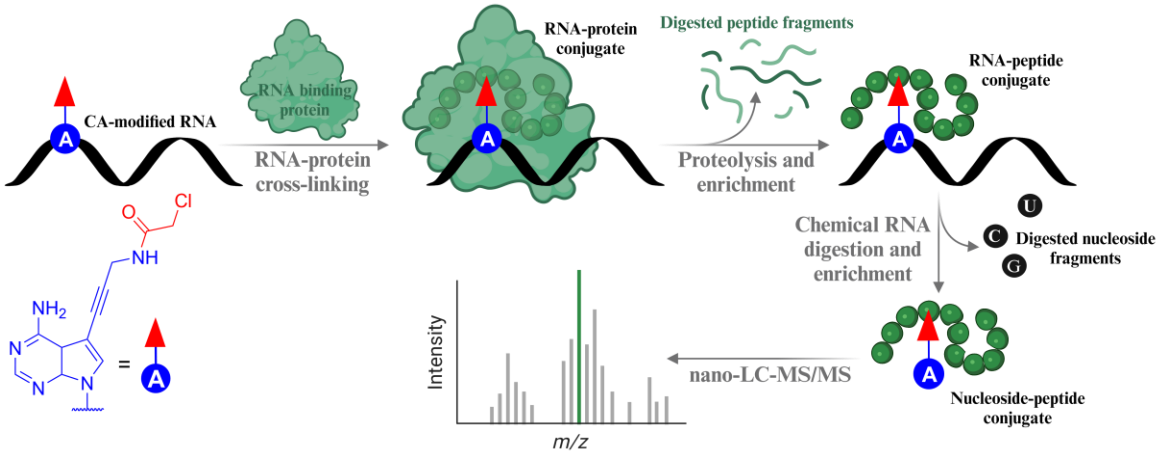



**Figure 47. Strategy for proteomic analysis of cross-linked amino acids in RBPs.**

In an effort to uncover the details of interactions between cross-linked RNA and proteins, to pinpoint targeted amino acid side-chain functionalities and to analyse locations of the cross-linked amino acids, proteomic analysis followed by nano-LC-MS/MS (Figure 47) was employed. For accurate assignment of reacted nucleobase, modified RNA with only one CA modification was used (**20RNA_1A$^{CA}$**). The covalently cross-linked RNA-protein complexes, prepared by reaction of **20RNA_1A$^{CA}$** with each of upper mentioned proteins, were first digested by proteases. A screening of proteases, including trypsin, chymotrypsin, proalanase, lysarginase, and α-lytic protease, was conducted to identify the most suitable enzymes for the efficient digestion of each studied protein, leading to maximum sequence coverage. Based on this, protein digestion of both **20RNA_1A$^{CA\text{-}hAgo2}$**

and **20RNA_1A**$^{\textbf{CA-HIV-RT}}$ complexes was achieved using trypsin, while chymotrypsin was employed for digestion of the **20RNA_1A**$^{\textbf{CA-HuR}}$ complex. The resulting RNA-peptide conjugates were enriched using silica-based spin columns to eliminate non-cross-linked peptides. Following thermal denaturation of the proteases I proceeded with RNA digestion. Mapping of RNA binding sites is highly challenging since conventional enzymatic digestion by nucleases is often incomplete due to steric hinderance, that leads to formation of complex RNA adducts (e.g., nucleosides, nucleotides, oligonucleotides). To address these limitations, I applied in this case a recently established chemical cleavage method. The efficient cutting of phosphodiester bonds within RNA-peptide conjugates was mediated by concentrated hydrofluoric acid (HF). Digestion into mono-nucleosides leaving free OH-group at 5'-end is crucial for improvement of peptide coverage, identification and resolution at single amino acid level[259]. Therefore, RNA-peptide conjugates were incubated overnight in concentrated 48% HF to generate peptides bearing mono-nucleoside residue originating from modified RNA. Upon removal of HF under stream of nitrogen in fume hood overnight, the samples were purified and again enriched from salts, digested nucleosides and traces of HF using C18 spin columns, prior to nano-LC-MS/MS (Figure 47). Cross-linked amino acids, identified by nano-LC-MS/MS (Table 10), were correlated with the available crystal structures of proteins to deduce any relationship between covalent capture and position of the residue within RBPs. In case of hAgo2, cross-linking reaction followed by complete protein and RNA digestion was replicated twice under identical conditions to enhance confidence in identifying amino acids. A third replicate with reduced 6 h incubation showcased cross-linking efficiency even in a shorter reaction period. We were able to identify five conjugated Cys out of total 22 Cys present in hAgo2 protein (Table 10, Figure 50, Figure 51, Figure 107-Figure 115). When compared to the available crystal structure of hAgo2-siRNA complex (PDB:4W5N), we were able to pinpoint three Cys residues in close proximity to interacting RNA [Cys362 (8.6 Å), Cys84 (7.0 Å) and Cys345 (6.7Å)]. The remaining two captured Cys were more distant, with Cys206 (18.3 Å) located in a disordered region, and Cys480 (15.8 Å) positioned on the edge of the RNA binding groove (Figure 51). A different case is the HuR protein, that is present as a dimer. Here we were not able to identify which of the Cys residues (Cys306-A, Cys306-B) originating from two homologous subunits is cross-linked to our modified RNA probe (Figure 48, Figure 52). Similar difficulty, as for HuR, arose when studying HIV-RT that is composed of two homologous subunits as well, but of different lengths. We identified two Cys (Cys38-A, Cys38-B, Cys280-A, Cys280-B) to be cross-linked, although the subunit identity was not possible to be determined (Figure 49, Figure 53, Figure 106). The overall acquired data (Table 10) represent the approximate distance for efficient cross-linking. However, since some Cys in proximity to RNA binding site were not captured, we hypothesise that suitable orientation as well as accessibility of the targeted amino acid in the RNA

recognition sequence are critical factors for efficient cross-linking. The application of proteomics aimed to provide a detailed understanding of RNA-protein interactions at amino acid level resolution. This information is crucial for mapping the precise RNA binding sites within the proteins involved in the cross-linking reactions.

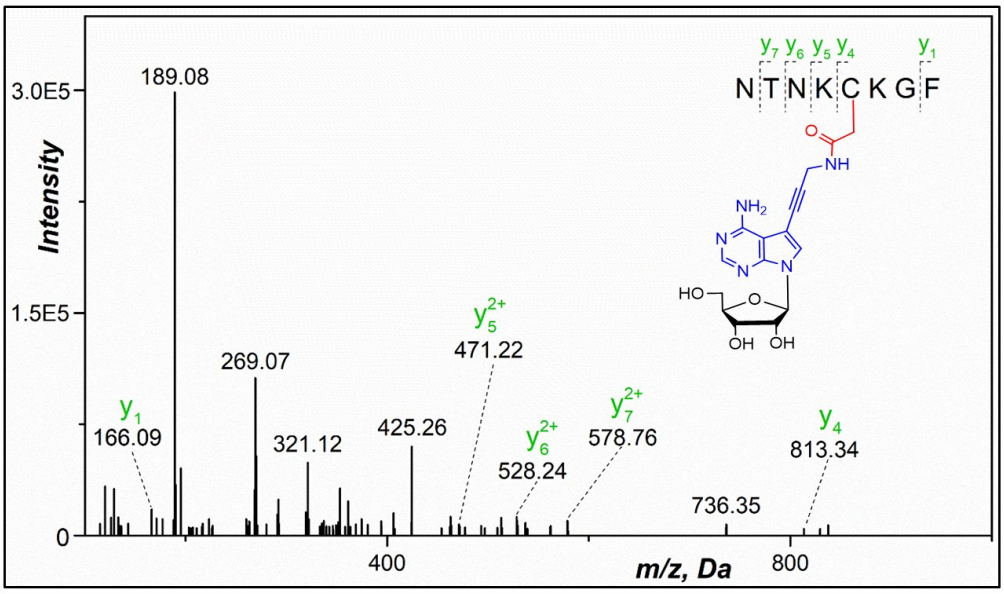



**Figure 48. Nano-ESI$^+$-MS/MS (fragmentation) spectrum of identified peptide from 20RNA_1A$^{CA\text{-}HuR}$ conjugate digest. Modified nucleoside (originating from RNA) is cross-linked to C306. *m/z* acquired: 424.1934 Da, z = 3+.**

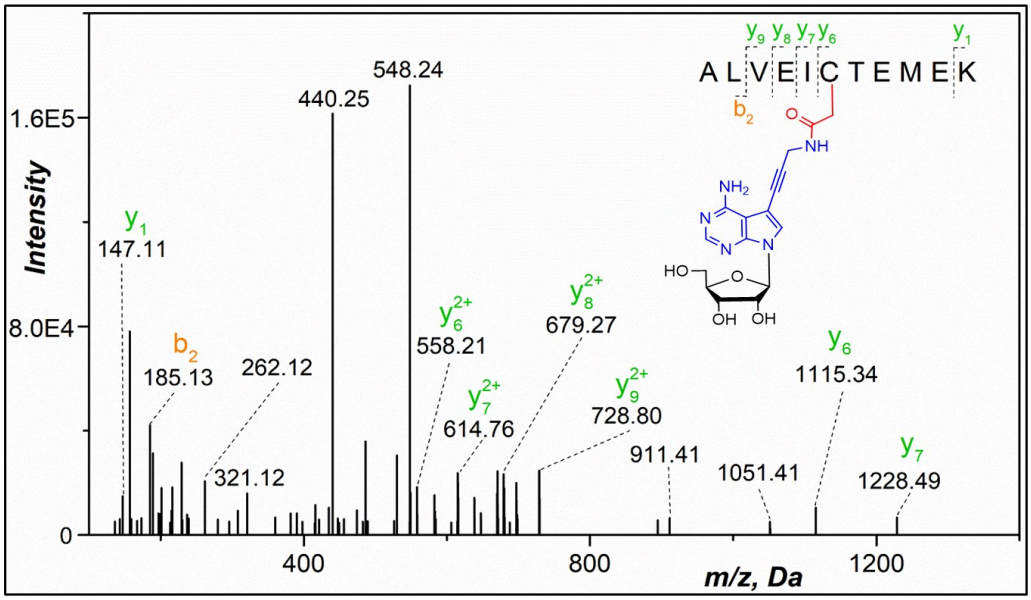



**Figure 49. Nano-ESI$^+$-MS/MS (fragmentation) spectrum of identified peptide from 20RNA_1A$^{CA\text{-}HIV\text{-}RT}$ conjugate digest. Modified nucleoside (originating from RNA) is cross-linked to C38. *m/z* acquired: 547.5820 Da, z = 3+.**

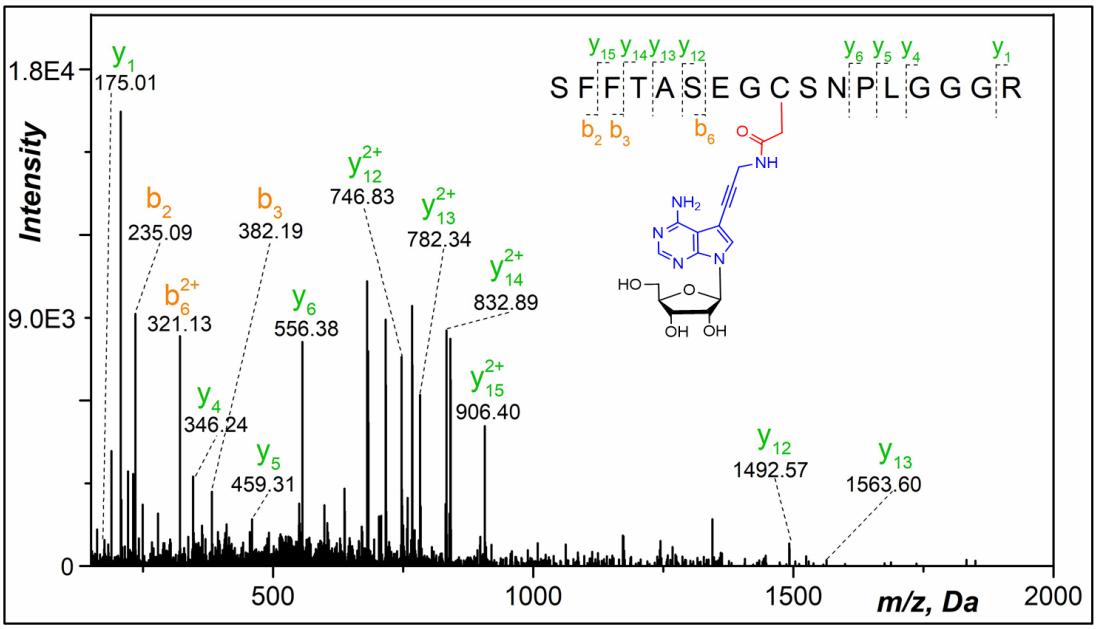



**Figure 50. Nano-ESI$^{+}$-MS/MS (fragmentation) spectrum of identified peptide from 20RNA_1A$^{CA\text{-}hAgo2}$ conjugate digest. Modified nucleoside (originating from RNA) is cross-linked to C206. *m/z* acquired: 682.6306 Da, z = 3+.**

**Table 10. Analysis of RNA-protein digests. Identified peptides with cross-linked amino acids.**

| Identified peptide* | Cross-linked amino acid | Peptide origin | Calculated mass | Measured mass |
|---|---|---|---|---|
| NTNK**C**KGF | C306 | HuR | 424.1932 Da for $[M+H]^{3+}$ | 424.1934 Da for $[M+H]^{3+}$ |
| ALVEI**C**TEMEK♦ | C38 | HIV-RT | 547.5819 Da for $[M+H]^{3+}$ | 547.5820 Da for $[M+H]^{3+}$ |
| QL**C**KLLR | C280 | HIV-RT | 411.8900 Da for $[M+H]^{3+}$ | 411.8861 Da for $[M+H]^{3+}$ |
| IDIYHYELDIKPEK**C**PR | C84 | hAgo2 | 623.5579 Da for $[M+H]^{4+}$ | 623.5577 Da for $[M+H]^{4+}$ |
| SFFTASEG**C**SNPLGGGR | C206 | hAgo2 | 682.6311 Da for $[M+H]^{3+}$ | 682.6306 Da for $[M+H]^{3+}$ |
| YP**H**LP**C**LQVGQEQK# | C345 or H342 | hAgo2 | 500.4933 Da for $[M+H]^{4+}$ | 500.4941 Da for $[M+H]^{4+}$ |
| Q**C**TEVHLK | C480 | hAgo2 | 659.3053 Da for $[M+H]^{2+}$ | 659.2984 Da for $[M+H]^{2+}$ |
| HTYLPLEV**C**NIVAGQR | C362 | hAgo2 | 724.6939 Da for $[M+H]^{3+}$ | 724.6937 Da for $[M+H]^{3+}$ |

* Cross-linked amino acid within identified peptide sequence is highlighted in bold and red. ♦ Oxidation of methionine to methionine sulfoxide. # Not distinguishable which of the two amino acids is cross-linked according to fragmentation spectra.

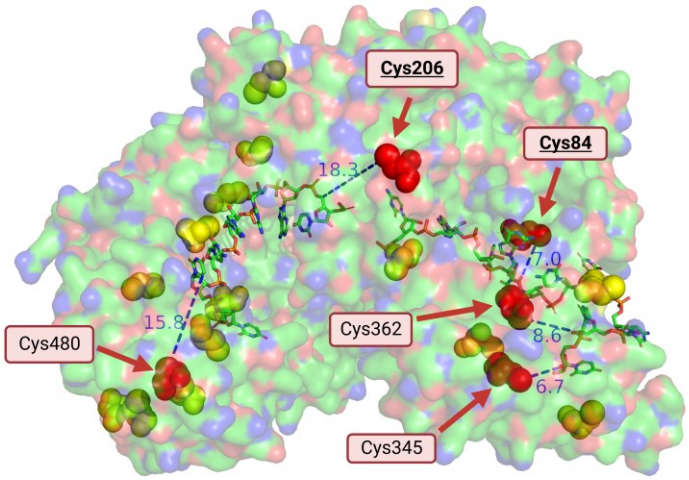



**Figure 51. Identified cysteines mapped on hAgo2-siRNA crystal structure (PDB: 4W5N). Residues are numbered according to amino acid position in our expression construct. Cys in red were detected in at least one cross-linking experiment from three experiments in total. Cys in red bold and underlined were detected in all three cross-linking experiments. Measured distance between sulphur atom and closest atom in RNA are in ångströms. Cys in yellow were not cross-linked.**

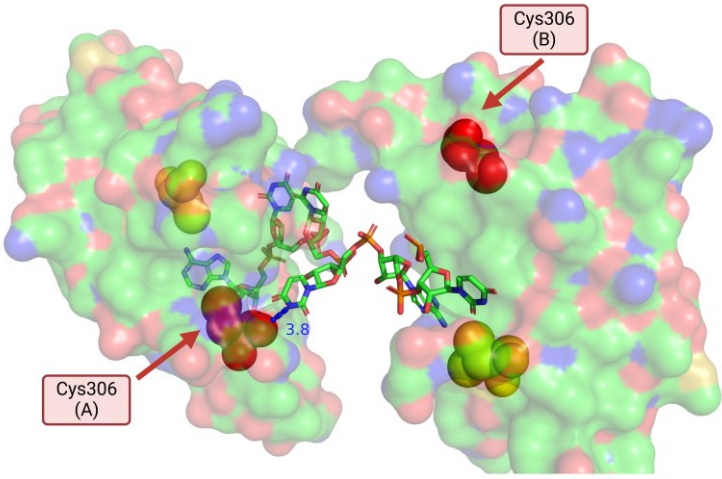



**Figure 52. Identified cysteines mapped on HuR (RRM3 domain)-RNA crystal structure (PDB: 6GD3). Residues are numbered according to amino acid position in our expression construct. Cys in red were detected in the cross-linking experiment. Measured distance for more closer Cys is in ångströms. Cys in yellow were not cross-linked.**

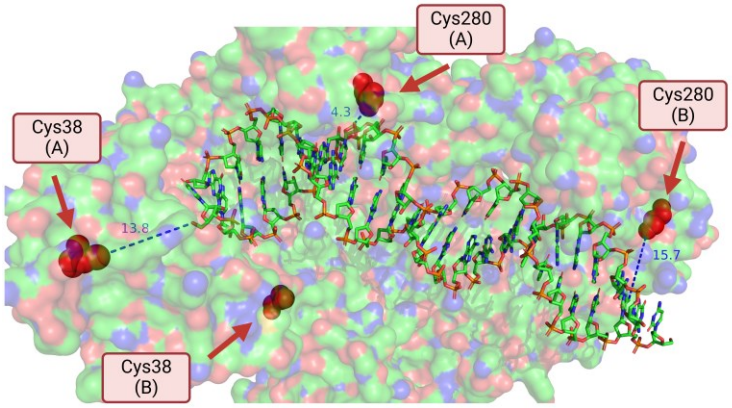



**Figure 53. Identified cysteines mapped on HIV-RT-RNA:DNA duplex crystal structure (PDB: 1HYS). Residues are numbered according to amino acid position in our expression construct. Cys in red were detected in the cross-linking experiment. Measured distance for more closer and easier accessible Cys is in ångströms.**

### 3.1.9 Cross-linking reactions in a complex protein mixture

To extend the applicability of cross-linking reactions with our CA-modified RNA probe to a more biologically relevant context, I performed reactions in a complex protein mixture. The cross-linking experiments were performed in presence of HeLa cell lysate with modified **36RNA_1A$^{CA}$-Cy5** and its natural congener **36RNA_1A-Cy5**, as control. SDS-PAGE analysis revealed presence of multiple captured proteins within the cell lysate in case of modified RNA (**36RNA_1A$^{CA}$-Cy5**), represented by bands with lower mobility when compared to unreacted nucleic acid standard (Figure 54). Conversely, the unmodified control probe (**36RNA_1A-Cy5**) was not capable to covalently target any of the proteins present in cell lysate in significant amounts. The formation of bands with slower mobility in trace amounts, is hypothesised to be primarily a result of incomplete denaturation of thermally stable proteins and their robust electrostatic interactions (Figure 54).

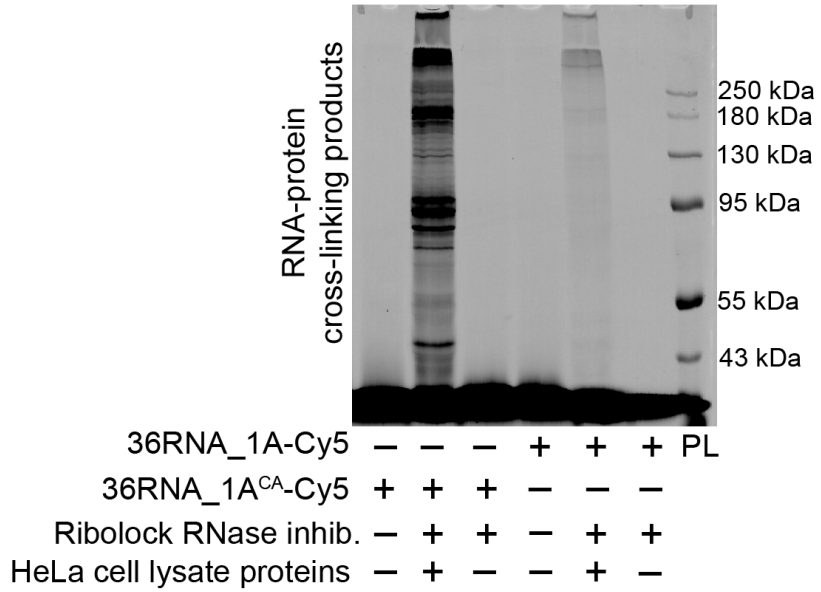



**Figure 54. 7% SDS-PAGE analysis of cross-linking reaction of either natural 36RNA_1A-Cy5 or modified 36RNA_1A$^{CA}$-Cy5 with proteins in HeLa cell lysate. Gel legend: (PL) pre-stained protein ladder. Cy5 channel scan.**

After obtaining promising outcomes from preliminary cell lysate experiments, I set out to undertake a more challenging task, selectively cross-linking protein of interest to its recognising sequence within a cellular lysate mixture. The goal was to explore the selective potential of sequence divergent CA-modified RNA probes to selectively cross-link with its recognising protein amid the presence of a diverse array of non-target proteins and cellular RNA molecules, that simulates the complexity found in biological systems. Since, I already verified the cross-linking efficacy of the modified probe with HuR in previously performed experiments (Figure 40), I picked this protein as a relevant target even for sequence-selective cross-linking within cell lysate system. HuR is regulatory protein implicated in mRNA stability and translation regulation and linked to various diseases including cancer[267]. It is known to preferentially bind to adenosine-uridine (A-U) rich elements, while binding to cytidine-guanosine (C-G) rich RNAs is significantly weaker[268]. To address this, I prepared a set of RNA probes, either rich in A-U nucleosides and containing the HuR binding motif, or a set of non-binding G-C rich RNA oligonucleotides (Table 2). For each biding or non-binding sequence, I prepared either CA-modified probes with the same amount of CA modifications or its natural counterparts (Table 2, Table 3). With all prepared modified and non-modified RNAs in hand, possessing the binding (**21RNA_3A$^{CA}$-bind**, **21RNA_3A-bind**) or non-binding motif (**21RNA_3A$^{CA}$-non-bind**, **21RNA_3A-non-bind**), I proceeded further to cross-linking reactions. All prepared probes were tested in cross-linking with HeLa cell lysate under identical reaction conditions (Figure 55). Reaction analysis was performed by SDS-PAGE prior to

WB detection using primary anti-HuR and secondary HRP-conjugated antibodies. This blot revealed the formation of a significantly slower-moving band exclusively in the case of modified RNA probe with the HuR binding motif (**21RNA_3A$^{CA}$-bind**). In contrast, the non-binding modified RNA (**21RNA_3A$^{CA}$-non-bind**) with the same number of CA modifications or the non-modified probes (**21RNA_3A-bind**, **21RNA_3A-non-bind**) did not form any covalently linked RNA-protein complexes (Figure 56). A notable shift of up to ≈ 250 kDa in case of **21RNA_3A$^{CA}$-bind** was observed, indicating the formation of RNA-HuR linear oligomers[269] (Figure 57). These results conclusively demonstrated that the HuR protein present in HeLa cell lysate was selectively targeted solely by the modified binding RNA probe, while the modified non-binding or non-modified RNAs remained unreactive (Figure 55). Thus, we were able to selectively cross-link HuR protein to modified binding AU-rich RNA sequence.

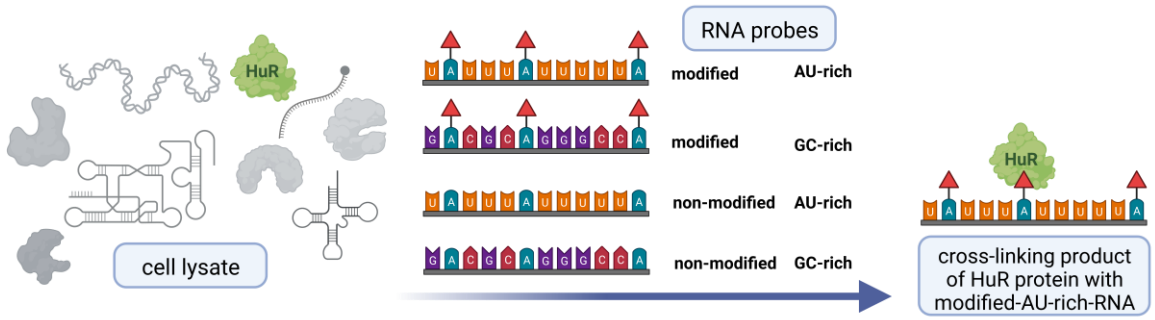



**Figure 55. Selective cross-linking reaction of HuR protein to modified AU-rich RNA probe.**

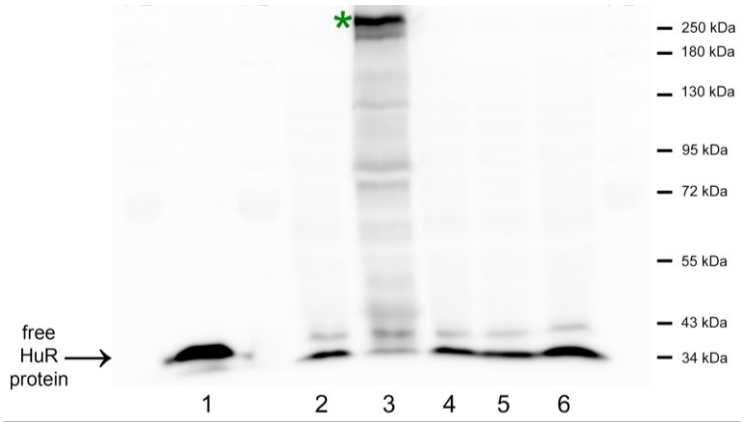



**Figure 56. Analysis by chemiluminescence blot imaging of cross-linking reaction with HuR protein in HeLa cell lysate. Blot legend: (1) standard, purified HuR protein; (2) natural 21RNA_3A-bind in reaction with HeLa cell lysate; (3) modified 21RNA_3A$^{CA}$-bind in reaction with HeLa cell lysate; (4) standard, HeLa cell lysate; (5) natural, 21RNA_3A-non-bind in reaction with HeLa cell lysate; (6) modified 21RNA_3A$^{CA}$-non-bind in reaction with HeLa cell lysate. (*) RNA-protein cross-linking product.**

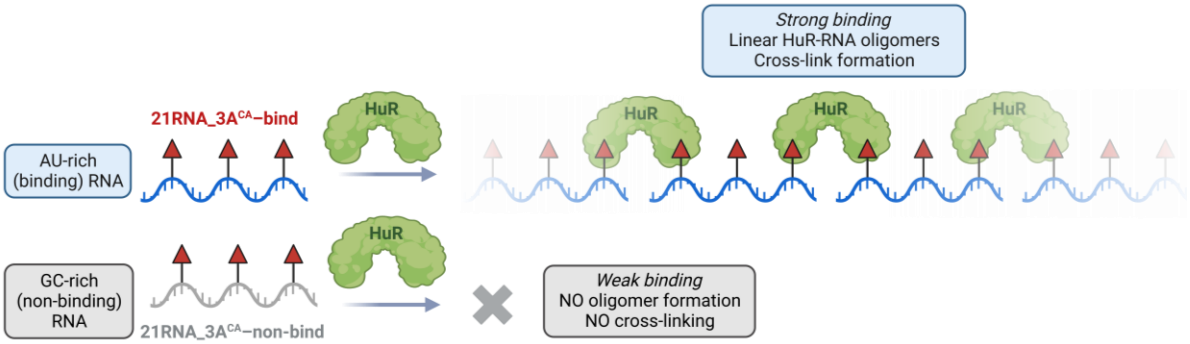



**Figure 57. Possible formation of RNA-HuR linear oligomers.**

### 3.1.10 Conclusion

In conclusion, I developed a straightforward synthetic pathway leading to CA-modified ribonucleoside triphosphate (**rA$^{CA}$TP**). Introduction of the reactive functionality was carried out by aqueous single-step Sonogashira cross-coupling reaction.

Subsequently, I demonstrated the suitability of **rA$^{CA}$TP** as a substrate for T7 RNAP IVT. This was proved by the successful synthesis of short RNA with one CA modification, as well as longer ones with up to seven CA modifications in a single strand.

The reactivity of CA-modified RNA probes was validated by bioconjugations with small (bio)molecules bearing thiol functionality and various peptides containing Cys and His residues. Moreover, efficient cross-linking of modified RNA was demonstrated by reaction with three distinct RBPs. Formation of a stable covalent bond in case of CA-modified RNA and proteins was proved not only by SDS-PAGE and WB analysis, but also by ESI-MS or LC-MS analysis of whole RNA-protein complexes. Crucially, precise RNA binding sites within all three studied proteins were identified by complete digestion of cross-linked RNA-protein conjugates, followed by nano-LC-MS/MS analysis. Furthermore, the efficiency of our CA-modified RNA probes in cross-linking was demonstrated not only with individually purified proteins, but also in the presence of a complex protein mixture.

## 3.2 Expedient production of site specifically nucleobase-labelled or hypermodified RNA with engineered thermophilic DNA polymerases

*This study was performed in collaboration with V. Havlíček, J. Matyašovský, M. Krömer, R. Pohl and L. Poštová Slavětínská. All experiments in this chapter were performed by me, unless otherwise stated. J. Matyašovský synthesised* ***rC$^{mBdp}$TP****. V. Havlíček performed synthesis of E-, Pent- and Ph-modified* ***rN$^{X}$TP****s. V. Havlíček and M. Krömer expressed SFM4-3 polymerase. M. Krömer performed in cellulo translation studies and LC-MS analysis of nucleoside digests. L. Poštová Slavětínská and R. Pohl performed NMR spectra acquisition and interpretation of all synthesised nucleosides and nucleotides.*

### 3.2.1 Introduction

Investigating RNA structure, dynamics, function, and deciphering effects of both natural and designer chemical modifications, as well as development of novel RNA therapies, warrant access to synthetic modified RNA with suitable functional groups (e.g., epitranscriptomic tags, fluorescent, spin or isotopic labels)[270,271].

Automated solid phase synthesis offers the advantage of producing larger quantities of RNA oligonucleotides; however, it is constrained for synthesising relatively short sequences, typically up to 60 nt[38,39]. This limitation becomes apparent when specific practical applications necessitate the generation of longer RNA sequences or incorporation of chemically sensitive moieties. Solid support chemical synthesis is unable to meet both requirements simultaneously. As an alternative approach, the synthesis of modified RNA oligonucleotides can be achieved through enzymatic methods. However, construction of larger modified RNAs by stepwise enzymatic ligation necessitates intricate design of fragments, might be hindered due to structural constraints, suffers from laborious purification and poor yields[152,154]. Site selective modifications can be installed with assistance of substrate promiscuous enzymes, e.g., methyltransferases, transglycosylases or ribozymes[145]. This route is restricted to terminal labelling or to predetermined consensus motifs[272]. Extensive research has been conducted on T7 bacteriophage RNAP due to its remarkable efficiency. Although it is capable to accommodate a wide range of modified nucleotides as substrates[101,111,113–119], it has been shown that this enzyme has limited tolerance to bulky or 7-deazaguanine modifications[43], requires optimal GGG trinucleotide initiation sequence[77,78] and is prone to termination when transcribing heavily structured RNA[79]. In addition, the inclusion of base-modified ribonucleoside triphosphates (**rN$^{X}$TP**s) in IVT generally results in RNA that is uniformly modified at all positions[43,101,111,143]. Notable exception is position-selective labelling by T7 RNAP (PLOR)[124,273], whereby omitting one nucleotide from IVT allows to pause synthesis at missing nucleotide site and to incorporate

modification at this position. The method, however, requires extensive optimisation of reaction parameters.

On the other hand, numerous investigations by researchers have consistently shown that thermostable DNA polymerases are highly efficient enzymes in construction of even heavily modified DNA from base-modified 2'-deoxyribonucleoside triphosphates (**dN$^{X}$TP**s)[127–129]. However, DNA polymerases are not amenable to incorporation of ribonucleoside triphosphates (rNTPs) to avoid disruption of genome integrity[274]. Nevertheless, this limitation was recently unlocked with power of directed evolution, and simultaneously TGK[88] and SFM4-3[91,92] polymerase were evolved. Both enzymes were proficient in synthesis of natural RNA or xenonucleic acids (XNA) with 2'-sugar modification, however, none of them has been examined for polymerisation of nucleobase-modified **rN$^{X}$TP**s leading to base-modified RNA.

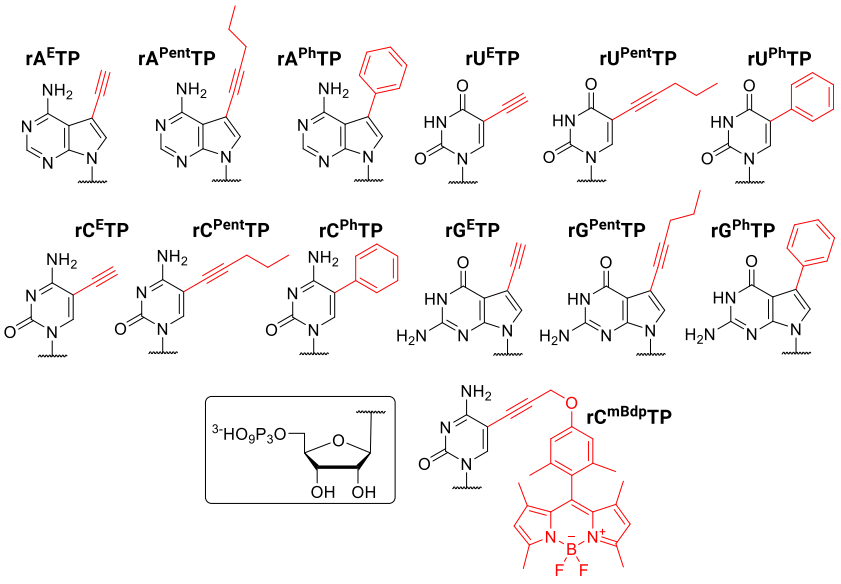
### 3.2.2 Base-modified nucleotide building blocks

To test the scope of modifications accepted by engineered mutant DNA polymerases[88,91,92] and to study the impact of modification characteristics on polymerase performance in PEX reaction, we designed a library of modified **rN$^{X}$TP**s. Modifications were attached to the position C-7 of 7-deazapurine or to position C-5 of pyrimidine bases, which makes them well tolerated by cognate DNA polymerases[69–73]. A full set of 12 unnatural **rN$^{X}$TP**s bearing clickable (ethynyl, E) or hydrophobic groups of increasing bulkiness, pentyn-1-yl (Pent) and phenyl (Ph), suitable for enhancement of protein binding[101] was synthesised (Scheme 12). Upper mentioned functional groups were attached to each pyrimidine or 7-deazapurine nucleotide (**rA$^{X}$TP**, **rU$^{X}$TP**, **rC$^{X}$TP**, **rG$^{X}$TP**). Synthesis of ethynyl-[43] and pentyne-modified building blocks was performed *via* Sonogashira cross-coupling reaction of iodinated nucleosides (rA$^{I}$, rU$^{I}$, rC$^{I}$, rG$^{I}$) and corresponding functionalised terminal alkynes. Next, triphosphorylation step of the base-modified nucleosides was performed under Yoshikawa conditions[44,45]. Reaction cascade with phosphorus(V) oxychloride in trimethylphosphate, followed by addition of tributylammonium pyrophosphate and tributylamine in acetonitrile was applied. Hydrolysis of cyclic triphosphate intermediate by 2 M TEAB buffer provided desired modified **rN$^{X}$TP**s. The phenyl functional group was tethered to the nucleobase by aqueous Suzuki cross-coupling reaction performed directly on iodinated nucleotides (**rA$^{I}$TP**, **rU$^{I}$TP**, **rC$^{I}$TP**). The exception is **rG$^{Ph}$TP**, that was prepared by coupling of iodinated nucleoside (rG$^{I}$) with phenyl boronic acid[43], followed by Yoshikawa phosphorylation procedure.

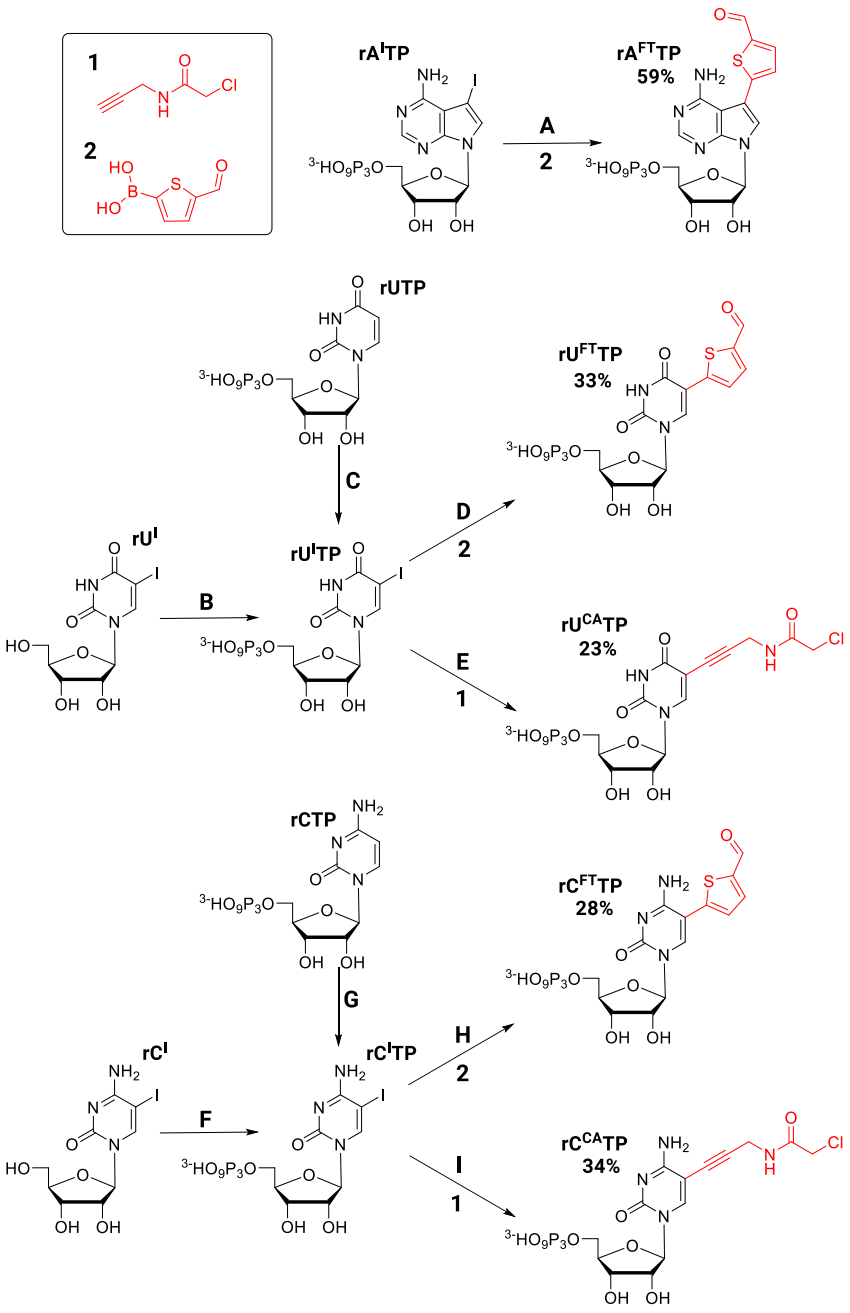
Moreover, a fluorescent **rC$^{mBdp}$TP** suitable for RNA labelling was prepared (Scheme 12). In the first step, alkyne[275] bearing the sterically demanding fluorescent tag (mBdp) was attached to cytosine nucleobase by Pd-catalysed cross-coupling, followed by triphosphorylation reaction.

I also extended the portfolio of tested **rN$^{X}$TP**s by triphosphates bearing highly reactive species, which might be useful for covalent capturing of RNA interacting proteins. I prepared a set of six modified nucleotide building blocks (**rA$^{X}$TP**, **rU$^{X}$TP**, **rC$^{X}$TP**) bearing chloroacetamide (CA) and formylthienyl (FT) functional groups. The synthesis of **rA$^{CA}$TP** was decribed in details in the aforementioned chapter (Section 3.1.2) and tethering the CA functionality to **rU$^{I}$TP** and **rC$^{I}$TP** was performed in analogy to this approach, by aqueous Pd-catalysed Sonogashira cross-coupling reaction with *N*-(propargyl)-chloroacetamide (**1**)[261] (Scheme 13). After reversed-phase and ion exchange HPLC purifications followed by conversion to sodium salt with Dowex 50WX8 in $Na^+$ cycle, the CA-modified nucleotides were isolated in good yields (23% for **rU$^{CA}$TP** and 34% for **rC$^{CA}$TP**). Introduction of FT group was achieved by one-step Suzuki cross-coupling of corresponding iodinated nucleotides (**rA$^{I}$TP**, **rU$^{I}$TP**, **rC$^{I}$TP**) and 5-formylthiophene-2-boronic acid (**2**). Couplings in presence of $Pd(OAc)_2$, TPPTS and $Cs_2CO_3$ were perfomed in a mixture of water and acetonitrile under inert atmosphere at 100 °C for 1 h. After two HPLC purifications the desired **rA$^{FT}$TP** (59%), **rU$^{FT}$TP**[276] (33%), **rC$^{FT}$TP** (28%) were isolated (Scheme 13). Identity and purity of all prepared nucleotide building blocks were proved by NMR analysis.

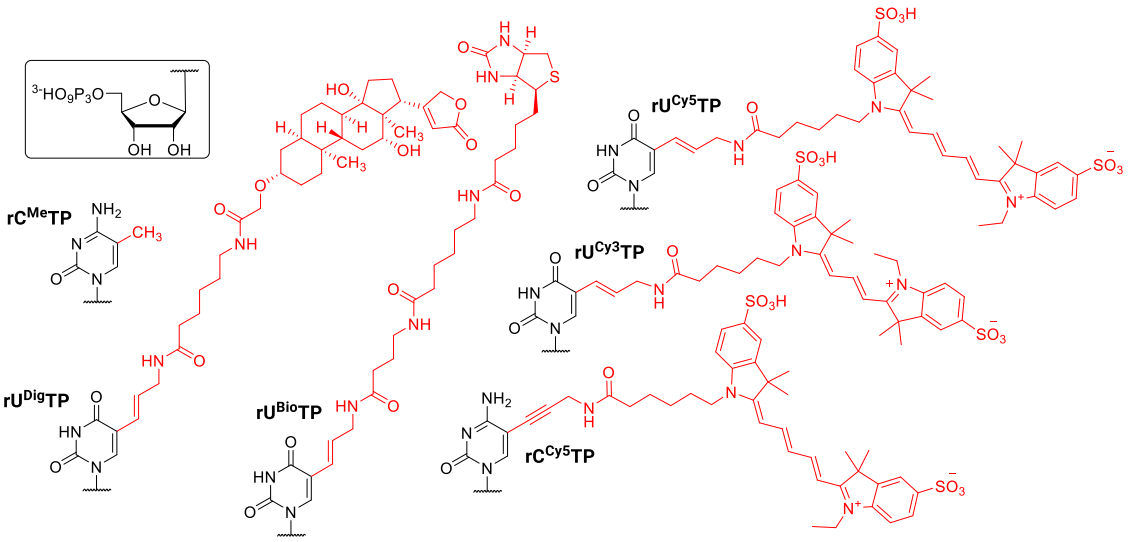
Additionally, the portfolio of 19 synthesised **rN$^{X}$TP**s was extended with commercially available compounds. Nucleotides bearing either cyanine dyes (**rU$^{Cy5}$TP**, **rU$^{Cy3}$TP**, **rC$^{Cy5}$TP**) useful for fluorescent studies, or biotin (Bio) and digoxigenin (Dig) affinity tags (**rU$^{Bio}$TP**, **rU$^{Dig}$TP**), as well as epitranscriptomic methyl (Me) modifications (**rC$^{Me}$TP**) were tested in PEX reactions (Scheme 14).




**Scheme 12. Chemical structures of hydrophobic E-, Pent-, Ph-modified rN$^{X}$TPs synthesised by V. Havlíček and fluorescent rC$^{mBdp}$TP prepared by J. Matyašovský.**




**Scheme 13. Synthetic pathway leading to CA- and FT-modified rN$^{X}$TPs. Detailed synthesis of *N*-(propargyl)-chloroacetamide (1) and rA$^{CA}$TP is shown in Scheme 10. Reaction conditions for triphosphorylation – B, F: i) $POCl_3$, $PO(OMe)_3$, 0°C, 4h, ii) $Bu_3N$, bis(tri-*n*-butylammonium)pyrophosphate, dry $CH_3CN$, 0 °C, 6 h, iii) 2 M TEAB. Reaction conditions for iodination – C, G: $NaN_3$, *N*-iodo-succinimide, $H_2O$, 40 °C, overnight. Reaction conditions for Suzuki cross-coupling – A, D, H: $Cs_2CO_3$, $Pd(OAc)_2$, TPPTS, 2, $H_2O/CH_3CN$, 100 °C, 1h. Reaction conditions for Sonogashira cross-coupling – E, I: CuI, $Pd(OAc)_2$, TPPTS, 1, DIPEA, $H_2O/CH_3CN$, 65 °C, 3h for rC$^{CA}$TP, 4 h for rU$^{CA}$TP.**



**Scheme 14. Chemical structures of commercially available rC$^{Me}$TP, rU$^{Dig}$TP, rU$^{Bio}$TP, rU$^{Cy5}$TP, rU$^{Cy3}$TP and rC$^{Cy5}$TP.**

### 3.2.3 Preparation of thermostable engineered DNA polymerases

Neither TGK nor SFM4-3 polymerase are commercially available, therefore recombinant expression was required. The pET-based plasmid for preparation of SFM4-3 polymerase was a kind gift from F. E. Romesberg group, whereas the pGDR11 vector encoding for TGK polymerase was obtained as a gift from J. C. Chaput laboratory. Plasmids contained gene for desired polymerase under either T7 promoter for SFM4-3 or T5 promoter for TGK. Both of them labelled with either C-terminal 6× his-tag in case of SFM4-3 or N-terminal 6× his-tag for TGK, crucial for efficient purification step. Thus, purification of the SFM4-3 polymerase was performed using nickel affinity chromatography on ÄKTA pure FPLC system. Conversely, the purification of the TGK polymerase followed a two-step procedure. Initially, it underwent purification through heparin-agarose, followed by a subsequent polishing step using nickel affinity chromatography (Figure 58). Purity of both prepared engineered DNA polymerases was analysed by SDS-PAGE (Figure 59) and UV-VIS absorbance spectrum measurements, that excluded any noticeable protein or DNA contamination. Prepared engineered mutant DNA polymerases were then tested in enzymatic synthesis of various nucleobase-modified RNAs with manifold of **rN$^{X}$TP**s.

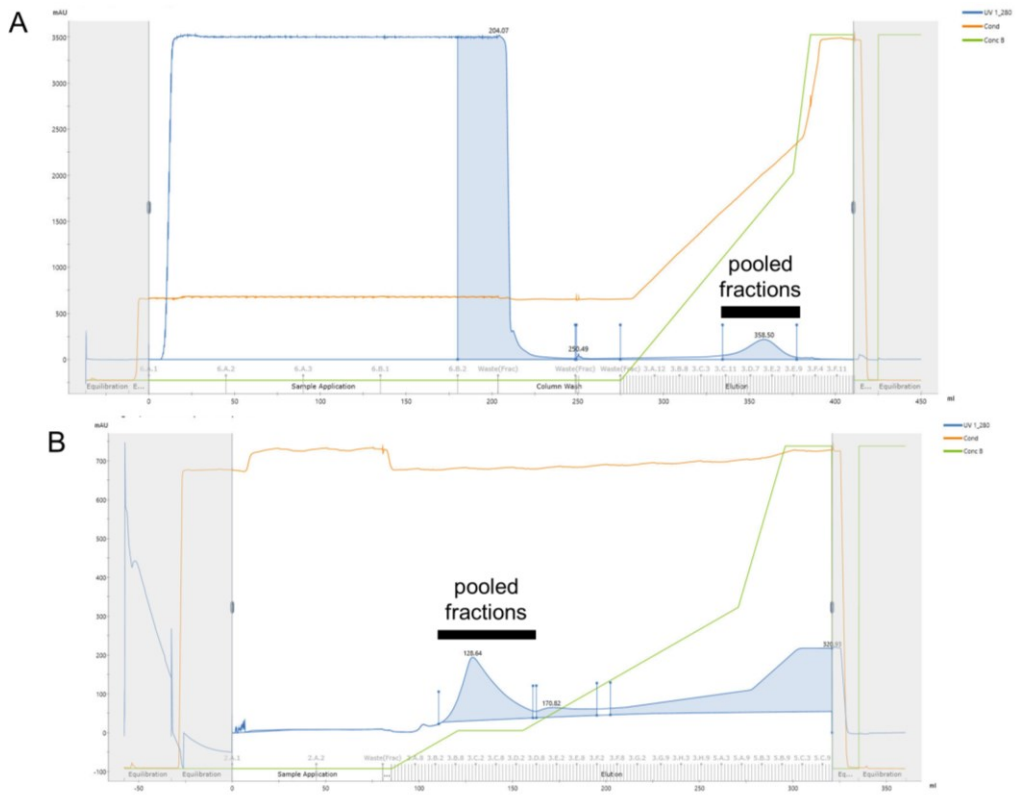



**Figure 58. A – HiTrap Heparin HP purification of TGK polymerase (1st separation). B – HisTrap HP purification of TGK polymerase (2nd purification).**

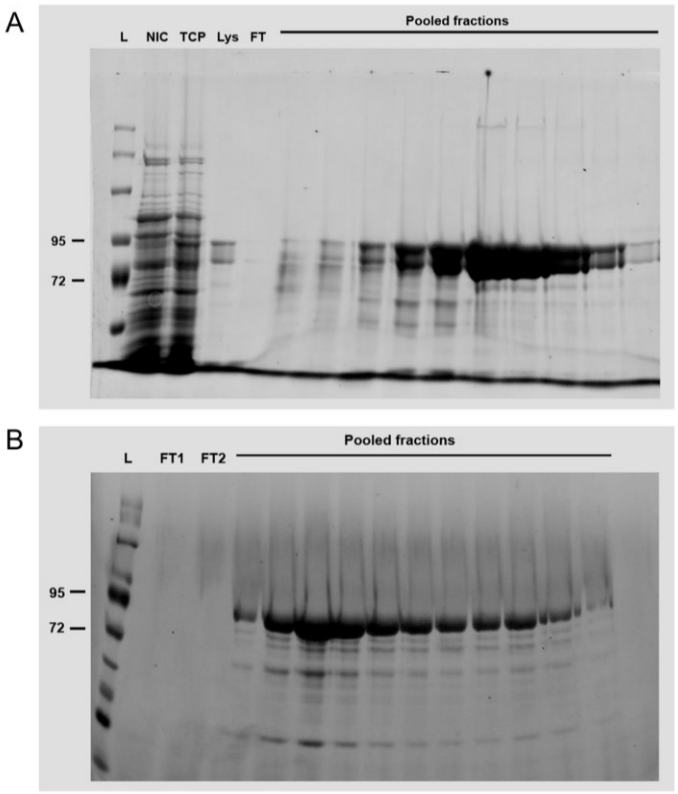



**Figure 59. Cy5 scan of 10% SDS-PAGE analysis of TGK polymerase purification. Post-staining of SDS-PAGE was performed with PageBlue protein staining solution. A – Analysis after HiTrap Heparin HP purification. B – Analysis after HisTrap HP purification. Gel legend: (L) pre-stained protein ladder; (NIC) non-induced control; (TCP) total cell protein; (Lys) heat-treated lysate; (FT) flow-through.**

### 3.2.4 Enzymatic construction of base-modified RNA by PEX reaction and mutant DNA polymerases

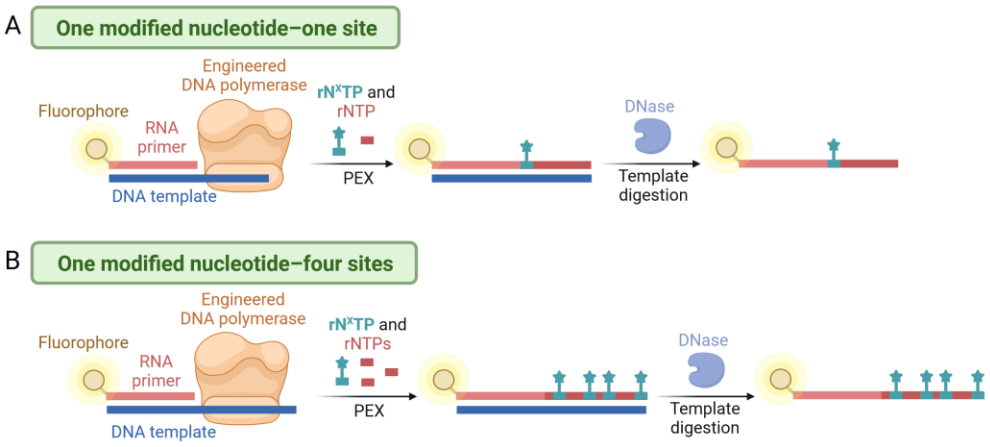



**Figure 60. PEX strategy for enzymatic synthesis of base-modified RNA through PEX reaction with engineered DNA polymerases. Templates requiring incorporation of single modification type at one (A) or four sites (B) were used.**

Initially, I tested incorporation of all aforementioned modified **rN$^{X}$TP**s in PEX requiring insertion of one modified nucleotide at single internal position, followed by three unmodified rNTPs (Figure 60 – A). PEX reactions were performed with fluorescently labelled 15 nt long RNA primer (**FAM-RNA-prim_15nt**, Table 11) and four various 19 nt ssDNA templates [**templ_19_nt_X (X = A, U, C, G)**] encoding for incorporation of each modified nucleotide (Table 12). PEXs were carried out at optimal temperature 60 °C in Thermopol buffer in presence of $Mg^{2+}$ with both engineered DNA polymerases. In each case of the tested **rN$^{X}$TP**s I observed formation of full-length products on dPAGE, confirming that the mutant polymerases are capable to incorporate a variety of modified substrates (Figure 61 – A, B; Figure 117 – Figure 131). Single-stranded modified RNA products were obtained from double-stranded DNA-RNA hybrids by mild template degradation using TurboDNase. All modified RNA probes, prepared either by TGK or SFM4-3 polymerase, were confirmed by MS analysis. All measured masses of each oligonucleotide corresponded to the theoretically predicted ones (Table 14). The only exceptions were MS spectra analysed for samples with CA-bearing **rA$^{CA}$TP** (Figure 62) and **rC$^{CA}$TP**. In these cases, the difference between predicted and analysed mass was ≈ 33 Da indicating possible side reaction of the reactive functionality. Since dPAGE analysis did not prove any possible cross-linking reaction with the polymerase (Figure 118, Figure 129), I assumed that the side conjugation reaction occurred intramolecularly. To prove this theory, I performed digestion of CA-modified RNA probes (**19RNA_A$^{CA}$**, **19RNA_C$^{CA}$**, Table 13) to single nucleosides. LC-MS analysis of nucleoside digests revealed cross-linking reaction of either **A$^{CA}$** (Figure 63, Figure 145) or **C$^{CA}$** to neighbouring guanine nucleobase (Figure 144, Figure 146). Therefore, I optimised the PEX reaction conditions either by reducing the reaction time (15 min) or even by lowering the reaction temperature to 37 °C to eliminate these side-reactions. Both of these changes mitigated formation of self-crosslinked by-products (Figure 61 – A, Figure 64, Figure 65, Table 14).

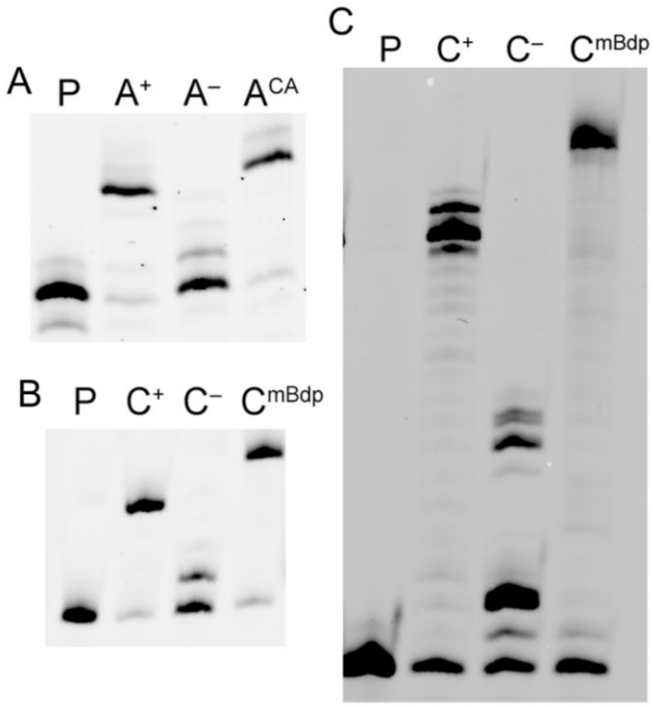



**Figure 61. Examples of dPAGE analysis of PEX reactions. A – PEX at optimised 37 °C with templ_19_nt_A encoding for incorporation of one rA$^{CA}$TP. B – PEX with templ_19_nt_C encoding for incorporation of one rC$^{mBdp}$TP. C – PEX with challenging 5'-(TINA)-templ_31nt encoding for incorporation of four rC$^{mBdp}$TP.**

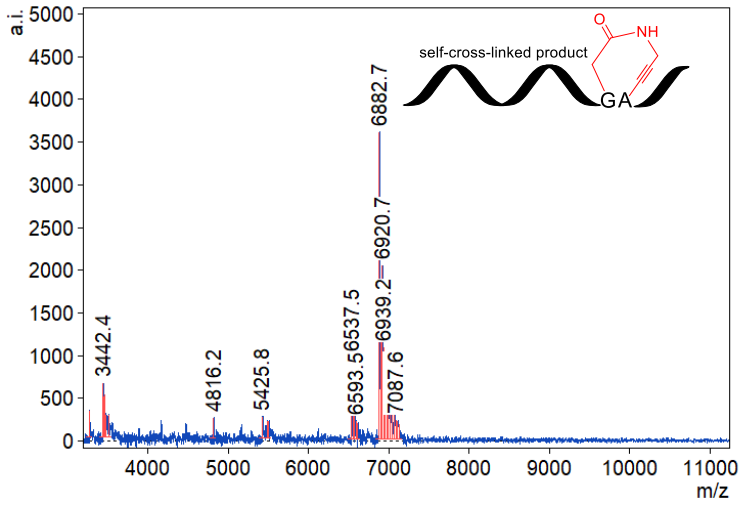



**Figure 62. MS-MALDI analysis of 19RNA_A$^{CA}$. Modified RNA prepared by standard conditions at 60 °C for 2 h. Self-cross-linked RNA probe analysed as major product. Calculated mass: 6915.6 Da; found mass: 6920.7 Da (product); found mass: 6882.7 Da (self-cross-linked product = product – HCl).**

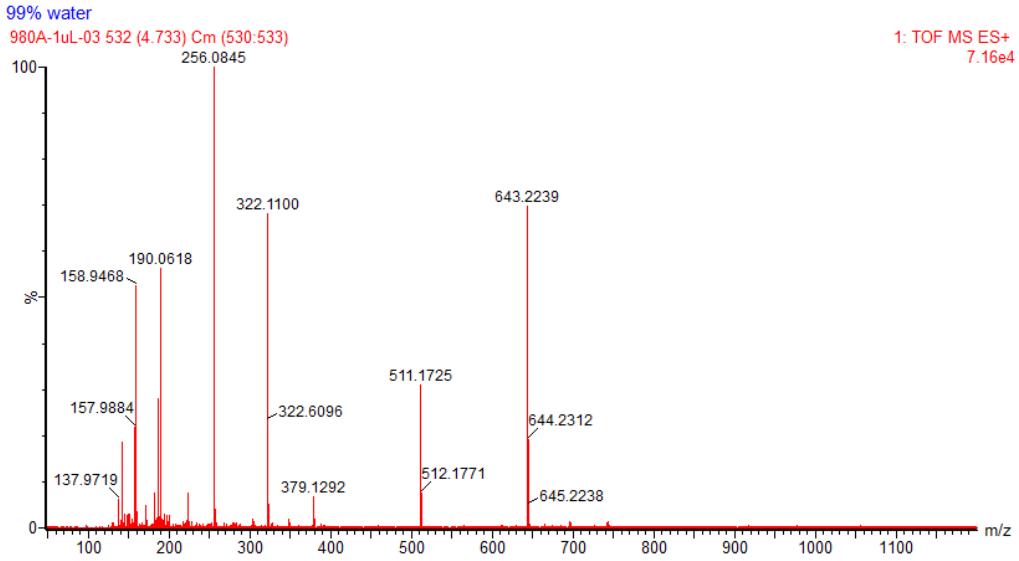



**Figure 63. LC-MS analysis of intramolecular cross-linked 19RNA_A$^{CA}$ digest. Raw spectrum of dinucleoside conjugate of rA$^{CA}$ and G; calculated mass: 643.2225 Da; found mass: 643.2239 Da (product).**

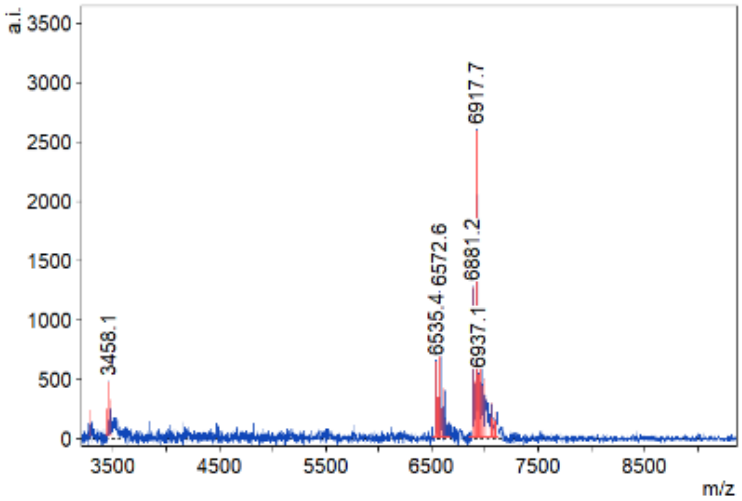



**Figure 64. MS-MALDI analysis of 19RNA_A$^{CA}$. Modified RNA prepared by optimised conditions with reduced reaction time (15 min at 60 °C). Significantly supressed formation of by-product. Desired RNA probe was present as major product. Calculated mass: 6915.6 Da; found mass: 6917.7 Da (product).**

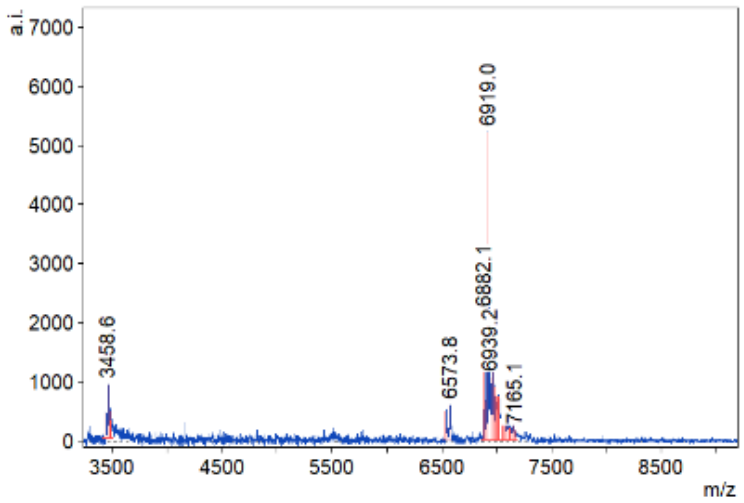



**Figure 65. MS-MALDI analysis of 19RNA_A$^{CA}$. Modified RNA prepared by optimised conditions with reduced temperature to 37 °C (for 2 h). Significantly supressed formation of by-product. Desired RNA probe was present as major product. Calculated mass: 6915.6 Da; found mass: 6919.0 Da (product).**

Further I investigated multiple incorporations of each modified nucleotide using more challenging 31 nt long ssDNA template (**5'-(TINA)-templ_31nt**, Table 12) encoding for 31 nt long RNA with four modification sites (Figure 60 – B). PEX of short RNA primer (**FAM-RNA-prim_15nt**) was successful with all tested **rN$^{X}$TP**s. Neither TGK (Figure 132 – Figure 139) nor SFM4-3 (Figure 61 – C, Figure 132 – Figure 139) polymerase have displayed any difficulties in PEX giving clean full-length products according to dPAGE analysis. Synthesis of modified RNA oligonucleotides was further investigated by MS analysis. Therefore, template removal was ensured with usage of TurboDNase followed by silica-based column purification. The only exceptions were ssRNA probes containing four hydrophobic **rC$^{mBdp}$** modifications, that were not compatible with standard column purification. These were prepared by PEX with dual-biotin-labelled ssDNA template (**5'-(dual-Bio)-templ_31nt**, Table 12), followed by purification with streptavidin magnetic beads and liberation of desired RNA product with thermal denaturation. Identities of all prepared modified ssRNAs were confirmed either by MS-MALDI (Figure 66) or LC-ESI-MS (Table 14).

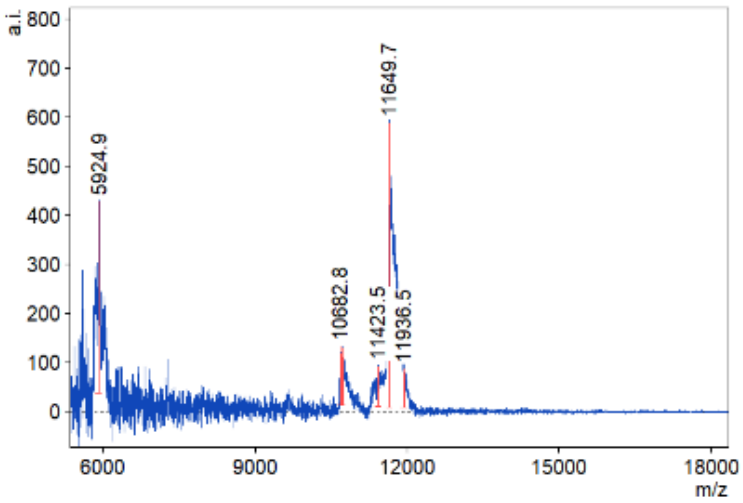



**Figure 66. MS-MALDI analysis of 31RNA_4C$^{mBdp}$. Calculated mass: 11646.1 Da; found mass: 11649.7 Da (product).**

**Table 11. List of RNA oligonucleotides used as primers in this study. For structures of 5'-end modifications see Figure 116.**

| RNA oligonucleotide name | Sequence (5'→3') | 5'-end modification | Length [nt] |
|---|---|---|---|
| **FAM-RNA-prim_15nt** | CAUGGGCGGCAUGGG | FAM | 15 |
| **Cy5-RNA-prim_15nt** | CAUGGGCGGCAUGGG | Cy5 | 15 |
| **RNA-prim_15nt** | CAUGGGCGGCAUGGG | – | 15 |
| **Cy5-mRNA-prim** | GACUCACUAUAGGGCCCCUCUCC | Cy5 | 23 |
| **FAM-RNA-prim_23nt** | GGGAAGAUAUAAUCCUAAUGAUA | FAM | 23 |
| **RNA-prim_23nt** | GGGAAGAUAUAAUCCUAAUGAUA | – | 23 |

**Table 12. List of DNA oligonucleotides used as templates in this study. For structures of 5'-end modifications see Figure 116.**

| DNA oligonucleotide name | Sequence (5'→3') | 5'-end modification | Length [nt] |
|---|---|---|---|
| **templ_16nt** | GCCCATGCCGCCCATG | – | 16 |
| **templ_19nt_A** | CCCTCCCATGCCGCCCATG | – | 19 |
| **templ_19nt_U** | CCCACCCATGCCGCCCATG | – | 19 |
| **templ_19nt_C** | CCCGCCCATGCCGCCCATG | – | 19 |
| **templ_19nt_G** | AAACCCCATGCCGCCCATG | – | 19 |
| **templ_19nt_mix** | CAGTCCCATGCCGCCCATG | – | 19 |
| **5'-(TINA)-templ_31nt** | CTAGCATGAGCTCAGTCCCATGCCGCCCATG | *ortho*-TINA | 31 |
| **5'-(dual-Bio)-templ_31nt** | CTAGCATGAGCTCAGTCCCATGCCGCCCATG | dual-Bio | 31 |
| **templ_35nt** | AGCTAGCATGAGCTCAGTAACCCATGCCGCCCATG | – | 35 |
| **templ_36nt** | ATCTCACATGCTATACACTATCCCATGCCGCCCATG | – | 36 |
| **templ_poly-U** | GTCTCGGCTGGGCGTTTCCCTAAAAAAAAACCCATGCCGCCCATG | – | 45 |
| **templ_50nt** | TCCCTATCCCAGCATACCTCTGTACTGATCCTCCCCCCATGCCGCCCATG | – | 50 |
| **templ_ribosw71_A** | AGGCTCTTGGTAGAAACTCCCAAACCATATCATTAGGATTATATCTTCCC | – | 50 |
| **templ_65nt** | GGACTACTTCTAATCTGTAAGAGCAGATCCCTGGCTAGCATGAGCTCAGTCCCATGCCGCCCATG | – | 65 |
| **templ_ribosw71_B** | GGGAAGATAATCAAGAGTTTAAGGCTCTTGGTAGAAACTCCCAAACCATATCATTAGGATTATATCTTCCC | – | 71 |
| **templ_98nt** | GACATCATGAGAGACATCGCCTCTGGGCTAATAGGACTACTTCTAATCTGTAAGAGCAGATCCCTGGCTAGCATGAGCTCAGTCCCATGCCGCCCATG | – | 98 |

**Table 13. List of modified RNA or RNA-DNA oligonucleotides prepared in this study.**

| Oligonucleotide name | Sequence (5'→ 3') | 5'-end modification | Length [nt] |
|---|---|---|---|
| **16RNA_C$^{X}$** | CAUGGGCGGCAUGGG**C$^{X}$** | Cy5 | 16 |
| **16DNA-RNA_C$^{X}$** | CATGGGCGGCATGGG**C$^{X}$** | Cy5 | 16 |
| **19RNA_A$^{X}$** | CAUGGGCGGCAUGGG**A$^{X}$**GGG | FAM | 19 |
| **19RNA_U$^{X}$** | CAUGGGCGGCAUGGG**U$^{X}$**GGG | FAM | 19 |
| **19RNA_C$^{X}$** | CAUGGGCGGCAUGGG**C$^{X}$**GGG | FAM | 19 |
| **19RNA_C$^{X}$** | CAUGGGCGGCAUGGG**C$^{X}$**GGG | Cy5 | 19 |
| **19RNA_G$^{X}$** | CAUGGGCGGCAUGGG**G$^{X}$**UUU | FAM | 19 |
| **19RNA_A$^{X}$U$^{X}$C$^{X}$G$^{X}$** | CAUGGGCGGCAUGGG**A$^{X}$C$^{X}$U$^{X}$G$^{X}$** | FAM | 19 |
| **31RNA_4A$^{X}$** | CAUGGGCGGCAUGGG**A$^{X}$**CUG**A$^{X}$**GCUC**A$^{X}$**UGCU**A$^{X}$**G | FAM | 31 |
| **31RNA_4U$^{X}$** | CAUGGGCGGCAUGGGAC**U$^{X}$**GAGC**U$^{X}$**CA**U$^{X}$**GC**U$^{X}$**AG | FAM | 31 |
| **31RNA_4C$^{X}$** | CAUGGGCGGCAUGGGA**C$^{X}$**UGAG**C$^{X}$**U**C$^{X}$**AUG**C$^{X}$**UAG | FAM | 31 |
| **31RNA_4C$^{X}$** | CAUGGGCGGCAUGGGA**C$^{X}$**UGAG**C$^{X}$**U**C$^{X}$**AUG**C$^{X}$**UAG | Cy5 | 31 |
| **31RNA_4G$^{X}$** | CAUGGGCGGCAUGGGACU**G$^{X}$**A**G$^{X}$**CUCAU**G$^{X}$**CUA**G$^{X}$** | FAM | 31 |
| **31RNA_A$^{X}$U$^{X}$C$^{X}$G$^{X}$** | CAUGGGCGGCAUGGG**A$^{X}$C$^{X}$U$^{X}$G$^{X}$A$^{X}$G$^{X}$C$^{X}$U$^{X}$C$^{X}$A$^{X}$U$^{X}$G$^{X}$C$^{X}$U$^{X}$ A$^{X}$G$^{X}$** | FAM | 31 |
| **33RNA_poly-U$^{X}$** | GGG**U$^{X}$U$^{X}$U$^{X}$U$^{X}$U$^{X}$U$^{X}$U$^{X}$U$^{X}$U$^{X}$U$^{X}$U$^{X}$**AGGGAAACGCCCAGCCGAGAC | ppp | 33 |
| **35DNA_dU_dT$^{FAM}$-RNA_A$^{X}$U$^{X}$C$^{X}$G$^{X}$** | CATGGGCGGCATGGG[dU][dT-FAM] **A$^{X}$C$^{X}$U$^{X}$G$^{X}$A$^{X}$G$^{X}$C$^{X}$U$^{X}$C$^{X}$A$^{X}$U$^{X}$G$^{X}$C$^{X}$U$^{X}$A$^{X}$G$^{X}$C$^{X}$U$^{X}$** | Cy5 | 35 |
| **dT$^{FAM}$-RNA_A$^{X}$U$^{X}$C$^{X}$G$^{X}$** | [dT-FAM]**A$^{X}$C$^{X}$U$^{X}$G$^{X}$A$^{X}$G$^{X}$C$^{X}$U$^{X}$C$^{X}$A$^{X}$U$^{X}$G$^{X}$C$^{X}$U$^{X}$A$^{X}$G$^{X}$C$^{X}$U$^{X}$** | p | 19 |
| **35RNA_A$^{X}$U$^{X}$C$^{X}$G** | GGG**A$^{X}$**GG**A$^{X}$U$^{X}$C$^{X}$A$^{X}$**G**U$^{X}$A$^{X}$C$^{X}$A$^{X}$**G**A$^{X}$**GG**U$^{X}$A$^{X}$U$^{X}$**G**C$^{X}$U$^{X}$**GGG**A$^{X}$ U$^{X}$A$^{X}$**GGG**A$^{X}$** | ppp | 35 |
| **36DNA_dU-RNA_C$^{X}$** | CATGGGCGGCATGGGA[dU]AGUGUAUAG**C$^{X}$**AUGUGAGAU | FAM | 36 |
| **RNA_C$^{X}$** | AGUGUAUAG**C$^{X}$**AUGUGAGAU | p | 19 |
| **36DNA_dU-RNA_A$^{X}$U$^{X}$C$^{X}$G$^{X}$** | CATGGGCGGCATGGGA[dU]**A$^{X}$G$^{X}$U$^{X}$G$^{X}$U$^{X}$A$^{X}$U$^{X}$A$^{X}$G$^{X}$C$^{X}$A$^{X}$U ${}^{X}$G$^{X}$U$^{X}$G$^{X}$A$^{X}$G$^{X}$A$^{X}$U$^{X}$** | FAM | 36 |
| **RNA_A$^{X}$U$^{X}$C$^{X}$G$^{X}$** | p**A$^{X}$G$^{X}$U$^{X}$G$^{X}$U$^{X}$A$^{X}$U$^{X}$A$^{X}$G$^{X}$C$^{X}$A$^{X}$U$^{X}$G$^{X}$U$^{X}$G$^{X}$A$^{X}$G$^{X}$A$^{X}$U$^{X}$** | p | 19 |
| **45RNA_poly-U$^{X}$** | CAUGGGCGGCAUGGG**U$^{X}$U$^{X}$U$^{X}$U$^{X}$U$^{X}$U$^{X}$U$^{X}$U$^{X}$U$^{X}$U$^{X}$**AGGGAAACG CCCAGCCGAGAC | FAM | 45 |
| **50RNA_A$^{X}$U$^{X}$C$^{X}$G** | CAUGGGCGGCAUGGGGGG**A$^{X}$**GG**A$^{X}$U$^{X}$C$^{X}$A$^{X}$**G**U$^{X}$A$^{X}$C$^{X}$A$^{X}$**G**A$^{X}$**G G**U$^{X}$A$^{X}$U$^{X}$**G**C$^{X}$U$^{X}$**GGG**A$^{X}$U$^{X}$A$^{X}$**GGG**A$^{X}$** | FAM | 50 |
| **65RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$** | CAUGGGCGGCAUGGG**A$^{X}$C$^{X}$U$^{X}$G$^{X}$A$^{X}$G$^{X}$C$^{X}$U$^{X}$C$^{X}$A$^{X}$U$^{X}$G$^{X}$C$^{X}$U$^{X}$A ${}^{X}$G$^{X}$C$^{X}$C$^{X}$A$^{X}$G$^{X}$G$^{X}$G$^{X}$A$^{X}$U$^{X}$C$^{X}$U$^{X}$G$^{X}$C$^{X}$U$^{X}$C$^{X}$U$^{X}$U$^{X}$A$^{X}$C$^{X}$A$^{X}$G$^{X}$A$^{X}$U$^{X}$ U$^{X}$A$^{X}$G$^{X}$A$^{X}$A$^{X}$G$^{X}$U$^{X}$A$^{X}$G$^{X}$U$^{X}$C$^{X}$C$^{X}$** | FAM | 65 |
| **98RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$** | CAUGGGCGGCAUGGG**A$^{X}$C$^{X}$U$^{X}$G$^{X}$A$^{X}$G$^{X}$C$^{X}$U$^{X}$C$^{X}$A$^{X}$U$^{X}$G$^{X}$C$^{X}$U$^{X}$A ${}^{X}$G$^{X}$C$^{X}$C$^{X}$A$^{X}$G$^{X}$G$^{X}$G$^{X}$A$^{X}$U$^{X}$C$^{X}$U$^{X}$G$^{X}$C$^{X}$U$^{X}$C$^{X}$U$^{X}$U$^{X}$A$^{X}$C$^{X}$A$^{X}$G$^{X}$A$^{X}$U$^{X}$ U$^{X}$A$^{X}$G$^{X}$A$^{X}$A$^{X}$G$^{X}$U$^{X}$A$^{X}$G$^{X}$U$^{X}$C$^{X}$C$^{X}$U$^{X}$A$^{X}$U$^{X}$U$^{X}$A$^{X}$G$^{X}$C$^{X}$C$^{X}$C$^{X}$A$^{X}$G$^{X}$A ${}^{X}$G$^{X}$G$^{X}$C$^{X}$G$^{X}$A$^{X}$U$^{X}$G$^{X}$U$^{X}$C$^{X}$U$^{X}$C$^{X}$U$^{X}$C$^{X}$A$^{X}$U$^{X}$G$^{X}$A$^{X}$U$^{X}$G$^{X}$U$^{X}$C$^{X}$** | FAM | 98 |

| | | | |
|---|---|---|---|
| **FAM-Cy5-Cy3-riboswitch_SNI-1** | GGGAAGAUAUAAUCCUAAUGAUA**U$^{X}$** | FAM | 24 |
| **FAM-Cy5-Cy3-riboswitch_PEX-1** | GGGAAGAUAUAAUCCUAAUGAUA**U$^{X}$**GGUUUGGGAGUUUCU ACCAAGAGCCU | FAM | 50 |
| **FAM-Cy5-Cy3-riboswitch_SNI-2** | GGGAAGAUAUAAUCCUAAUGAUA**U$^{X}$**GGUUUGGGAGUUUCU ACCAAGAGCCU**U$^{X}$** | FAM | 51 |
| **FAM-Cy5-Cy3-riboswitch** | GGGAAGAUAUAAUCCUAAUGAUA**U$^{X}$**GGUUUGGGAGUUUCU ACCAAGAGCCU**U$^{X}$**AAACUCUUGAUUAUCUUCCC | FAM | 71 |
| **Cy5-Cy3-riboswitch** | GGGAAGAUAUAAUCCUAAUGAUA**U$^{X}$**GGUUUGGGAGUUUCU ACCAAGAGCCU**U$^{X}$**AAACUCUUGAUUAUCUUCCC | – | 71 |

p = Monophosphate residue; ppp = Triphosphate residue; X = Any nucleobase modification; Modified nucleotides are highlighted in bold. Primer region is underlined.

**Table 14. MS data of prepared RNA oligonucleotides in this study.**

| Analysed RNA | Calculated mass | Measured mass (prepared by TGK pol.) | Measured mass (prepared by SFM4-3 pol.) |
|---|---|---|---|
| **19RNA_A$^{E}$** | 6810.8 Da | 6813.4 Da | 6811.9 Da |
| **19RNA_A$^{Pent}$** | 6852.9 Da | 6853.9 Da | 6853.9 Da |
| **19RNA_A$^{Ph}$** | 6862.9 Da | 6864.1 Da | 6864.5 Da |
| **19RNA_A$^{CA}$** | 6915.6 Da | 6882.7 Da$^{*}$ | 6883.4 Da$^{*}$ |
| **19RNA_A$^{FT}$** | 6896.9 Da | 6901.4 Da | 6899.4 Da |
| **19RNA_A$^{CA}$** | 6915.6 Da | 6917.7 Da$^{\Delta}$ | N/A |
| **19RNA_A$^{CA}$** | 6915.6 Da | 6919.0 Da$^{\bullet}$ | 6919.5 Da$^{\bullet}$ |
| **19RNA_U$^{E}$** | 6788.7 Da | 6789.90 Da | 6791.0 Da |
| **19RNA_U$^{Pent}$** | 6830.8 Da | 6831.7 Da | 6831.2 Da |
| **19RNA_U$^{Ph}$** | 6840.8 Da | 6841.9 Da | 6841.8 Da |
| **19RNA_U$^{Bio}$** | 7244.3 Da | 7248.3 Da | 7245.2 Da |
| **19RNA_U$^{Dig}$** | 7363.5 Da | 7366.8 Da | 7364.4 Da |
| **19RNA_U$^{CA}$** | 6894.3 Da | 6895.5 Da | 6896.9 Da |
| **19RNA_U$^{FT}$** | 6874.8 Da | 6876.3 Da | 6879.5 Da |
| **19RNA_C$^{Me}$** | 6240.8 Da | 6242.0 Da | N/A |
| **19RNA_C$^{E}$** | 6787.8 Da | 6789.7 Da | 6788.8 Da |
| **19RNA_C$^{Pent}$** | 6829.9 Da | 6832.2 Da | 6831.1 Da |
| **19RNA_C$^{Ph}$** | 6839.9 Da | 6844.0 Da | 6841.0 Da |
| **19RNA_C$^{mBdp}$** | 7164.9 Da | 7165.4 Da | 7166.9 Da |
| **19RNA_C$^{Cy5}$** | 7455.6 Da | 7460.7 Da | 7459.0 Da |
| **19RNA_C$^{CA}$** | 6893.3 Da | 6860.0 Da$^{*}$ | 6859.9 Da$^{*}$ |
| **19RNA_C$^{FT}$** | 6873.9 Da | 6875.2 Da | 6875.7 Da |
| **19RNA_C$^{CA}$** | 6893.3 Da | 6894.3 Da$^{\Delta}$ | N/A |

| | | | |
|---|---|---|---|
| **19RNA_C$^{CA}$** | 6893.3 Da | 6894.6 Da• | 6894.6 Da• |
| **19RNA_G$^{E}$** | 6709.7 Da | 6712.1 Da | 6715.6 Da |
| **19RNA_G$^{Pent}$** | 6751.8 Da | 6754.1 Da | 6752.6 Da |
| **19RNA_G$^{Ph}$** | 6761.8 Da | 6764.8 Da | 6763.7 Da |
| **31RNA_4A$^{E}$** | 10658.1 Da | 10659.5 Da | 10659.4 Da |
| **31RNA_4A$^{Pent}$** | 10826.4 Da | 10829.3 Da | 10827.6 Da |
| **31RNA_4A$^{Ph}$** | 10866.4 Da | 10869.3 Da | 10869.6 Da |
| **31RNA_4U$^{E}$** | 10662.1 Da | 10663.4 Da | 10663.8 Da |
| **31RNA_4U$^{Pent}$** | 10830.4 Da | 10831.9 Da | 10832.5 Da |
| **31RNA_4U$^{Ph}$** | 10870.4 Da | 10873.3 Da | 10871.1 Da |
| **31RNA_4U$^{Bio}$** | 12484.6 Da | 12485.8 Da | 12485.4 Da |
| **31RNA_4U$^{Dig}$** | 12961.2 Da | 12963.1 Da | 12962.3 Da |
| **31RNA_4C$^{Me}$** | 10085.1 Da | 10086.2 Da | N/A |
| **31RNA_4C$^{E}$** | 10662.1 Da | 10662.9 Da | 10663.5 Da |
| **31RNA_4C$^{Pent}$** | 10830.4 Da | 10831.2 Da | 10831.4 Da |
| **31RNA_4C$^{Ph}$** | 10870.4 Da | 10871.0 Da | 10872.6 Da |
| **31RNA_4C$^{mBdp}$** | 11646.1 Da | 11990.7 Da֍ | 11649.7 Da |
| **31RNA_4G$^{E}$** | 10658.16 Da | 10659.2 Da | 10657.4 Da |
| **31RNA_4G$^{Pent}$** | 10826.5 Da | 10831.1 Da | 10830.8 Da |
| **31RNA_4G$^{Ph}$** | 10866.4 Da | 10869.7 Da | 10869.6 Da |
| **19RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$** | 6815.6 Da | 6815.0 Da | N/A |
| **31RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$** | 12604.6 Da | 12604.0 Da | N/A |
| **65RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$** | 29179.8 Da | 29531.0 Da□ | N/A |
| **98RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$** | 45836.3 Da | 45804.0 Da | N/A |
| **RNA_C$^{Cy5}$** | 6898.6 Da | 6898.1 Da | N/A |
| **RNA_A$^{E}$U$^{Bio}$C$^{Cy5}$G$^{Pent}$** | 10305.3 Da | 10304.6 Da | N/A |
| **dT$^{FAM}$-RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$** | 9719.4 Da | 9719.4 Da | N/A |
| **FAM-Cy5-Cy3-riboswitch_SNI-1** | 8940.5 Da | 8939.5 Da | N/A |
| **FAM-Cy5-Cy3-riboswitch_PEX-1** | 17323.5 Da | 17321.2 Da | N/A |
| **FAM-Cy5-Cy3-riboswitch_SNI-2** | 18297.4 Da | 18294.4 Da | N/A |
| **FAM-Cy5-Cy3-riboswitch_PEX-2** | 24569.1 Da | 24564.4 Da | N/A |
| **Cy5-Cy3-riboswitch** | 24032.1 Da | 24027.7 Da | N/A |

N/A = Not assessed; * = Self-cross-linked product; ● = Reduced temperature; Δ = Reduced reaction time; ֍ = Product + rGMP; □ = Product + **rA$^{E}$MP**.

### 3.2.5 Comparison of T7 RNAP and thermostable engineered DNA polymerases in synthesis of base-modified RNA

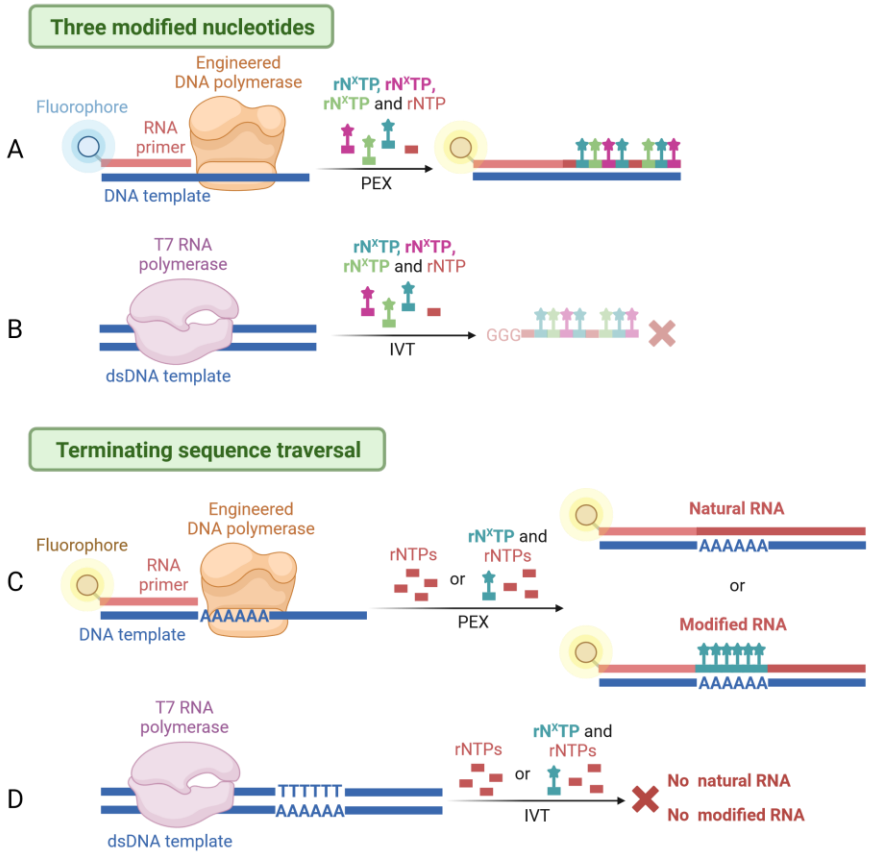



**Figure 67. Incorporation of three variously modified rN$^{X}$TP by either PEX (A) of IVT (B). Incorporation of rU$^{Bio}$TP into homopolymeric poly-U sequence by PEX (C) or IVT (D).**

To benchmark engineered polymerases against T7 RNAP I designed challenging either ssDNA template (**templ_50nt**, Table 12) for PEX or dsDNA template (**ds-templ_52nt**, Table 16) for T7 IVT, where the synthesised part of RNA is identical (Figure 67 – A, B). Then I tested simultaneous incorporation of several different modified nucleotides. I used either a mixture of three phenyl-modified (**rA$^{Ph}$TP**, **rU$^{Ph}$TP**, **rC$^{Ph}$TP**, Mix-1), three distinct modified (**rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP**, Mix-2) or sterically demanding (**rA$^{Ph}$TP**, **rU$^{Bio}$TP**, **rC$^{mBdp}$TP**, Mix-3) nucleotides, all of them in combination with natural rGTP. Only in IVT a mixture of rGTP and [α-$^{32}$P]-rGTP was used for RNA labelling. As it was expected, the T7 RNAP substantially failed in synthesis of modified RNA with most modifications (Mix-1, Mix-3), while some full-length product was observed for Mix-2 (Figure 68 – A). SFM4-3 synthesised RNA with Mix-1 and Mix-2 with only modest yield and entirely did

not accept Mix-3 (Figure 69 – B). On the other hand, TGK polymerase was successful in accepting all **rN$^{X}$TP** mixtures giving high yield of full-length products in each case (Figure 69 – A).

Additionally, the efficiency of the mutant polymerases was compared to T7 RNAP in synthesis of homopolymeric termination-prone poly-U sequence (Figure 67 – C, D). I used either ssDNA template (**templ_poly-U**, Table 12) in PEX or dsDNA template (**ds-templ_poly-U**, Table 16) for IVT, containing internal poly-U coding sequence. The poly-U part was in both cases modified by bulky **rU$^{Bio}$** nucleotides. Only, TGK polymerase delivered significant amount of modified ssRNA, whereas both SFM4-3 (Figure 70) and T7 RNAP (Figure 68 – B) completely failed in synthesis of modified RNA and did not succeed even in production of natural congener.

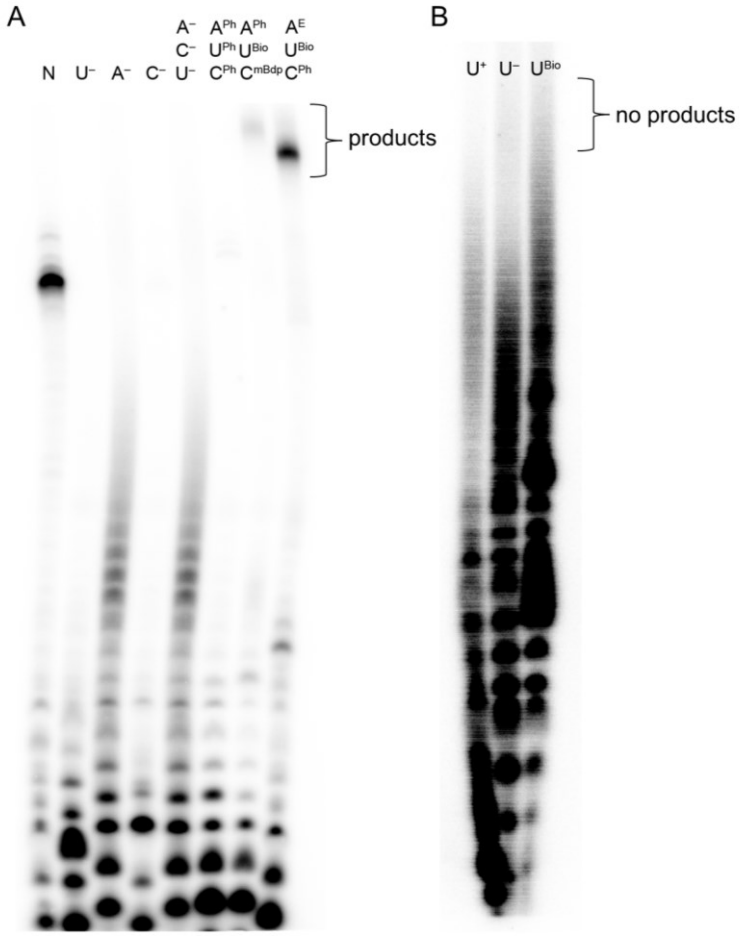



**Figure 68. Phosphor imaging of 12.5% dPAGE analysis of IVT reaction. Figure A – IVT with various mixtures of rN$^{X}$TPs, Figure B – IVT with rU$^{Bio}$TP and poly-U template. Gel legend – A: (N) positive control, all natural rNTPs; (U$^{-}$) negative control, mixture of rATP, rCTP, rGTP/[α-$^{32}$P]-rGTP and $H_2O$; (A$^{-}$) negative control, mixture of rUTP, rCTP, rGTP/[α-$^{32}$P]-rGTP and $H_2O$; (C$^{-}$) negative control, mixture of rATP, rUTP, rGTP/[α-$^{32}$P]-rGTP and $H_2O$; (A$^{-}$, U$^{-}$, C$^{-}$) negative control, mixture of rGTP/[α-$^{32}$P]-rGTP and $H_2O$; (A$^{Ph}$, U$^{Ph}$, C$^{Ph}$) modification, mixture of rA$^{Ph}$TP, rU$^{Ph}$TP, rC$^{Ph}$TP (Mix-1) and rGTP/[α-$^{32}$P]-rGTP; (A$^{Ph}$, U$^{Bio}$, C$^{mBdp}$) modification, mixture of rA$^{Ph}$TP, rU$^{Bio}$TP, rC$^{mBdp}$TP (Mix-3) and rGTP/[α-$^{32}$P]-rGTP; (A$^{E}$, U$^{Bio}$, C$^{Ph}$) modification, mixture of rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Ph}$TP (Mix-2) and rGTP/[α-$^{32}$P]-rGTP. Gel legend – B: (U$^{+}$) positive control, all natural rNTPs; (U$^{-}$) negative control, mixture of rATP, rCTP, rGTP/[α-$^{32}$P]-rGTP and $H_2O$; (U$^{Bio}$) modification, mixture of rATP, rU$^{Bio}$TP, rCTP, rGTP/[α-$^{32}$P]-rGTP.**

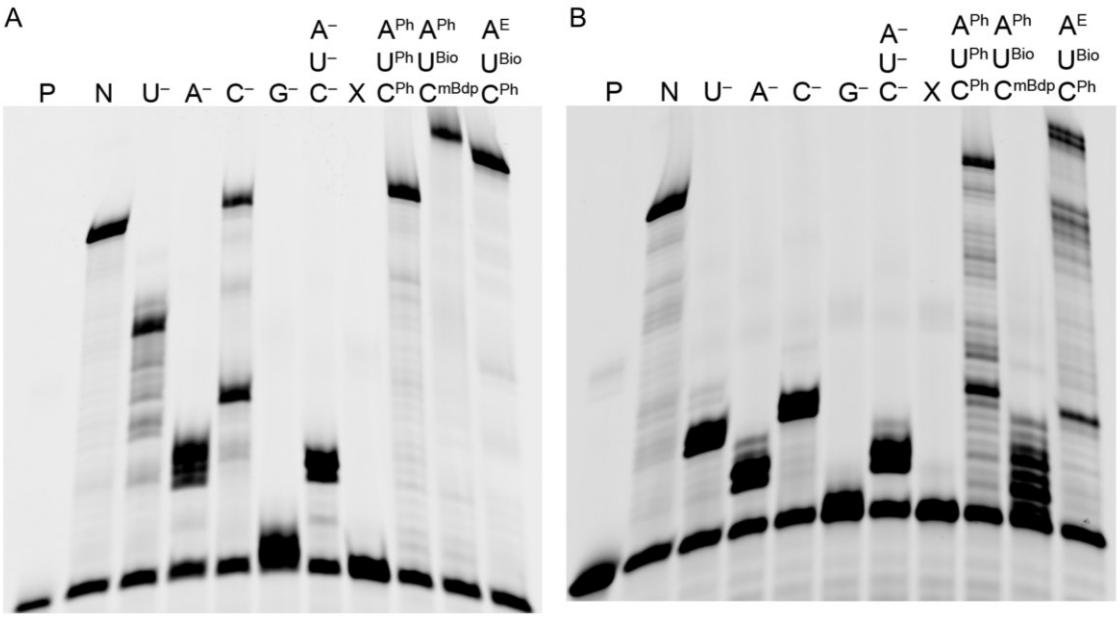



**Figure 69. Cy5 scan of 12.5% dPAGE analysis of PEX reaction with various mixtures of rN$^{X}$TPs. Figure A – PEX with TGK polymerase, Figure B – PEX with SFM4-3 polymerase. Gel legend: (P) RNA primer; (N) positive control, all natural rNTPs; (U$^{-}$) negative control, mixture of rATP, rCTP, rGTP and $H_2O$; (A$^{-}$) negative control, mixture of rUTP, rCTP, rGTP and $H_2O$; (C$^{-}$) negative control, mixture of rATP, rUTP, rGTP and $H_2O$; (G$^{-}$) negative control, mixture of rATP, rUTP, rCTP and $H_2O$; (A$^{-}$, U$^{-}$, C$^{-}$) negative control, mixture of GTP and $H_2O$; (X) negative control, no rNTPs; (A$^{Ph}$, U$^{Ph}$, C$^{Ph}$) modification, mixture of rA$^{Ph}$TP, rU$^{Ph}$TP, rC$^{Ph}$TP (Mix-1) and GTP; (A$^{Ph}$, U$^{Bio}$, C$^{mBdp}$) modification, mixture of rA$^{Ph}$TP, rU$^{Bio}$TP, rC$^{mBdp}$TP (Mix-3) and GTP; (A$^{E}$, U$^{Bio}$, C$^{Ph}$) modification, mixture of rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Ph}$TP (Mix-2) and GTP.**

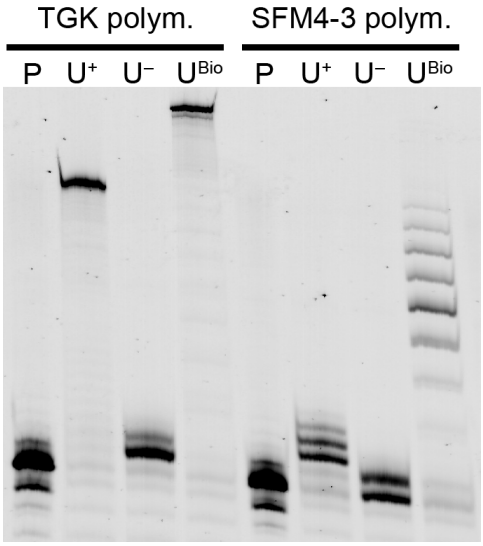



**Figure 70. FAM scan of 12.5% dPAGE analysis with rU$^{Bio}$TP and poly-U template. Gel legend: (P) RNA primer; (U+) positive control, all natural rNTPs; (U–) negative control, mixture of rATP, rCTP, rGTP and $H_2O$; (U$^{Bio}$) modification, mixture of rATP, rU$^{Bio}$TP, rCTP, rGTP.**

To assess the enzymatic activity of the engineered DNA polymerases comprehensively, I performed detailed kinetic studies. Therefore, 15 nt long either RNA (**Cy5-RNA-prim_15nt**, Table

11) or DNA primer (**Cy5-DNA-prim_15nt**, Table) complementary to 16 nt DNA template (**templ_16nt**, Table 12) was used for SNI of **rC$^{mBdp}$TP** at different time periods. Interestingly, TGK polymerase provided higher conversion than SFM4-3 for both primer types in virtually all time points as revealed by dPAGE and subsequent ImageJ quantification (**Figure 71**). Analogously, with same RNA primer (**Cy5-RNA-prim_15nt**) and ssDNA template (**templ_16nt**), I monitored SNI of **rC$^{mBdp}$TP** with varying concentrations of SFM4-3 or TGK polymerase at fixed time. Results gained by dPAGE analysis further supported higher activity of TGK over SFM4-3 polymerase (**Figure 71**).

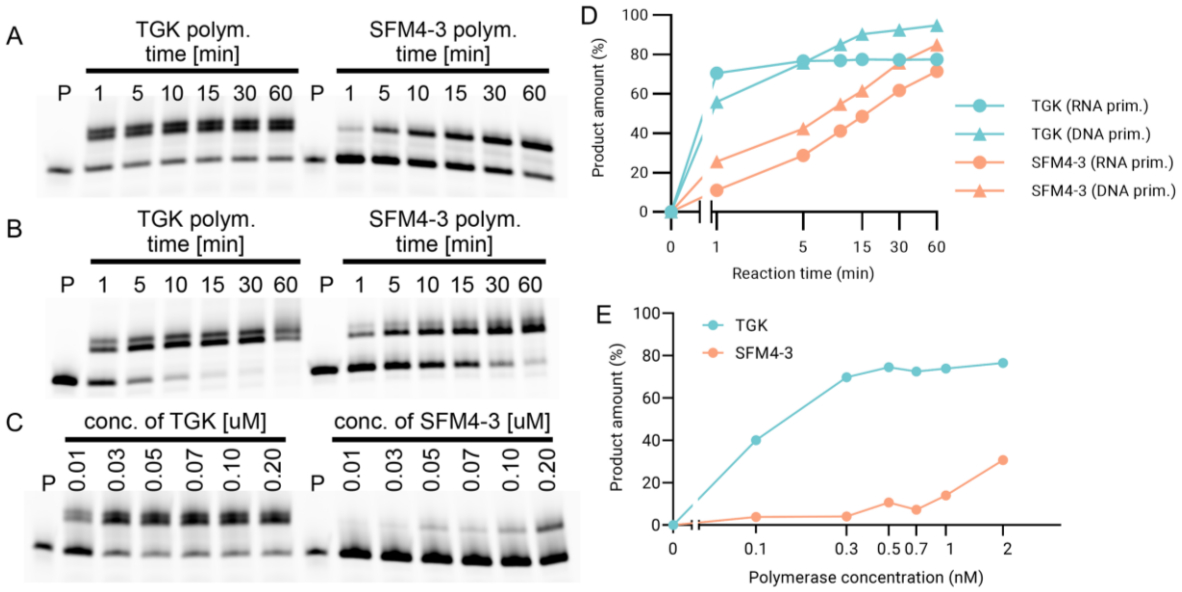



**Figure 71. Activity profiling of engineered thermostable DNA polymerases in synthesis of 16RNA_C$^{mBdp}$ (with RNA primer) or 16DNA-RNA_C$^{mBdp}$ (with DNA primer). Figures A, B, C – Cy5 scan of 22.5% dPAGE analysis of SNI of rC$^{mBdp}$TP. Figure A – SNI with RNA primer at different time periods, Figure B – SNI with DNA primer at different time periods, Figure C – SNI with RNA primer and varying polymerase concentrations. Figure D – Time course insertion of single nucleotide was monitored for SFM4-3 and TGK polymerase as percentage of primer conversion to product, Figure E – Efficiency of SNI with rC$^{mBdp}$TP and various concentrations of either TGK or SFM4-3 polymerase represented by percentage of primer conversion to product.**

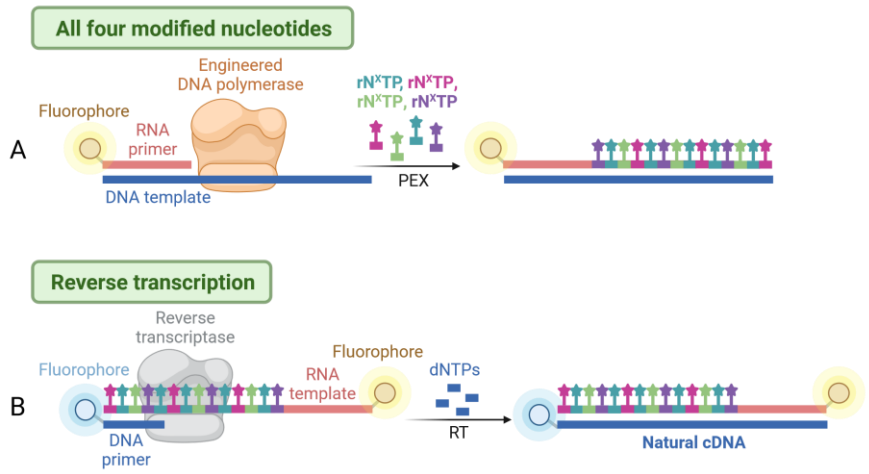



**Figure 72. A – PEX with four variously modified rN$^{X}$TP. B – Reverse transcription of highly modified RNA template into complementary DNA strand.**

Next, I tested the ability of engineered polymerases to incorporate a mixture of all four different base-modified **rN$^{X}$TPs** (**rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP**, **rG$^{Pent}$TP**) in PEX (Figure 72 – A). I tested variously long ssDNA templates (19 nt – **templ_19nt_mix**, 31 nt – **5'-(TINA)-templ_31nt**, 65 nt – **templ_65nt** or even 98 nt long – **templ_98nt**, Table 12) encoding for incorporation of up to 83 consecutive modified nucleotides displaying four different functional groups. Successful extension of RNA primer (**FAM-RNA-prim_15nt**, Table 11) with all tested ssDNA templates was achieved only by TGK polymerase. Formation of full-length hypermodified RNA products (**19RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$**, **31RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$**, **65RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$**, **98RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$**, Table 13) was confirmed by dPAGE (Figure 73, Figure 140 – Figure 143). Template digestion by DNase provided desired ssRNA probes, that were corroborated by ESI-MS (Figure 74, Table 14). On the other hand, SFM4-3 polymerase was inefficient in synthesis of shorter hypermodified products and completely failed in PEX of longer sequences (Figure 140 – Figure 143).

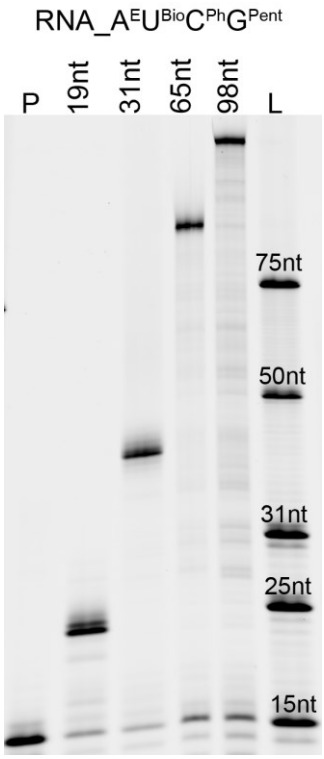



**Figure 73. FAM scan of 12.5% dPAGE analysis of PEX reaction with rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Ph}$TP, rG$^{Pent}$TP and variously long ssDNA templates. Gel legend: (P) RNA primer; (19nt) modified RNA product 19RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$; (31nt) modified RNA product 31RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$; (65nt) modified RNA product 65RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$; (98nt) modified RNA product 98RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$; (L) RNA ladder composed of FAM-labelled RNA oligonucleotides of indicated length.**

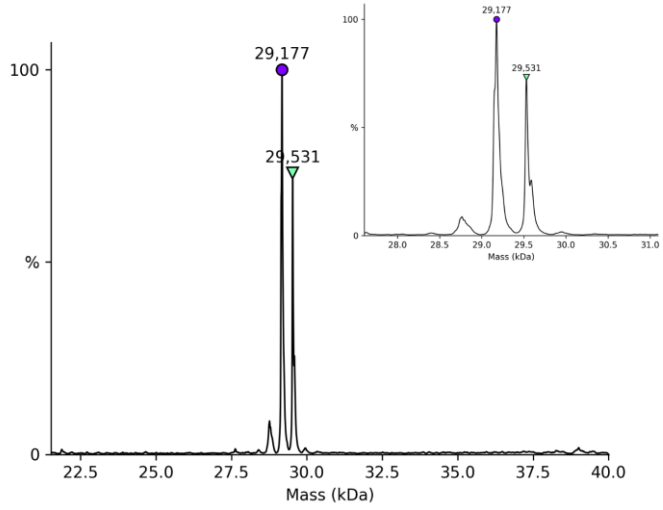



**Figure 74. LC-ESI-MS analysis of 65RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$. Deconvoluted spectrum; calculated mass: 29179.8 Da; found mass: 29177.0 Da (product); found mass: 29531.0 (product + rA$^{E}$MP).**

To demonstrate retrieval of cDNA from all-nucleobase modified RNA (**FAM-98RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$**) prepared by TGK, I carried out reverse transcription (Figure 72 – B). The Cy5-labelled DNA primer (**Cy5-DNA-prim_20nt**) complementary to FAM-labelled 98 nt long modified RNA template (**FAM-98RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$**) was successfully prolonged by SuperScript IV reverse transcriptase in presence of natural 2'-deoxyribonucleoside triphosphates (dNTPs). A band of full-length 98 nt long Cy5-labelled cDNA (**98cDNA**) was observed on dPAGE (Figure 75).

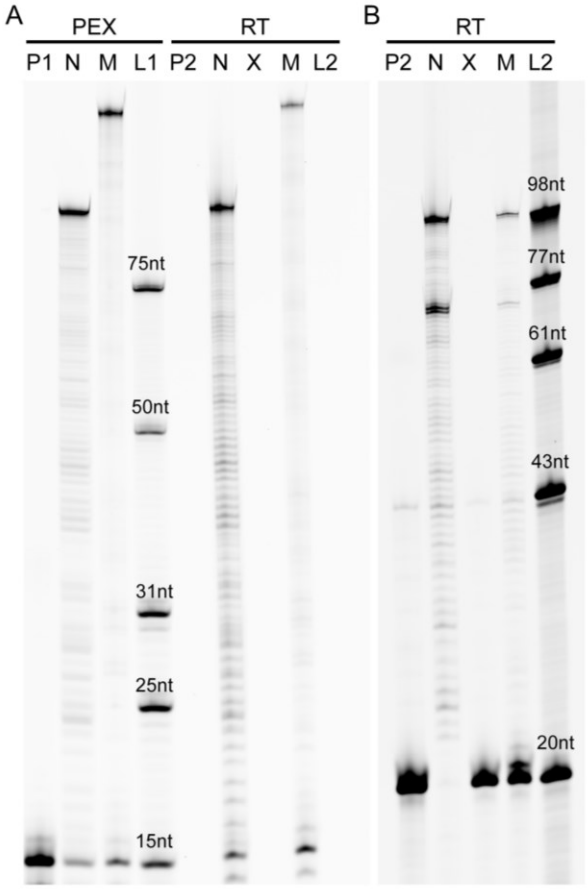



**Figure 75. 12.5% dPAGE analysis. Gel legend: (P1) RNA primer; (N) natural RNA, FAM-98RNA_AUCG; (M) modified RNA, FAM-98RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$; (L1) RNA ladder composed of FAM-labelled RNA oligonucleotides of indicated length; (P2) DNA primer; (N) cDNA from natural RNA, 98cDNA; (X) negative control, no RNA; (M) cDNA from modified RNA, 98cDNA; (L2) DNA ladder composed of Cy5-labelled DNA oligonucleotides of indicated length. A – FAM scan (visualisation of RNA ladder, RNA primer and RNA templates); B – Cy5 scan (visualisation of DNA ladder, DNA primer and cDNA products).**

Herein I elaborated a novel route to nucleobase-modified RNA that is based on previously reported[88,91,92] engineered thermophilic primer-dependent DNA polymerases. While by means of IVT with T7 RNAP it has been possible to install smaller, dispersed modifications along whole RNA length[43,101,111,143], I demonstrated superior performance of engineered thermophilic polymerases in several aspects. TGK polymerase, and to less extent SFM4-3 polymerase, incorporated bulky and

hydrophobic groups at high density, as well as highly reactive CA or aldehyde (FT) moieties. I proved that TGK polymerase, but not SFM4-3 can traverse precarious regions, such as U-rich homopolymers, where T7 RNAP terminates. Facile incorporation of modified **rG$^{X}$TP**s, which is generally inefficient with T7 RNAP and requires unmodified initiator nucleotide[43], was achieved even for sterically demanding **rG$^{Ph}$TP** modification. Previously elusive, high density all-four-nucleobase hypermodified RNA was synthesised by TGK polymerase in up to 98 nt length with respectable efficiency.

These findings are are in accordance with previous structural research on cognate thermophilic DNA polymerases. It revealed that B-family class, where TGK polymerase belongs to, is well adapted to accommodating major groove modifications attached to nucleobases[71]. On the other hand, SFM4-3 derived from A-family Taq polymerase, had substantially narrower scope, limited to smaller modifications in less dense arrangement. Since the performance of TGK polymerase generally surpassed SFM4-3, I decided to further proceed only with TGK polymerase.

**Table 15. List of DNA oligonucleotides used as primers in this study. For structures of 5'-end modifications see Figure 116.**

| DNA oligonucleotide name | Sequence (5'→3') | 5'-end or internal modification | Length [nt] |
|---|---|---|---|
| **FAM-DNA-prim_15nt** | CATGGGCGGCATGGG | 5'-FAM | 15 |
| **Cy5-DNA-prim_15nt** | CATGGGCGGCATGGG | 5'-Cy5 | 15 |
| **FAM-DNA-prim_dU** | CATGGGCGGCATGGGA[dU] | 5'-FAM; internal [dU] | 17 |
| **Cy5-DNA-prim_dU_dT-FAM** | CATGGGCGGCATGGG[dU][dT] | 5'-Cy5; internal [dU][dT-FAM] | 17 |
| **Cy5-DNA-prim_20nt** | GACATCATGAGAGACATCGC | 5'-Cy5 | 20 |
| **DNA-REV-prim_21nt** | ACACCAAAATACATTCACACA | – | 21 |
| **DNA-REV-prim_24nt** | CTCAGACCTTCATACGGGATGATG | – | 24 |
| **DNA-FOR-prim_26nt** | AATATGATGATAATATGGTTATAATT | – | 26 |
| **DNA-REV-prim_26nt** | ACCATGGTTGTGGCCATATTATCATC | – | 26 |
| **5'-(dual-Bio)-DNA-FOR-prim_29nt** | CCCTTTCGTCTAATACGACTCACTATAGG | 5'-dual-Bio | 29 |
| **DNA-REV-prim_48nt** | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTACGCCAGAATGCGTTC | – | 48 |

**Table 16. List of DNA oligonucleotides for generation of dsDNA templates used in this study.**

| DNA oligonucleotide name | Sequence (5'→ 3') | 5'-end modification | Length [nt] |
|---|---|---|---|
| **templ_poly-U-s** | *TAATACGACTCACTATA*GGGTTTTTTTTTAGGGAAA CGCCCAGCCGAGAC | – | 50 |
| **templ_poly-U-as** | [mG][mU]CTCGGCTGGGCGTTTCCCTAAAAAAAAA CCC*TATAGTGAGTCGTATTA* | 2'-*O*-Me-G, 2'-*O*-Me-U | 50 |
| **templ_52bp-s** | *TAATACGACTCACTATA*GGGAGGATCAGTACAGAGG TATGCTGGGATAGGGA | – | 52 |
| **templ_52bp-as** | [mU][mC]CCTATCCCAGCATACCTCTGTACTGATC CTCCC*TATAGTGAGTCGTATTA* | 2'-*O*-Me-U 2'-*O*-Me-C | 52 |

s = Sense strand used for generation of double stranded DNA template. as = Anti-sense strand used for generation of double stranded DNA template. T7 RNAP promoter region is highlighted in italics.

### 3.2.6 Synthesis of all-nucleobase modified RNA through primer degradation

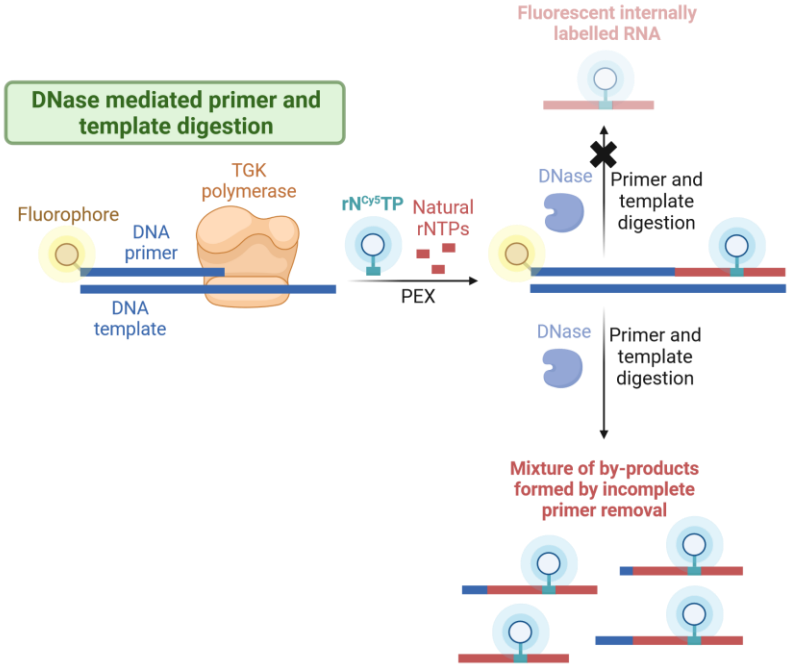



**Figure 76. One-step DNA primer and template removal strategy. Upon PEX with a mixture of rATP, rUTP, rC$^{Cy5}$TP, rGTP the DNA primer and template are removed by usage of TurboDNase. Regrettably, cleavage of the DNA primer is incomplete, leaving one or several deoxyribonucleotides behind at the 5'-end of the generated RNA fragment.**

While I have previously demonstrated the feasibility of elongating an RNA primer using a mixture of all four unnatural ribonucleoside triphosphates (**rN$^{X}$TPs**), the enzymatic synthesis of hypermodified RNA oligonucleotides, wherein all nucleobases are fully modified, has remained an unmet and unexplored challenge.

To achieve this, my first attempt was to generate hypermodified RNA with usage of DNA primer and template in PEX with **rN$^{X}$TP**s, followed by one-step degradation of both DNA segments

with TurboDNase (Figure 76). Thus, I carried out PEX with FAM-labelled ssDNA primer (**FAM-DNA-prim_15nt**, Table 15) complementary to 36 nt long ssDNA template (**templ_36nt**, Table 12) and mix of rATP, rUTP, **rC$^{Cy5}$TP**, rGTP to visualise the RNA segment after DNA primer removal. Positive control (natural RNA) was prepared by reaction of all four natural rNTPs, while in the negative control a mixture of only rATP, rUTP, rGTP was used. Unfortunately, the usage of TurboDNase for DNA primer degradation led to formation of multiple products, even with high DNase concentration or prolonged reaction time, as revealed by dPAGE (Figure 77).

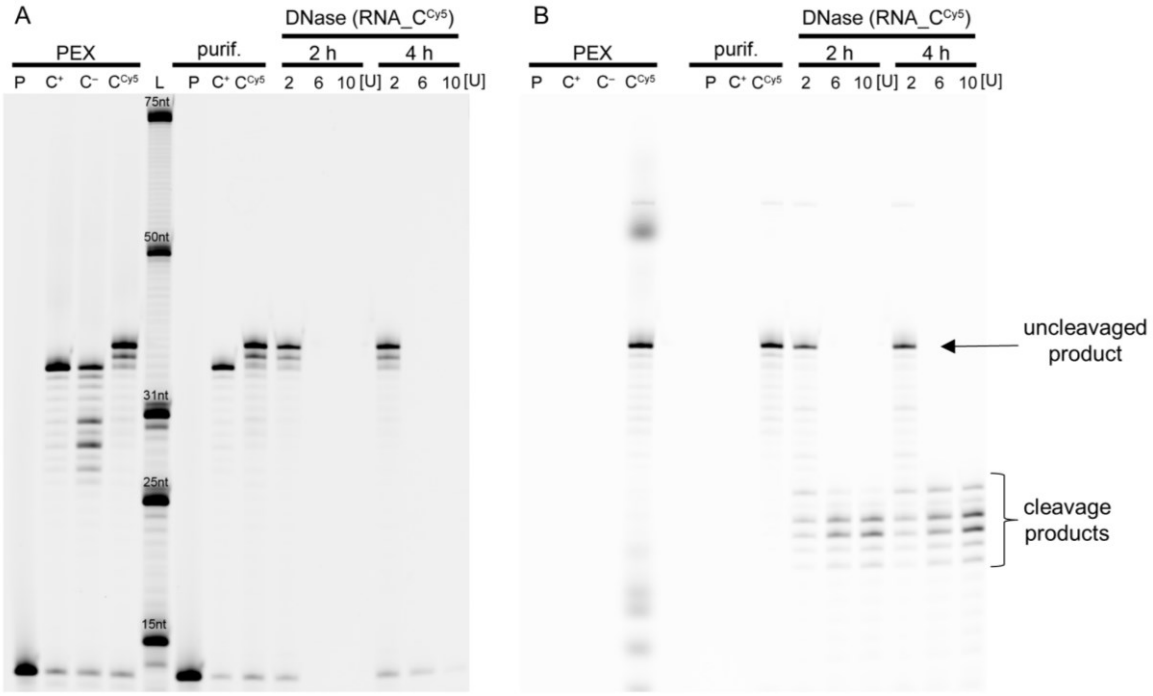



**Figure 77. 12.5% dPAGE analysis. Gel legend: (P) DNA primer; (C$^+$) positive control, all natural rNTPs; (C$^-$) negative control, mixture of rATP, rUTP, rGTP and $H_2O$; (C$^{Cy5}$) modification, mixture of rATP, rUTP, rC$^{Cy5}$TP, rGTP; (L) RNA ladder composed of FAM-labelled RNA oligonucleotides of indicated length. TurboDNase cleavage of modified 36DNA-RNA_C$^{Cy5}$ with 2 U, 6 U, 10 U of the enzyme for 2 h or 4 h incubation time. A – FAM scan; B – Cy5 scan (visualisation of Cy5-modification).**

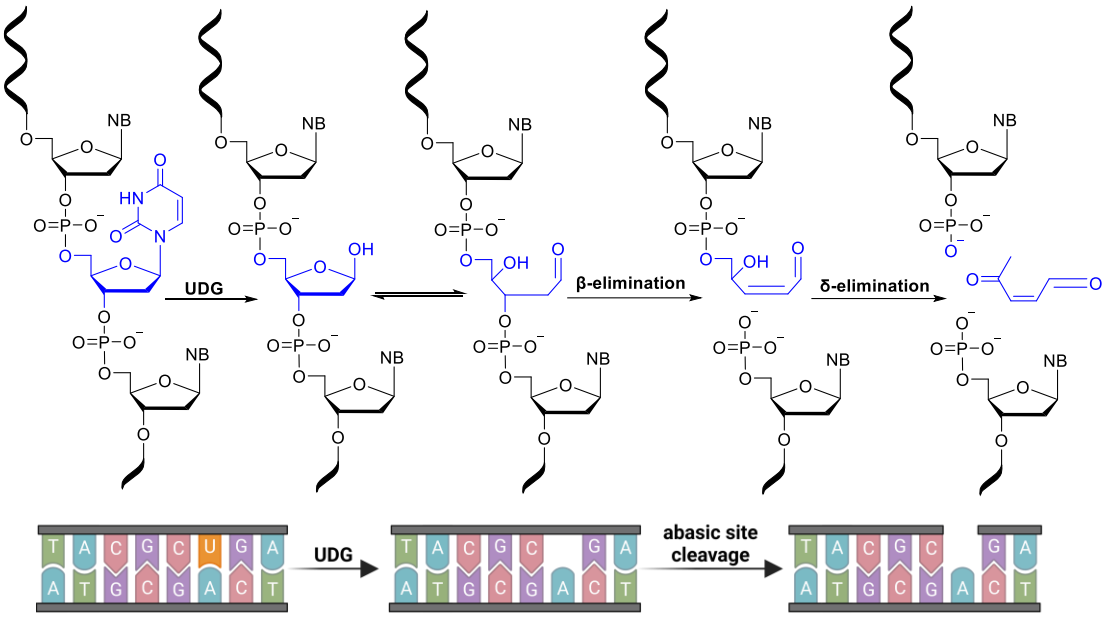



**Figure 78. Mechanism of UDG-mediated 2'-deoxyuridine cleavage.**

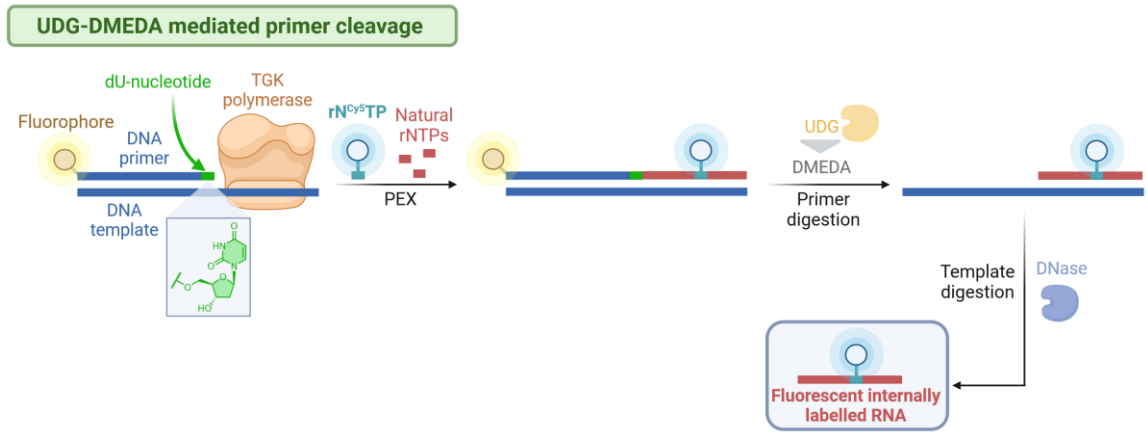



**Figure 79. Pilot DNA primer removal strategy. 5'-FAM labelled DNA primer contains single 2'-dU modification at 3'-end terminus. Upon PEX with a mixture of rATP, rUTP, rC$^{Cy5}$TP, rGTP, dU is cleaved by UDG to generate abasic site, that is subsequently eliminated by DMEDA under mild basic conditions. DNA template and rest of the primer are degraded with DNase to produce internal fluorescently labelled RNA.**

Therefore, I slightly modified this protocol inspired by naturally occurring repair processes with uracil-DNA glycosylase (UDG)[277]. This enzyme specifically recognises and facilitates the removal of the uracil nucleobase by cleaving the *N*-glycosidic bond in uracil-containing single or double-stranded DNAs, leading to the formation of an abasic site. The resulting unstable ribose, lacking the glycosidic bond, promptly transforms into its open-chain form. Subsequent cleavage, either through specific endonucleases or chemical treatment under basic conditions, leads to the breaking of the phosphodiester bond and the creation of an $\alpha$, $\beta$-unsaturated aldehyde. The final

elimination process results in the formation of a single nucleotide gap in the cleaved DNA strand (Figure 78). First, to test this approach I used ssDNA primer containing single dU modification at 3'-end (**FAM-DNA-prim_dU**, Table 15) and ssDNA template (**templ_36nt**, Table 12) for PEX with a mixture of rATP, rUTP, **rC$^{Cy5}$TP**, rGTP (Figure 79). Control reactions were performed under same condition with either natural rNTPs or by omitting rCTP/**rC$^{Cy5}$TP**. Generated modified PEX product (**36DNA_dU-RNA_C$^{Cy5}$**) was treated in the first step by UDG to form abasic site, followed by cleavage under mild basic conditions using *N,N'*-dimethylethylenediamine (DMEDA) in the second step[278]. Final DNA template degradation was performed with TurboDNase (Figure 80). Successful formation of the desired cleavage product (**RNA_C$^{Cy5}$**) was confirmed by dPAGE (Figure 80) and MS analysis (Figure 81).

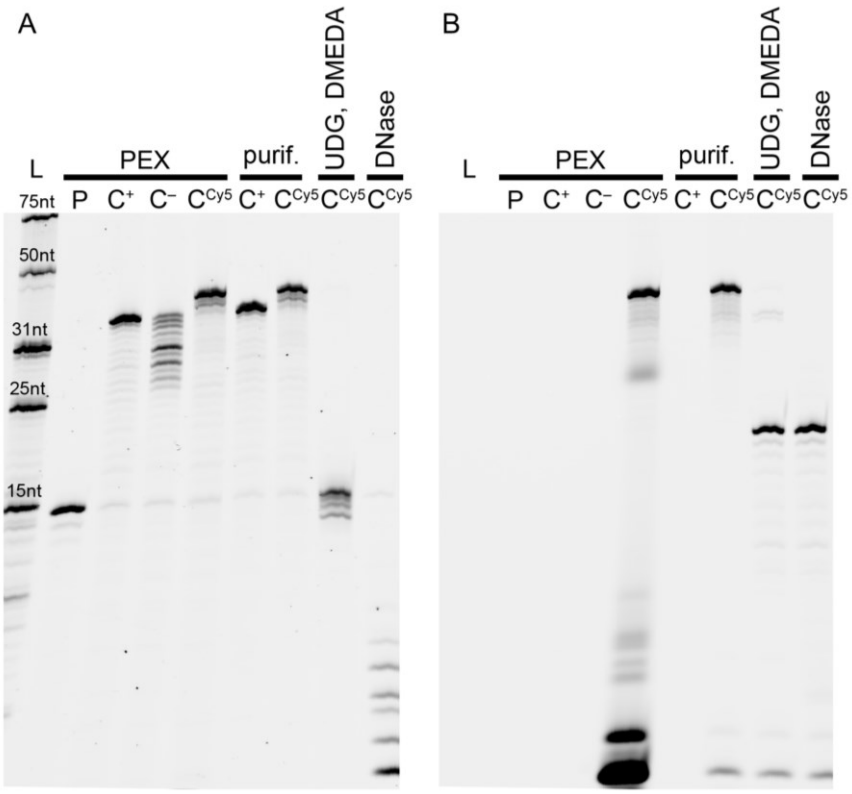



**Figure 80. 22.5% dPAGE analysis. Gel legend: (L) RNA ladder composed of FAM-labelled RNA oligonucleotides of indicated length; (P) DNA primer; (C$^+$) positive control, all natural rNTPs; (C$^-$) negative control, mixture of rATP, rUTP, rGTP and $H_2O$; (C$^{Cy5}$) modification, mixture of rATP, rUTP, rC$^{Cy5}$TP, rGTP. (UDG, DMEDA) Sample analysis after UDG followed by DMEDA cleavage. (DNase) Sample analysis after TurboDNase cleavage. A – FAM scan; B – Cy5 scan (visualisation of Cy5-modification).**

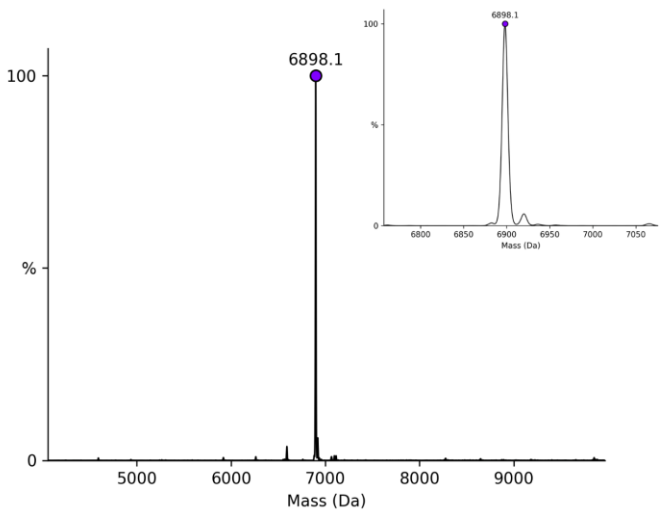



**Figure 81. LC-ESI-MS analysis of RNA_C$^{Cy5}$. Deconvoluted spectrum; calculated mass: 6898.6 Da; found mass: 6898.1 Da (product).**

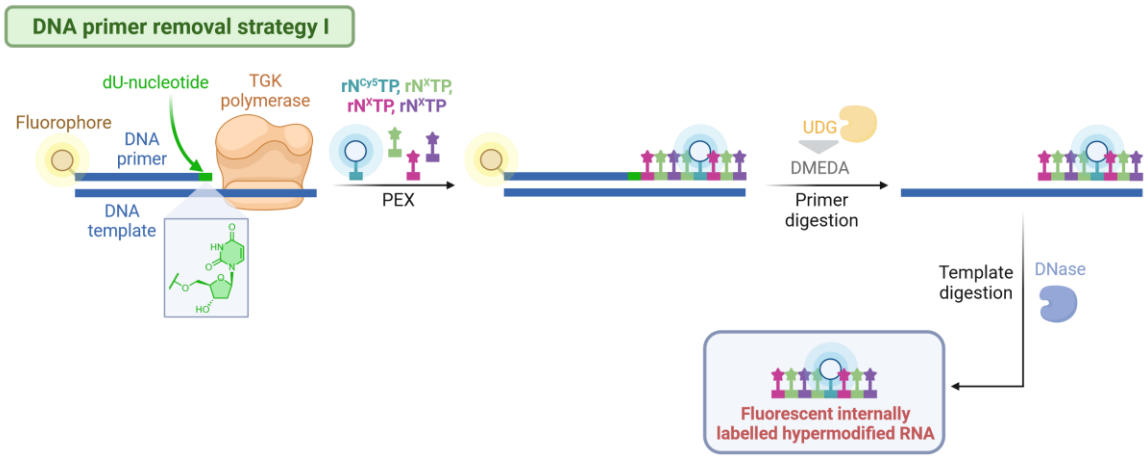



**Figure 82. DNA primer removal strategy. 5'-FAM labelled DNA primer contains single 2'-dU modification at 3'-end terminus. Upon PEX with four rN$^{X}$TPs, one of them labelled with Cy5, dU is cleaved by UDG to generate abasic site, that is subsequently eliminated by DMEDA under mild basic conditions. DNA template and rest of the primer are degraded with DNase to produce fully modified RNA.**

With the optimised cleavage procedure in my hands, I redirected my attention towards the ultimate objective – the production of hypermodified RNA polymers. First, for PEX reaction I used a mixture of all four base-modified **rN$^{X}$TP**s (**rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Cy5}$TP**, **rG$^{Pent}$TP**), one of them fluorescently labelled for visualisation of RNA cleavage product (Figure 82). The dU-containing DNA primer (**FAM-DNA-prim_dU**) was effectively extended along complementary DNA template (**templ_36nt**) even by four nucleobase-modified **rN$^{X}$TP**s producing the final PEX product (**36DNA_dU-RNA_A$^{E}$U$^{Bio}$C$^{Cy5}$G$^{Pent}$**, Table 13). Five negative control reactions were carried out, either with absence of all nucleotides or by omitting one of them in each reaction. Positive control

was performed with all four natural rNTPs (Figure 83). Next, the modified PEX product (**36DNA_dU-RNA_A$^{E}$U$^{Bio}$C$^{Cy5}$G$^{Pent}$**) was used in reaction cascade with UDG followed by DMEDA cleavage and final DNase template degradation. Identity of desired all-nucleobase modified product (**RNA_A$^{E}$U$^{Bio}$C$^{Cy5}$G$^{Pent}$**) was confirmed by dPAGE (Figure 84) and MS (Figure 85).

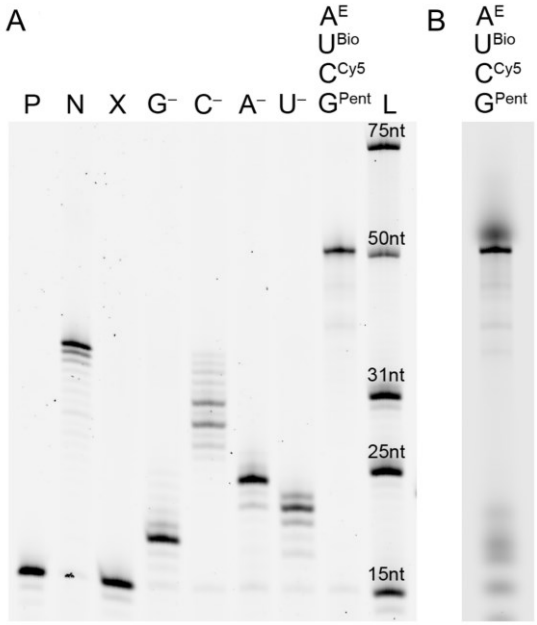



**Figure 83. 12.5% dPAGE analysis of PEX reaction. Gel legend: (P) DNA primer; (N) positive control, all natural rNTPs; (X) negative control, no rNTPs; (G$^{-}$) negative control, mixture of rATP, rUTP, rCTP and $H_2O$; (C$^{-}$) negative control, mixture of rATP, rUTP, rGTP and $H_2O$; (A$^{-}$) negative control, mixture of rUTP, rCTP, rGTP and $H_2O$; (U$^{-}$) negative control, mixture of rATP, rCTP, rGTP and $H_2O$; (A$^{E}$, U$^{Bio}$, C$^{Cy5}$, G$^{Pent}$) modification, mixture of rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Cy5}$TP, rG$^{Pent}$TP; (L) RNA ladder composed of FAM-labelled RNA oligonucleotides of indicated length. A – FAM scan; B – Cy5 scan (visualisation of Cy5-modification).**

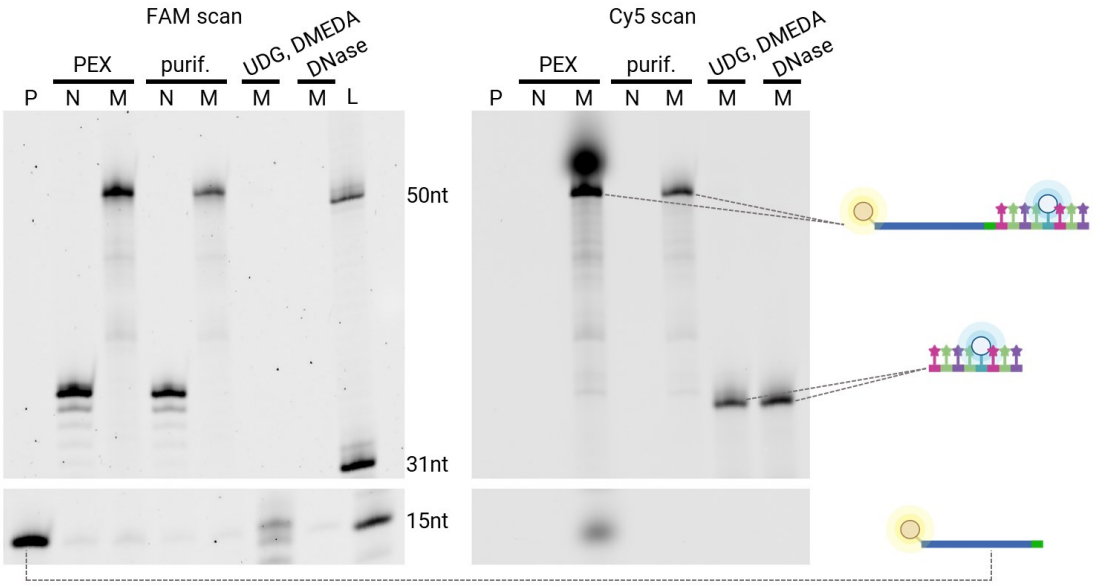



**Figure 84. FAM and Cy5 scan of dPAGE of DNA primer removal strategy. Cy5 scan allows to monitor RNA part of the molecule, while FAM scan is tracking fate of DNA component. Gel legend: (P) DNA primer; (N) positive control, PEX with natural rNTPs; (M) modified RNA, PEX with a mixture of rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Cy5}$TP, rG$^{Pent}$TP. (L) Ladder composed of FAM-labelled RNA oligonucleotides of indicated length.**

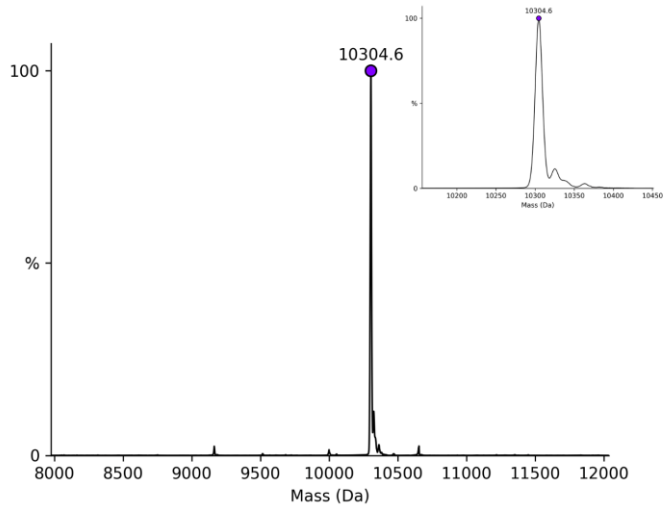



**Figure 85. LC-ESI-MS analysis of RNA_A$^{E}$U$^{Bio}$C$^{Cy5}$G$^{Pent}$. Deconvoluted spectrum; calculated mass: 10305.3 Da; found mass: 10304.6 Da (product).**

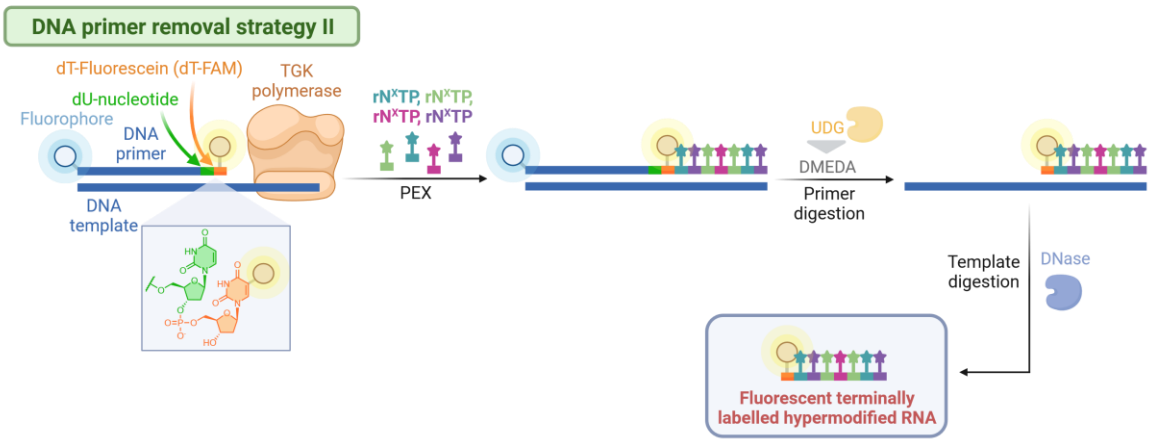



**Figure 86. DNA primer removal strategy. 5'-Cy5 labelled primer contains single 2'-dU followed by fluorescein tagged 2'-dT at 3'-end terminus. Upon PEX with four rN$^{X}$TPs, dU is cleaved by UDG to generate abasic site, that is subsequently eliminated by DMEDA under mild basic conditions. However herein 5'-(fluorescein-dT) label is retained at RNA part allowing its detection. DNA template and rest of the primer are degraded with DNase to produce fully modified RNA.**

When slightly modified, this procedure allows to obtain fluorescent terminally labelled hypermodified RNA (Figure 86). In this case, I used DNA primer carrying dU immediately followed by fluorescein-linked dT nucleotide (**Cy5-DNA-prim_dU_dT-FAM**, Table 15). Also, in this case the PEX reaction with combination of four nucleobase-modified **rN$^{X}$TPs** (**rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP**, **rG$^{Pent}$TP**) yielded the desired DNA-RNA hybrid product (**35DNA_dU_dT$^{FAM}$-RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$**) as confirmed by dPAGE (Figure 87). Upon UDG-DMEDA mediated cleavage the fluorescently labelled primer-originating-dT nucleotide is retained and facilitates tracking of the residual 5'-terminally labelled fluorescent all-base modified RNA (**dT$^{FAM}$-RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$**). Formation of the fluorescent hypermodified polymer was confirmed by dPAGE (Figure 88) and ESI-MS (Figure 89).

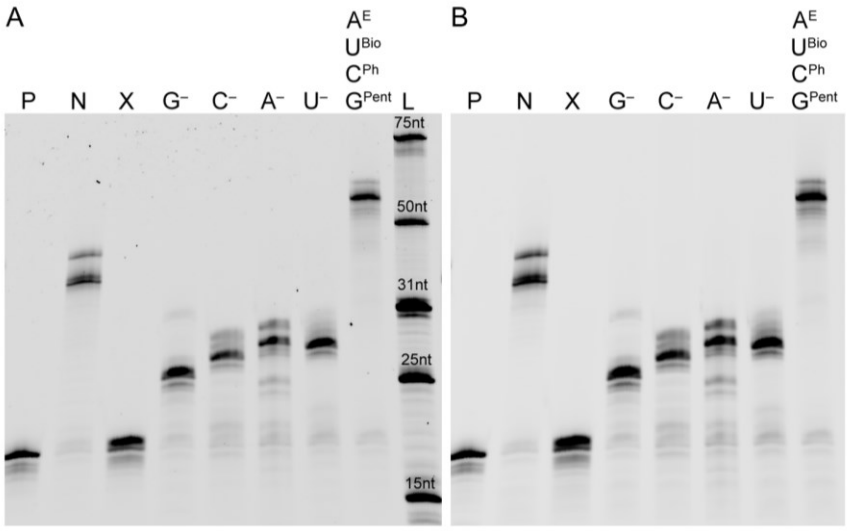



**Figure 87. 22.5% dPAGE analysis of PEX reaction. Gel legend: (P) DNA primer; (N) positive control, all natural rNTPs; (X) negative control, no rNTPs; ($G^-$) negative control, mixture of rATP, rUTP, rCTP and $H_2O$; ($C^-$) negative control, mixture of rATP, rUTP, rGTP and $H_2O$; ($A^-$) negative control, mixture of rUTP, rCTP, rGTP and $H_2O$; ($U^-$) negative control, mixture of rATP, rCTP, rGTP and $H_2O$; ($A^E$, $U^{Bio}$, $C^{Ph}$, $G^{Pent}$) modification, mixture of $rA^ETP$, $rU^{Bio}TP$, $rC^{Ph}TP$, $rG^{Pent}TP$; (L) RNA ladder composed of FAM-labelled RNA oligonucleotides of indicated length. A – FAM-scan; B – Cy5-scan.**

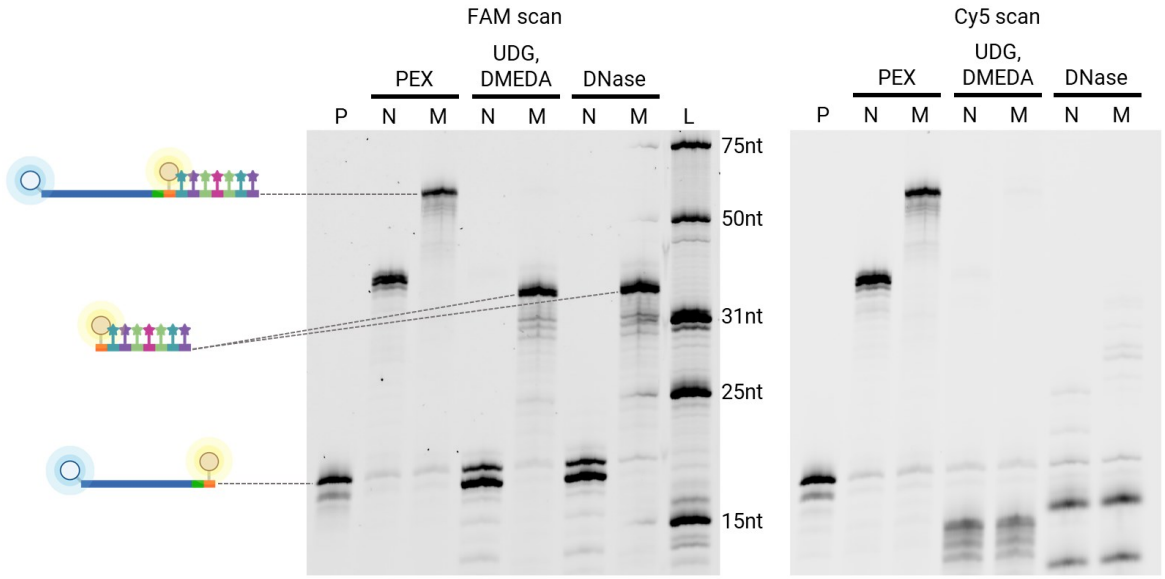



**Figure 88. FAM and Cy5 scan of dPAGE of DNA primer removal strategy. FAM scan allows to monitor RNA part of the molecule, while Cy5 scan is tracking fate of DNA component. Gel legend: (P) DNA primer; (N) positive control, PEX with natural rNTPs; (M) modified RNA, PEX with a mixture of $rA^ETP$, $rU^{Bio}TP$, $rC^{Ph}TP$, $rG^{Pent}TP$. (L) Ladder composed of FAM-labelled RNA oligonucleotides of indicated length.**

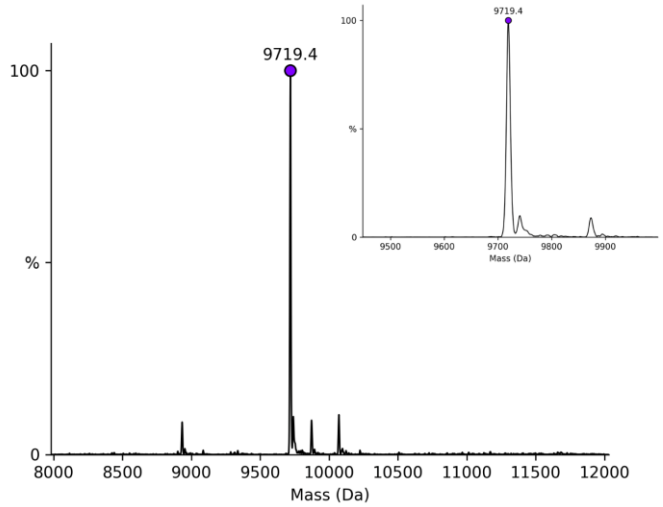



**Figure 89. LC-ESI-MS analysis of dT$^{FAM}$-RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$. Deconvoluted spectrum; calculated mass: 9719.4 Da; found mass: 9719.4 Da (product).**

### 3.2.7 Site-specific RNA labelling for secondary structure analysis by FRET

Decorating RNA at a designated internal position is largely unmet need that prevents effective study of RNA structure, interactions, dynamics and epitranscriptome[271]. IVT with T7 RNAP and modified **rN$^{X}$TP**s usually leads to pervasive RNA labelling. Moreover, incorporating multiple distinct or bulky modifications can be inefficient due to T7 RNAP's limited tolerance. Single labelling using PLOR within homopolymeric sequences is not possible, and method requires programming of robotic platform for automation[124,273]. To fill this methodological gap, I deployed improved version of stepwise SNI reaction previously developed for DNA labelling[137], that would enable site-specific insertion of modifications at predefined positions also in RNA.

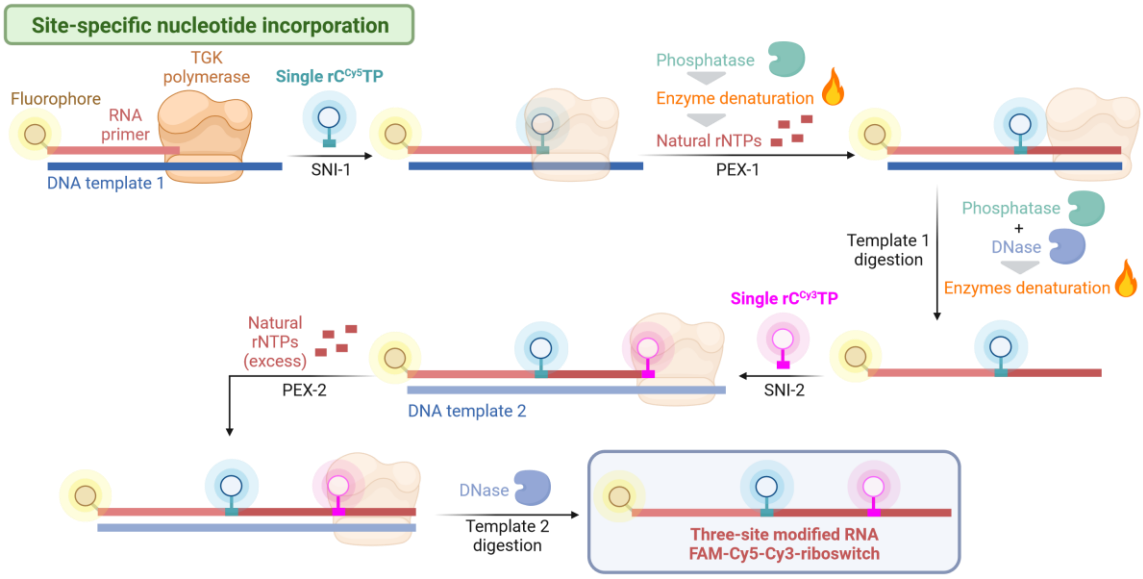



**Figure 90. Overall strategy of single-tube site-specific RNA labelling at two distinct positions by combination of SNI and PEX reactions.**

For this occasion, I decided to modify the well-studied aptamer domain of the adenine riboswitch at specific positions of loop-1 and loop-2 with fluorescent tags to study the structure and dynamics using Förster (fluorescence) resonance energy transfer (FRET) measurements[279]. The enzymatic synthesis of FAM-Cy5-Cy3-triple-labelled riboswitch (**FAM-Cy5-Cy3-riboswitch**, Table 13) was performed with combination of SNI and PEX (Figure 90) using FAM-labelled RNA primer (**FAM-RNA-prim_23nt**, Table 11) and two differently long ssDNA templates (**templ_ribosw71_A**, **templ_ribosw71_B**, Table 12). Initially, RNA primer complementary to the first DNA template (**templ_ribosw71_A**) was extended by only one **rU$^{Cy5}$** nucleotide by SNI. Degradation of the free unincorporated **rU$^{Cy5}$TP** was ensured by shrimp alkaline phosphatase (rSAP). Following heat denaturation of enzyme and addition of abundant amount of four natural rNTP (rATP, rUTP, rCTP, rGTP) used in the second step led to complete elongation of the primer along DNA template (**templ_ribosw71_A**). Degradation of unreacted natural rNTPs was performed again by usage of rSAP with concurrent template degradation using TurboDNase. Subsequently, both enzymes were denatured by increased temperature. Addition of second, longer DNA template (**templ_ribosw71_B**) in presence of **rU$^{Cy3}$TP** led to second SNI. Finally, subsequent extension of the primer with large excess of all four natural rNTPs produced the desired full-length 71 nt long **FAM-Cy5-Cy3-riboswitch** (Figure 90). The progress of whole cascade was monitored by dPAGE (Figure 91) and confirmed by ESI-MS of each intermediate (Table 14) and final full-length product (Figure 92, Table 14). This "one-pot" approach occurring in a single tube is virtually nondependent on desired modified position and allowed to instal simultaneously two bulky Cy5 and Cy3 fluorophores. Since additional

tag was brought in as a part of synthetic primer, up to three labels were collectively introduced in the riboswitch molecule. Whole procedure allowed access to RNA labelled at user-defined positions, even inside homopolymeric sequence, with respectable yield (13%).

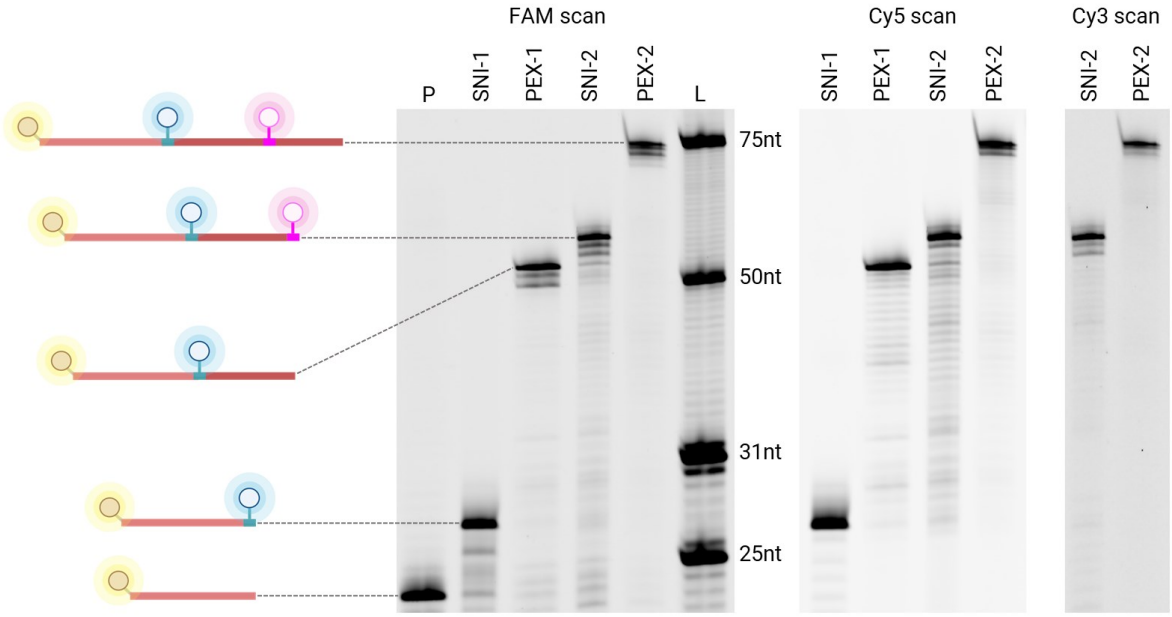



**Figure 91. FAM, Cy5 and Cy3 scan of dPAGE analysis of reaction intermediates and final FAM-Cy5-Cy3-riboswitch. Gel legend: (P) RNA primer; (SNI-1) SNI of rC$^{Cy5}$TP; (PEX-1) PEX with natural rNTPs; (SNI-2) SNI of rC$^{Cy3}$TP; (PEX-2) PEX with natural rNTPs. (L) Ladder composed of FAM-labelled RNA oligonucleotides of indicated length.**

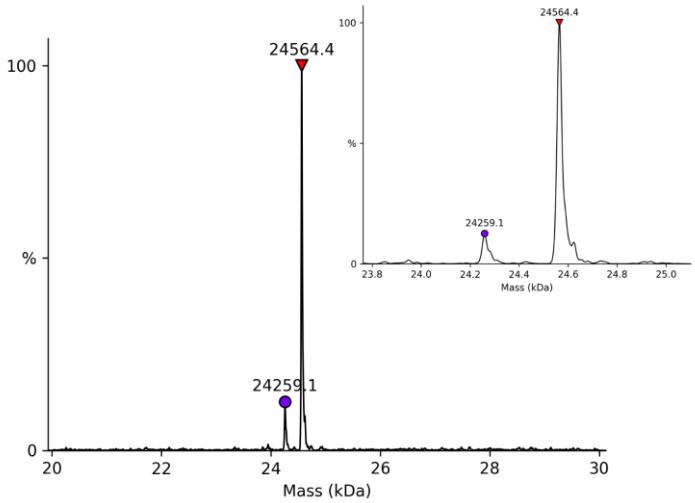



**Figure 92. LC-ESI-MS analysis of FAM-Cy5-Cy3-riboswitch. Deconvoluted spectrum; calculated mass: 24569.1 Da; found mass: 24564.4 Da (product); found mass: 24259.1 Da (product - rCMP).**

For conformational studies, I prepared in a similar way the Cy5-Cy3-labelled version of the riboswitch (Figure 93, Figure 94) and performed simple FRET measurements. These measurements involved the detection and quantification of energy transfer between two fluorophores in proximity. If one fluorophore known as the donor is excited and the second fluorophore (the acceptor) is in proximity to the donor, energy is transferred to the acceptor, resulting in its fluorescence. FRET measurements were carried out in presence of 4 M urea and ($Mg^{2+}$)[280] with increasing concentration of the adenine ligand. Two, Cy5 and Cy3 labelled arms of riboA71 aptaswitch (**Cy5-Cy3-riboswitch**) are brought together upon adenine docking. Thus, the prepared FRET probe exhibited increased emission at 670 nm, when irradiated at 530 nm, while decreased emissions at 570 nm (Figure 94). This indicates conformational changes in riboswitch arms tertiary structure to occur in corelation with previous observations[279]. Negative FRET control measurements were carried out under same conditions, but with two separate and distinct DNA oligonucleotides, one of them labelled with Cy5 and the other one bearing the Cy3 tag (Table 17). As expected, no enhanced FRET signal was observed in this case (Figure 94).

**Table 17. DNA oligonucleotides used for negative control FRET measurements. For structures of 5'-end modifications see Figure 116.**

| DNA oligonucleotide name | Sequence (5'→3') | 5'-end modification | Length [nt] |
|---|---|---|---|
| **Cy5-DNA-oligo** | CATGGGCGGCATGGG | Cy5 | 15 |
| **Cy3-DNA-oligo** | TGGCGCGAAGGCTGTCATTG | Cy3 | 20 |

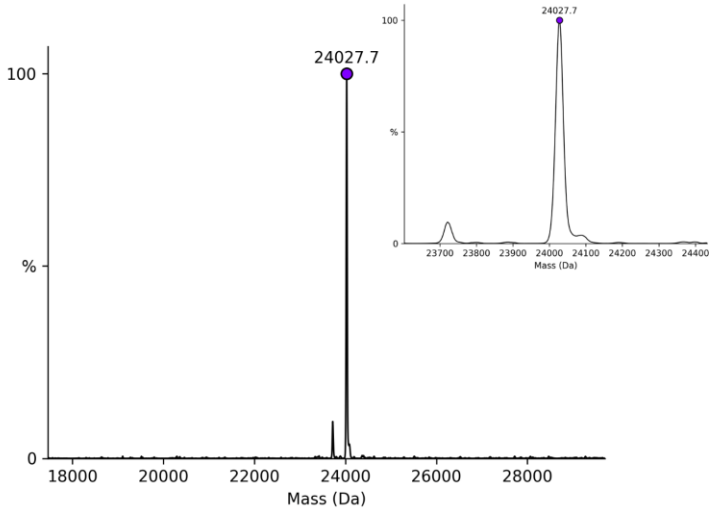



**Figure 93. LC-ESI-MS analysis of Cy5-Cy3-riboswitch. Deconvoluted spectrum; calculated mass: 24032.1 Da; found mass: 24027.7 Da (product).**

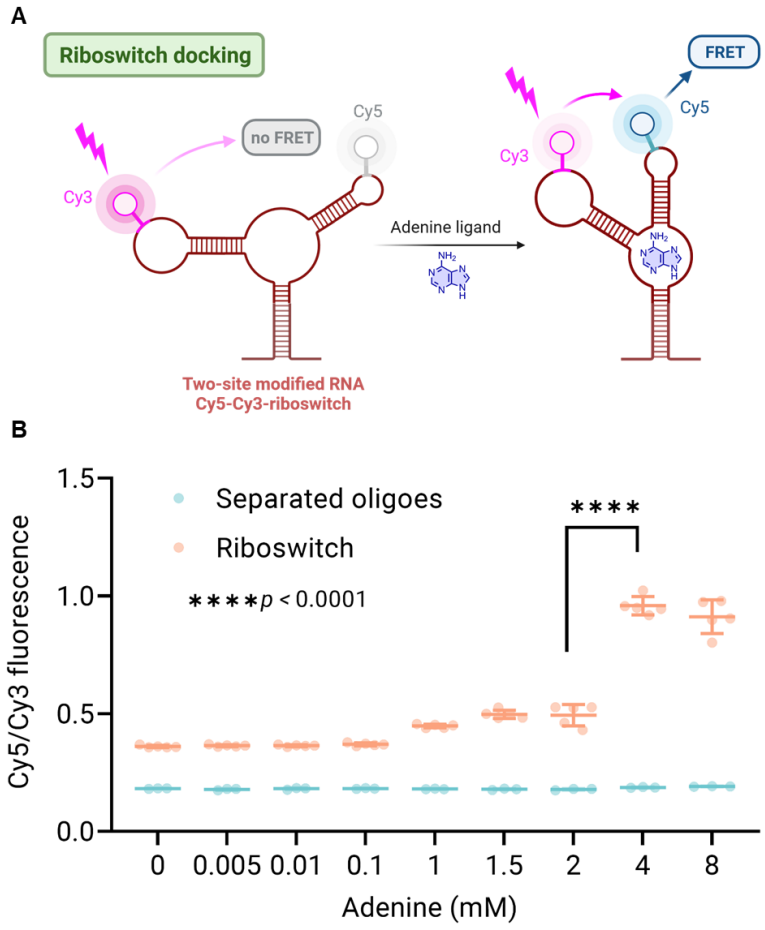



**Figure 94. A – Two, Cy5 and Cy3 labelled, arms of riboA71 aptaswitch (Cy5-Cy3-riboswitch) are brought together upon adenine docking. Induced tertiary structural changes lead to augmented FRET signal. B – Normalised FRET signals measured at various adenine concentrations. FRET measurements were carried out in independent replicates, *n* = 5 for riboswitch, *n* = 3 for negative controls.**

### 3.2.8 Enzymatic synthesis of modified messenger RNAs

The established methods for synthesis of mRNA harness T7 RNAP IVT leading to replacement of natural nucleotides with modified counterparts across the entire sequence[245]. Conversely, PLOR's site-specific labelling encounters diminished yields with an increased number of pause-restart cycles[124]. Furthermore, splint ligation exhibits inefficiencies and necessitates laborious gel purification[152], rendering all these methodologies unsuitable for the comprehensive assessment of sequence- and region-dependent impacts of nucleobase modifications in lengthy mRNA constructs.

Therefore, I endeavoured to implement our novel approach for the targeted and segmented incorporation of functional groups at specified positions within mRNA molecules. Since commercial chemical synthesis of RNA primers containing canonical 5'-end cap structure is not available, I used EMCV-IRES mediated translation initiation[281,282]. The internal ribosome entry site (IRES) from

encephalomyocarditis virus (EMCV) is a highly structured RNA element that initiates 5'-end cap-independent protein synthesis[283]. The designed constructs fluorescently labelled at 5'-end contained internal ribosomal entry side (IRES), gene-encoding sequence for nanoluciferase (nLuc) and long poly(A) tail to regulate mRNA stability. I selected the nLuc reporter system for its small size, enhanced stability, high sensitivity, and low background crucial for achieving a high signal-to-noise ratio, when compared to conventional ones[284]. I devised five mRNA constructs, natural as a control and either fully modified or with modifications only in the gene coding part, eventually two mRNAs with single modification at defined position (Figure 95). I opted to incorporate 5-methylcytidine nucleoside triphosphate (**rC$^{Me}$TP**), a naturally occurring building block that has been previously demonstrated to enhance the performance of mRNA translation[29].

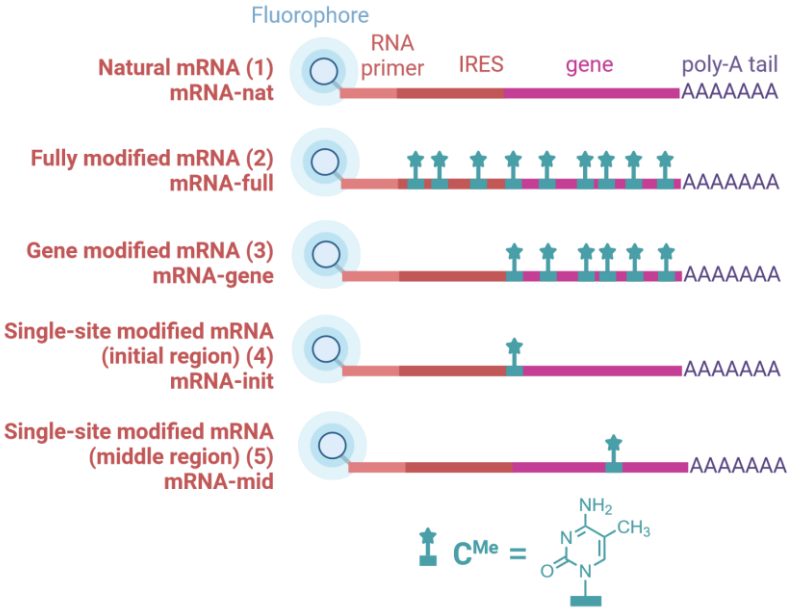



**Figure 95. Designed either natural or variously modified mRNA constructs containing 5'-end fluorescent tag, IRES, gene coding part and poly(A) tail.**

For synthesis of mRNAs constructs, I needed to prepare the necessary ssDNA templates (**templ_IRES**, **templ_IRES-prolonged**, **templ_IRES-nLuc**, Table 18). These were first amplified by PCR reaction from source custom synthesised plasmid (Figure 149) using two short DNA primers complementary to the plasmid (**DNA-REV-prim_26nt** for **templ_IRES**, **DNA-REV-prim_24nt** for **templ_IRES-prolonged**, **DNA-REV-prim_48nt** for **templ_IRES-nLuc**, Table 15), one of them dual-biotin labelled (**5'-(dual-Bio)-DNA-FOR-prim_29nt**, Table 15), and high-fidelity Q5 DNA polymerase. Generation of desired ssDNA templates from dsDNA constructs was enabled by streptavidin magnetic beads, followed by strand separation in alkaline conditions with 50 mM NaOH.

Successful PCR amplification and strand separation of all desired templates (Table 18) was verified by agarose gel (Figure 96).

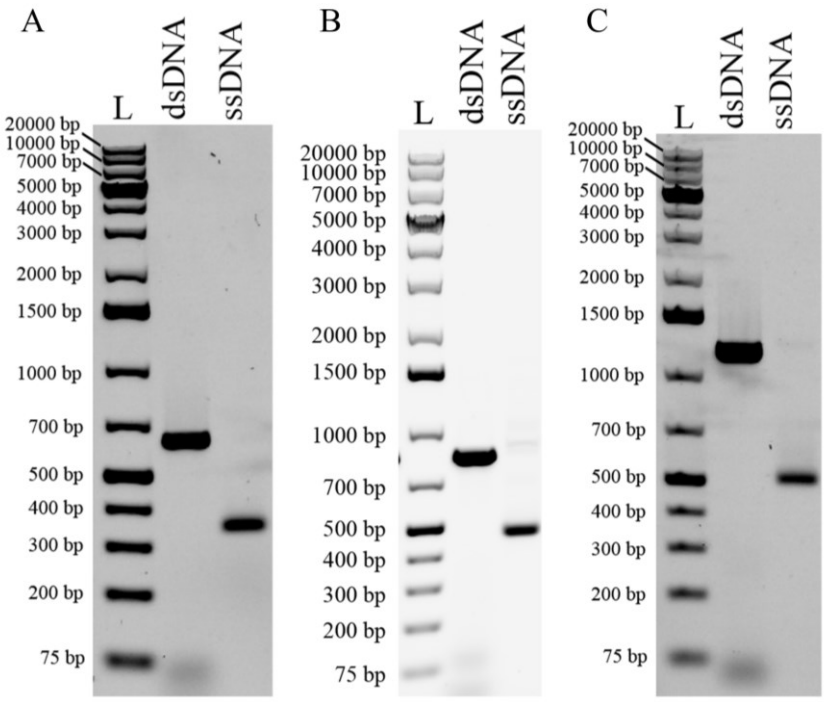



**Figure 96. Native 1% agarose gel analysis of dsDNA and ssDNA templates after magnetoseparation. Gel legend: (L) GeneRuler 1 kb Plus DNA ladder; (dsDNA) aliquot of dsDNA template; (ssDNA) aliquot of ssDNA template. A – analysis of templ_IRES, B – analysis of templ_IRES-prolonged, C – analysis of templ_IRES-nLuc.**

**Table 18. List of prepared DNA oligonucleotides used as templates in this study.**

| DNA oligonucleotide name | Sequence (5'→3') | 5'-end modification | Length [nt] |
|---|---|---|---|
| **templ_IRES** | ACCATGGTTGTGGCCATATTATCATCGTGTTTTT<br>CAAAGGAAAACCACGTCCCCGTGGTTCGGGGGGC<br>CTAGACGTTTTTTTAACCTCGACTAAACACATGT<br>AAAGCATGTGCACCGAGGCCCCAGATCAGATCCC<br>ATACAATGGGGTACCTTCTGGGCATCCTTCAGCC<br>CCTTGTTGAATACGCTTGAGGAGAGCCATTTGAC<br>TCTTTCCACAACTATCCAACTCACAACGTGGCAC<br>TGGGGTTGTGCCGCCTTTGCAGGTGTATCTTATA<br>CACGTGGCTTTTGGCCGCAGAGGCACCTGTCGCC<br>AGGTGGGGGGTTCCGCTGCCTGCAAAGGGTCGCT<br>ACAGACGTTGTTTGTCTTCAAGAAGCTTCCAGAG<br>GAACTGCTTCCTTCACGACATTCAACAGACCTTG<br>CATTCCTTTGGCGAGAGGGGAAAGACCCCTAGGA<br>ATGCTCGTCAAGAAGACAGGGCCAGGTTTCCGGG<br>CCCTCACATTGCCAAAAGACGGCAATATGGTGGA<br>AAATAACATATAGACAAACGCACACCGGCCTTAT<br>TCCAAGCGGCTTCGGCCAGTAACGTTAGGGGGGG<br>GGGAGGGAGAGGGGCCCTATAGTGAGTCGTATTA<br>GACGAAAGGG | – | 622 |

| | | | |
|---|---|---|---|
| **templ_IRES-prolonged** | CTCAGACCTTCATACGGGATGATGACATGGATGT<br>CGATCTTCAGCCCATTTTCACCGCTCAGGACAAT<br>CCTTTGGATCGGAGTTACGGACACCCCGAGATTC<br>TGAAACAAACTGGACACACCTCCCTGTTCAAGGA<br>CTTGGTCCAGGTTGTAGCCGGCTGTCTGTCGCCA<br>GTCCCCAACGAAATCTTCGAGTGTGAAGACCATG<br>GTTGTGGCCATATTATCATCGTGTTTTTCAAAGG<br>AAAACCACGTCCCCGTGGTTCGGGGGGCCTAGAC<br>GTTTTTTTAACCTCGACTAAACACATGTAAAGCA<br>TGTGCACCGAGGCCCCAGATCAGATCCCATACAA<br>TGGGGTACCTTCTGGGCATCCTTCAGCCCCTTGT<br>TGAATACGCTTGAGGAGAGCCATTTGACTCTTTC<br>CACAACTATCCAACTCACAACGTGGCACTGGGGT<br>TGTGCCGCCTTTGCAGGTGTATCTTATACACGTG<br>GCTTTTGGCCGCAGAGGCACCTGTCGCCAGGTGG<br>GGGGTTCCGCTGCCTGCAAAGGGTCGCTACAGAC<br>GTTGTTTGTCTTCAAGAAGCTTCCAGAGGAACTG<br>CTTCCTTCACGACATTCAACAGACCTTGCATTCC<br>TTTGGCGAGAGGGGAAAGACCCCTAGGAATGCTC<br>GTCAAGAAGACAGGGCCAGGTTTCCGGGCCCTCA<br>CATTGCCAAAAGACGGCAATATGGTGGAAAATAA<br>CATATAGACAAACGCACACCGGCCTTATTCCAAG<br>CGGCTTCGGCCAGTAACGTTAGGGGGGGGGGAGG<br>GAGAGGGGCCCTATAGTGAGTCGTATTAGACGAA<br>AGGG | – | 820 |
| **templ_IRES-nLuc** | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAC<br>GCCAGAATGCGTTCGCACAGCCGCCAGCCGGTCA<br>CTCCGTTGATGGTTACTCGGAACAGCAGGGAGCC<br>GTCGGGGTTGATCAGGCGCTCGTCGATAATTTTG<br>TTGCCGTTCCACAGGGTCCCTGTTACAGTGATCT<br>TTTTGCCGTCGAACACGGCGATGCCTTCATACGG<br>CCGTCCGAAATAGTCGATCATGTTCGGCGTAACC<br>CCGTCGATTACCAGTGTGCCATAGTGCAGGATCA<br>CCTTAAAGTGATGATCATCCACAGGGTACACCAC<br>CTTAAAAATTTTTTCGATCTGGCCCATTTGGTCG<br>CCGCTCAGACCTTCATACGGGATGATGACATGGA<br>TGTCGATCTTCAGCCCATTTTCACCGCTCAGGAC<br>AATCCTTTGGATCGGAGTTACGGACACCCCGAGA<br>TTCTGAAACAAACTGGACACACCTCCCTGTTCAA<br>GGACTTGGTCCAGGTTGTAGCCGGCTGTCTGTCG<br>CCAGTCCCCAACGAAATCTTCGAGTGTGAAGACC<br>ATGGTTGTGGCCATATTATCATCGTGTTTTTCAA<br>AGGAAAACCACGTCCCCGTGGTTCGGGGGGCCTA<br>GACGTTTTTTTAACCTCGACTAAACACATGTAAA<br>GCATGTGCACCGAGGCCCCAGATCAGATCCCATA<br>CAATGGGGTACCTTCTGGGCATCCTTCAGCCCCT<br>TGTTGAATACGCTTGAGGAGAGCCATTTGACTCT<br>TTCCACAACTATCCAACTCACAACGTGGCACTGG<br>GGTTGTGCCGCCTTTGCAGGTGTATCTTATACAC<br>GTGGCTTTTGGCCGCAGAGGCACCTGTCGCCAGG<br>TGGGGGGTTCCGCTGCCTGCAAAGGGTCGCTACA<br>GACGTTGTTTGTCTTCAAGAAGCTTCCAGAGGAA<br>CTGCTTCCTTCACGACATTCAACAGACCTTGCAT<br>TCCTTTGGCGAGAGGGGAAAGACCCCTAGGAATG<br>CTCGTCAAGAAGACAGGGCCAGGTTTCCGGGCCC<br>TCACATTGCCAAAAGACGGCAATATGGTGGAAAA<br>TAACATATAGACAAACGCACACCGGCCTTATTCC<br>AAGCGGCTTCGGCCAGTAACGTTAGGGGGGGGGG<br>AGGGAGAGGGGCCCTATAGTGAGTCGTATTAGAC<br>GAAAGGG | – | 1163 |

With optimised procedure for template generation in hands, I turned to mRNA synthesis. First, I prepared two control mRNAs, either natural or **rC$^{Me}$**-modified in whole length. The PEX reaction (Figure 98) was conducted with generated ssDNA template (**templ_IRES-nLuc**, Table 18, Figure 96) and fluorescently labelled RNA primer (**Cy5-mRNA-prim**, Table 11) to enable accurate RNA quantification and convenient detection in gel electrophoresis. PEX reaction for synthesis of non-modified mRNA (**mRNA-nat**, Figure 98, Table 21) was performed with combination of all four natural rNTPs (rATP, rUTP, rCTP, rGTP), while for generation of fully modified mRNA (**mRNA-full**, Table 21) combination of rATP, rUTP, **rC$^{Me}$TP**, rGTP was used (Figure 98). For segmental modifying gene-coding part, while keeping the IRES region unmodified in mRNA (gene modified mRNA, **mRNA-gene**), I performed a two-step PEX reaction (Figure 98). First, the non-modified IRES (**Cy5-IRES-RNA**, Figure 97, Table 21) was prepared by PEX reaction with short RNA primer – **Cy5-mRNA-prim** and generated ssDNA template (**templ_IRES**, Figure 96, Table 18) using combination of all four natural rNTPs. Second, this RNA intermediate (**Cy5-IRES-RNA**, Figure 97) after template degradation and silica-based column purification prepared in 73% yield was used as RNA megaprimer for second PEX reaction using template encoding for full-length mRNA (**templ_IRES-nLuc)** and a mix of rATP, rUTP, **rC$^{Me}$TP**, rGTP. For preparation of point-modified mRNAs (Figure 98), with single modification either in the initial (**mRNA-init**, Table 21) or middle (**mRNA-mid**, Table 21) region of gene coding part (Figure 98), I applied a methodology combining SNI and PEX reaction with two differently long prepared ssDNA templates. First, using **Cy5-IRES-RNA** as RNA megaprimer and ssDNA template – **templ_IRES-nLuc** I performed SNI with **rC$^{Me}$TP**. Degradation of unincorporated nucleotide was ensured by rSAP followed by heat denaturation of the enzyme. Next, subsequent PEX reaction with a mixture of all four natural rNTPs generated full-length mRNA with modification in the initial part of the gene encoding sequence (**mRNA-init**, Figure 98). For preparation of modified mRNA in the middle part of the gene encoding sequence (**mRNA-mid**, Figure 98) I applied similar methodology. First, the RNA megaprimer (**Cy5-IRES-RNA_prolonged**, Figure 97, Table 21), with unmodified IRES region and part of the gene coding sequence, was prepared by PEX reaction of **Cy5-mRNA-prim** and generated ssDNA template – **templ_IRES-prolonged** (Figure 96, Table 18) in presence of natural rNTPs. After DNA template removal by TurboDNase followed by purification step, the desired second RNA megaprimer (**Cy5-IRES-RNA_prolonged,** Figure 97) was isolated in 76% yield. This RNA intermediate, upon annealing to ssDNA template (**templ_IRES-nLuc**), was used in SNI with **rC$^{Me}$TP**. Subsequent degradation of free nucleotide and heat enzyme denaturation followed by final PEX reaction with rNTPs generated the desired mRNA product (**mRNA-mid**, Figure 98). The final full-length mRNA products, containing poly(A) tail, were isolated by means of poly(T) magnetic beads capture. Liberation of the

products was ensured by mild thermal denaturation at 50 °C and all mRNA products were verified by urea agarose gel analysis (Figure 99). All desired mRNAs were isolated in satisfactory amounts (15-82%), already exceeding yields achieved by state-of-the-art (splint ligation)[152]. Moreover, unlike conventional T7 RNAP IVT, our method obviates need to HPLC purify mRNA, since no dsRNA by-products can be generated[250].

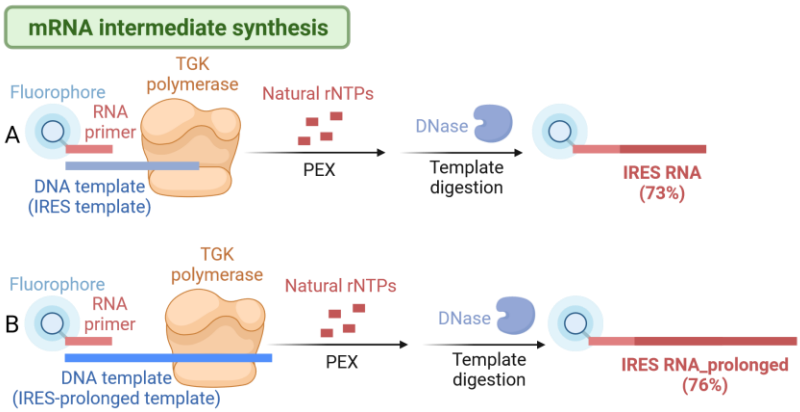



**Figure 97. Construction of RNA intermediates (Cy5-IRES-RNA, Cy5-IRES-RNA_prolonged) by PEX used as RNA megaprimers for mRNA synthesis.**

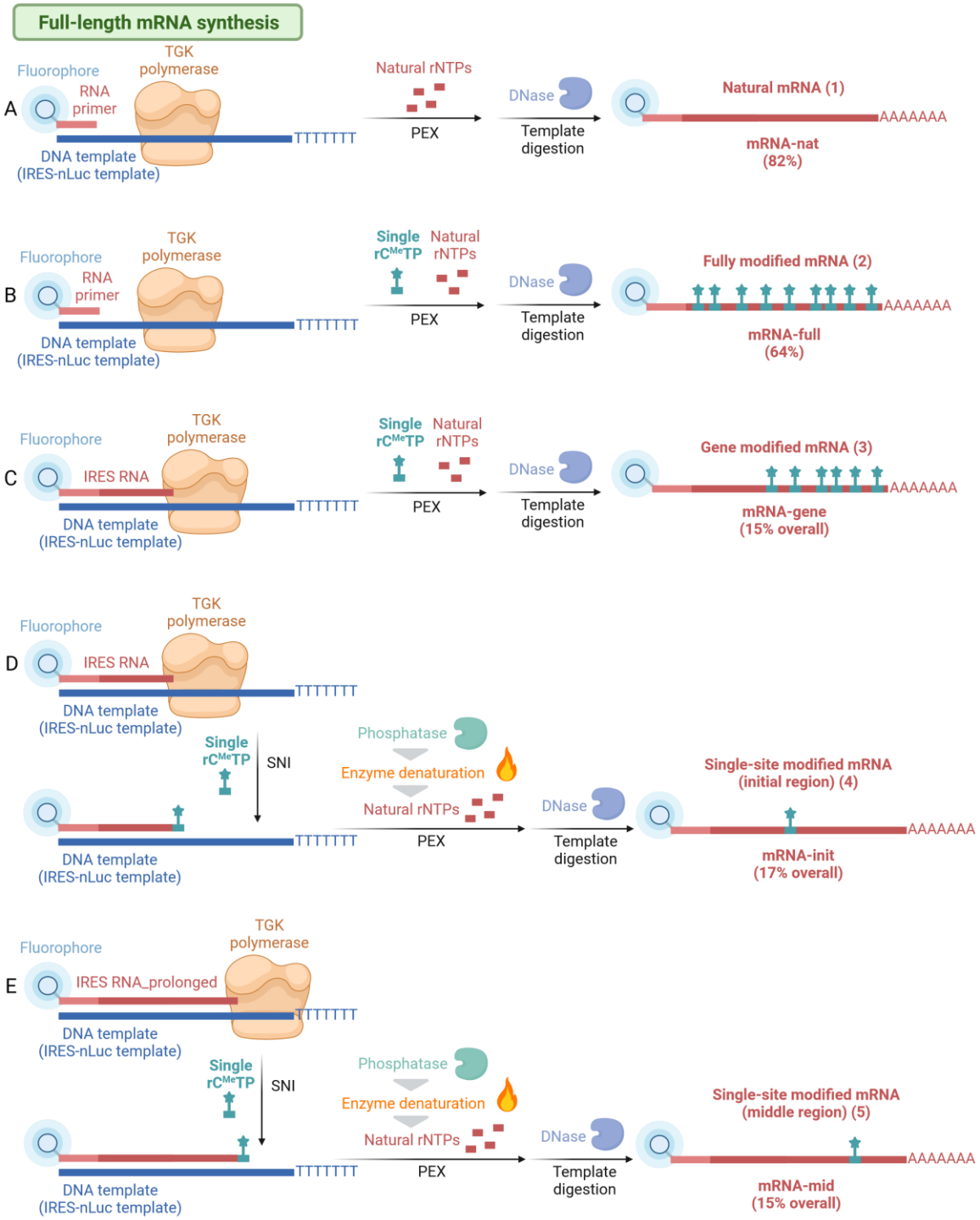



**Figure 98. Overview of enzymatic synthesis of mRNA. mRNA constructs were generally prepared by PEX with TGK polymerase and 5'-Cy5-labelled RNA primer and ssDNA template encoding for primer annealing site, IRES, gene-coding sequence and poly(A) tail.**

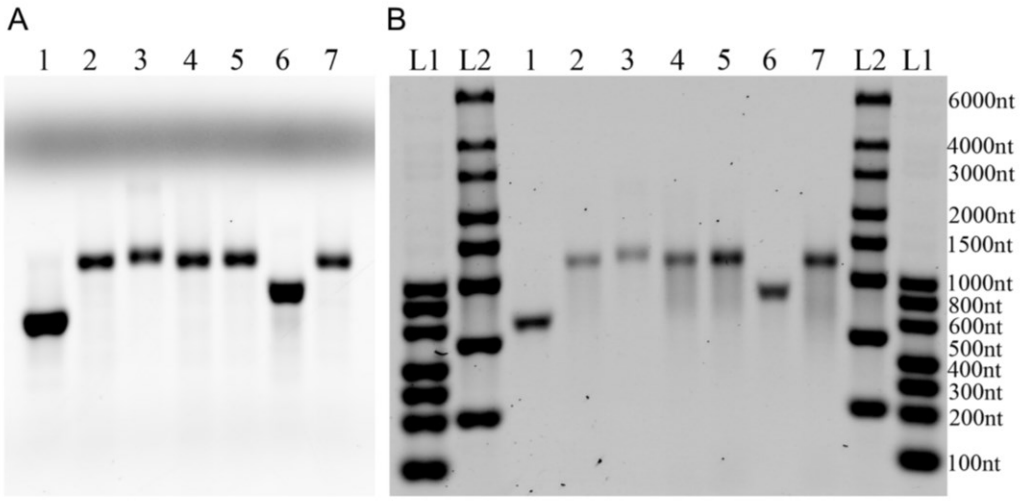



**Figure 99. 1.5% urea agarose gel analysis of mRNA intermediates and full-length products. Gel legend: (1) IRES-RNA; (2) mRNA-nat; (3) mRNA-full; (4) mRNA-gene; (5) mRNA-init; (6) IRES-RNA_prolonged; (7) mRNA-mid. (L1) RiboRuler Low Range RNA ladder; (L2) RiboRuler High Range RNA ladder. A – Cy5 scan; B – GelRed scan.**

To prove introduction of single **rC$^{Me}$** modification at designed position in both single-site modified mRNAs (**mRNA-init**, **mRNA-mid**), I performed bisulfite sequencing. Although, this method is well-established for DNA[285], it remains challenging for RNA due to its lower stability and reactivity leading to incomplete bisulfite conversions. RNA treatment with bisulfite causes chemical deamination and conversion of unmethylated cytosines to uracils, while leaving base-methylated cytosines unaffected[286]. Both point-modified mRNAs were upon bisulfite treatment amplified by one-pot RT-PCR. Thus, in the first step the converted mRNA was reverse transcribed into corresponding cDNA, followed by PCR amplification with forward and reverse DNA primers (**DNA-FOR-prim_26nt**, **DNA-REV-prim_21nt**, Table 15). Analysis by Sanger sequencing (Figure 100, Figure 101) revealed presence of single **rC$^{Me}$** modification in analysed mRNAs (**mRNA-init**, **mRNA-mid**).

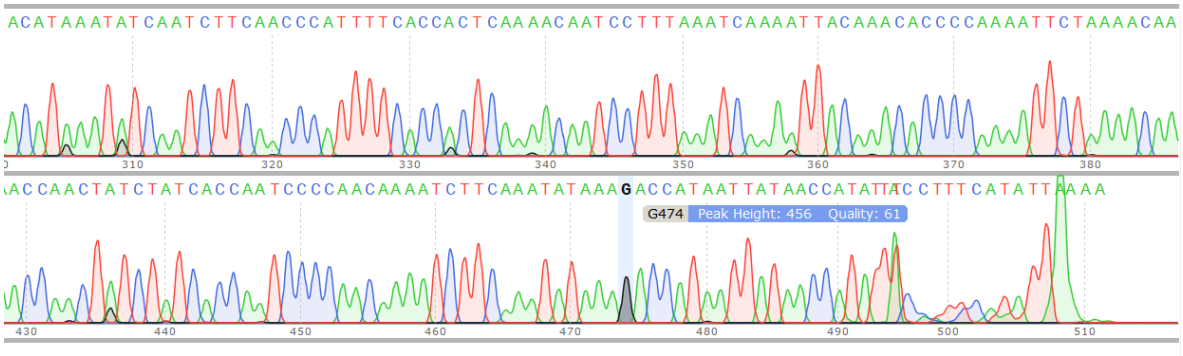



**Figure 100. Data from Sanger sequencing of bisulfite converted mRNA (mRNA-init). Magnified area of interest. Highlighted peak (G474) revealed presence of $C^{Me}$ modification. Upon bisulfite treatment all cytosines were converted to uracils and detected as adenines by RT-PCR and sequencing. The methylated cytosine was not affected by bisulfite treatment therefore it was detected as guanine by RT-PCR and sequencing. For full data see Figure 147.**

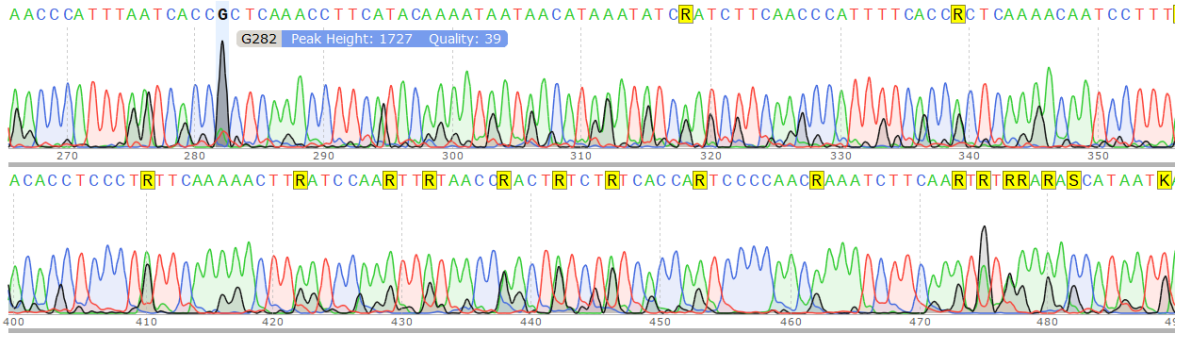



**Figure 101. Data from Sanger sequencing of bisulfite converted mRNA (mRNA-mid). Magnified area of interest. Highlighted peak (G282) revealed presence of $C^{Me}$ modification. Upon bisulfite treatment all cytosines were converted to uracils and detected as adenines by RT-PCR and sequencing. The methylated cytosine was not affected by bisulfite treatment therefore it was detected as guanine by RT-PCR and sequencing. For full data see Figure 148.**

### 3.2.9 Effects of variously base-modified mRNAs on translation efficiency

To evaluate translation efficiency of prepared mRNA constructs, I subjected them to *in vitro* translation studies using rabbit reticulocyte lysate system and probed for nanoluciferase activity. *In vitro* translations of fully modified (**mRNA-full**), gene-modified (**mRNA-gene**) and both single-site modified (**mRNA-init**, **mRNA-mid**) mRNAs were performed at 30 °C for 1.5 h. As control reactions, I performed translations with natural mRNA (**mRNA-nat**) or with absence of any RNA template. Translation reactions were stop by addition of cycloheximide and upon addition of furimazine substrate, intensity of glow-type luminescence was measured. Intriguingly, both point-modified mRNAs (**mRNA-init**, **mRNA-mid**) enhanced translation efficiency in comparison with natural mRNA (**mRNA-nat**), while the gene-modified mRNA (**mRNA-gene**) led to moderate decrease in potency. As expected, high density of 5-methylcytosine substitutions in whole mRNA length (**mRNA-full**) were detrimental to translation, most likely due to IRES tertiary structure disruption

and loss of ribosome binding (Figure 102). The best performing candidates – two mRNAs with single methyl-modification (**mRNA-init**, **mRNA-mid**) and natural mRNA (**mRNA-nat**) were tested in *in cellulo* translation studies with HEK293T cells, along with two negative controls. Control A represents signal from cells treated only with Lipofectamine MessengerMAX and NanoLuc substrate. Control B represents signal from non-treated cells (neither transfection agent nor mRNA) upon NanoLuc substrate addition. Upon 4 h incubation luciferase signal was measured. Protein yields from *in cellulo* translations followed the same pattern as in *in vitro* experiments (Figure 102). Single modifications introduced in either initial (**mRNA-init**) or middle (**mRNA-mid**) region of the gene-coding part of mRNA elicit significant translation enhancement in comparison to natural mRNA (**mRNA-nat**). Here, for the first time, I revealed that single 5-methylcytosine modifications within gene-coding region of mRNA led to significant enhancement of translation efficiency when compared either to unmodified (**mRNA-nat**) or fully modified mRNA (**mRNA-full**).

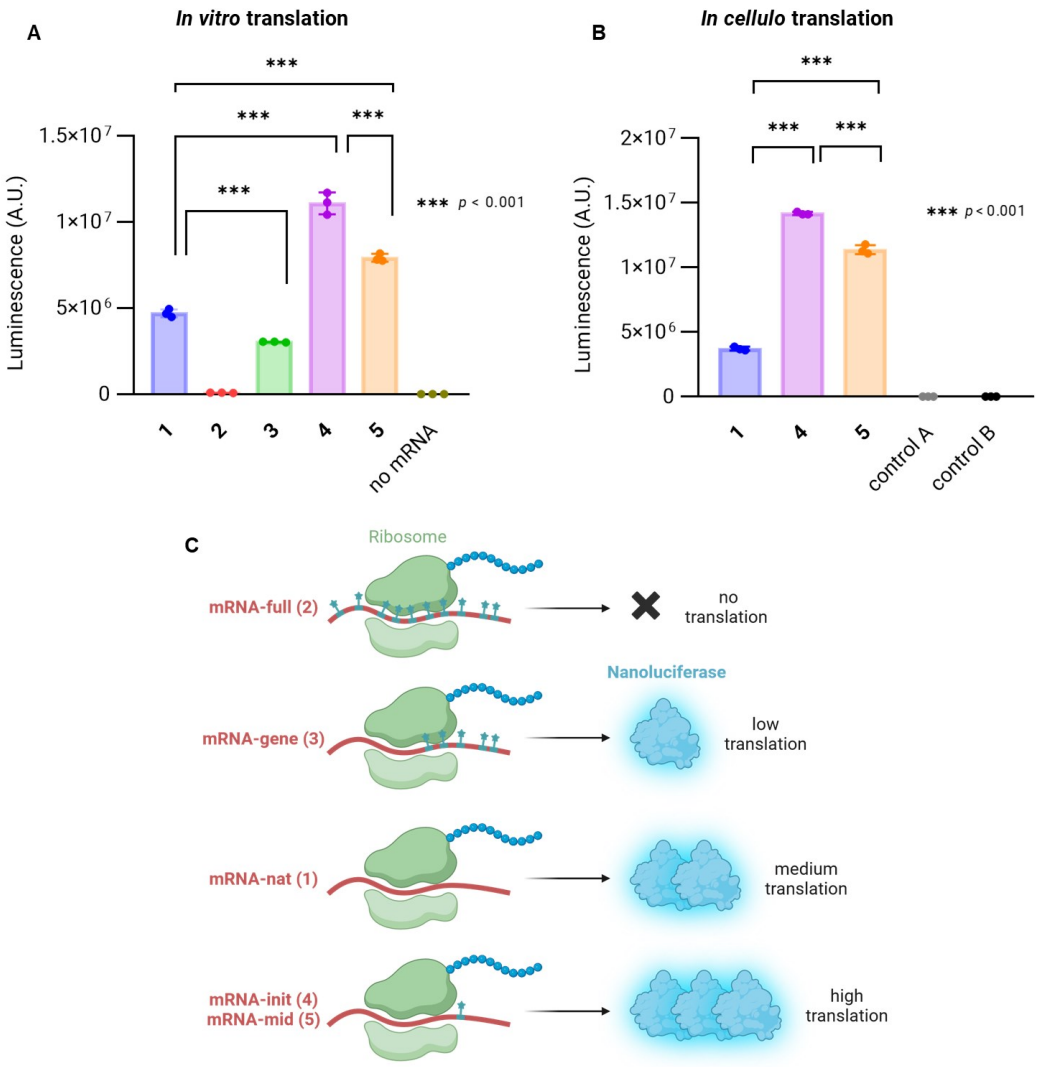



**Figure 102. A – mRNA constructs tested in *in vitro* translation studies with rabbit reticulocyte lysate system. B – mRNA constructs tested in *in cellulo* translation studies upon transfection into HEK293T cells. Graph legend: (1) mRNA-nat; (2) mRNA-full; (3) mRNA-gene; (4) mRNA-init; (5) mRNA-mid; (no mRNA) absence of any mRNA; (control A) cells treated only with Lipofectamine MessengerMAX; (control B) non-treated cells (neither transfection agent nor mRNA). C – Summary of luciferase translation assays outcome.**

### 3.2.10 Conclusion

In summary, we successfully synthesised a set of hydrophobic modifications of increasing bulkiness, including clickable ethynyl functionality, attached to each 7-deazapurine and pyrimidine nucleobase, six nucleotides decorated with reactive CA and FT groups, and one fluorescently tagged nucleoside triphosphate. Apart from the synthesised ones, I expanded the range of tested nucleobase-modified triphosphates with commercial **rN$^{X}$TP**s bearing fluorescent or affinity tags and even epitranscriptomic modifications.

I disclosed a novel strategy for synthesis of on-demand modified RNA based on employing engineered thermostable primer-dependent DNA polymerases. I demonstrated superior performance of these enzymes over conventional T7 RNAP in accepting manifold of nucleobase-modified **rN$^{X}$TP**s. Furthermore, I conducted multisite-specific labelling of RNA, even in previously unattainable homopolymeric sequences, with distinct tags for structural studies.

In analogy to principles of enzymatic DNA synthesis, this novel approach delivers flexible choice of primer chemistry and 5'-terminus labelling. Taking advantage of primer chemistry flexibility, I devised a method whereby synthetic DNA primer can be trimmed off to scarlessly leave behind RNA entirely built of modified nucleotides.

Finally, I probed limits of engineered DNA polymerase for long-range mRNA synthesis. By combining SNI and PEX techniques I successfully incorporated a single **rC$^{Me}$** modification at internal position of long mRNA sequence.

# 4 Conclusion

The primary objective of the first part of this thesis was to design and synthesise a novel CA-modified ribonucleoside triphosphate, that would serve as an important building block for construction of base-modified RNA probes used for selective Cys or His targeting within RBPs. Synthesis based on one-step aqueous Pd-catalysed Sonogashira cross-coupling reaction resulted in the desired **rA$^{CA}$TP** with satisfactory yield.

The modified nucleotide building block served as a good substrate for IVT with T7 RNAP. The IVT enabled successful synthesis of short RNA probes containing single CA modification, as well as longer sequences with up to seven modifications in one strand. Any potential inhibition of RNA polymerase resulting from undesired cross-linking with either **rA$^{CA}$TP** itself or during incorporation into RNA was excluded. Successful formation of modified RNA products was confirmed by dPAGE and MS analysis. Moreover, hazardous radioactive RNA labelling was replaced by a fluorescent labelling method facilitated by ligation reaction.

The reactivity of CA-modified RNA probes was first demonstrated on model bioconjugation reactions with small (bio)molecules and peptides. The CA-modified RNA selectively reacted with Cys and His residues in peptides, excluding any possible reaction with other nucleophilic amino acids, like Lys or Arg. As expected, greater reactivity was noted with Cys-containing peptide when compared to the one with less nucleophilic His residue. Next by EMSA analysis I provided evidence that the presence of CA moiety in RNA does not interfere with binding of proteins. I demonstrated enhanced cross-linking efficiency of CA-modified RNA probe with three distinct model RBPs, through significantly increased reaction conversions (22-30%), surpassing those attained by conventional methods. Additionally, formation of stable RNA-protein conjugates was confirmed by intact ESI-MS or western blot analysis. By digestion of RNA-protein conjugates, first by proteases and in the next step by application of HF-mediated chemical digestion, I was able to pinpoint exact RNA binding sites within proteins and precisely analyse the cross-linked amino acid residues. In addition, I revealed that our CA-modified RNA probe is capable of sequence-selective targeting protein of interest within a complex protein mixture. Thus, it proved to be effective for cross-linking reactions and identification of RBPs from cell lysate. Furthermore, I showed preserved selectivity towards RBPs by demonstrating negligible cross-linking efficiency when applied to weak- and/or non-RBPs. This observation highlights the specificity of the method, as it exhibited reduced cross-linking yields with proteins with no affinity for RNA.

By utilising a highly efficient CA chemical cross-linker, we have surpassed the limitations associated with conventional cross-linking methods. This innovative method lays the first stone

towards construction of reactive probes designed to selectively target Cys and His amino acids within RBPs. It sets the stage for extending the method to identify previously unexplored regulatory factors, such as poly(A) tail binding proteins. Additionally, the approach holds promise for in-cell RNA cross-linking studies and advances in the field of proteomics. This marks a significant contribution to the exploration and understanding of cellular regulatory mechanisms.

In the second part of my thesis, I developed a novel method for alternative enzymatic synthesis of modified RNA leveraging engineered thermostable primer-dependent DNA polymerases. To assess the range of modifications accepted by engineered mutant DNA polymerases (TGK or SFM4-3), I tested 25 variously sterically demanding modified **rN$^{X}$TP**s bearing hydrophobic, fluorescent, reactive and epitranscriptomic modifications along with affinity tags.

Both enzymes exhibited notable proficiency in accepting a diverse array of nucleobase-modified **rN$^{X}$TP**s, facilitating the synthesis of RNA sequences encoding for one or several modified nucleotides. Although, in case of PEX with **rA$^{CA}$TP** and **rC$^{CA}$TP** formation of undesirable cross-linking products was observed due to long incubation times at high temperature, this was effectively solved by either reducing the reaction time or the temperature. This highlights the adaptability of mutant polymerases to function proficiently even at lower temperatures, thereby enabling incorporation of sensitive (reactive) functional groups. Successful synthesis of all modified RNA probes was proved by dPAGE analysis. Generation of ssRNA probes, ensured by DNA template removal either under mild conditions with TurboDNase or by magnetoseparation, was proved by MS analysis. More importantly, I revealed the superior performance of these engineered enzymes compared to conventional T7 RNAP in accommodating a variety of **rN$^{X}$TP**s. By multiple combinations thereof, I synthesised even previously elusive highly functionalized 98 nt long RNA composed of four different nucleobase-modified triphosphates with four distinct functionalities, that was successfully reverse transcribed into cDNA, a critical step for preserving the flow of genetic information. Moreover, our methodology provides a versatile range of options for selecting primer chemistry (RNA or DNA) and enables the labelling of the 5'-terminus with numerous functionalities (FAM, Cy5, biotin, etc.). By utilising the flexibility of primer chemistry, I devised a novel method where synthetic DNA primer containing single dU modification can be precisely trimmed off from synthesised RNA-DNA hybrids by assistance of UDG and DMEDA. The outcome is the scarless removal of the synthetic DNA primer, leaving behind RNA entirely composed of modified nucleotides. In addition to these achievements, I conducted multisite-specific labelling of RNA, even in previously inaccessible homopolymeric sequences. Through the combination of SNI and PEX, I successfully prepared dual or even triple-labelled RNA molecules at specific positions, employing distinct fluorescent tags for in-depth structural studies using FRET. To the best of my knowledge,

only one publication has explored the ligation of synthetic and enzymatically generated fragments, offering access to mRNA containing a single 2'-*O*-methyl nucleotide[152]. However, this method encountered challenges such as low yields (3-7%) and laboruis gel purifications. Additionally, the technique constrained the labelling position to be in close proximity to the mRNA termini. Leveraging our novel approach, I unravelled the importance of the precise positioning of single **rC$^{Me}$** modifications within mRNA. Furthermore, in comparison to the current state-of-the-art ligation method, this approach yielded modified mRNA in notably high yields and excluded formation of harmful dsRNA by-products, as common in conventional IVT reactions. Point-modified mRNA significantly improved translation compared to its natural RNA counterpart either in *in vitro* or in *cellulo* studies. This clearly emphasises the importance of determining not only the type, but also the sequence context of modifications within mRNA.

Thus, I delineate the evolution of several novel innovative methods and protocols designed for the synthesis of variously base-modified RNA. These advancements not only represent significant contributions to the nucleic acids field, but also lay the first stone for potential future applications and explorations in RNA biochemistry, biotechnology, or RNA therapeutics.

# 5 Experimental Section

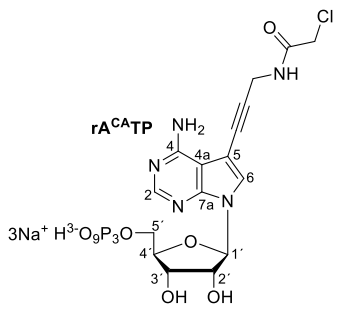
## 5.1 Chloroacetamide-modified nucleotide and RNA for bioconjugations and cross-linking with cysteine- or histidine-containing peptides and RNA-binding proteins

### 5.1.1 General remarks – Synthetic part

All solvents and reagents were purchased from commercial suppliers (Fluorochem, Sigma Aldrich, Lach-Ner) and used as received without further purification, unless otherwise specified. Phosphoryl chloride ($POCl_3$) and trimethyl phosphate [$PO(OMe)_3$] were distilled prior to use. Reactions were performed in heat gun-dried glassware under argon atmosphere. Milli-Q water was used in the synthetic part. Reactions were monitored by thin layer chromatography (TLC) on TLC Silica gel 60 F254 (Merck) and detected by UV (254 nm) and by Advion Expression Compact Mass Spectrometer connected with Plate Express® TLC Plate Reader using electrospray ionization (ESI). NMR spectra were measured on a Bruker AVANCE III HD 500 spectrometer (1H at 500.0 MHz, 13C at 125.7 MHz, 31P at 202.4 MHz) in $D_2O$ (referenced to tert-butanol) at 25°C. Chemical shifts are given in ppm (δ-scale) and coupling constants (J) are in Hz. Multiplicity of the peaks is as followed: s = singlet, d = doublet, t = triplet, dd= doublet of doublets, ddd = doublet of doublets of doublets, bpent = broad pentet. Complete assignment of all NMR signals was achieved by using a combination of H,H-COSY, H,C-HSQC and H,C-HMBC experiments. Labelling of NMR signals assignments corresponds to a numbering depicted in compound formulas. High-resolution mass spectra were measured on LTQ Orbitrap XL (Thermo Fisher Scientific). All mass spectra were acquired by the MS service at IOCB. Column chromatography was performed using silica gel (40-63 µm, Fluorochem) by flash liquid chromatography system (FLC) Teledyne ISCO Combi Flash Rf 200 or 300. Purification of the ribonucleoside triphosphate was performed using HPLC (Waters modular HPLC system) on a Phenomenex Kinetex 5 µm EVO C18 100 Å, AXIA Packed LC column (250×21.2 mm) and POROS HQ 50 column (packed in-house, 26×120 mm). Purity of all final compounds was determined by NMR spectroscopy.

### 5.1.2 Chemical synthesis of 7-{3-[N-(2-chloroacetamido)]-prop-2-yn-1-yl}-7-deazaadenosine-5'-*O*-triphosphate (rA$^{CA}$TP)

Iodo-modified ribonucleoside triphosphate in sodium form **rA$^{I}$TP**[260] (30 mg, 43.0 µmol, 1 equiv.), *N*-(propargyl)chloroacetamide[261] (14.1 mg, 107.2 µmol, 2.5 equiv.), CuI (0.82 mg, 4.3 µmol, 10 mol%), $Pd(OAc)_2$ (0.48 mg, 2.1 µmol, 5 mol%) and TPPTS (2.44 mg, 4.3 µmol, 10 mol%) were




dissolved in a mixture of $H_2O$/acetonitrile (2:1, 1.5 mL) in a microwave vial containing a stir bar under argon atmosphere. *N*,*N*′-diisopropylethylamine (75 µL, 0.43 mmol, 10 equiv.) was added *via* syringe and the mixture was stirred under argon atmosphere at 60 °C for 4 h. After the reaction, the mixture was cooled down, the solvents were removed under reduced pressure and the residue was again dissolved in 5 mL of $H_2O$ and filtrated through 0.22 µm HPLC filter prior to purification. The HPLC separation was performed on C18 reversed phase column using linear gradient from 0.1 M TEAB (aq) to 0.1 M TEAB in 50% MeOH followed by several co-evaporations with $H_2O$. The second purification was performed on POROS HQ 50 column using linear gradient from $H_2O$ to 400 mM TEAB (aq) followed by several co-evaporations with $H_2O$. The pure product was isolated as a triethylammonium salt and converted to sodium salt using Dowex 50WX8 in $Na^+$ cycle. The solvent was evaporated, pure product was again dissolved in small amount of $H_2O$ and freeze-dried overnight. The desired product **rA$^{CA}$TP** (as sodium salt) was obtained as a white solid powder (7.3 mg, 24%).

**$^1$H NMR (500 MHz, $D_2O$, ref(*t*BuOH) = 1.25 ppm):** 4.16 (ddd, 1H, $J_{gem}$ = 11.7 Hz, $J_{5'a,P}$ = 4.6 Hz, $J_{5'a,4'}$ = 3.2 Hz, H-5'a); 4.20 (s, 2H, $CH_2Cl$); 4.28 (ddd, 1H, $J_{gem}$ = 11.7 Hz, $J_{5'b,P}$ = 6.6 Hz, $J_{5'b,4'}$ = 3.0 Hz, H-5'b); 4.29 (s, 2H,C≡CC**H$_2$**); 4.36 (bpent, 1H, $J_{4',3'} = J_{4',5'a} = J_{4',5'b} = J_{4',P}$ = 2.9 Hz, H-4'); 4.58 (dd, 1H, $J_{3',2'}$ = 5.3 Hz, $J_{3',4'}$ = 2.8 Hz, H-3'); 4.68 (dd, 1H, $J_{2',1'}$ = 6.8 Hz, $J_{2',3'}$ = 5.3 Hz, H-2'); 6.24 (d, 1H, $J_{1',2'}$ = 6.8 Hz, H-1'); 7.78 (s, 1H, H-6); 8.15 (s, 1H, H-2).

**$^{13}$C NMR (125.7 MHz, $D_2O$, ref(*t*BuOH) = 31.60 ppm):** 32.28 (C≡C**C**H$_2$); 44.13 ($CH_2Cl$); 67.40 (d, $J_{C,P}$ = 5.5 Hz, $CH_2$-5'); 72.54 (CH-3'); 75.85 (CH-2'); 77.30 (**C**≡CCH$_2$); 85.84 (d, $J_{C,P}$ = 9.1 Hz, CH-4'); 87.78 (CH-1'); 89.75 (C≡**C**CH$_2$); 98.46 and 105.14 (C-5,4a); 128.39 (CH-6); 151.35 (C-7a); 154.39 (CH-2); 159.46 (C-4); 171.89 (NHCO).

**$^{31}$P NMR (202.4 MHz, $D_2O$):** –21.63 (t, *J* = 19.8 Hz, $P_\beta$); –10.46 (d, *J* = 19.8 Hz, $P_\alpha$); –7.14 (d, *J* = 19.8 Hz, $P_\gamma$).

**HR MS (ESI$^-$): $C_{16}H_{18}O_8N_5(^{35}Cl)P$** calculated: 474.05870, found: 474.05803; **$C_{16}H_{19}O_{11}N_5(^{35}Cl)P_2$** calculated: 554.02503, found: 554.02420; **$C_{16}H_{18}O_{11}N_5(^{35}Cl)NaP_2$:** calculated: 576.00670, found: 576.00594.

### 5.1.3 General remarks – Biochemical part

All PAGE gels were analysed by fluorescence and/or phosphor imaging using Typhoon FLA 9500 (GE Healthcare Life Sciences). Mass spectra of oligonucleotides were measured on UltrafleXtreme MALDI-TOF/TOF (Bruker) mass spectrometer with 1 kHz smartbeam II laser technology. The matrix consisted of 3-hydroxypicolinic acid (HPA)/picolinic acid (PA)/ammonium tartrate in ratio 9:1:1. The matrix (1 µL) was applied on the target (ground steel) and dried down at room temperature. The sample (1 µL) and the matrix (1 µL) were mixed and added on the top of the dried matrix preparation spot and dried down at room temperature. Samples were concentrated on CentriVap vacuum concentrator system (Labconco) and lyophilized on FreeZone 2.5 L freeze dryer (Labconco). Fluorescence spectra were measured in a 100 µL quartz cuvette at room temperature on a Fluoromax 4 spectrofluorometer (HORIBA Scientific). Single stranded DNA oligonucleotides for preparation of double stranded DNA templates were purchased from Generi Biotech. DNase I, T7 RNA polymerase with the corresponding transcription buffer, RiboLock RNase Inhibitor, proteases and C18 spin columns were obtained from Thermo Fisher Scientific. The natural ribonucleoside triphosphates (rNTPs), protein ladder [Color Prestained Protein Standard, Broad Range (10-250 kDa)], T4 RNA ligase 1 and the corresponding ligase buffer and additives for ligation reaction and Monarch RNA purification kits (10 µg and 50 µg) were purchased from New England Biolabs. ZR small-RNA PAGE recovery kit was purchased from Zymo research. Microspin G-25 columns and Amicon ultra-0.5 centrifugal concentrator (10 kDa MWCO) were purchased from Merck. [α-32P]-GTP (111 TBq/mmol, 370 MBq/mL) was obtained from MGP. The pCp-Cy5 (cytidine-5'-phosphate-3'-(6-aminohexyl)phosphate, labelled with Cy5) and pCp-biotin (cytidine-5'-phosphate-3'-(6-aminohexyl)phosphate, labelled with biotin) were purchased from Jena Bioscience. Fluorescein-thiol was purchased from BioActs and Biotin-thiol from Polypure. Peptides were prepared by automated solid phase synthesis according to standard procedures, HPLC purified and characterized by MS-MALDI-TOF at IOCB. RNase/DNase free solutions for biochemical reactions were prepared using Milli-Q water, that was treated with DEPC and sterilized by autoclaving. Concentrations of the prepared RNA solutions were calculated using extinction coefficients obtained from on-line tool at https://www.atdbio.com/tools/oligo-calculator and A260 values measured on Nanodrop 1000 (Thermo Fischer Scientific). Intact ESI-MS analysis was carried out on AQUITY UPLC I-Class system (Waters) coupled to mass spectrometer Synapt G2 (Waters). The nano-LC-MS/MS analysis was performed on UltiMate 3000 RSLCnano system (Thermo Fisher Scientific) coupled to a mass spectrometer Orbitrap Fusion Lumos Tribrid (Thermo Fisher Scientific). LC-ESI-MS spectra were acquired on Agilent 1290 Infinity II Bio system with DAD detector and mass spectrometer MSD XT. LC-ESI-MS analysis of oligonucleotides were carried out according to standard procedures using

mobile phases A (12.2 mM $Et_3N$, 300 mM HFIP in $H_2O$) and B (12.2 mM $Et_3N$, 300 mM HFIP in $H_2O$ in 100% MeOH) by 10 min gradient from 5% B to 100% B in A using bioZen 1.7 μm oligo column 2.1×50 mm (Phenomenex) on Agilent UHPLC Bio system. Deconvolutions of LC-ESI-MS spectra were carried out using UniDec program.

### 5.1.4 General procedures

**General purification procedure I (silica spin columns)**

The samples (adjusted with $H_2O$ to final volume 50 μL) were purified using the Monarch RNA purification kit (50 μg), according to standard supplier's protocol. Elution of the samples was performed into DNase/RNase free tube with 50 μL of $H_2O$.

**General purification procedure II (gel filtration spin columns)**

The Microspin G-25 columns were pre-washed with 2X 500 μL of $H_2O$. The analysed sample (adjusted with $H_2O$ to final volume 50 μL) was loaded on the column and purified according to standard supplier's protocol. Elution of the samples was performed into DNase/RNase free tube.

**General purification procedure III (gel extraction)**

The samples after preparative scale transcription reaction (400 μL) were mixed with 400 μL of 2X stop solution (95% [v/v] formamide, 0.5 mM EDTA, 0.025% [w/v] bromophenol blue, 0.025% [w/v] SDS in $H_2O$) and denatured by heating at 95 °C for 2 min followed by cooling on ice. Samples were loaded on preparative 22.5% dPAGE (19:1 mono:bis acrylamide) containing 1X TBE buffer (pH 8) and urea (7 M). Separation was performed at 20 mA for 2-3 h (until the dye migrated to the bottom third of the gel). The RNA transcripts were visualized by UV shadowing (254 nm) and the purification was performed according to standard supplier's protocol using ZR small-RNA PAGE recovery kit (Zymo research). Elution of the RNA was performed into DNase/RNase free tube with 2X 20 μL of $H_2O$.

**Preparation and quantification of HeLa cell lysate**

Cells ($3\times10^7$) were thawed on ice and resuspended in 150 μL of lysis buffer (50 mM Tris, 150 mM NaCl, 1% Triton X-100, 5 mM EDTA, pH 7.4) containing HaltProtease inhibitor (1X, Thermo Fisher Scientific). Cell disruption was carried out by sonication by 3X 30 sec of 100% power bursts on ice (Hielscher digital ultrasonic generator UP200St) with tip probe. Cellular debris was removed at

20000X g for 20 min at 4 °C. Proteins were quantified by QuantiPro BCA assay kit (Sigma Aldrich) with BSA as a standard.

**Preparation dsDNA templates for transcription reaction**

A solution of complementary ssDNA oligonucleotides (10 μL, 100 μM of each) in annealing buffer (10 mM Tris, 50 mM NaCl, 1 mM EDTA, pH 7.8) was heated up to 95 °C for 5 min in a thermal cycler (with heated lid to 105 °C) and then slowly cooled down to 25 °C (0.02 °C/$s^{-1}$).

**General procedure for dPAGE analysis of IVT reaction**

The samples after transcription reaction (10 μL) were stopped by mixing with 2X stop solution (10 μL) containing 95% [v/v] formamide, 0.5 mM EDTA, 0.025% [w/v] bromophenol blue, 0.025% [w/v] SDS in $H_2O$. Samples were denatured by heating at 65 °C for 10 min and then immediately cooled on ice. Aliquots of the samples (10 μL) were subjected to vertical gel electrophoresis on dPAGE (19:1 mono:bis acrylamide) containing 1X TBE buffer (pH 8) and urea (7 M) at 42 mA for 30-40 min (until the dye migrated to the bottom third of the gel). The gel was then autoradiographed at least for 1 h and then visualized by phosphor imager.

**General procedure for dPAGE analysis of bioconjugations**

The reaction (10 μL) mixture was combined with 10 μL of 2X stop solution (95% [v/v] formamide, 0.5 mM EDTA, 0.025% [w/v] bromophenol blue, 0.025% [w/v] SDS in H2O) and denatured by heating at 95 °C for 2 min and immediately cooled on ice. Aliquots of the samples (10 μL) were subjected to vertical gel electrophoresis on 22.5% dPAGE (19:1 mono:bis acrylamide) containing 1X TBE buffer (pH 8) and urea (7 M) at 20 mA for 2-3 h (until the dye migrated to the bottom third of the gel). The gel was visualized by a fluorescent scanner using Cy5 channel scan.

**General procedure for EMSA analysis**

The samples (10 μL) were combined with 3 μL of sterile glycerol and analysed by 5% (0.5X TBE) native-PAGE (37.5:1 mono:bis acrylamide) in 0.5X TBE running buffer at 60 V for 2 h with cooling (4 °C). The gel was visualized by a fluorescent scanner using Cy5 channel scan.

**General procedure for RNA-protein cross-linking analysis by SDS-PAGE**

Samples (10 μL) were combined with 2.5 μL of 5X SDS stop solution (500 mM DTT, 150 mM Tris, 20 mM EDTA, 10% SDS, 50% glycerol, pH 6.8) denatured at 95 °C for 5 min and analysed by

denaturing Tris-Gly-SDS-PAGE (37.5:1 mono:bis acrylamide, 375 mM Tris, 0.1% SDS) in 1X Tris-Gly-SDS running buffer at 180 V for 1.5 h at room temperature. The gel was visualized by a fluorescent scanner using Cy5 channel scan.

**General procedure for intact ESI-MS analysis of RNA-protein conjugates**

The sample was injected onto a MassPREP Micro desalting column (20 µm, 5×2.1 mm ID, Waters), desalted and eluted by fast gradient (4 min). Mobile phase A (10 mM ammonium acetate, aq, pH 9.0) and mobile phase B (acetonitrile) were used for elution. Separation was carried out by AQUITY UPLC I-Class system on-line coupled to Mass Spectrometer Synapt G2 to acquire mass spectra using electrospray ionization in positive mode. TOF mass range was set from m/z = 500 to 4000. Raw spectrum was subtracted and deconvoluted (MaxEnt1, Waters) to produce the final spectrum.

**General procedure I for nano-LC-MS/MS analysis of proteolytic digests**

The sample was dissolved in 15 µL of 0.1% trifluoroacetic acid (TFA, aq) and 5 µL aliquot of the sample was injected on an UltiMate 3000 RSLC nano system coupled to a mass spectrometer Orbitrap Fusion Lumos Tribrid. The peptides were trapped on a PepMap100 column (5 µm, 5 mm by 300 µm internal diameter (ID), Thermo Fisher Scientific) and desalted with 2% acetonitrile in 0.1% formic acid (FA, aq) at a low rate of 5 µL/min. Eluted peptides were separated using an EASY-Spray PepMap100 C18 analytical column (2 µm, 50 cm by 75 µm ID, Thermo Fisher Scientific). The 30 min elution gradient at a constant flow rate of 300 nL/min was set to start at 5% phase B (0.1% FA in 99.9% acetonitrile) and 95% phase A (0.1% FA, aq). Then the content of acetonitrile was increased gradually up to 50% of phase B. The Orbitrap mass range was set from *m*/*z* = 350 to 2000 in the MS mode and for ions with a charge state 2-6 the fragmentation spectra were acquired. A Proteome Discoverer 2.5 (Thermo Fisher Scientific) was used for peptide and protein identification using Sesquest HS and MS Amanda as search engines and databases of protein sequences and common contaminants.

### 5.1.5 Analysis of cross-linking of modified 20RNA_1A$^{CA}$ to T7 RNA polymerase by denaturing SDS-PAGE

*In vitro* transcription reactions were performed in total volume of 20 µL in transcription reaction buffer (5X, 4 µL) containing either natural rATP (2 mM) or modified **rA$^{CA}$TP** (2 mM), three natural rNTPs (2 mM, rCTP, rUTP, rGTP), $MgCl_2$ (25 mM), RiboLock RNase inhibitor (1 U/µL), Triton X-100 (0.1%), dsDNA template (1.5 µM, **20DNA_1A**), T7 RNA polymerase (5 U/µL) and [α-$^{32}$P]-GTP (111 TBq/mmol, 370 MBq/mL, 0.2 µL). Transcription reactions were performed at 37 °C for 2 h in

a thermal cycler with heated lid (75 °C). After this time, 10 μL aliquots of natural and modified RNAs were combined with 2.5 μL of 5X SDS stop solution (500 mM DTT, 150 mM Tris, 20 mM EDTA, 10% SDS, 50% glycerol, pH 6.8) and denatured at 95 °C for 5 min. The remaining 10 μL aliquots of natural and modified RNAs were combined with DNase I (1 μL, 1 U/μL) and further incubated at 37 °C for 15 min. After that, the samples were purified using Monarch RNA purification kit (10 μg), according to standard supplier's protocol and eluted in 10 μL of $H_2O$. Eluted samples were combined with 2.5 μL of 5X SDS stop solution (500 mM DTT, 150 mM Tris, 20 mM EDTA, 10% SDS, 50% glycerol, pH 6.8) and denatured at 95 °C for 5 min. All purified and non-purified samples of natural and modified RNAs were analysed by 10% denaturing Tris-Gly-SDS-PAGE (37.5:1 mono:bis acrylamide, 375 mM Tris, 0.1% SDS) in 1X Tris-Gly-SDS running buffer at 180 V for 1 h at room temperature. The gel was then autoradiographed at least for 1 h and then visualized by phosphor imager.

### 5.1.6 Denaturing PAGE analysis of inhibition of *in vitro* transcription reaction by $rA^{CA}TP$

*In vitro* transcription reactions were performed in total volume of 10 μL in transcription reaction buffer (5X, 2 μL) containing all four natural rNTPs (2 mM, rATP, rCTP, rUTP, rGTP), $MgCl_2$ (25 mM), RiboLock RNase inhibitor (1 U/μL), Triton X-100 (0.1%), dsDNA template (1.5 μM, **20DNA_1A**), T7 RNA polymerase (5 U/μL) and [α-$^{32}$P]-GTP (111 TBq/mmol, 370 MBq/mL, 0.1 μL) either with or without addition of **$rA^{CA}TP$** (2 mM). Transcription reactions were performed at 37 °C for 2 h in a thermal cycler with heated lid (75 °C). Samples were analysed by 20% dPAGE.

### 5.1.7 Incorporation of $rA^{CA}TP$ using 20DNA_1A template in analytical scale for dPAGE analysis

*In vitro* transcription reactions were performed in total volume of 10 μL in transcription reaction buffer (5X, 2 μL) containing modified **$rA^{CA}TP$** (2 mM), three natural rNTPs (2 mM, rCTP, rUTP, rGTP), $MgCl_2$ (25 mM), RiboLock RNase inhibitor (1 U/μL), Triton X-100 (0.1%), dsDNA template (1.5 μM, **20DNA_1A**), T7 RNA polymerase (5 U/μL) and [α-$^{32}$P]-GTP (111 TBq/mmol, 370 MBq/mL, 0.1 μL). The negative control experiment was performed under the same conditions with $H_2O$ used instead of the solution of modified **$rA^{CA}TP$** and with natural rATP (2 mM) in case of positive control. Transcription reactions were performed at 37 °C for 2 h in a thermal cycler with heated lid (75 °C). Samples were analysed by 20% dPAGE.

### 5.1.8 Incorporation of rA$^{CA}$TP using 35DNA_1A template in analytical scale for dPAGE analysis

*In vitro* transcription reactions were performed in total volume of 10 μL in transcription reaction buffer (5X, 2 μL) containing modified **rA$^{CA}$TP** (2 mM), three natural rNTPs (2 mM, rCTP, rUTP, rGTP), $MgCl_2$ (15 mM), RiboLock RNase inhibitor (1 U/μL), Triton X-100 (0.1%), dsDNA template (1.25 μM, **35DNA_1A**), T7 RNA polymerase (4 U/μL) and [α-$^{32}$P]-GTP (111 TBq/mmol, 370 MBq/mL, 0.1 μL). The negative control experiment was performed under the same conditions with $H_2O$ used instead of the solution of modified **rA$^{CA}$TP** and with natural rATP (2 mM) in case of positive control. Transcription reactions were performed at 37 °C for 2 h in a thermal cycler with heated lid (75 °C). Samples were analysed by 12.5% dPAGE.

### 5.1.9 Incorporation of rA$^{CA}$TP using 35DNA_3A and/or 35DNA_7A template in analytical scale for dPAGE analysis

*In vitro* transcription reactions were performed in total volume of 10 μL in transcription reaction buffer (5X, 2 μL) containing modified **rA$^{CA}$TP** (3 mM), three natural rNTPs (2 mM, rCTP, rUTP, rGTP), $MgCl_2$ (15 mM), RiboLock RNase inhibitor (1 U/μL), Triton X-100 (0.1%), dsDNA template (1.25 μM, **35DNA_3A** or **35DNA_7A**), T7 RNA polymerase (6 U/μL) and [α-$^{32}$P]-GTP (111 TBq/mmol, 370 MBq/mL, 0.1 μL). The negative control experiment was performed under the same conditions with $H_2O$ used instead of the solution of modified **rA$^{CA}$TP** and with natural rATP (3 mM) in case of positive control. Transcription reactions were performed at 37 °C for 2 h in a thermal cycler with heated lid (75 °C). Samples were analysed by 12.5% dPAGE.

### 5.1.10 Enzymatic synthesis of 20RNA_1A or 20RNA_1A$^{CA}$ in semi-preparative scale

*In vitro* transcription reactions were performed in total volume of 50 μL in transcription reaction buffer (5X, 10 μL) containing either natural rATP (2 mM) or modified **rA$^{CA}$TP** (2 mM), three natural rNTPs (2 mM, rCTP, rUTP, rGTP), $MgCl_2$ (25 mM), RiboLock RNase inhibitor (1 U/μL), Triton X-100 (0.1%), dsDNA template (1.5 μM, **20DNA_1A**) and T7 RNA polymerase (5 U/μL). Transcription reactions were performed at 37 °C for 2 h in a thermal cycler with heated lid (75 °C). After this time, DNase I (5 μL, 1 U/μL) was added to the mixture and the solution was further heated for 30 min at 37 °C in a thermal cycler. The samples were purified (General purification procedure I).

### 5.1.11 Enzymatic synthesis of 35RNA_1A or 35RNA_1A$^{CA}$ in semi-preparative scale

*In vitro* transcription reactions were performed in total volume of 30 μL in transcription reaction buffer (5X, 6 μL) containing either natural rATP (2 mM) or modified **rA$^{CA}$TP** (2 mM), three natural rNTPs (2 mM, rCTP, rUTP, rGTP), $MgCl_2$ (15 mM), RiboLock RNase inhibitor (1 U/μL), Triton X-100 (0.1%), dsDNA template (1.25 μM, **35DNA_1A**) and T7 RNA polymerase (4 U/μL). Transcription reactions were performed at 37 °C for 2 h in a thermal cycler with heated lid (75 °C). After this time, DNase I (3 μL, 1 U/μL) was added to the mixture and the solution was further heated for 30 min at 37 °C in a thermal cycler. The samples were purified (General purification procedure I).

### 5.1.12 Enzymatic synthesis of 35RNA_3A$^{CA}$ and 35RNA_7A$^{CA}$ in semi-preparative scale

*In vitro* transcription reactions were performed in total volume of 30 μL in transcription reaction buffer (5X, 6 μL) containing modified **rA$^{CA}$TP** (3 mM), three natural rNTPs (2 mM, rCTP, rUTP, rGTP), $MgCl_2$ (15 mM), RiboLock RNase inhibitor (1 U/μL), Triton X-100 (0.1%), dsDNA template (1.25 μM, **35DNA_3A** or **35DNA_7A**) and T7 RNA polymerase (6 U/μL). Transcription reactions were performed at 37 °C for 2 h in a thermal cycler with heated lid (75 °C). After this time, DNase I (3 μL, 1 U/μL) was added to the mixture and the solution was further heated for 30 min at 37 °C in a thermal cycler. The samples were purified (General purification procedure I).

### 5.1.13 Enzymatic synthesis of 21RNA_3A-bind, 21RNA_3A$^{CA}$-bind and 21RNA_3A-non-bind, 21RNA_3A$^{CA}$-non-bind in preparative scale

*In vitro* transcription reactions were performed in total volume of 50 μL in transcription reaction buffer (5X, 10 μL) containing modified **rA$^{CA}$TP** (2 mM), three natural rNTPs (2 mM, rCTP, rUTP, rGTP), $MgCl_2$ (25 mM), RiboLock RNase inhibitor (1 U/μL), Triton X-100 (0.1%), dsDNA template (1.5 μM, either **21DNA_3A-bind** or **21DNA_3A-non-bind**) and T7 RNA polymerase (5 U/μL). The positive control experiment was performed under the same conditions with natural rATP (2 mM) used instead of the solution of modified **rA$^{CA}$TP**. Transcription reactions were performed at 37 °C for 2 h in a thermal cycler with heated lid (75 °C). After this time, DNase I (5 μL, 1 U/μL) was added to the mixture and the solution was further heated for 30 min at 37 °C in a thermal cycler. The samples were purified (General purification procedure I).

### 5.1.14 Enzymatic synthesis of 20RNA_1A or 20RNA_1A$^{CA}$ in preparative scale

*In vitro* transcription reactions were performed in total volume of 400 μL in transcription reaction buffer (5X, 80 μL) containing either natural rATP or modified **rA$^{CA}$TP** (2 mM), three natural rNTPs (2 mM, rCTP, rUTP, rGTP), $MgCl_2$ (25 mM), RiboLock RNase inhibitor (1 U/μL), Triton X-100 (0.1%), dsDNA template (1.5 μM, **20DNA_1A**1) and T7 RNA polymerase (5 U/μL). Transcription reactions were performed at 37 °C for 2 h in a thermal cycler with heated lid (75 °C). After this time, DNase I (80 μL, 1 U/μL) was added to the mixture and the solution was further heated for 30 min at 37 °C in a thermal cycler. The samples were purified (General purification procedure III).

### 5.1.15 Enzymatic synthesis of 35RNA_1A or 35RNA_1A$^{CA}$ in preparative scale

*In vitro* transcription reactions were performed in total volume of 400 μL in transcription reaction buffer (5X, 80 μL) containing either natural rATP or modified **rA$^{CA}$TP** (2 mM), three natural rNTPs (2 mM, rCTP, rUTP, rGTP), $MgCl_2$ (15 mM), RiboLock RNase inhibitor (1 U/μL), Triton X-100 (0.1%), dsDNA template (1.25 μM, **35DNA_1A**) and T7 RNA polymerase (4 U/μL). Transcription reactions were performed at 37 °C for 2 h in a thermal cycler with heated lid (75 °C). After this time, DNase I (80 μL, 1 U/μL) was added to the mixture and the solution was further heated for 30 min at 37 °C in a thermal cycler. The samples were purified (General purification procedure III).

### 5.1.16 Enzymatic synthesis of 35RNA_3A$^{CA}$ or 35RNA_7A$^{CA}$ in preparative scale

*In vitro* transcription reactions were performed in total volume of 400 μL in transcription reaction buffer (5X, 80 μL) containing modified **rA$^{CA}$TP** (3 mM), three natural rNTPs (2 mM, rCTP, rUTP, rGTP), $MgCl_2$ (15 mM), RiboLock RNase inhibitor (1 U/μL), Triton X-100 (0.1%), dsDNA template (1.25 μM, either **35DNA_3A** or **35DNA_7A**) and T7 RNA polymerase (6 U/μL). Transcription reactions were performed at 37 °C for 2 h in a thermal cycler with heated lid (75 °C). After this time, DNase I (80 μL, 1 U/μL) was added to the mixture and the solution was further heated for 30 min at 37 °C in a thermal cycler. The samples were purified (General purification procedure III).

### 5.1.17 Preparation of 21RNA_1A-Cy5 and 21RNA_1A$^{CA}$-Cy5 using pCp-Cy5

The ligation reaction was performed in total volume of 50 μL in T4 RNA ligase buffer (10X, 5 μL), DMSO (10%) and PEG 8000 (5%) with either natural **20RNA_1A** or modified **20RNA_1A$^{CA}$** (10 μM), rATP (1 mM), **pCp-Cy5** (50 μM) and T4 RNA ligase 1 (6 U/μL) in presence of RiboLock RNase inhibitor (1 U/μL). The mixture was incubated at 16 °C in a thermal cycler (with heated lid

65 °C) for 18 h. The mixture was freeze-dried, dissolved in 50 μL of $H_2O$ and purified (General purification procedure I).

### 5.1.18 Preparation of 36RNA_1A-Cy5 and 36RNA_1A$^{CA}$-Cy5 using pCp-Cy5

The ligation reaction was performed in total volume of 50 μL in T4 RNA ligase buffer (10X, 5 μL), DMSO (10%) and PEG 8000 (5%) with either natural **35RNA_1A** or modified **35RNA_1A$^{CA}$** (3 μM), rATP (1 mM), **pCp-Cy5** (50 μM) and T4 RNA ligase 1 (6 U/μL) in presence of RiboLock RNase inhibitor (1 U/μL). The mixture was incubated at 16 °C in a thermal cycler (with heated lid 65 °C) for 18 h. The mixture was freeze-dried, dissolved in 50 μL of $H_2O$ and purified (General purification procedure I).

### 5.1.19 Preparation of 36RNA_3A$^{CA}$-Cy5 and 36RNA_7A$^{CA}$-Cy5 using pCp-Cy5

The ligation reaction was performed in total volume of 50 μL in T4 RNA ligase buffer (10X, 5 μL), DMSO (10%) and PEG 8000 (5%) with either **35RNA_3A$^{CA}$** or **35RNA_7A$^{CA}$** (3 μM), rATP (1 mM), **pCp-Cy5** (50 μM) and T4 RNA ligase 1 (6 U/μL) in presence of RiboLock RNase inhibitor (1 U/μL). The mixture was incubated at 16 °C in a thermal cycler (with heated lid 65 °C) for 18 h. The mixture was freeze-dried, dissolved in 50 μL of $H_2O$ and purified (General purification procedure I).

### 5.1.20 Preparation of 21RNA_1A-Bio and 21RNA_1A$^{CA}$-Bio using pCp-Bio

The ligation reaction was performed in total volume of 50 μL in T4 RNA ligase buffer (10X, 5 μL), DMSO (10%) and PEG 8000 (5%) with either **natural 20RNA_1A** or modified **20RNA_1A$^{CA}$** (10 μM), rATP (1 mM), **pCp-Bio** (50 μM) and T4 RNA ligase 1 (6 U/μL) in presence of RiboLock RNase inhibitor (1 U/μL). The mixture was incubated at 16 °C in a thermal cycler (with heated lid 65 °C) for 18 h. The mixture was freeze-dried, dissolved in 50 μL of $H_2O$ and purified (General purification procedure I).

### 5.1.21 Bioconjugation of natural or modified RNA with either glutathione (GSH), pept-(+)-H or biotin-thiol in analytical scale for dPAGE analysis

The reaction mixture (10 μL) containing labelled RNA: **21RNA_1A-Cy5** or **21RNA_1A$^{CA}$-Cy5** (0.5 μM) and the corresponding peptide (5 mM, 10 000 equiv. of GSH or Ac-KRPRGRPKGSKH-$NH_2$, **pept-(+)-H**, dissolved in $H_2O$) or Bio-thiol (5 mM, 10 000 equiv., dissolved in $H_2O$) was incubated in TEAA (triethylammonium acetate buffer, 0.3 M, pH 8.0 at 25 °C) at 25 °C in a thermal cycler (with heated lid 65 °C) for 48 h. Samples were analysed on dPAGE.

### 5.1.22 Bioconjugation of natural or modified RNA with either pept-(+)-C or pept-(–)-C in analytical scale for dPAGE analysis

The reaction mixture (10 µL) containing labelled RNA: **21RNA_1A-Cy5** or **21RNA_1A$^{CA}$-Cy5** (0.5 µM) and the corresponding peptide (either 50 µM, 100 equiv. or 1.25 mM, 2500 equiv. of Ac-KRPRGRPKGSKC-$NH_2$, **pept-(+)-C** or 1.25 mM, 2 500 equiv. of Ac-DEPEGEPDGSDC-$NH_2$**, pept-(–)-C**, dissolved in $H_2O$) was incubated in TEAA (triethylammonium acetate buffer, 0.3 M, pH 8.0 at 25 °C) at 25 °C in a thermal cycler (with heated lid 65 °C) for 48 h. Samples were analysed on dPAGE.

### 5.1.23 Bioconjugation of natural or modified RNA with either pept-K or pept-R in analytical scale for dPAGE analysis

The reaction mixture (10 µL) containing labelled RNA: **21RNA_1A-Cy5** or **21RNA_1A$^{CA}$-Cy5** (0.5 µM) and the corresponding peptide (5 mM, 10 000 equiv. of Ac-SGYTARAQSG-$NH_2$, **pept-R** or Ac-SGYTAKAQSG-$NH_2$, **pept-K**, dissolved in $H_2O$) was incubated in TEAA (triethylammonium acetate buffer, 0.3 M, pH 8.0 at 25 °C) at 25 °C in a thermal cycler (with heated lid 65 °C) for 48 h. Samples were analysed on dPAGE.

### 5.1.24 Bioconjugation of natural or modified RNA with fluorescein-thiol in analytical scale for dPAGE analysis

The reaction mixture (10 µL) containing labelled RNA: **21RNA_1A-Cy5** or **21RNA_1A$^{CA}$-Cy5** (0.5 µM) and FL-thiol (5 mM, 10 000 equiv., stock in DMSO) was incubated in a solution of 5% DMSO in TEAA (triethylammonium acetate buffer, 0.3 M, pH 8.0 at 25 °C) at 25 °C in a thermal cycler (with heated lid 65 °C) for 48 h. Samples were analysed on dPAGE.

### 5.1.25 Bioconjugation of modified RNA with increasing concentration of pept-(+)-C in analytical scale for dPAGE analysis

The reaction mixture (10 µL) containing labelled RNA: **21RNA_1A$^{CA}$-Cy5** (0.5 µM) and the corresponding peptide [5 µM (10 equiv.), 50 µM (100 equiv.), 250 µM (500 equiv.), 500 µM (1000 equiv.), 1.25 mM, (2500 equiv.), Ac-KRPRGRPKGSKC-$NH_2$, **pept-(+)-C**, dissolved in $H_2O$] was incubated in TEAA (triethylammonium acetate buffer, 0.3 M, pH 8.0 at 25 °C) at 25 °C in a thermal cycler (with heated lid 65 °C) for 48 h. Samples were analysed on dPAGE.

### 5.1.26 Bioconjugation of modified RNA either with glutathione (GSH) or pept-(+)-H in semi-preparative scale for MS-MALDI-TOF analysis

The reaction mixture (50 μL) containing modified **20RNA_1A$^{CA}$** (5 μM) and the corresponding peptide (50 mM, 10000 equiv., **GSH** or Ac-KRPRGRPKGSKH-$NH_2$, **pept-(+)-H**, dissolved in $H_2O$) was incubated in TEAA (triethylammonium acetate buffer, 0.3 M, pH 8.0 at 25 °C) at 25 °C in a thermal cycler (with heated lid 65 °C) for 48 h. The mixture was purified (in case of reaction with **pept-(+)-H**, General purification procedure I was used and in case of **GSH**, General purification procedure II was used).

### 5.1.27 Bioconjugation of modified RNA with pept-(+)-C in semi-preparative scale for LC-MS analysis

The reaction mixture (50 μL) containing modified **20RNA_1A$^{CA}$** (5 μM) and the corresponding peptide (12.5 mM, 2500 equiv., Ac-KRPRGRPKGSKC-$NH_2$, **pept-(+)-C**, dissolved in $H_2O$) was incubated in TEAA (triethylammonium acetate buffer, 0.3 M, pH 8.0 at 25 °C) at 25 °C in a thermal cycler (with heated lid 65 °C) for 48 h. The reaction mixture was purified (General purification procedure I). The sample was evaporated to dryness and dissolved in 20 μL of $H_2O$. 20 μL of the purified reaction mixture were injected on Waters ACQUITY Premier CSH C18 1.7 μm, 2.1×150 mm column. Separation was performed at 60 °C column temperature with flow rate 0.25 mL/min with mobile phases: A (12.2 mM $Et_3N$, 300 mM HFIP in water) and B (12.2 mM $Et_3N$, 300 mM HFIP in 100% MeOH).

**Table 19. Separation gradient.**

| Time (min) | Mobile phase A (%) | Mobile phase B (%) |
|---|---|---|
| 0.00 | 95 | 5 |
| 30.00 | 0 | 100 |
| 32.00 | 0 | 100 |

MS settings were as follows: capillary voltage –3 kV, drying gas flow 12 L/min, nebulizer pressure 35 psig, drying gas temperature 350 °C, fragmentor 70 V, mass range 500 – 3000 m/z, 0.1 Da step size, 1.46 sec per scan cycle.

### 5.1.28 Bioconjugation of modified RNA with biotin-thiol in semi-preparative scale for MS-MALDI-TOF analysis

The reaction mixture (50 μL) containing modified **20RNA_1A$^{CA}$** (5 μM) and Bio-thiol (50 mM, 10000 equiv., dissolved in $H_2O$) was incubated in TEAA (triethylammonium acetate buffer, 0.3 M,

pH 8.0 at 25 °C) at 25 °C in a thermal cycler (with heated lid 65 °C) for 48 h. The reaction mixture was purified (General purification procedure II).

### 5.1.29 Bioconjugation of modified RNA with fluorescein-thiol in semi-preparative scale for MS-MALDI-TOF analysis

The reaction mixture (10 μL) containing modified **20RNA_1A$^{CA}$** (12.5 μM) and FL-thiol (50 mM, 10000 equiv., stock in DMSO) was incubated in a solution of 10% DMSO in TEAA (triethylammonium acetate buffer, 0.3 M, pH 8.0 at 25 °C) at 25 °C in a thermal cycler (with heated lid 65 °C) for 48 h. The reaction mixture was purified (General purification procedure I).

### 5.1.30 Bioconjugation and fluorescence measurements of either natural or modified RNA with fluorescein-thiol

The reaction mixture (10 μL) containing either natural **20RNA_1A** or modified **20RNA_1A$^{CA}$** (12.5 μM) and FL-thiol (50 mM, 10000 equiv., stock in DMSO) was incubated in a solution of 10% DMSO in TEAA (triethylammonium acetate buffer, 0.3 M, pH 8.0 at 25 °C) at 25 °C in a thermal cycler (with heated lid 65 °C) for 48 h. The reaction mixture was purified (General purification procedure I) and adjusted with $H_2O$ to final volume of 100 μL and transferred to a quartz cuvette. Emission fluorescent spectra were measured at 25 °C and the excitation wavelength was set to 490 nm.

### 5.1.31 EMSA of natural or modified RNA with HuR

The reaction mixture (10 μL) containing labelled RNA (**21RNA_1A-Cy5** or **21RNA_1A$^{CA}$-Cy5**, 0.5 μM, 1 equiv.) in 30 mM HEPES-NaOH buffer (pH 7.45 at 25 °C) was incubated with HuR protein (0.5 μM, 1 equiv.) at 37 °C for 18 h in a thermal cycler (with heated lid at 65 °C). Samples were analysed by native-PAGE.

### 5.1.32 EMSA of natural or modified RNA with HIV-RT

The reaction mixture (10 μL) containing labelled RNA (**21RNA_1A-Cy5** or **21RNA_1A$^{CA}$-Cy5**, 0.5 μM, 1 equiv.) in 30 mM HEPES-NaOH buffer (pH 7.45 at 25 °C) was incubated with HIV-RT (1 μM, 2 equiv.) at 20 °C for 21 h in a thermal cycler (with heated lid at 65 °C). Samples were analysed by native-PAGE.

### 5.1.33 EMSA of natural or modified RNA with hAgo2

The reaction mixture (10 μL) containing labelled RNA (**21RNA_1A-Cy5** or **21RNA_1A$^{CA}$-Cy5**, 0.5 μM, 1 equiv.) in 30 mM HEPES-NaOH buffer (pH 7.45 at 25 °C) was incubated with hAgo2 protein

(1 μM, 2 equiv.) at 37 °C for 21 h in a thermal cycler (with heated lid at 65 °C). Samples were analysed by native-PAGE.

### 5.1.34 Kinetic study of cross-linking reaction of modified RNA with HIV-RT

The reaction mixture (10 μL) containing labelled **21RNA_1A$^{CA}$-Cy5** (0.5 μM, 1 equiv.) in 30 mM HEPES-NaOH buffer (pH 7.45 at 25 °C) was incubated with HIV-RT (1 μM, 2 equiv.) at 20 °C in a thermal cycler (with heated lid at 65 °C) for different time periods: 0.5 h, 1 h, 6 h, 10 h and 24 h. Samples were analysed by SDS-PAGE.

### 5.1.35 Cross-linking reaction of natural or modified RNA with RBPs

The reaction mixture (10 μL) containing labelled RNA (**21RNA_1A-Cy5** or **21RNA_1A$^{CA}$-Cy5**, 0.5 μM, 1 equiv.) in 30 mM HEPES-NaOH buffer (pH 7.45 at 25 °C) was incubated either with HuR protein (1 μM, 2 equiv.), HIV-RT (1 μM, 2 equiv.) or with hAgo2 protein (1 μM, 2 equiv.). Reactions with HuR and hAgo2 proteins were incubated at 37 °C and the reaction with HIV-RT was incubated at 20 °C in a thermal cycler (with heated lid at 65 °C) for 20 h. Samples were analysed by SDS-PAGE.

### 5.1.36 Cross-linking reactions of modified RNA with weakly- or non-RBPs

The reaction mixture (10 μL) containing labelled **21RNA_1A$^{CA}$-Cy5** (0.5 μM, 1 equiv.) in 30 mM HEPES-NaOH buffer (pH 7.45 at 25 °C) was incubated either with HuR protein (1 μM, 2 equiv.) or with various weakly- or non-RBPs: BSA (1 μM, 2 equiv.), SSB (1 μM, 2 equiv.), Gal1 (1 μM, 2 equiv.), lysozyme (1 μM, 2 equiv.) and histone H2A (1 μM, 2 equiv.) at 37 °C for 18 h in a thermal cycler (with heated lid at 65 °C). Samples were analysed by SDS-PAGE.

### 5.1.37 Cross-linking reactions of either natural 36RNA_1A-Cy5 or modified 36RNA-1A$^{CA}$-Cy5, 36RNA-3A$^{CA}$-Cy5 and 36RNA-7A$^{CA}$-Cy5 with HuR protein

The reaction mixture (10 μL) containing labelled RNA (**36RNA_1A-Cy5** or **36RNA_1A$^{CA}$-Cy5**, **36RNA_3A$^{CA}$-Cy5**, **36RNA_7A$^{CA}$-Cy5**, 0.5 μM, 1 equiv.) in 30 mM HEPES-NaOH buffer (pH 7.45 at 25 °C) in presence of RiboLock RNase inhibitor (1 μL, 40 U/μL) was incubated with HuR protein (1 μM, 2 equiv.) at 37 °C for 15 h in a thermal cycler (with heated lid at 65 °C). Samples were analysed by SDS-PAGE.

### 5.1.38 Cross-linking reactions of either natural 36RNA_1A-Cy5 or modified 36RNA-1A$^{CA}$-Cy5, 36RNA-3A$^{CA}$-Cy5 and 36RNA-7A$^{CA}$-Cy5 with BSA

The reaction mixture (10 μL) containing labelled RNA (**36RNA_1A-Cy5** or **36RNA_1A$^{CA}$-Cy5**, **36RNA_3A$^{CA}$-Cy5**, **36RNA_7A$^{CA}$-Cy5**, 0.5 μM, 1 equiv.) in 30 mM HEPES-NaOH buffer (pH 7.45 at 25 °C) in presence of RiboLock RNase inhibitor (1 μL, 40 U/μL) was incubated with BSA (1 μM, 2 equiv.) at 37 °C for 17 h in a thermal cycler (with heated lid at 65 °C). Samples were analysed by SDS-PAGE.

### 5.1.39 Cross-linking reactions of either natural 36RNA_1A-Cy5 or modified 36RNA-1A$^{CA}$-Cy5 with HeLa cell lysate proteins

The reaction mixture (10 μL) containing labelled RNA (**36RNA_1A-Cy5** or **36RNA_1A$^{CA}$-Cy5**, 0.5 μM, 1 equiv.) in 30 mM HEPES-NaOH buffer (pH 7.45 at 25 °C) in presence of RiboLock RNase inhibitor (1 μL, 40 U/μL) was incubated with HeLa cell lysate (5 μL, 2.72 μg/μL) at 37 °C for 16 h in a thermal cycler (with heated lid at 65 °C). Samples were analysed by SDS-PAGE.

### 5.1.40 WB analysis of cross-linking reaction of natural or modified RNA with HuR protein

The reaction mixture (10 μL) containing **20RNA_1A** or **20RNA_1A$^{CA}$** (3 μM, 1.5 equiv.) in 30 mM HEPES-NaOH buffer (pH 7.45 at 25 °C) was incubated with HuR protein (2 μM, 1 equiv.) at 37 °C for 21 h in a thermal cycler (with heated lid at 65 °C). Samples were analysed by SDS-PAGE. Afterwards, WB was performed using the WET/TANK blotting system (Mini Trans-Blot® Cell) and PVDF transfer membrane. After protein transfer, the membrane was incubated in blocking buffer (5.5% casein buffer) at room temperature for 1 h and then with the fluorescent (Alexa Fluor 488) mouse monoclonal $IgG_1$ anti-HuR/ELAVL1 antibody (1:200 diluted) and subsequently with fluorescent secondary antibody Alexa Fluor 488 goat anti-mouse IgG (H+L) cross-adsorbed secondary antibody (1:400 diluted) overnight at 4 °C. The membrane was visualized by a fluorescent scanner using Alexa Fluor 488 channel scan.

### 5.1.41 WB analysis of cross-linking reaction of natural or modified RNA with HIV-RT

The reaction mixture (20 μL) containing **20RNA_1A** or **20RNA_1A$^{CA}$** (3 μM, 1.5 equiv.) in 30 mM HEPES-NaOH buffer (pH 7.45 at 25 °C) was incubated with HIV-RT (2 μM, 1 equiv.) at 20 °C for 20 h in a thermal cycler (with heated lid at 65 °C). Samples were analysed by SDS-PAGE. Afterwards, WB was performed using WET/TANK blotting system (Mini Trans-Blot® Cell) and PVDF transfer membrane. After protein transfer, the membrane was incubated in blocking buffer (5.5% casein

buffer) at room temperature for 1 h and then with rabbit anti-HIV 1 reverse transcriptase polyclonal antibody (1:500 diluted) and subsequently with fluorescent secondary antibody Alexa Fluor 488 affinity pure goat anti-rabbit IgG (1:500 diluted) overnight at 4 °C. The membrane was visualized by a fluorescent scanner using Alexa Fluor 488 channel scan.

### 5.1.42 WB analysis of cross-linking reaction of biotinylated natural or modified RNA with hAgo2 protein

The reaction mixture (10 μL) containing **21RNA_1A-Bio** or **21RNA_1A$^{CA}$-Bio** (7 μM, 3.5 equiv.) in 30 mM HEPES-NaOH buffer (pH 7.45 at 25 °C) was incubated with hAgo2 protein (2 μM, 1 equiv.) at 37 °C for 19 h in a thermal cycler (with heated lid at 65 °C). Samples were analysed by SDS-PAGE. Afterwards, WB was performed using WET/TANK blotting system (Mini Trans-Blot® Cell) and PVDF transfer membrane. After protein transfer, the membrane was incubated in blocking buffer (5.5% casein buffer) at room temperature for 1 h and then with rabbit anti-Argonaute-2 polyclonal antibody (1:500 diluted), following by fluorescent secondary antibody Alexa Fluor 488 affinity pure goat anti-rabbit IgG (1:500 diluted) and subsequently with eBioscience Streptavidin APC conjugate (1:500 diluted) overnight at 4 °C. The membrane was visualized by a fluorescent scanner using Alexa Fluor 488 channel and Cy5 channel scans.

### 5.1.43 Selective targeting of HuR protein by cross-linking reaction in HeLa cell lysate

The reaction mixture (20 μL) containing either binding sequences for HuR (natural **21RNA_3A-bind** or modified **21RNA_3A$^{CA}$-bind**, 20 μM) or non-binding sequences for HuR (natural **21RNA_3A-non-bind** or modified **21RNA_3A$^{CA}$-non-bind**, 20 μM) was incubated with HeLa cell lysate (9.5 μL, 15 μg/μL) in presence of RiboLock RNase inhibitor (0.5 μL, 40 U/μL) at 37 °C for 16 h in a thermal cycler (with heated lid at 65 °C). Samples were analysed by SDS-PAGE. Afterwards, WB was performed using WET/TANK blotting system (Mini Trans-Blot® Cell) and PVDF transfer membrane. After protein transfer, the membrane was incubated in blocking buffer (5.5% casein buffer) for 48 h at 4 °C and then with the fluorescent (Alexa Fluor 488) mouse monoclonal $IgG_1$ anti-HuR/ELAVL1 antibody (1:200 diluted) for 2.5 h at room temperature and subsequently with horse anti-mouse IgG, HRP-linked secondary antibody (1:2 500 diluted) for 2 h at room temperature. The membrane was incubated with chemiluminescent substrate Radiance ECL for 2 min. The membrane was visualized by chemiluminescence blot imaging by a chemiluminescent scanner (Syngene G:BOX Chemi XRQ).

### 5.1.44 Intact ESI-MS analysis of 20RNA_1A$^{CA\text{-}HuR}$ conjugate

The reaction mixture (30 μL) containing **20RNA_1A$^{CA}$** (9 μM, 1.53 equiv.) in 30 mM HEPES-NaOH buffer (pH 7.45 at 25 °C) was incubated with HuR protein (5.88 μM, 1 equiv.) at 25 °C for 20 h in a thermal cycler (with heated lid at 65 °C). 25 μL were used for ESI-MS analysis.

### 5.1.45 LC-MS analysis of conjugation mixture of 20RNA_1A$^{CA}$ with HIV-RT protein

The reaction mixture (50 μL) containing **20RNA_1A$^{CA}$** (15 μM, 1.5 equiv.) in 25 mM HEPES-NaOH buffer (pH 7.45 at 25 °C) was incubated with HIV-RT (10 μM, 1 equiv.) at 20 °C for 22 h in a thermal cycler (with heated lid at 65 °C). After this time, the sample was concentrated to 20 μL using Amicon Ultra-0.5 centrifugal concentrator (10 kDa MWCO). LC-MS analysis was carried out on Agilent 1290 Infinity II Bio system. 20 μL of the purified reaction mixture were injected on Waters ACQUITY Premier CSH C18 1.7 μm, 2.1×150 mm column. Separation was performed at 45 °C column temperature with flow rate 0.25 mL/min with mobile phases: A (36 mM TEAB + 0.25 % $Et_3N$ in water) and B (36 mM TEAB + 0.25 % $Et_3N$ in 90% acetonitrile).

**Table 20. Separation gradient.**

| Time (min) | Mobile phase A (%) | Mobile phase B (%) |
|---|---|---|
| 0.00 | 95 | 5 |
| 5.00 | 95 | 5 |
| 8.00 | 65 | 35 |
| 28.00 | 35 | 65 |
| 28.01 | 0 | 100 |
| 32.00 | 0 | 100 |

MS settings were as follows: capillary voltage –5 kV, drying gas flow 12 L/min, nebulizer pressure 35 psig, drying gas temperature 350 °C, fragmentor 70 V, mass range 500 – 3000 m/z, 0.1 Da step size, 1.46 sec per scan cycle.

### 5.1.46 Preparation of 20RNA_1A$^{CA\text{-}HuR}$ conjugate digest for nano-LC-MS/MS analysis

The reaction mixture (100 μL) containing **20RNA_1A$^{CA}$** (7.5 μM, 1.23 equiv.) in 30 mM HEPES-NaOH buffer (pH 7.45 at 25 °C) was incubated with HuR protein (6.08 μM, 1 equiv.) at 37 °C for 25 h in a thermal cycler (with heated lid at 65 °C). After this time, urea (8 M) and DTT (10 mM) were added to the mixture and incubated at 37 °C for 1 h with shaking (800 rpm) in a thermal block. Then 9.67 μL of freshly prepared saturated solution of iodoacetamide (dissolved in $H_2O$) were added to the

mixture and further incubated at 37 °C for 1 h with shaking (800 rpm) in a thermal block. The solution was diluted with 700 μL of $H_2O$ and 175 μL of the cleavage buffer (500 mM Tris, 10 mM $CaCl_2$, pH 8.0 at 25 °C). Then 0.96 μg of chymotrypsin protease (1:25 protease/protein ratio) were added to the mixture and incubated at 37 °C with gentle rotation on Hula-mixer overnight. The protease was denatured by heating at 70 °C for 10 min and the sample was concentrated on vacuum concentrator. Then the solution was purified (General purification procedure I) and eluted in 50 μL of $H_2O$ followed by evaporation to dryness. The solid sample was dissolved in 200 μL of concentrated hydrofluoric acid (HF) and incubated overnight at 4 °C. The HF was evaporated, and the sample was purified on C18 spin columns according to standard supplier's protocol. Elution was performed with 3X 50 μL of 70% acetonitrile and the sample was evaporated to dryness prior to mass analysis.

### 5.1.47 Preparation of 20RNA_1A$^{CA\text{-}HIV\text{-}RT}$ conjugate digest for nano-LC-MS/MS analysis

The reaction mixture (25 μL) containing **20RNA_1A$^{CA}$** (30 μM, 1.13 equiv.) in 30 mM HEPES-NaOH buffer (pH 7.45 at 25 °C) was incubated with HIV-RT (26.64 μM, 1 equiv.) at 20 °C for 25 h in a thermal cycler (with heated lid at 65 °C). After this time, 75 μL of $H_2O$ were added to the mixture followed by urea (8 M) and DTT (10 mM). The mixture was incubated at 37 °C for 1 h with shaking (800 rpm) in a thermal block. Then 9.67 μL of freshly prepared saturated solution of iodoacetamide (dissolved in $H_2O$) were added to the mixture and further incubated at 37 °C for 1 h with shaking (800 rpm) in a thermal block. The solution was diluted with 700 μL of $H_2O$ and 175 μL of the cleavage buffer (500 mM ammonium bicarbonate, pH 8.0 at 25 °C). Then 3.12 μg of trypsin protease (1:25 protease/protein ratio) were added to the mixture and incubated at 37 °C with gentle rotation on Hula-mixer overnight. The protease was denatured by heating at 70 °C for 10 min and the sample was concentrated on vacuum concentrator. Then the solution was purified (General purification procedure I) and eluted in 50 μL of $H_2O$ followed by evaporation to dryness. The solid sample was dissolved in 200 μL of concentrated hydrofluoric acid (HF) and incubated overnight at 4 °C. The HF was evaporated, and the sample was purified on C18 spin columns according to standard supplier's protocol. Elution was performed with 3X 50 μL of 70% acetonitrile and the sample was evaporated to dryness prior to mass analysis.

### 5.1.48 Preparation of 20RNA_1A$^{CA\text{-}hAgo2}$ conjugate digest for nano-LC-MS/MS analysis

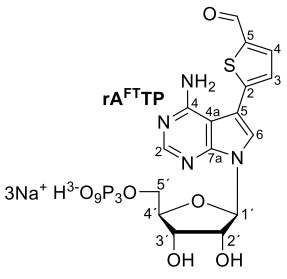
The reaction mixture (105 μL) containing **20RNA_1A$^{CA}$** (7.86 μM, 1.23 equiv.) in 30 mM HEPES-NaOH buffer (pH 7.45 at 25 °C) was incubated with hAgo2 protein (6.38 μM, 1 equiv.) at 37 °C for 24 h in a thermal cycler (with heated lid at 65 °C). After this time, urea (8 M) and DTT (10 mM) were added to the mixture and incubated at 37 °C for 1 h with shaking (800 rpm) in a thermal block. Then 10.15 μL of freshly prepared saturated solution of iodoacetamide (dissolved in $H_2O$) were added to the mixture and further incubated at 37 °C for 1 h with shaking (800 rpm) in a thermal block. The solution was diluted with 729 μL of $H_2O$ and 216 μL of the cleavage buffer (500 mM ammonium bicarbonate, pH 8.0 at 25 °C). Then 2.68 μg of trypsin protease (1:25 protease/protein ratio) were added to the mixture and incubated at 37 °C with gentle rotation on Hula-mixer overnight. The protease was denatured by heating at 70 °C for 10 min and the sample was concentrated on vacuum concentrator. Then the solution was purified (General purification procedure I) and eluted in 50 μL of $H_2O$ followed by evaporation to dryness. The solid sample was dissolved in 200 μL of concentrated hydrofluoric acid (HF) and incubated overnight at 4 °C. The HF was evaporated, and the sample was purified on C18 spin columns according to standard supplier's protocol. Elution was performed with 3X 50 μL of 70% acetonitrile and the sample was evaporated to dryness prior to mass analysis. The sample was prepared in two more replicates. One following the same procedure with 24 h incubation time for RNA-protein cross-linking and the other one following the same procedure just with shorter 6 h incubation time for RNA-protein cross-linking reaction.

## 5.2 Expedient production of site specifically nucleobase-labelled or hypermodified RNA with engineered thermophilic DNA polymerases

### 5.2.1 General remarks – Synthetic part

All solvents and reagents were purchased from commercial suppliers (Fluorochem, Sigma Aldrich, Lach-Ner) and used as received without further purification, unless otherwise specified. POCl3 and PO(OMe)3 were distilled prior to use. Reactions were performed in heat gun-dried glassware under inert argon atmosphere. Reactions were monitored by TLC Silica gel 60 F254 and detected by UV (254 nm) and by Advion Expression Compact Mass Spectrometer connected with Plate Express® TLC Plate Reader using electrospray ionization. NMR spectra were measured on a Bruker AVANCE III HD 500 spectrometer (1H at 500.0 MHz, 11B at 160.4 MHz, 13C at 125.7 MHz, 19F at 470.4 MHz and 31P at 202.4 MHz) in D2O (referenced to either to tert-butanol or dioxane) or DMSO-d6 (referenced to δ(CHD2SO2CD3) = 2.50 ppm and δ(CD3SO2CD3) = 39.7 ppm) at 25°C. Chemical shifts are given in ppm (δ-scale) and coupling constants (J) are in Hz. Complete assignment of all NMR signals was achieved by using a combination of H,H-COSY, H,C-HSQC and H,C-HMBC experiments. Labelling of NMR signals assignments corresponds to a numbering depicted in compound formulas. High-resolution mass spectra [HR MS ($ESI^-$)] were measured on LTQ Orbitrap XL (Thermo Fisher Scientific) and acquired by the MS service at IOCB. Column chromatography was performed using silica gel (40-63 μm, Fluorochem) by FLC Teledyne ISCO Combi Flash Rf 200 or 300. Purification of triphosphates was performed using HPLC (Waters modular HPLC system) on a Phenomenex Kinetex 5 μm EVO C18 100 Å, AXIA Packed LC column (250×21.2 mm) and POROS HQ 50 column (packed in-house, 26×120 mm). Purity of all final compounds was determined by NMR spectroscopy. Milli-Q water was used in the synthetic part. 5-Iodocytidine was purchased from Biosynth. 5-Formyl-2-thienylboronic acid was obtained from Sigma Aldrich.

### 5.2.2 Chemical synthesis of 7-(5-formylthien-2-yl)-7-deazaadenosine-5'-*O*-triphosphate ($rA^{FT}TP$)




Triphosphate (**$rA^{I}TP$**, 1 equiv.)[260], 5-formylthiophene-2-boronic acid (1.8 equiv.), $Cs_2CO_3$ (3 equiv.), $Pd(OAc)_2$ (5 mol%) and TPPTS (25 mol%) were dissolved in a mixture of $H_2O$/acetonitrile (2:1, 1.5 mL) under argon atmosphere. The mixture was stirred at 100 °C for 1 h. Afterwards, solvents were evaporated under reduced pressure and the residue was again dissolved in $H_2O$ prior to purification. The HPLC separation was performed on C18 reversed phase column using linear gradient from 0.1 M TEAB

(aq.) to 0.1 M TEAB in 50% MeOH followed by several co-evaporations with $H_2O$. The pure product was isolated as a triethylammonium salt and converted to sodium salt using Dowex 50WX8 in $Na^+$ cycle. The solvent was evaporated, pure product was again dissolved in small amount of $H_2O$ and freeze-dried overnight. The desired product **rA$^{FT}$TP** (as sodium salt) was obtained as a yellow solid powder in 59% yield.

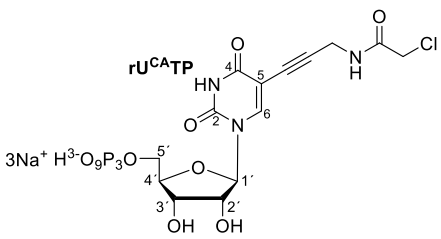
**$^1$H NMR (500 MHz, $D_2O$, ref(*t*BuOH) = 1.25 ppm):** 4.16 (ddd, 1H, $J_{gem}$ = 11.7, $J_{5'a,P}$ = 4.9, $J_{5'a,4'}$ = 3.5, H-5'a); 4.26 (ddd, 1H, $J_{gem}$ = 11.7, $J_{5'b,P}$ = 6.7, $J_{5'b,4'}$ = 3.4, H-5'b); 4.35 (qd, 1H, $J_{4',3'}$ = $J_{4',5'a}$ = $J_{4',5'b}$ = 3.2, $J_{4',P}$ = 2.0, H-4'); 4.58 (dd, 1H, $J_{3',2'}$ = 5.4, $J_{3',4'}$ = 2.9, H-3'); 4.72 (dd, 1H, $J_{2',1'}$ = 6.8, $J_{2',3'}$ = 5.4, H-2'); 6.25 (d, 1H, $J_{1',2'}$ = 6.8, H-1'); 7.28 (d, 1H, $J_{3,4}$ = 3.9, H-3-thienyl); 7.77 (s, 1H, H-6); 7.96 (d, 1H, $J_{4,3}$ = 3.9, H-4-thienyl); 8.13 (s, 1H, H-2); 9.74 (s, 1H, CHO).

**$^{13}$C NMR (125.7 MHz, $D_2O$, ref(*t*BuOH) = 31.60 ppm):** 67.44 (d, $J_{C,P}$ = 5.7, $CH_2$-5'); 72.45 (CH-3'); 75.65 (CH-2'); 85.71 (d, $J_{C,P}$ = 8.9, CH-4'); 87.79 (CH-1'); 102.46 (C-4a); 112.20 (C-5); 124.66 (CH-6); 130.26 (CH-3-thienyl); 142.56 (CH-4-thienyl); 143.56 (C-5-thienyl); 148.69 (C-2-thienyl); 152.71 (C-7a); 153.80 (CH-2); 158.84 (C-4); 188.49 (CHO).

**$^{31}$P NMR (202.4 MHz, $D_2O$):** -21.49 (t, $J$ =19.6, $P_\beta$); -10.46 (d, $J$ = 19.6, $P_\alpha$); -7.10 (d, $J$ = 19.6, $P_\gamma$).

**HR MS (ESI):** **$C_{16}H_{18}O_{14}N_4P_3S$** calculated 614.97531, found 614.97546; **$C_{16}H_{17}O_{14}N_4P_3SNa$** calculated 636.95780, found 636.95772; **$C_{16}H_{16}O_{14}N_4P_3SNa_2$** calculated 658.93974, found 658.93890.

### 5.2.3 Chemical synthesis of 5-{3-[*N*-(2-chloroacetamido)]-prop-2-yn-1-yl}-uridine-5'-*O*-triphosphate (rU$^{CA}$TP)




Triphosphate (**rU$^{I}$TP**[276], 1 equiv.), *N*-(propargyl)chloroacetamide[261] (3 equiv.), CuI (10 mol%), $Pd(OAc)_2$ (5 mol%) and TPPTS (25 mol%) were dissolved in a mixture of water/acetonitrile (2:1, 1 mL) under argon atmosphere. After addition of DIPEA (10 eq.) the mixture was stirred at 65 °C for 1 h. Additional amount of $Pd(OAc)_2$ (5 mol%) was added and the reaction was stirred for another 2 h at 65 °C. Afterwards, solvents were removed under reduced pressure and the residue was dissolved in $H_2O$ prior to purification. HPLC separation was performed using reverse phase column with gradient starting from 0.1 M TEAB (aq.) up to 0.1 M TEAB in 50% MeOH. The second purification was performed on POROS HQ 50 column using linear gradient from $H_2O$ to 400 mM TEAB (aq.) followed by several co-evaporations with $H_2O$. The pure product was isolated as a triethylammonium salt and converted to sodium salt using Dowex 50WX8 in $Na^+$ cycle.

The solvent was evaporated, pure product was again dissolved in small amount of $H_2O$ and freeze-dried overnight. The desired product **rU$^{CA}$TP** (as sodium salt) was obtained as a white solid powder in 23% yield.

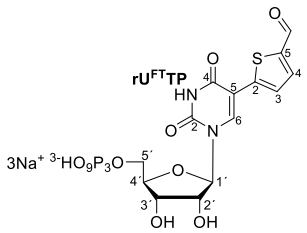
**$^1$H NMR (500 MHz, $D_2O$, ref(*t*BuOH) = 1.25 ppm):** 4.20 (s, 2H, $CH_2Cl$); 4.23 – 4.34 (m, 5H, H-4',5',C≡CC**H$_2$**); 4.39 (dd, 1H, $J_{2',3'}$ = 5.1, $J_{2',1'}$ = 4.7, H-2'); 4.47 (bt, 1H, $J_{3',4'}$ = $J_{3',2'}$ = 5.0, H-3'); 5.96 (d, 1H, $J_{1',2'}$ = 4.7, H-1'); 8.24 (s, 1H, H-6).

**$^{13}$C NMR (125.7 MHz, $D_2O$, ref(*t*BuOH) = 31.60 ppm):** 31.13 (C≡C**C**$H_2$); 44.26 ($CH_2Cl$); 66.63 (d, $J_{C,P}$ = 5.3, $CH_2$-5'); 71.04 (CH-3'); 75.95 (**C**≡C$CH_2$); 76.01 (CH-2'); 85.32 (d, $J_{C,P}$ = 9.2, CH-4'); 90.81 (CH-1'); 91.23 (C≡**C**$CH_2$); 101.20 (C-5); 146.83 (CH-6); 152.9 (C-2); 166.8 (C-4); 171.59 (NHCO).

**$^{31}$P NMR (202.4** MHz**, $D_2O$):** -21.72 (t, $J$ = 19.6, $P_\beta$); -10.71 (d, $J$ = 20.0, $P_\alpha$); -6.58 (bs, 1P, $P_\gamma$).

**HR MS (ESI$^-$): $C_{14}H_{18}O_{16}N_3P_3(^{35}Cl)$** calculated: 611.95939, found: 611.95880; **$C_{14}H_{17}O_{16}N_3P_3(^{35}Cl)Na$** calculated: 633.94134, found: 633.94089; **$C_{14}H_{16}O_{16}N_3P_3(^{35}Cl)Na_2$** calculated: 655.92328, found: 655.92267.

### 5.2.4 Chemical synthesis of 5-(5-formylthien-2-yl)-uridine-5'-*O*-triphosphate (rU$^{FT}$TP)




Triphosphate (**rU$^I$TP**[276] 1 equiv.), 5-formylthiophene-2-boronic acid (2.0 equiv.), $Cs_2CO_3$ (2.0 equiv.), $Pd(OAc)_2$ (5 mol%) and TPPTS (25 mol%) were dissolved in a mixture of $H_2O$/acetonitrile (2:1, 1.0 mL) under argon atmosphere. The mixture was stirred at 100 °C for 1.0 h. Afterwards, solvents were evaporated under reduced pressure and the residue was again dissolved in $H_2O$ prior to purification. The HPLC separation was performed on C18 reversed phase column using linear gradient from 0.1 M TEAB (aq.) to 0.1 M TEAB in 50% MeOH followed by several co-evaporations with $H_2O$. The pure product was isolated as a triethylammonium salt and converted to sodium salt using Dowex 50WX8 in $Na^+$ cycle. The solvent was evaporated, pure product was again dissolved in small amount of $H_2O$ and freeze-dried overnight. The desired product **rU$^{FT}$TP**[276] (as sodium salt) was obtained as a yellow solid powder in 33% yield.

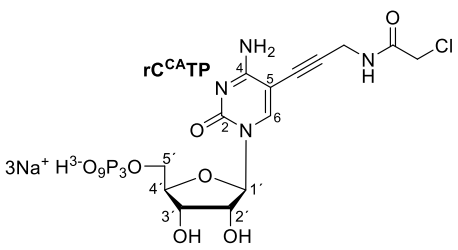
**$^1$H NMR (500 MHz, $D_2O$, ref(*t*BuOH) = 1.25 ppm):** 4.24 – 4.41 (m, 3H, H-4',5'); 4.50 – 4.57 (m, 2H, H-2',3'); 6.07 (d, 1H, $J_{1',2'}$ = 5.0 Hz, H-1'); 7.77 (d, 1H, $J_{3,4}$ = 4.2 Hz, H-3-thienyl); 8.02 (d, 1H, $J_{4,3}$ = 4.2 Hz, H-4-thienyl); 8.51 (s, 1H, H-6); 9.81 (s, 1H, CHO).

**$^{13}$C NMR (125.7 MHz, $D_2O$, ref(*t*BuOH) = 31.60 ppm):** 66.99 (d, $J_{C,P}$ = 5.5 Hz, CH2-5'); 71.60 (CH-3'); 76.27 (CH-2'); 85.88 (d, $J_{C,P}$ = 9.0 Hz, CH-4'); 90.80 (CH-1'); 111.07 (C-5); 127.48 (CH-3-thienyl); 140.71 (CH-6); 141.78 (CH-4-thienyl); 143.33 (C-5-thienyl); 146.42 (C-2-thienyl); 153.00 (C-2); 165.29 (C-4); 189.23 (CHO).

**$^{31}$P NMR (202.4 MHz, $D_2O$):** -21.41 (t, J = 19.8 Hz, $P_\beta$); -10.72 (d, J = 19.8 Hz, $P_\alpha$); -6.13 (m, $P_\gamma$).

**HR MS ($ESI^-$): $C_{14}H_{16}O_{16}N_2P_3S$** calculated: 592.94389, found: 592.94411; **$C_{14}H_{15}O_{16}N_2P_3SNa$** calculated: 614.92583, found: 614.92632; **$C_{14}H_{14}O_{16}N_2P_3SNa_2$** calculated: 636.90778, found: 636.90792.

### 5.2.5 Chemical synthesis of 5-{3-[*N*-(2-chloroacetamido)]-prop-2-yn-1-yl}-cytidine-5'-*O*-triphosphate (rC$^{CA}$TP)




Triphosphate (**rC$^{I}$TP**[287], 1 equiv.), *N*-(propargyl)chloroacetamide[261] (2.5 equiv.), CuI (10 mol%), $Pd(OAc)_2$ (5 mol%) and TPPTS (20 mol%) were dissolved in a mixture of water/acetonitrile (2:1, 1 mL) under argon atmosphere. After addition of DIPEA (10 equiv.) the mixture was stirred at 65 °C for 3 h. Afterwards, solvents were removed under reduced pressure and the residue was dissolved in $H_2O$ prior to purification. HPLC separation was performed using reverse phase column with gradient starting from 0.1 M TEAB (aq.) up to 0.1 M TEAB in 50% MeOH. The second purification was performed on POROS HQ 50 column using linear gradient from $H_2O$ to 400 mM TEAB (aq.) followed by several co-evaporations with $H_2O$. The pure product was isolated as a triethylammonium salt and converted to sodium salt using Dowex 50WX8 in $Na^+$ cycle. The solvent was evaporated, pure product was again dissolved in small amount of $H_2O$ and freeze-dried overnight. The desired product **rC$^{CA}$TP** (as sodium salt) was obtained as a white solid powder in 34% yield.

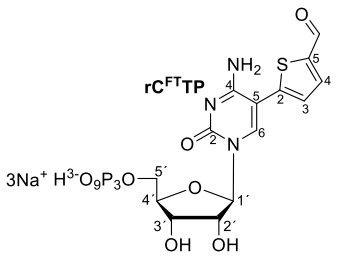
**$^{1}$H NMR (500 MHz, $D_2O$, ref(*t*BuOH) = 1.25 ppm):** 4.20 (s, 2H, $CH_2Cl$); 4.25 – 4.35 (m, 5H, H-4',5',C≡CC**H$_2$**); 4.32 (dd, 1H, $J_{2',3'}$ = 5.2, $J_{2',1'}$ = 3.7, H-2'); 4.42 (bdd, 1H, $J_{3',4'}$ = 5.9, $J_{3',2'}$ = 5.2, H-3'); 5.95 (d, 1H, $J_{1',2'}$ = 3.7, H-1'); 8.24 (s, 1H, H-6).

**$^{13}$C NMR (125.7 MHz, $D_2O$, ref(*t*BuOH) = 31.6 ppm):** 32.17 (C≡C**C**$H_2$); 44.23 ($CH_2Cl$); 66.33 (d, $J_{C,P}$ = 5.4, $CH_2$-5'); 70.54 (CH-3'); 75.55 (**C**≡C$CH_2$); 76.42 (CH-2'); 84.66 (d, $J_{C,P}$ = 9.1, CH-4'); 91.70 (CH-1'); 93.11 (C≡**C**$CH_2$); 94.14 (C-5); 147.06 (CH-6); 158.21 (C-2); 167.07 (C-4); 171.73 (NHCO).

**$^{31}$P NMR (202.4 MHz, $D_2O$):** -21.35 (t, $J$ = 19.7, $P_\beta$); -10.52 (d, $J$ = 19.7, $P_\alpha$); -6.04 (bd, $J$ = 19.7, $P_\gamma$).

**HR MS ($ESI^-$): $C_{14}H_{17}O_9N_4P(^{35}Cl)$** calculated: 451.04272, found: 451.04224; **$C_{14}H_{18}O_{12}N_4P_2(^{35}Cl)$** calculated: 531.00850, found: 531.00835; **$C_{14}H_{17}O_{12}N_4P_2(^{35}Cl)Na$** calculated: 552.99044, found: 552.99013.

### 5.2.6 Chemical synthesis of 5-(5-formylthien-2-yl)-cytidine-5'-*O*-triphosphate (rC$^{FT}$TP)




Triphosphate (**rC$^{I}$TP**[287], 1 equiv.), 5-formylthiophene-2-boronic acid (1.8 equiv.), $Cs_2CO_3$ (3 equiv.), $Pd(OAc)_2$ (5 mol%) and TPPTS (25 mol%) were dissolved in a mixture of $H_2O$/acetonitrile (2:1, 1.5 mL) under argon atmosphere. The mixture was stirred at 100 °C for 1 h. Afterwards, solvents were evaporated under reduced pressure and the residue was again dissolved in $H_2O$ prior to purification. The HPLC separation was performed on C18 reversed phase column using linear gradient from 0.1 M TEAB (aq.) to 0.1 M TEAB in 50% MeOH followed by several co-evaporations with $H_2O$. The pure product was isolated as a triethylammonium salt and converted to sodium salt using Dowex 50WX8 in $Na^+$ cycle. The solvent was evaporated, pure product was again dissolved in small amount of $H_2O$ and freeze-dried overnight. The desired product **rC$^{FT}$TP** (as sodium salt) was obtained as a white solid powder in 28% yield.

**$^{1}$H NMR (500 MHz, $D_2O$, ref(*t*BuOH) = 1.25 ppm):** 4.20 – 4.32 (m, 3H, H-4',5'); 4.39 – 4.46 (m, 2H, H-2',3'); 6.01 (d, 1H, $J_{1',2'}$ = 4.3, H-1'); 7.43 (d, 1H, $J_{3,4}$ = 3.9, H-3-thienyl); 8.03 (d, 1H, $J_{4,3}$ = 3.9, H-4-thienyl); 8.06 (s, 1H, H-6); 9.84 (s, 1H, CHO).

**$^{13}$C NMR (125.7 MHz, $D_2O$, ref(*t*BuOH) = 31.60 ppm):** 66.71 (d, $J_{C,P}$ = 5.5, $CH_2$-5'); 71.12 (CH-3'); 76.18 (CH-2'); 85.05 (d, $J_{C,P}$ = 9.1, CH-4'); 91.59 (CH-1'); 104.88 (C-5); 132.52 (CH-3-thienyl); 141.95 (CH-4-thienyl); 143.74 (CH-6); 145.14 (C-5-thienyl); 146.01 (C-2-thienyl); 158.91 (C-2); 165.97 (C-4); 188.89 (CHO).

**$^{31}$P NMR (202.4 MHz, $D_2O$):** -21.68 (t, $J$ = 19.9, $P_\beta$); -10.84 (d, $J$ = 19.9, $P_\alpha$); -7.26 (m, $P_\gamma$).

**HR MS (ESI): $C_{14}H_{17}O_{15}N_3P_3S$** calculated: 591.95987, found: 591.95897; **$C_{14}H_{16}O_{15}N_3P_3SNa$** calculated: 613.94181, found: 613.94141; **$C_{14}H_{15}O_{15}N_3P_3SNa_2$** calculated: 635.92376, found: 635.92274.

### 5.2.7 General remarks – Biochemical part

All PAGE and agarose gels were analysed by fluorescence and/or phosphor imaging using Typhoon FLA 9500 (GE Healthcare Life Sciences). Quantification from dPAGE gels was performed by ImageJ. Mass spectra of oligonucleotides were measured on UltrafleXtreme MALDI-TOF/TOF (Bruker) mass spectrometer with 1 kHz smartbeam II laser technology by MS service at IOCB. LC-ESI-MS spectra of oligonucleotides were acquired on Agilent 1290 Infinity II Bio system with DAD detector and mass spectrometer MSD XT. LC-ESI-MS analysis of oligonucleotides were carried out according to standard procedures using mobile phases A (12.2 mM $Et_3N$, 300 mM HFIP in H2O) and B (12.2 mM $Et_3N$, 300 mM HFIP in H2O in 100% MeOH) by 10 min gradient from 5% B to 100% B in A using ACQUITY Premier CSH C18 column 1.7 μm 2.1×150 mm (Waters) on Agilent UHPLC Bio system. Deconvolutions of LC-ESI-MS spectra were carried out using UniDec program. Fluorescence spectra were measured in a 100 μL quartz cuvette at room temperature on a Fluoromax 4 spectrofluorometer (HORIBA Scientific). RNase/DNase free solutions for biochemical reactions were prepared using Milli-Q water, that was treated with DEPC and sterilized by autoclaving. Concentrations of the prepared RNA solutions were calculated using extinction coefficients obtained from on-line tool at https://www.atdbio.com/tools/oligo-calculator and $A_{260}$ values measured on Nanodrop 1000 (Thermo Fischer Scientific), except of Cy5-labelled mRNA samples, that were quantified by fluorescence measurements against standard curve from Cy5-labelled oligonucleotides on Tecan Spark Multimode Reader. DNA oligonucleotides were purchased from Eurofins, Generi Biotech and Biomers. RNA oligonucleotides were obtained from Biomers. HiTrap Heparin HP, HisTrap FF and HisTrap HP affinity columns were purchased from GE Healthcare. Dynabeads MyOne Streptavidin C1, Ribolock RNase inhibitor, TurboDNase, T7 RNA polymerase with the corresponding transcription buffer, RiboRuler High Range RNA ladder, RiboRuler Low Range RNA ladder, TriTrack DNA loading dye, GeneRuler 1 kb Plus DNA ladder, Uracil-DNA Glycosylase, SuperScript™ IV reverse transcriptase, Lipofectamine MessengerMAX, OptiMEM medium were purchased from Thermo Fisher Scientific. Q5 High-Fidelity DNA polymerase, deoxyribonucleotide solution mix, ribonucleotide solution mix, ThermoPol buffer, nucleoside digestion mix, Proteinase K, Shrimp Alkaline Phosphatase, Monarch RNA cleanup kit, LunaScript Multiplex one-step RT-PCR kit, magnetic mRNA isolation kit, NEB 10-beta competent E. coli (high efficiency) were bought from New England Biolabs. Microspin G-25 columns were obtained from Sigma Aldrich. mRNA template encoding plasmid was designed and purchased from VectorBuilder. QIAquick nucleotide removal kit, QIAquick gel extraction kit, QIAquick PCR purification kit were bought from QIAGEN. EZ RNA methylation kit and RNA Clean & Concentrator-25 kit were purchased from Zymo Research. BL21(DE3) pLysS cells, rabbit

reticulocyte lysate, amino acid mixture, Nano-Glo luciferase assay system were obtained from Promega. Aminoallyl-UTP-Cy3 (**rU$^{Cy3}$TP**), Aminoallyl-UTP-Cy5 (**rU$^{Cy5}$TP**), 5-Propargylamino-CTP-Cy5 (**rC$^{Cy5}$TP**), Biotin-16-UTP (**rU$^{Bio}$TP**), Digoxigenin-11-UTP (**rU$^{Dig}$TP**), 5-Methyl-CTP (**rC$^{Me}$TP**) were purchased from Jena Bioscience. NucleoSpin gel and PCR clean-up midi kit was purchased from Macherey-Nagel.

### 5.2.8 Protocol for TGK polymerase (expression and purification)

NEB 10-beta cells were transformed by pGDR11-TGLK plasmid (kind gift from J. C. Chaput) and outgrown in 20 mL 2X YT medium containing 100 µg/mL ampicillin, overnight. Next day, cells were harvested and resuspended in 10 mL of fresh 2X YT medium, which was used to inoculate 2X 1 L flasks with 2X YT medium containing 100 µg/mL ampicillin. Cultures were shaken at 37 °C until $OD_{600}$ = 1.0. Expression was triggered with IPTG (1 mM) for 5 h at 37 °C. Cells were harvested at 8 000 g, 4 °C, 10 min and the pellet was stored at -80 °C. Pellet was then resuspended in buffer containing TRIS (10 mM), NaCl (200 mM), glycerol (20%), pH 8.5 and lysed at 80 °C for 1 h, followed by cooling on ice for 20 min. Precipitates were removed at 100 000 g, 4 °C, 30 min. First purification was carried out on HiTrap Heparin HP 5 mL column by gradient from 200 to 1000 mM NaCl. Pure fractions were combined and repurified on HisTrap HP 5 mL column by elution with gradient of imidazole (from 0 to 300 mM) in 50 mM TRIS, 500 mM NaCl, 10% glycerol, pH 8.5 buffer. Pure fractions were combined, concentrated and buffer exchanged on 30 kDa MWCO Vivaspin concentrators with buffer containing HEPES (40 mM), KCl (250 mM), EDTA (1 mM), pH 7.4, quantified by Nanodrop and stored as 50% glycerol solution at -80 °C[88,89].

### 5.2.9 General procedures

**Sample preparation protocol for dPAGE analysis**

Samples were diluted with $H_2O$ to final 0.2 µM concentration. 10 µL of this solution were combined with 10 µL of 2X stop solution (95% [v/v] formamide, 0.5 mM EDTA, 0.025% [w/v] bromophenol blue, 0.025% [w/v] SDS in $H_2O$), denatured by heating at 95 °C for 2 min and then immediately cooled on ice. Aliquots of the denatured samples (10 µL) were subjected to vertical gel electrophoresis on either 12.5% or 22.5% (as indicated by gels) dPAGE (19:1 mono:bis acrylamide) containing TBE buffer (1X, pH 8.0) and urea (7 M) at 25 mA. Gels were visualised by fluorescence imaging.

**Urea agarose gel analysis**

Aliquots of **Cy5-IRES-RNA** or **Cy5-IRES-RNA_prolonged** used as standards and each of the quantified, fluorescently labelled mRNA samples (150 ng) were combined with 2X stop solution (95% [v/v] formamide, 0.5 mM EDTA, 0.025% [w/v] bromophenol blue, 0.025% [w/v] SDS in $H_2O$) and loaded on gel. 5 µL of each RNA ladders (RiboRuler High Range, RiboRuler Low Range), prepared according to manufacturer's procedure was loaded on gel. All samples were analysed on 1.5% agarose gel containing 2 M urea, TAE buffer (1X) and GelRed (1X) in TAE running buffer (1X) with addition of 2 M urea at 100 V at room temperature. The gel was afterwards visualised by a fluorescent scanner.

**Standard conditions for analytical scale PEX reaction using templ_19nt_X (X = A, U, C, G)**

Reaction mixture was heated up to 95 °C for 30 sec followed by incubation at 60 °C for 2 h in a thermal cycler with heated lid (100 °C). Afterwards, TurboDNase (2 U) was added, and the mixture was further incubated at 37 °C for 15 min, prior to dPAGE analysis.

**Modified conditions for analytical scale PEX reaction using templ_19nt_X (X = A, U, C, G)**

Both primer and template in presence of ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C $s^{-1}$). After addition of remaining components, the mixture was incubated at 60 °C for 2 h in a thermal cycler with heated lid (100 °C). After this time, TurboDNase (2 U) was added, and the mixture was further incubated at 37 °C for 15 min prior to dPAGE analysis.

**Procedure for analytical scale PEX reaction at mild conditions using templ_19nt_X (X = A, U, C, G)**

Both primer and template in presence of ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C $s^{-1}$). After addition of remaining components, the mixture was incubated at 37 °C for 2 h in a thermal cycler with heated lid (65 °C). After this time, TurboDNase (2 U) was added, and the mixture was further incubated at 37 °C for 15 min prior to dPAGE analysis.

**Standard conditions for analytical scale PEX reaction using 5'-(TINA)-templ_31nt**

Reaction mixture was heated up to 95 °C for 30 sec followed by incubation at 60 °C for 2 h in a thermal cycler with heated lid (100 °C). After this time, TurboDNase (4 U) was added, and the mixture was further incubated at 37 °C for 30 min prior to dPAGE analysis.

**Standard conditions for analytical scale PEX reaction using four different base-modified rN$^{X}$TPs and variously long templates**

Both primer and template in presence of ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C $s^{-1}$). After addition of remaining components, the mixture was incubated at 60 °C for 1 h (**templ_19nt_mix**), 2 h (**5'-(TINA)-templ_31nt**), 4 h (**templ_65nt**) or 6 h (**templ_98nt**) in a thermal cycler with heated lid (100 °C). After this time, TurboDNase (2 U for template – **templ_19nt_mix**; 4 U for templates – **5'-(TINA)-templ_31nt**; **templ_65nt**; **templ_98nt**) was added, and the mixture was further incubated at 37 °C for 30 min prior to dPAGE analysis.

**Standard conditions for semi-preparative scale PEX reaction using templ_19nt_X (X = A, U, C, G) followed by spin column purification**

Reaction mixture was heated up to 95 °C for 30 sec followed by incubation at 60 °C for 2 h in a thermal cycler with heated lid (100 °C). After this time, TurboDNase (10 U) was added and the mixture was further incubated at 37 °C for 30 min. Sample was purified using Monarch RNA cleanup kit (50 µg), according to standard supplier's protocol and eluted into DNase/RNase free tube with 50 µL of $H_2O$. Sample was evaporated to dryness and diluted again in 10 µL of $H_2O$ prior to mass spectrometry analysis.

**Modified conditions I for semi-preparative scale PEX reaction using templ_19nt_X (X = A, U, C, G) followed by spin column purification**

Both primer and template in presence of ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C $s^{-1}$). After addition of remaining components, the mixture was incubated at 60 °C for 15 min in a thermal cycler with heated lid (100 °C). After this time, TurboDNase (10 U) was added, and the mixture was further incubated at 37 °C for 30 min. The sample was purified using Monarch RNA cleanup kit (50 µg), according to standard supplier's protocol and eluted into DNase/RNase free tube with 50 µL of $H_2O$. Sample was evaporated to dryness and diluted again in 10 µL of $H_2O$ prior to mass spectrometry analysis.

**Modified conditions II for semi-preparative scale PEX reaction using templ_19nt_X (X = A, U, C, G) followed by spin column purification**

Both primer and template in presence of ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C $s^{-1}$). After addition of remaining components, the mixture was incubated at 60 °C for 2 h in a thermal cycler with heated lid (100 °C). After this time, TurboDNase (10 U) was added, and the mixture was further incubated at 37 °C for 30 min. The sample was purified

using Monarch RNA cleanup kit (50 µg), according to standard supplier's protocol and eluted into DNase/RNase free tube with 50 µL of $H_2O$. Sample was evaporated to dryness and diluted again in 10 µL of $H_2O$ prior to mass spectrometry analysis.

**Procedure for semi-preparative scale PEX reaction at mild conditions using templ_19nt_X (X = A, U, C, G) followed by spin column purification**

Both primer and template in presence of ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C $s^{-1}$). After addition of remaining components, the mixture was incubated at 37 °C for 2 h in a thermal cycler with heated lid (65 °C). After this time, TurboDNase (10 U) was added, and the mixture was further incubated at 37 °C for 30 min. The sample was purified using Monarch RNA cleanup kit (50 µg), according to standard supplier's protocol and eluted into DNase/RNase free tube with 50 µL of $H_2O$. Sample was evaporated to dryness and diluted again in 10 µL of $H_2O$ prior to mass spectrometry analysis.

**Conditions for semi-preparative scale PEX reaction using 5'-(TINA)-templ_31nt followed by spin column purification**

Reaction mixture was heated up to 95 °C for 30 sec followed by incubation at 60 °C for 2 h in a thermal cycler with heated lid (100 °C). Afterwards, TurboDNase (20 U) was added, and the mixture was further incubated at 37 °C for 30 min. Sample was purified using Monarch RNA cleanup kit (50 µg), according to standard supplier's protocol. Elution was performed into DNase/RNase free tube with 50 µL of $H_2O$. Sample was evaporated to dryness and diluted again in 10 µL of $H_2O$ prior to mass spectrometry analysis.

**Conditions for semi-preparative scale PEX reaction using 5'-(dual-Bio)-templ_31nt followed by magnetoseparation**

Both primer and template in presence of ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C $s^{-1}$). After addition of remaining components, the mixture was incubated at 60 °C for 6 h in a thermal cycler with heated lid (100 °C). After PEX reaction, the ssRNA strand was generated by magnetoseparation. Briefly, pre-washed SMB (100 µL) were combined with crude PEX reaction and 100 µL of binding buffer [TRIS (10 mM), EDTA (1 mM), NaCl (100 mM), pH 7.5] and incubated at 25 °C for 1 h at 800 rpm in a thermal mixer. Then, SMB were washed successively with wash buffer [TRIS (10 mM), EDTA (1 mM), NaCl (500 mM), pH 7.5] and water. Strand separation was performed in 50 µL of water by heating up the mixture to 70 °C for 5 min. Isolated product was cooled down, additionally combined with TurboDNase (10 U) and incubated at

37 °C for 1 h. Sample was evaporated to dryness and diluted again in 10 µL of $H_2O$ prior to mass spectrometry analysis.

**Conditions for semi-preparative scale PEX reaction using four different base-modified rN$^X$TPs and variously long templates**

Both primer and template in presence of ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C $s^{-1}$). After addition of remaining components, the mixture was incubated at 60 °C for 1 h (**templ_19nt_mix**), 2 h (**5'-(TINA)-templ_31nt**), 4 h (**templ_65nt**) or 6 h (**templ_98nt**) in a thermal cycler with heated lid (100 °C). After this time, TurboDNase (10 U for template – **templ_19nt_mix**; 20 U for templates – **5'-(TINA)-templ_31nt**; **templ_65nt**; **templ_98nt**) was added, and the mixture was further incubated at 37 °C for 30 min. Crude sample was further purified using QIAquick nucleotide removal kit according to standard supplier's protocol. Elution was performed into DNase/RNase free tube with 50 µL of $H_2O$. Sample was evaporated to dryness and diluted again in 10 µL of $H_2O$ prior to mass spectrometry analysis.

**Standard protocol for RNA digestion and mass spectrometry analysis**

Samples prepared either by PEX reaction with **rA$^{CA}$TP** (**19RNA_A$^{CA}$**) and/or by PEX reaction with **rC$^{CA}$TP** (**19RNA_C$^{CA}$**) were purified, lyophilised and again diluted in 17 µL of $H_2O$. Samples were combined with nucleoside digestion mix reaction buffer (2 µL, 10X) and nucleoside digestion mix (1 µL) followed by incubation at 37 °C for 1 h in a thermal cycler with heated lid (65 °C). 1 µL of the reaction was then directly analysed on Waters UPLC-H system coupled to Xevo G2-XS mass spectrometer in ESI+ mode with internal calibration standard Leucine Enkephalin (556.2771 Da). Separation was performed on Luna C18 PS 2.1×150 mm column using mobile phases: (A) 0.1% formic acid in $H_2O$ and (B) 0.1% formic acid in acetonitrile. Gradient was set as follows: 1.5 min hold 1% B then 7.5 min gradient from 1 to 50% B. Mass spectrometer settings were collection mass range 50-1200 Da, scan duration 0.5 s, 3.0 kV capillary voltage, 40 V sampling cone, 100 °C source temperature, 250 °C desolvation temperature, 50 L/h cone gas flow, 600 L/h desolvation gas flow.

**Standard magnetoseparation procedure for generation of ssDNA templates after PCR**

Generation of ssDNA templates was ensured by magnetoseparation procedure. Briefly, pre-washed SMB (100 µL per 6.5 µg dsDNA template) were combined with purified PCR product and binding buffer [TRIS (10 mM), EDTA (1 mM), NaCl (100 mM), TWEEN 20 (0.1%), pH 7.5] and incubated overnight at 25 °C on Hula-mixer. Then, SMB were successively washed with wash buffer [TRIS (10 mM), EDTA (1 mM), NaCl (500 mM), TWEEN 20 (0.1%), pH 7.5] and water. Strand separation was

performed by addition of NaOH (50 mM), followed by neutralisation with TRIS (1 M, pH 7.5) and HCl (1 M). The generated ssDNA product was additionally purified using QIAquick nucleotide removal kit, according to standard supplier's protocol.

**Standard conditions for PCR reaction**

PCR reaction was performed in a thermal cycler as follows: denaturation at 98 °C for 30 sec, then 25 cycles of denaturation at 98 °C for 15 sec, annealing at 68 °C for 30 sec and extension at 72 °C for 1 min, followed by final elongation step at 72 °C for 2 min.

**Standard work up protocol for either IRES-RNA or IRES-RNA_prolonged**

After PEX, the mixture (10 µL) was combined with TurboDNase (2 U) and incubated at 37 °C for 30 min followed by Proteinase K (0.8 U) digestion at 37 °C for 30 min. Final RNA was purified according to standard supplier's protocol using RNA Clean & Concentrator-25 kit and quantified by fluorescence (Cy5) using Tecan Spark Multimode Reader. For 10X scaled-up reactions, the amount of TurboDNase (20 U) and Proteinase K (8 U) was increased accordingly.

**Standard work up protocol for mRNA**

After PEX, the mixture (10 µL) was combined with TurboDNase (2 U) and incubated at 37 °C for 30 min followed by Proteinase K (0.8 U) digestion at 37 °C for 30 min, prior to urea gel analysis. Scaled-up reactions (30 µL) were treated with TurboDNase (6 U) and Proteinase K (2.4 U) following same conditions. Final mRNA mixtures were purified using magnetic mRNA isolation kit with a slightly modified protocol. Briefly, 500 µL of oligo-$dT_{25}$-MB pre-washed with binding buffer were incubated with mRNA (30 µL) at 25 °C for 1.5 h at 300 rpm in a thermal shaker. Afterwards, the oligo-$dT_{25}$-MB were washed with wash buffer 1, followed by wash buffer 2 and low salt buffer (each 1X 1 mL). The elution was performed with 100 µL of elution buffer at 50 °C for 3 min at 300 rpm in a thermal shaker. After lyophilisation to dryness, the mRNA was again diluted in 20 µL of $H_2O$, quantified by fluorescence (Cy5) using Tecan Spark Multimode Reader and used in *in vitro* and/or *in cellulo* translation studies.

**Standard reaction conditions for RT-PCR**

Reaction was treated as follows. For RT incubation at 50 °C for 10 min, then at 55 °C for 10 min was performed, followed by enzyme denaturation at 98 °C for 1 min. For PCR, 35 cycles of denaturation at 98 °C for 10 sec, annealing at 53 °C for 30 sec and extension at 72 °C for 40 sec were applied, followed by final elongation step at 72 °C for 5 min.

### 5.2.10 Analytical scale PEX reaction with templ_19nt_X (X = A, U, C, G) and TGK polymerase (incorporation of 1 modification)

**Incorporation of $rA^{E}TP$, $rA^{Pent}TP$, $rA^{Ph}TP$**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_A** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.5 µM), rGTP (0.1 mM) and one of the modified **$rA^{E}TP$**, **$rA^{Pent}TP$** or **$rA^{Ph}TP$** (0.1 mM). Positive control was performed under same conditions with natural rATP (0.1 mM) instead of modified $rA^{X}TP$s. For negative control reaction $H_2O$ was used instead of rATP. Reactions were carried out following the standard protocol.

**Incorporation of $rA^{CA}TP$, $rA^{FT}TP$**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_A** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.25 µM), rGTP (0.1 mM) and one of the modified **$rA^{CA}TP$** or **$rA^{FT}TP$** (0.05 mM). Positive control was performed under same conditions with natural rATP (0.05 mM) instead of modified $rA^{X}TP$s. For negative control reaction $H_2O$ was used instead of rATP. Reactions were carried out following the standard protocol.

**Incorporation of $rA^{CA}TP$ at mild conditions**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_A** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.5 µM), rGTP (0.2 mM) and the modified **$rA^{CA}TP$** (0.05 mM). Positive control was performed under same conditions with natural rATP (0.05 mM) instead of modified **$rA^{CA}TP$**. For negative control reaction $H_2O$ was used instead of rATP. Reactions were carried out following the standard protocol.

**Incorporation of $rU^{E}TP$, $rU^{Pent}TP$, $rU^{Ph}TP$**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_U** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.75 µM), rGTP (0.05 mM) and one of the modified **$rU^{E}TP$**, **$rU^{Pent}TP$** or **$rU^{Ph}TP$** (0.1 mM). Positive control was performed under same conditions with natural rUTP (0.1 mM) instead of modified $rU^{X}TP$s. For negative control reaction $H_2O$ was used instead of rUTP. Reactions were carried out following the standard protocol.

**Incorporation of rU$^{Bio}$TP, rU$^{Dig}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_U** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (1.0 µM), rGTP (0.05 mM) and one of the modified **rU$^{Bio}$TP** or **rU$^{Dig}$TP** (0.1 mM). Positive control was performed under same conditions with natural rUTP (0.1 mM) instead of modified rU$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rUTP. Reactions were carried out following the standard protocol.

**Incorporation of rU$^{CA}$TP, rU$^{FT}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_U** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.5 µM), rGTP (0.05 mM) and one of the modified **rU$^{CA}$TP** or **rU$^{FT}$TP** (0.05 mM). Positive control was performed under same conditions with natural rUTP (0.05 mM) instead of modified rU$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rUTP. Reactions were carried out following the standard protocol.

**Incorporation of rC$^{Me}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.5 µM), rGTP (0.1 mM) and the modified **rC$^{Me}$TP** (0.1 mM). Positive control was performed under same conditions with natural rCTP (0.1 mM) instead of modified **rC$^{Me}$TP**. For negative control reaction $H_2O$ was used instead of rCTP. Reactions were carried out following the standard protocol.

**Incorporation of rC$^{E}$TP, rC$^{Pent}$TP, rC$^{Ph}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.5 µM), rGTP (0.2 mM) and one of the modified **rC$^{E}$TP**, **rC$^{Pent}$TP** or **rC$^{Ph}$TP** (0.1 mM). Positive control was performed under same conditions with natural rCTP (0.1 mM) instead of modified rC$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rCTP. Reactions were carried out following the standard protocol.

**Incorporation of rC$^{mBdp}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 µM), 5'-(Cy5)-labelled RNA primer – **Cy5-RNA-prim_15nt** (4.0 µM), TGK polymerase (2.0 µM), rGTP (0.05 mM) and the modified **rC$^{mBdp}$TP** (0.1 mM). Positive control was performed under same conditions with natural rCTP (0.1 mM) instead of modified **rC$^{mBdp}$TP**. For negative control reaction $H_2O$ was used instead of rCTP. Reactions were carried out following the standard protocol.

**Incorporation of rC$^{Cy5}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (2.0 µM), rGTP (0.05 mM) and the modified **rC$^{Cy5}$TP** (0.1 mM). Positive control was performed under same conditions with natural rCTP (0.1 mM) instead of modified **rC$^{Cy5}$TP**. For negative control reaction $H_2O$ was used instead of rCTP. Reactions were carried out following the standard protocol.

**Incorporation of rC$^{CA}$TP, rC$^{FT}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.5 µM), rGTP (0.2 mM) and one of the modified **rC$^{CA}$TP** or **rC$^{FT}$TP** (0.05 mM). Positive control was performed under same conditions with natural rCTP (0.05 mM) instead of modified rC$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rCTP. Reactions were carried out following the standard protocol.

**Incorporation of rC$^{CA}$TP at mild conditions**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (1.5 µM), rGTP (0.4 mM) and the modified **rC$^{CA}$TP** (0.05 mM). Positive control was performed under same conditions with natural rCTP (0.05 mM) instead of modified **rC$^{CA}$TP**. For negative control reaction $H_2O$ was used instead of rCTP. Reactions were carried out following the standard protocol.

**Incorporation of rG$^{E}$TP, rG$^{Pent}$TP, rG$^{Ph}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA **templ_19nt_G** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (1.0 µM), rUTP (0.2 mM) and one of the modified **rG$^{E}$TP**, **rG$^{Pent}$TP** or **rG$^{Ph}$TP** (0.1 mM). Positive control was performed under same conditions with natural rGTP (0.1 mM) instead of modified rG$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rGTP. Reactions were carried out following the standard protocol.

### 5.2.11 Analytical scale PEX reaction with templ_19nt_X (X = A, U, C, G) and SFM4-3 polymerase (incorporation of 1 modification)

**Incorporation of rA$^{E}$TP, rA$^{Pent}$TP, rA$^{Ph}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_A** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (0.5 µM), rGTP (0.1 mM) and one of the modified **rA$^{E}$TP**, **rA$^{Pent}$TP** or **rA$^{Ph}$TP** (0.1 mM). Positive control was performed under same conditions with natural rATP (0.1 mM) instead of modified rA$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rATP. Reactions were carried out following the standard protocol.

**Incorporation of rA$^{CA}$TP, rA$^{FT}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_A** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (0.5 µM), rGTP (0.2 mM) and one of the modified **rA$^{CA}$TP** or **rA$^{FT}$TP** (0.05 mM). Positive control was performed under same conditions with natural rATP (0.05 mM) instead of modified rA$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rATP. Reactions were carried out following the standard protocol.

**Incorporation of rA$^{CA}$TP at mild conditions**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_A** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (1.0 µM), rGTP (0.4 mM) and the modified **rA$^{CA}$TP** (0.05 mM). Positive control was performed under same conditions with natural rATP (0.05 mM) instead of modified **rA$^{CA}$TP**. For negative control reaction $H_2O$ was used instead of rATP. Reactions were carried out following the standard protocol.

**Incorporation of rU$^{E}$TP, rU$^{Pent}$TP, rU$^{Ph}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_U** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (0.5 µM), rGTP (0.1 mM) and one of the modified **rU$^{E}$TP**, **rU$^{Pent}$TP** or **rU$^{Ph}$TP** (0.1 mM). Positive control was performed under same conditions with natural rUTP (0.1 mM) instead of modified rU$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rUTP. Reactions were carried out following the standard protocol.

**Incorporation of rU$^{Bio}$TP, rU$^{Dig}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_U** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (2.0 µM), rGTP (0.05 mM) and one of the modified **rU$^{Bio}$TP** or **rU$^{Dig}$TP** (0.1 mM). Positive control was performed under same conditions with natural rUTP (0.1 mM) instead of modified rU$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rUTP. Reactions were carried out following the standard protocol.

**Incorporation of rU$^{CA}$TP, rU$^{FT}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_U** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (1.0 µM), rGTP (0.1 mM) and one of the modified **rU$^{CA}$TP** or **rU$^{FT}$TP** (0.05 mM). Positive control was performed under same conditions with natural rUTP (0.05 mM) instead of modified rU$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rUTP. Reactions were carried out following the standard protocol in section.

**Incorporation of rC$^{E}$TP, rC$^{Pent}$TP, rC$^{Ph}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (0.5 µM), rGTP (0.1 mM) and one of the modified **rC$^{E}$TP**, **rC$^{Pent}$TP** or **rC$^{Ph}$TP** (0.1 mM). Positive control was performed under same conditions with natural rCTP (0.1 mM) instead of modified rC$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rCTP. Reactions were carried out following the standard protocol.

**Incorporation of rC$^{mBdp}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 µM), 5'-(Cy5)-labelled RNA primer – **Cy5-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (2.0 µM), rGTP (0.05 mM) and the modified **rC$^{mBdp}$TP** (0.1 mM). Positive control was performed under same conditions with natural rCTP (0.1 mM) instead of modified **rC$^{mBdp}$TP**. For negative control reaction $H_2O$ was used instead of rCTP. Reactions were carried out following the standard protocol.

**Incorporation of rC$^{Cy5}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA templare – **templ_19nt_C** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (2.0 µM), rGTP (0.05 mM) and the modified **rC$^{Cy5}$TP** (0.1 mM). Positive control was performed under same conditions with natural rCTP (0.1 mM) instead of modified **rC$^{Cy5}$TP**. For negative control reaction $H_2O$ was used instead of rCTP. Reactions were carried out following the standard protocol.

**Incorporation of rC$^{CA}$TP, rC$^{FT}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (0.5 µM), rGTP (0.2 mM) and one of the modified **rC$^{CA}$TP** or **rC$^{FT}$TP** (0.05 mM). Positive control was performed under same conditions with natural rCTP (0.05 mM) instead of modified rC$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rCTP. Reactions were carried out following the standard protocol.

**Incorporation of rC$^{CA}$TP at mild conditions**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (1.0 µM), rGTP (0.4 mM) and the modified **rC$^{CA}$TP** (0.05 mM). Positive control was performed under same conditions with natural rCTP (0.05 mM) instead of modified **rC$^{CA}$TP**. For negative control reaction $H_2O$ was used instead of rCTP. Reactions were carried out following the standard protocol.

**Incorporation of rG$^{E}$TP, rG$^{Pent}$TP, rG$^{Ph}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_G** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (1.5 µM), rUTP (0.4 mM) and one of the modified **rG$^{E}$TP**, **rG$^{Pent}$TP** or **rG$^{Ph}$TP** (0.1 mM). Positive control was performed under same conditions with natural rGTP (0.1 mM) instead of modified rG$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rGTP. Reactions were carried out following the standard protocol.

### 5.2.12 Semi-preparative scale PEX reaction with templ_19nt_X (X = A, U, C, G) and TGK polymerase (incorporation of 1 modification)

**Incorporation of rA$^{E}$TP, rA$^{Pent}$TP, rA$^{Ph}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_A** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.5 µM), rGTP (0.1 mM) and one of the modified **rA$^{E}$TP**, **rA$^{Pent}$TP** or **rA$^{Ph}$TP** (0.1 mM). Reactions were carried out following the standard protocol.

**Incorporation of rA$^{CA}$TP, rA$^{FT}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_A** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.25 µM), rGTP (0.1 mM) and one of the modified **rA$^{CA}$TP** or **rA$^{FT}$TP** (0.05 mM). Reactions were carried out following the standard protocol.

**Incorporation of rA$^{CA}$TP at reduced time**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_A** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.25 µM), rGTP (0.1 mM) and the modified **rA$^{CA}$TP** (0.05 mM). Reaction was carried out following the standard protocol.

**Incorporation of rA$^{CA}$TP at mild conditions**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_A** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.5 µM), rGTP (0.2 mM) and the modified **rA$^{CA}$TP** (0.05 mM). Reaction was carried out following the standard protocol.

**Incorporation of rU$^{E}$TP, rU$^{Pent}$TP, rU$^{Ph}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_U** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.75 µM), rGTP (0.05 mM) and one of the modified **rU$^{E}$TP**, **rU$^{Pent}$TP** or **rU$^{Ph}$TP** (0.1 mM). Reactions were carried out following the standard protocol.

**Incorporation of rU$^{Bio}$TP, rU$^{Dig}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_U** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (1.0 µM), rGTP (0.05 mM) and one of the modified **rU$^{Bio}$TP** or **rU$^{Dig}$TP** (0.1 mM). Reactions were carried out following the standard protocol.

**Incorporation of rU$^{CA}$TP, rU$^{FT}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_U** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.5 µM), rGTP (0.05 mM) and one of the modified **rU$^{CA}$TP** or **rU$^{FT}$TP** (0.05 mM). Reactions were carried out following the standard protocol.

**Incorporation of rC$^{Me}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 µM), RNA primer – **RNA-prim_15nt** (4.0 µM), TGK polymerase (0.5 µM), rGTP (0.1 mM) and the modified **rC$^{Me}$TP** (0.1 mM). Reaction was carried out following the standard protocol.

**Incorporation of rC$^{E}$TP, rC$^{Pent}$TP, rC$^{Ph}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.5 µM), rGTP (0.2 mM) and one of the modified **rC$^{E}$TP**, **rC$^{Pent}$TP** or **rC$^{Ph}$TP** (0.1 mM). Reactions were carried out following the standard protocol.

**Incorporation of rC$^{mBdp}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (2.4 µM), 5'-(Cy5)-labelled RNA primer – **Cy5-RNA-prim_15nt**

(2.0 μM), TGK polymerase (2.0 μM), rGTP (0.05 mM) and the modified **rC$^{mBdp}$TP** (0.1 mM). Reaction was carried out following the standard protocol.

**Incorporation of rC$^{Cy5}$TP**

Reaction was performed in total volume of 50 μL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 μM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 μM), TGK polymerase (2.0 μM), rGTP (0.05 mM) and the modified **rC$^{Cy5}$TP** (0.1 mM). Reaction was carried out following the standard protocol.

**Incorporation of rC$^{CA}$TP, rC$^{FT}$TP**

Reaction was performed in total volume of 50 μL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 μM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 μM), TGK polymerase (0.5 μM), rGTP (0.2 mM) and one of the modified **rC$^{CA}$TP** or **rC$^{FT}$TP** (0.05 mM). Reactions were carried out following the standard protocol.

**Incorporation of rC$^{CA}$TP at reduced time**

Reaction was performed in total volume of 50 μL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 μM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 μM), TGK polymerase (0.5 μM), rGTP (0.2 mM) and the modified **rC$^{CA}$TP** (0.05 mM). Reaction was carried out following the standard protocol.

**Incorporation of rC$^{CA}$TP at mild conditions**

Reaction was performed in total volume of 50 μL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 μM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 μM), TGK polymerase (1.5 μM), rGTP (0.4 mM) and the modified **rC$^{CA}$TP** (0.05 mM). Reaction was carried out following the standard protocol.

**Incorporation of rG$^{E}$TP, rG$^{Pent}$TP, rG$^{Ph}$TP**

Reaction was performed in total volume of 50 μL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_G** (4.8 μM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 μM), TGK polymerase (1.0 μM), rUTP (0.2 mM) and one of the modified **rG$^{E}$TP**, **rG$^{Pent}$TP** or **rG$^{Ph}$TP** (0.1 mM). Reactions were carried out following the standard protocol.

### 5.2.13 Semi-preparative scale PEX reaction with templ_19nt_X (X = A, U, C, G) and SFM4-3 polymerase (incorporation of 1 modification)

**Incorporation of rA$^{E}$TP, rA$^{Pent}$TP, rA$^{Ph}$TP using SFM4-3 polymerase**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_A** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (0.5 µM), rGTP (0.1 mM) and one of the modified **rA$^{E}$TP**, **rA$^{Pent}$TP** or **rA$^{Ph}$TP** (0.1 mM). Reactions were carried out following the standard protocol.

**Incorporation of rA$^{CA}$TP, rA$^{FT}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_A** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (0.5 µM), rGTP (0.2 mM) and one of the modified **rA$^{CA}$TP** or **rA$^{FT}$TP** (0.05 mM). Reactions were carried out following the standard protocol.

**Incorporation of rA$^{CA}$TP at mild conditions**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_A** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (1.0 µM), rGTP (0.4 mM) and the modified **rA$^{CA}$TP** (0.05 mM). Reaction was carried out following the standard protocol.

**Incorporation of rU$^{E}$TP, rU$^{Pent}$TP, rU$^{Ph}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_U** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (0.5 µM), rGTP (0.1 mM) and one of the modified **rU$^{E}$TP**, **rU$^{Pent}$TP** or **rU$^{Ph}$TP** (0.1 mM). Reactions were carried out following the standard protocol.

**Incorporation of rU$^{Bio}$TP, rU$^{Dig}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_U** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (2.0 µM), rGTP (0.05 mM) and one of the modified **rU$^{Bio}$TP** or **rU$^{Dig}$TP** (0.1 mM). Reactions were carried out following the standard protocol.

**Incorporation of rU$^{CA}$TP, rU$^{FT}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_U** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (1.0 µM), rGTP (0.1 mM) and one of the modified **rU$^{CA}$TP** or **rU$^{FT}$TP** (0.05 mM). Reactions were carried out following the standard protocol.

**Incorporation of rC$^{E}$TP, rC$^{Pent}$TP, rC$^{Ph}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (0.5 µM), rGTP (0.1 mM) and one of the modified **rC$^{E}$TP**, **rC$^{Pent}$TP** or **rC$^{Ph}$TP** (0.1 mM). Reactions were carried out following the standard protocol.

**Incorporation of rC$^{mBdp}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 µM), 5'-(Cy5)-labelled RNA primer – **Cy5-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (2.0 µM), rGTP (0.05 mM) and the modified **rC$^{mBdp}$TP** (0.1 mM). Reaction was carried out following the standard protocol.

**Incorporation of rC$^{Cy5}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (2.0 µM), rGTP (0.05 mM) and the modified **rC$^{Cy5}$TP** (0.1 mM). Reaction was carried out following the standard protocol.

**Incorporation of rC$^{CA}$TP, rC$^{FT}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (0.5 µM), rGTP (0.2 mM) and one of the modified **rC$^{CA}$TP** or **rC$^{FT}$TP** (0.05 mM). Reactions were carried out following the standard protocol.

**Incorporation of rC$^{CA}$TP at mild conditions**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_C** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0

μM), SFM4-3 polymerase (1.0 μM), rGTP (0.4 mM) and the modified **rC$^{CA}$TP** (0.05 mM). Reaction was carried out following the standard protocol.

**Incorporation of rG$^{E}$TP, rG$^{Pent}$TP, rG$^{Ph}$TP**

Reaction was performed in total volume of 50 μL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_G** (4.8 μM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 μM), SFM4-3 polymerase (1.5 μM), rUTP (0.4 mM) and one of the modified **rG$^{E}$TP**, **rG$^{Pent}$TP** or **rG$^{Ph}$TP** (0.1 mM). Reactions were carried out following the standard protocol.

### 5.2.14 Analytical scale PEX reaction with 5'-(TINA)-templ_31nt and TGK polymerase (incorporation of 4 modifications)

**Incorporation of rA$^{E}$TP, rA$^{Pent}$TP, rA$^{Ph}$TP**

Reaction was performed in total volume of 10 μL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 μM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 μM), TGK polymerase (0.25 μM), mixture of rUTP, rCTP, rGTP (0.4 mM) and one of the modified **rA$^{E}$TP**, **rA$^{Pent}$TP** or **rA$^{Ph}$TP** (0.2 mM). Positive control was performed under same conditions with natural rATP (0.2 mM) instead of modified rA$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rATP. Reactions were carried out following the standard protocol.

**Incorporation of rU$^{E}$TP, rU$^{Pent}$TP, rU$^{Ph}$TP**

Reaction was performed in total volume of 10 μL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 μM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 μM), TGK polymerase (0.5 μM), mixture of rATP, rCTP, rGTP (0.4 mM) and one of the modified **rU$^{E}$TP**, **rU$^{Pent}$TP** or **rU$^{Ph}$TP** (0.2 mM). Positive control was performed under same conditions with natural rUTP (0.2 mM) instead of modified rU$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rUTP. Reactions were carried out following the standard protocol.

**Incorporation of rU$^{Bio}$TP, rU$^{Dig}$TP**

Reaction was performed in total volume of 10 μL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 μM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 μM), TGK polymerase (1.0 μM), mixture of rATP, rCTP, rGTP (0.2 mM) and one of the modified **rU$^{Bio}$TP** or **rU$^{Dig}$TP** (0.2 mM). Positive control was performed under same conditions with natural rUTP (0.2 mM) instead of modified rU$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rUTP. Reactions were carried out following the standard protocol.

**Incorporation of rC$^{Me}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **5'-(TINA)-templ_31nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.5 µM), mixture of rATP, rUTP, rGTP (0.4 mM) and the modified **rC$^{Me}$TP** (0.2 mM). Positive control was performed under same conditions with natural rCTP (0.2 mM) instead of modified **rC$^{Me}$TP**. For negative control reaction $H_2O$ was used instead of rCTP. Reactions were carried out following the standard protocol.

**Incorporation of rC$^{E}$TP, rC$^{Pent}$TP, rC$^{Ph}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.5 µM), mixture of rATP, rUTP, rGTP (0.4 mM) and one of the modified **rC$^{E}$TP**, **rC$^{Pent}$TP** or **rC$^{Ph}$TP** (0.2 mM). Positive control was performed under same conditions with natural rCTP (0.2 mM) instead of modified rC$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rCTP. Reactions were carried out following the standard protocol.

**Incorporation of rC$^{mBdp}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (2.4 µM), 5'-(Cy5)-labelled RNA primer – **Cy5-RNA-prim_15nt** (2.0 µM), TGK polymerase (2.0 µM), mixture of rATP, rUTP, rGTP (0.4 mM) and the modified **rC$^{mBdp}$TP** (0.8 mM). Positive control was performed under same conditions with natural rCTP (0.4 mM) instead of modified **rC$^{mBdp}$TP**. For negative control reaction $H_2O$ was used instead of rCTP. Reactions were carried out following the standard protocol.

**Incorporation of rG$^{E}$TP, rG$^{Pent}$TP, rG$^{Ph}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (1.5 µM), mixture of rATP, rUTP, rCTP (0.8 mM) and one of the modified **rG$^{E}$TP**, **rG$^{Pent}$TP** or **rG$^{Ph}$TP** (0.4 mM). Positive control was performed under same conditions with natural rGTP (0.4 mM) instead of modified rG$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rGTP. Reactions were carried out following the standard protocol.

### 5.2.15 Analytical scale PEX reaction with 5'-(TINA)-templ_31nt and SFM4-3 polymerase (incorporation of 4 modifications)

**Incorporation of rA$^{E}$TP, rA$^{Pent}$TP, rA$^{Ph}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (0.25 µM), mixture of rUTP, rCTP, rGTP (0.4 mM) and one of the modified **rA$^{E}$TP**, **rA$^{Pent}$TP** or **rA$^{Ph}$TP** (0.2 mM). Positive control was performed under same conditions with natural rATP (0.2 mM) instead of modified rA$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rATP. Reactions were carried out following the standard protocol.

**Incorporation of rU$^{E}$TP, rU$^{Pent}$TP, rU$^{Ph}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (0.75 µM), mixture of rATP, rCTP, rGTP (0.4 mM) and one of the modified **rU$^{E}$TP**, **rU$^{Pent}$TP** or **rU$^{Ph}$TP** (0.2 mM). Positive control was performed under same conditions with natural rUTP (0.2 mM) instead of the modified rU$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rUTP. Reactions were carried out following the standard protocol.

**Incorporation of rU$^{Bio}$TP, rU$^{Dig}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (1.0 µM), mixture of rATP, rCTP, rGTP (0.2 mM) and one of the modified **rU$^{Bio}$TP** or **rU$^{Dig}$TP** (0.2 mM). Positive control was performed under same conditions with natural rUTP (0.2 mM) instead of modified rU$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rUTP. Reactions were carried out following the standard protocol.

**Incorporation of rC$^{E}$TP, rC$^{Pent}$TP, rC$^{Ph}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing labelled ssDNA template **5'-(TINA)-templ_31nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (0.5 µM), mixture of rATP, rUTP, rGTP (0.4 mM) and one of the modified **rC$^{E}$TP**, **rC$^{Pent}$TP** or **rC$^{Ph}$TP** (0.2 mM). Positive control was performed under same conditions with natural rCTP (0.2 mM) instead of modified rC$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rCTP. Reactions were carried out following the standard protocol.

**Incorporation of rC$^{mBdp}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (2.4 µM), 5'-(Cy5)-labelled RNA primer – **Cy5-RNA-prim_15nt** (2.0 µM), SFM4-3 polymerase (2.0 µM), mixture of rATP, rUTP, rGTP (0.4 mM) and the modified **rC$^{mBdp}$TP** (0.8 mM). Positive control was performed under same conditions with natural rCTP (0.4 mM) instead of modified **rC$^{mBdp}$TP**. For negative control reaction $H_2O$ was used instead of rCTP. Reactions were carried out following the standard protocol.

**Incorporation of rG$^{E}$TP, rG$^{Pent}$TP, rG$^{Ph}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (2.0 µM), mixture of rATP, rUTP, rCTP (1.0 mM) and one of the modified **rG$^{E}$TP**, **rG$^{Pent}$TP** or **rG$^{Ph}$TP** (0.4 mM). Positive control was performed under same conditions with natural rGTP (0.4 mM) instead of modified rG$^{X}$TPs. For negative control reaction $H_2O$ was used instead of rGTP. Reactions were carried out following the standard protocol.

### 5.2.16 Semi-preparative scale PEX reaction with 5'-(TINA)-templ_31nt or 5'-(dual-Bio)-templ_31nt and TGK polymerase (incorporation of 4 modifications)

**Incorporation of rA$^{E}$TP, rA$^{Pent}$TP, rA$^{Ph}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.25 µM), mixture of rUTP, rCTP, rGTP (0.4 mM) and one of the modified **rA$^{E}$TP**, **rA$^{Pent}$TP** or **rA$^{Ph}$TP** (0.2 mM). Reactions were carried out following the standard protocol.

**Incorporation of rU$^{E}$TP, rU$^{Pent}$TP, rU$^{Ph}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.5 µM), mixture of rATP, rCTP, rGTP (0.4 mM) and one of the modified **rU$^{E}$TP**, **rU$^{Pent}$TP** or **rU$^{Ph}$TP** (0.2 mM). Reactions were carried out following the standard protocol.

**Incorporation of rU$^{Bio}$TP, rU$^{Dig}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (1.0 µM), mixture of rATP, rCTP, rGTP (0.2 mM) and one of the modified **rU$^{Bio}$TP** or **rU$^{Dig}$TP** (0.2 mM). Reactions were carried out following the standard protocol.

**Incorporation of rC$^{Me}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **5'-(TINA)-templ_31nt** (4.8 µM), RNA primer – **RNA-prim_15nt** (4.0 µM), TGK polymerase (0.5 µM), mixture of rATP, rUTP, rGTP (0.4 mM) and the modified **rC$^{Me}$TP** (0.2 mM). Reaction was carried out following the standard protocol.

**Incorporation of rC$^{E}$TP, rC$^{Pent}$TP, rC$^{Ph}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (0.5 µM), mixture of rATP, rUTP, rGTP (0.4 mM) and one of the modified **rC$^{E}$TP**, **rC$^{Pent}$TP** or **rC$^{Ph}$TP** (0.2 mM). Reactions were carried out following the standard protocol.

**Incorporation of rC$^{mBdp}$TP**

Reaction was performed in total volume of 100 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(dual-Bio)-templ_31nt** (2.4 µM), RNA primer – **RNA-prim_15nt** (2.0 µM), TGK polymerase (2.0 µM), mixture of rATP, rUTP, rGTP (0.4 mM) and the modified **rC$^{mBdp}$TP** (0.8 mM). Reaction was carried out following the standard protocol.

**Incorporation of rG$^{E}$TP, rG$^{Pent}$TP, rG$^{Ph}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (1.5 µM), mixture of rATP, rUTP, rCTP (0.8 mM) and one of the modified **rG$^{E}$TP**, **rG$^{Pent}$TP** or **rG$^{Ph}$TP** (0.4 mM). Reactions were carried out following the standard protocol.

**Semi-preparative scale PEX reaction with with 5'-(TINA)-templ_31nt or 5'-(dual-Bio)-templ_31nt and SFM4-3 polymerase (incorporation of 4 modifications)**

**Incorporation of rA$^{E}$TP, rA$^{Pent}$TP, rA$^{Ph}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing labelled ssDNA template **5'-(TINA)-templ_31nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (0.25 µM), mixture of rUTP, rCTP, rGTP (0.4 mM) and one of the modified **rA$^{E}$TP**, **rA$^{Pent}$TP** or **rA$^{Ph}$TP** (0.2 mM). Reactions were carried out following the standard protocol.

**Incorporation of rU$^{E}$TP, rU$^{Pent}$TP, rU$^{Ph}$TP**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (0.75 µM), mixture of rATP, rCTP, rGTP (0.4 mM) and one of the modified **rU$^{E}$TP**, **rU$^{Pent}$TP** or **rU$^{Ph}$TP** (0.2 mM). Reactions were carried out following the standard protocol.

**Incorporation of rU$^{Bio}$TP, rU$^{Dig}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (1.0 µM), mixture of rATP, rCTP, rGTP (0.2 mM) and one of the modified **rU$^{Bio}$TP** or **rU$^{Dig}$TP** (0.2 mM). Reactions were carried out following the standard protocol.

**Incorporation of rC$^{E}$TP, rC$^{Pent}$TP, rC$^{Ph}$TP**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (0.5 µM), mixture of rATP, rUTP, rGTP (0.4 mM) and one of the modified **rC$^{E}$TP**, **rC$^{Pent}$TP** or **rC$^{Ph}$TP** (0.2 mM). Reactions were carried out following the standard protocol.

**Incorporation of rC$^{mBdp}$TP**

Reaction was performed in total volume of 100 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(dual-Bio)-templ_31nt** (2.4 µM), RNA primer – **RNA-prim_15nt** (2.0 µM),

SFM4-3 polymerase (2.0 μM), mixture of rATP, rUTP, rGTP (0.4 mM) and the modified **rC$^{mBdp}$TP** (0.8 mM). Reaction was carried out following the standard protocol.

**Incorporation of rG$^{E}$TP, rG$^{Pent}$TP, rG$^{Ph}$TP**

Reaction was performed in total volume of 50 μL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 μM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 μM), SFM4-3 polymerase (2.0 μM), mixture of rATP, rUTP, rCTP (1.0 mM) and one of the modified **rG$^{E}$TP**, **rG$^{Pent}$TP** or **rG$^{Ph}$TP** (0.4 mM). Reactions were carried out following the standard protocol.

### 5.2.17 Analytical scale PEX reaction with combination of four different base-modified rN$^{X}$TPs and variously long templates

**PEX with rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Ph}$TP, rG$^{Pent}$TP and templ_19nt_mix using TGK polymerase**

Reaction was performed in total volume of 10 μL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_mix** (4.8 μM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 μM), TGK polymerase (1.0 μM), mixture of **rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP**, **rG$^{Pent}$TP** (0.1 mM). Positive control was performed under same conditions with a mixture of natural rNTPs (0.1 mM) instead of modified **rN$^{X}$TP**s. Five different negative control reactions were performed under same conditions using in each reaction a different mixture of three natural rNTPs [0.1 mM; r(AUC), r(AUG), r(ACG), r(UCG)] and $H_2O$ or with complete replacement of all rNTPs by $H_2O$. Reactions were carried out following the standard protocol.

**PEX with rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Ph}$TP, rG$^{Pent}$TP and 5'-(TINA)-templ_31nt using TGK polymerase**

Reaction was performed in total volume of 10 μL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 μM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 μM), TGK polymerase (2.0 μM), mixture of **rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP**, **rG$^{Pent}$TP** (0.2 mM). Positive control was performed under same conditions with a mixture of natural rNTPs (0.2 mM) instead of modified **rN$^{X}$TP**s. Five different negative control reactions were performed under same conditions using in each reaction a different mixture of three natural rNTPs [0.2 mM; r(AUC), r(AUG), r(ACG), r(UCG)] and $H_2O$ or with complete replacement of all rNTPs by $H_2O$. Reactions were carried out following the standard protocol.

**PEX with rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Ph}$TP, rG$^{Pent}$TP and templ_65nt using TGK polymerase**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_65nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (2.5 µM), mixture of **rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP**, **rG$^{Pent}$TP** (0.4 mM). Positive control was performed under same conditions with a mixture of natural rNTPs (0.4 mM) instead of modified **rN$^{X}$TP**s. Five different negative control reactions were performed under same conditions using in each reaction a different mixture of three natural rNTPs [0.4 mM; r(AUC), r(AUG), r(ACG), r(UCG)] and $H_2O$ or with complete replacement of all rNTPs by $H_2O$. Reactions were carried out following the standard protocol.

**PEX with rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Ph}$TP, rG$^{Pent}$TP and templ_98nt using TGK polymerase**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_98nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (5.0 µM), mixture of **rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP**, **rG$^{Pent}$TP** (0.4 mM). Positive control was performed under same conditions with a mixture of natural rNTPs (0.4 mM) instead of modified **rN$^{X}$TP**s. Five different negative control reactions were performed under same conditions using in each reaction a different mixture of three natural rNTPs [0.4 mM; r(AUC), r(AUG), r(ACG), r(UCG)] and $H_2O$ or with complete replacement of all rNTPs by $H_2O$. Reactions were carried out following the standard protocol.

**PEX with rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Ph}$TP, rG$^{Pent}$TP and templ_19nt_mix using SFM4-3 polymerase**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_mix** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), SFM4-3 polymerase (1.0 µM), mixture of **rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP**, **rG$^{Pent}$TP** (0.1 mM). Positive control was performed under same conditions with a mixture of natural rNTPs (0.1 mM) instead of modified **rN$^{X}$TP**s. Five different negative control reactions were performed under same conditions using in each reaction a different mixture of three natural rNTPs [0.1 mM; r(AUC), r(AUG), r(ACG), r(UCG)] and $H_2O$ or with complete replacement of all rNTPs by $H_2O$. Reactions were carried out following the standard protocol.

**PEX with rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Ph}$TP, rG$^{Pent}$TP and 5'-(TINA)-templ_31nt using SFM4-3 polymerase**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-**

**RNA-prim_15nt** (4.0 μM), SFM4-3 polymerase (2.0 μM), mixture of **rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP**, **rG$^{Pent}$TP** (0.2 mM). Positive control was performed under same conditions with a mixture of natural rNTPs (0.2 mM) instead of modified **rN$^{X}$TP**s. Five different negative control reactions were performed under same conditions using in each reaction a different mixture of three natural rNTPs [0.2 mM; r(AUC), r(AUG), r(ACG), r(UCG)] and $H_2O$ or with complete replacement of all rNTPs by $H_2O$. Reactions were carried out following the standard protocol.

**PEX with rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Ph}$TP, rG$^{Pent}$TP and templ_65nt using SFM4-3 polymerase**

Reaction was performed in total volume of 10 μL in ThermoPol buffer (1X) containing ssDNA template – **templ_65nt** (4.8 μM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 μM), SFM4-3 polymerase (2.5 μM), mixture of **rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP**, **rG$^{Pent}$TP** (0.4 mM). Positive control was performed under same conditions with a mixture of natural rNTPs (0.4 mM) instead of modified **rN$^{X}$TP**s. Five different negative control reactions were performed under same conditions using in each reaction a different mixture of three natural rNTPs [0.4 mM; r(AUC), r(AUG), r(ACG), r(UCG)] and $H_2O$ or with complete replacement of all rNTPs by $H_2O$. Reactions were carried out following the standard protocol.

**PEX with rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Ph}$TP, rG$^{Pent}$TP and templ_98nt using SFM4-3 polymerase**

Reaction was performed in total volume of 10 μL in ThermoPol buffer (1X) containing ssDNA template – **templ_98nt** (4.8 μM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 μM), SFM4-3 polymerase (5.0 μM), mixture of **rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP**, **rG$^{Pent}$TP** (0.4 mM). Positive control was performed under same conditions with a mixture of natural rNTPs (0.4 mM) instead of modified **rN$^{X}$TP**s. Five different negative control reactions were performed under same conditions using in each reaction a different mixture of three natural rNTPs [0.4 mM; r(AUC), r(AUG), r(ACG), r(UCG)] and $H_2O$ or with complete replacement of all rNTPs by $H_2O$. Reactions were carried out following the standard protocol.

### 5.2.18 Semi-preparative scale PEX reaction with combination of four different base-modified rN$^{X}$TPs and variously long templates

**PEX with rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Ph}$TP, rG$^{Pent}$TP and templ_19nt_mix using TGK polymerase**

Reaction was performed in total volume of 50 μL in ThermoPol buffer (1X) containing ssDNA template – **templ_19nt_mix** (4.8 μM), RNA primer – **RNA-prim_15nt** (4.0 μM), TGK polymerase (1.0 μM), mixture of **rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP**, **rG$^{Pent}$TP** (0.1 mM). Reaction was carried out following the standard protocol.

**PEX with rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Ph}$TP, rG$^{Pent}$TP and 5'-(TINA)-templ_31nt using TGK polymerase**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing labelled ssDNA template – **5'-(TINA)-templ_31nt** (4.8 µM), RNA primer – **RNA-prim_15nt** (4.0 µM), TGK polymerase (2.0 µM), mixture of **rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP**, **rG$^{Pent}$TP** (0.2 mM). Reaction was carried out following the standard protocol.

**PEX with rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Ph}$TP, rG$^{Pent}$TP and templ_65nt using TGK polymerase**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_65nt** (4.8 µM), RNA primer – **RNA-prim_15nt** (4.0 µM), TGK polymerase (2.5 µM), mixture of **rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP**, **rG$^{Pent}$TP** (0.4 mM). Reaction was carried out following the standard protocol.

**PEX with rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Ph}$TP, rG$^{Pent}$TP and templ_98nt using TGK polymerase**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_98nt** (4.8 µM), RNA primer – **RNA-prim_15nt** (4.0 µM), TGK polymerase (5.0 µM), mixture of **rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP**, **rG$^{Pent}$TP** (0.4 mM). Reaction was carried out following the standard protocol.

**PEX with rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Ph}$TP, rG$^{Pent}$TP and templ_98nt for RT analysis**

Reactions were performed in total volume of 50 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_98nt** (4.8 µM), 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (4.0 µM), TGK polymerase (5.0 µM) and a mixture of either natural rNTPs (0.4 mM) or modified **rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP**, **rG$^{Pent}$TP** (0.4 mM). Reactions were carried out following the standard protocol.

### 5.2.19 RT analysis of natural or base-modified RNA

Reaction was performed in total volume of 10 µL in SSIV reaction buffer (1X). Briefly, 5'-(FAM)-labelled base-modified RNA template (0.4 µM; **98RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$**) and 5'-(Cy5)-labelled DNA primer – **Cy5-DNA-prim_20nt** (0.2 µM) in presence of SSIV reaction buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C s$^{-1}$). After addition of SSIV RT (200 U), DTT (5 mM) and dNTPs (0.5 mM), the mixture was incubated at 55 °C for 5 h in a thermal cycler with heated lid (95 °C). Positive control was performed under same conditions with 5'-(FAM)-labelled natural RNA template (**98RNA_AUCG**) and negative control was performed in absence of RNA template. After reaction, the samples were purified using QIAquick nucleotide removal kit

according to standard supplier's protocol, eluted in 10 µL of $H_2O$ and prepared for analysis according to protocol.

### 5.2.20 Comparison of TGK polymerase, SFM4-3 polymerase and T7 RNAP in enzymatic synthesis of RNA

**Analysis of PEX reaction with RNA primer and various amounts of TGK or SFM4-3 polymerase**

Reactions were performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_16nt** (0.24 µM), 5'-(Cy5)-labelled RNA primer – **Cy5-RNA-prim_15nt** (0.2 µM), the modified **rC$^{mBdp}$TP** (0.1 mM) and either TGK or SFM4-3 polymerase at different concentrations (0.01 µM, 0.03 µM, 0.05 µM, 0.07 µM, 0.1 µM, 0.2 µM). Reaction mixtures were heated up to 95 °C for 30 sec, followed by incubation at 60 °C for 1 h in a thermal cycler with heated lid (100 °C).

**Kinetic studies of PEX reaction with RNA primer and TGK or SFM4-3 polymerase**

Reactions were performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_16nt** (0.24 µM), 5'-(Cy5)-labelled RNA primer – **Cy5-RNA-prim_15nt** (0.2 µM), the modified **rC$^{mBdp}$TP** (0.1 mM) and either TGK (0.1 µM) or SFM4-3 (0.1 µM) polymerase. Reaction mixtures were heated up to 95 °C for 30 sec, followed by incubation at 60 °C for different time periods (1 min, 5 min, 10 min, 15 min, 30 min, 1 h) in a thermal cycler with heated lid (100 °C).

**Kinetic studies of PEX reaction with DNA primer and TGK or SFM4-3 polymerase**

Reactions were performed in total volume of 10 µL in ThermoPol buffer (1X) containing ssDNA template – **templ_16nt** (0.24 µM), 5'-(Cy5)-labelled DNA primer – **Cy5-DNA-prim_15nt** (0.2 µM), the modified **rC$^{mBdp}$TP** (0.1 mM) and either TGK (0.1 µM) or SFM4-3 (0.1 µM) polymerase. Reaction mixtures were heated up to 95 °C for 30 sec, followed by incubation at 60 °C for different time periods (1 min, 5 min, 10 min, 15 min, 30 min, 1 h) in a thermal cycler with heated lid (100 °C).

**PEX with templ_poly-U and TGK or SFM4-3 polymerase**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X). ssDNA template – **templ_poly-U** (0.24 µM) and 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_15nt** (0.2 µM) in presence of ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C $s^{-1}$). After addition of a mixture of rATP, rCTP, rGTP (0.1 mM), the modified **rU$^{Bio}$TP** (0.1 mM) and either TGK (0.1 µM) or SFM4-3 (0.1 µM) polymerase, the reaction mixtures were incubated at 60 °C for 1.5 h in a thermal cycler with heated lid (100 °C). Positive control was performed under

same conditions with natural rUTP (0.1 mM) instead of the modified **rU$^{Bio}$TP**. For negative control reaction $H_2O$ was used instead of rUTP.

**PEX reaction with templ_50nt and TGK or SFM4-3 polymerase**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X). ssDNA template – **templ_50nt** (0.24 µM) and 5'-(Cy5)-labelled RNA primer – **Cy5-RNA-prim_15nt** (0.2 µM) in presence of ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C $s^{-1}$). After addition of rGTP (0.1 mM), a mixture of the modified **rA$^{Ph}$TP**, **rU$^{Ph}$TP**, **rC$^{Ph}$TP** (0.1 mM, Mix-1); **rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP** (0.1 mM, Mix-2); **rA$^{Ph}$TP**, **rU$^{Bio}$TP**, **rC$^{mBdp}$TP** (0.1 mM, Mix-3) and either TGK (0.1 µM) or SFM4-3 (0.1 µM) polymerase, the reaction mixtures were incubated at 60 °C for 2 h in a thermal cycler with heated lid (100 °C). Positive control was performed under same conditions with a mixture of natural rNTPs (0.1 mM). Six different negative control reactions were performed under same conditions using in each reaction either a different mixture of natural rNTPs [0.1 mM; r(AUC); r(AUG); r(ACG); r(UCG)] and $H_2O$ or a mixture of rGTP (0.1 mM) and $H_2O$ or the natural rNTPs were completely replaced by $H_2O$. Reaction mixtures were combined with TurboDNase (2 U) and incubated at 37 °C for 15 min, followed by incubation with Proteinase K (0.8 U) at 37 °C for 15 min.

**IVT reaction with ds-templ_poly-U and T7 RNAP**

Reaction was performed in total volume of 10 µL in transcription reaction buffer (1X) containing dsDNA template – **ds-templ_poly-U** (0.5 µM), mixture of rATP, rCTP, rGTP (0.1 mM), the modified **rU$^{Bio}$TP** (0.1 mM), DTT (10 mM), $MgCl_2$ (15 mM), RiboLock RNase inhibitor (10 U), Triton X-100 (0.1%), T7 RNAP (15 U) and [α-$^{32}$P]-GTP (111 TBq/mmol, 370 MBq/mL, 0.1 µL). Positive control was performed under same conditions with natural rUTP (0.1 mM) instead of the modified **rU$^{Bio}$TP**. For negative control reaction $H_2O$ was used instead of rUTP. Transcription reactions were performed at 37 °C for 2 h in a thermal cycler with heated lid (65 °C). Crude reaction mixtures were combined with 10 µL of 2X stop solution, denatured by heating at 65 °C for 10 min and then immediately cooled on ice. Aliquots of the samples (10 µL) were subjected to vertical gel electrophoresis.

**IVT reaction with ds-templ_52bp and T7 RNAP**

Reaction was performed in total volume of 10 µL in transcription reaction buffer (1X) containing dsDNA template – **ds-templ_52bp** (0.5 µM, rGTP (0.1 mM), a mixture of the modified **rA$^{Ph}$TP**, **rU$^{Ph}$TP**, **rC$^{Ph}$TP** (0.1 mM, Mix-1); **rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP** (0.1 mM, Mix-2); **rA$^{Ph}$TP**, **rU$^{Bio}$TP**, **rC$^{mBdp}$TP** (0.1 mM, Mix-3), DTT (10 mM), $MgCl_2$ (15 mM), RiboLock RNase inhibitor (10 U),

Triton X-100 (0.1%), T7 RNAP (15 U) and [α-$^{32}$P]-GTP (111 TBq/mmol, 370 MBq/mL, 0.1 µL). Positive control was performed under same conditions with a mixture of natural rNTPs (0.1 mM). Four different negative control reactions were performed under same conditions using in each reaction either a different mixture of natural rNTPs [0.1 mM; r(AUG); r(ACG); r(UCG)] and $H_2O$ or a mixture of rGTP (0.1 mM) and $H_2O$. Transcription reactions were performed at 37 °C for 2 h in a thermal cycler with heated lid (65 °C). Crude reaction mixtures were combined with 10 µL of 2X stop solution, denatured by heating at 65 °C for 10 min and then immediately cooled on ice. Aliquots of the samples (10 µL) were subjected to vertical gel electrophoresis.

### 5.2.21 Synthesis of base-modified RNA with cleavable DNA primer

**Analytical PEX reaction with rC$^{Cy5}$TP and fluorescently labelled DNA primer**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X). ssDNA template – **templ_36nt** (4.8 µM) and 5'-(6-FAM)-labelled DNA primer – **FAM-DNA-prim_15nt** (4.0 µM) in ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C s$^{-1}$). After addition of TGK polymerase (2.0 µM), mixture of rATP, rUTP, rGTP (0.1 mM) and the modified **rC$^{Cy5}$TP** (0.1 mM), the reaction was incubated at 60 °C for 2 h in a thermal cycler with heated lid (100 °C). Positive control was performed under same conditions with natural rCTP (0.1 mM) instead of modified **rC$^{Cy5}$TP**. For negative control reaction $H_2O$ was used instead of rCTP. After PEX, the crude reaction mixture was purified using QIAquick nucleotide removal kit, followed by elution with 10 µL of $H_2O$. For optimalisation of DNA primer cleavage, the PEX reaction with **rC$^{Cy5}$TP** was 3X scaled-up and the generated modified **36DNA-RNA_C$^{Cy5}$** was purified using QIAquick nucleotide removal kit, followed by elution with 10 µL of $H_2O$. The purified mixture (10 µL) was combined with various amounts of TurboDNase (2 U, 6 U, 10 U) and incubated at 37 °C for 2 h to generate the desired cleavage product (**RNA_C$^{Cy5}$**). After this time, aliquots for gel analysis were removed and the reaction was further incubated at 37 °C for another 2 h (4 h in total).

**Analytical PEX reaction with rC$^{Cy5}$TP and fluorescently labelled DNA primer with internal dU modification**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X). ssDNA template – **templ_36nt** (4.8 µM) and 5'-(6-FAM)-labelled DNA primer with internal dU modification – **FAM-DNA-prim_dU** (4.0 µM) in ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C s$^{-1}$). After addition of TGK polymerase (1.5 µM), mixture of rATP, rUTP, rGTP (0.1 mM) and the modified **rC$^{Cy5}$TP** (0.1 mM), the reaction was incubated at 60 °C for 2 h in a thermal cycler with heated lid (100 °C). Positive control was performed under same conditions with natural

rCTP (0.1 mM) instead of the modified **rC$^{Cy5}$TP**. For negative control reaction $H_2O$ was used instead of rCTP. Crude reaction mixture was purified using QIAquick nucleotide removal kit, followed by elution with 10 µL of $H_2O$. For cleavage of the DNA primer, the purified mixture (**36DNA_dU-RNA_C$^{Cy5}$**) was combined with UDG buffer (1 µL, 10X), UDG (2 U) and further incubated at 37 °C for 30 min. Solution of DMEDA (1.2 µL, 1 M, pH 8.5) was added and again incubated at 37 °C for further 30 min to generate the desired cleavage product (**RNA_C$^{Cy5}$**). After that, TurboDNase (2 U) was added and incubated at 37 °C for 1 h.

**Analytical PEX reaction with rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Cy5}$TP, rG$^{Pent}$TP and fluorescently labelled DNA primer with internal dU modification**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X). ssDNA template – **templ_36nt** (4.8 µM) and 5'-(6-FAM)-labelled DNA primer with internal dU modification – **FAM-DNA-prim_dU** (4.0 µM) in ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C s$^{-1}$). After addition of TGK polymerase (1.5 µM) and a mixture of **rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Cy5}$TP**, **rG$^{Pent}$TP** (0.1 mM), the reaction was incubated at 60 °C for 2 h in a thermal cycler with heated lid (100 °C). Positive control was performed under same conditions with a mixture of natural rNTPs (0.1 mM) instead of the modified **rN$^{X}$TP**s. Five different negative control reactions were performed under same conditions using in each reaction a different mixture of three natural rNTPs [0.1 mM; r(AUC); r(AUG); r(ACG); r(UCG)] and $H_2O$ or by complete replacement of natural rNTPs by $H_2O$. Crude reaction mixture was purified using QIAquick nucleotide removal kit, followed by elution with 10 µL of $H_2O$. For cleavage of the DNA primer, the purified mixture (**36DNA_dU-RNA_A$^{E}$U$^{Bio}$C$^{Cy5}$G$^{Pent}$**) was combined with UDG buffer (1 µL, 10X), UDG (2 U) and further incubated at 37 °C for 30 min to generate the desired cleavage product (**RNA_A$^{E}$U$^{Bio}$C$^{Cy5}$G$^{Pent}$**). Solution of DMEDA (1.2 µL, 1 M, pH 8.5) was added and again incubated at 37 °C for further 30 min. After that, TurboDNase (2 U) was added and incubated at 37 °C for 1 h.

**Analytical PEX reaction with rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Ph}$TP, rG$^{Pent}$TP and dual fluorescently labelled DNA primer with internal dU modification**

Reaction was performed in total volume of 10 µL in ThermoPol buffer (1X). ssDNA template – **templ_35nt** (4.8 µM) and dually labelled DNA primer with 5'-(Cy5) and internal dU and dT-(FAM) modification – **Cy5-DNA-prim_dU_dT-FAM** (4.0 µM) in ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C s$^{-1}$). After addition of TGK polymerase (2.0 µM) and a mixture of **rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP**, **rG$^{Pent}$TP** (0.2 mM), the reaction was incubated at 60 °C for 2 h in a thermal cycler with heated lid (100 °C). Positive control was performed under same

conditions with a mixture of natural rNTPs (0.2 mM) instead of the modified **rN$^{X}$TP**s. Five different negative control reactions were performed under same conditions using in each reaction a different mixture of three natural rNTPs [0.2 mM; r(AUC), r(AUG), r(ACG), r(UCG)] and $H_2O$ or by complete replacement of natural rNTPs by $H_2O$. For cleavage of the DNA primer, the crude mixture (**35DNA_dU_dT$^{FAM}$-RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$**) was combined with UDG buffer (1 μL, 10X), UDG (2 U) and further incubated at 37 °C for 30 min. Solution of DMEDA (1.2 μL, 1 M, pH 8.5) was added and again incubated at 37 °C for further 30 min to generate the desired cleavage product (**dT$^{FAM}$-RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$**). After that, TurboDNase (2 U) was added and incubated at 37 °C for 1 h.

**Semi-preparative PEX reaction with rC$^{Cy5}$TP and fluorescently labelled DNA primer with internal dU modification**

Reaction was performed in total volume of 50 μL in ThermoPol buffer (1X). ssDNA template – **templ_36nt** (4.8 μM) and 5'-(6-FAM)-labelled DNA primer with internal dU modification – **FAM-DNA-prim_dU** (4.0 μM) in ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C s$^{-1}$). After addition of TGK polymerase (1.5 μM), mixture of rATP, rUTP, rGTP (0.1 mM) and the modified **rC$^{Cy5}$TP** (0.1 mM), the reaction was incubated at 60 °C for 2 h in a thermal cycler with heated lid (100 °C). Crude reaction mixture was purified using QIAquick nucleotide removal kit, followed by elution with 50 μL of $H_2O$. For cleavage of the DNA primer, the purified mixture (**36DNA_dU-RNA_C$^{Cy5}$**) was combined with UDG buffer (5 μL, 10X), UDG (10 U) and further incubated at 37 °C for 30 min. Solution of DMEDA (6 μL, 1 M, pH 8.5) was added and again incubated at 37 °C for further 30 min. After that, TurboDNase (10 U) was added and incubated at 37 °C for 1 h. Sample was purified using Microspin G-25 columns, evaporated to dryness, again diluted in 10 μL of water prior to analysis.

**Semi-preparative PEX reaction with rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Cy5}$TP, rG$^{Pent}$TP and fluorescently labelled DNA primer with internal dU modification**

Reaction was performed in total volume of 50 μL in ThermoPol buffer (1X). ssDNA template – **templ_36nt** (4.8 μM) and 5'-(6-FAM)-labelled DNA primer with internal dU modification – **FAM-DNA-prim_dU** (4.0 μM) in ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C s$^{-1}$). After addition of TGK polymerase (1.5 μM) and a mixture of **rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Cy5}$TP**, **rG$^{Pent}$TP** (0.1 mM), the reaction was incubated at 60 °C for 2 h in a thermal cycler with heated lid (100 °C). Crude reaction mixture was purified using QIAquick nucleotide removal kit, followed by elution with 50 μL of $H_2O$. For cleavage of the DNA primer, the purified mixture (**36DNA_dU-RNA_A$^{E}$U$^{Bio}$C$^{Cy5}$G$^{Pent}$**) was combined with UDG buffer (5 μL, 10X), UDG (10 U) and

further incubated at 37 °C for 30 min. Solution of DMEDA (6 µL, 1 M, pH 8.5) was added and again incubated at 37 °C for further 30 min. After that, TurboDNase (10 U) was added and incubated at 37 °C for 1 h. Sample was purified using Microspin G-25 columns, evaporated to dryness, again diluted in 10 µL of water prior to analysis.

**Semi-preparative PEX reaction with rA$^{E}$TP, rU$^{Bio}$TP, rC$^{Ph}$TP, rG$^{Pent}$TP and dual fluorescently labelled DNA primer with internal dU modification**

Reaction was performed in total volume of 50 µL in ThermoPol buffer (1X). ssDNA template – **templ_35nt** (4.8 µM) and dually labelled DNA primer with 5'-(Cy5) and internal dU and dT-(FAM) modification – **Cy5-DNA-prim_dU_dT-FAM** (4.0 µM) in ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C s$^{-1}$). After addition of TGK polymerase (2.0 µM) and a mixture of **rA$^{E}$TP**, **rU$^{Bio}$TP**, **rC$^{Ph}$TP**, **rG$^{Pent}$TP** (0.2 mM), the reaction was incubated at 60 °C for 2 h in a thermal cycler with heated lid (100 °C). The crude mixture containing the desired PEX product (**35DNA_dU_dT$^{FAM}$-RNA_A$^{E}$U$^{Bio}$C$^{Ph}$G$^{Pent}$**) was combined with UDG buffer (5 µL, 10X), UDG (10 U) and further incubated at 37 °C for 30 min. Solution of DMEDA (6 µL, 1 M, pH 8.5) was added and again incubated at 37 °C for further 30 min. After that, TurboDNase (10 U) was added and incubated at 37 °C for 1 h. Sample was purified using Microspin G-25 columns, evaporated to dryness, again diluted in 10 µL of water prior to analysis.

### 5.2.22 Selective fluorescent RNA labelling at specific position for structural studies

**Analytical scale PEX reaction for preparation of FAM-Cy5-Cy3-riboswitch**

Reaction mixture containing ssDNA template – **templ_ribosw71_A** (0.48 µL, 100 µM) and 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_23nt** (0.4 µL, 100 µM) in presence of ThermoPol buffer (1 µL, 10X) was heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C s$^{-1}$). SNI-1: After additon of **rU$^{Cy5}$TP** (1 µL, 1 mM) and TGK polymerase (0.4 µL, 50 µM), the final reaction mixture (7 µL) was incubated at 45 °C for 15 min. Mixture was combined with rSAP (1 U) and incubated at 37 °C for 15 min, followed by heat denaturation at 65 °C for 10 min. PEX-1: Reaction was combined with a mixture of natural rNTPs (2.0 µL, 4 mM), additional amount of TGK polymerase (1.0 µL, 10 µM) and incubated at 45 °C for 15 min, followed by incubation at 60 °C for 15 min. Then the mixture was combined with rSAP (1 U), TurboDNase (2 U) and incubated at 37 °C for 30 min, followed by heat denaturation at 75 °C for 15 min. After addition of second ssDNA template – **templ_ribosw71_B** (0.5 µL, 100 µM), the mixture was heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C s$^{-1}$). SNI-2: After addition of **rU$^{Cy3}$TP (**1 µL, 1 mM) and TGK polymerase (1.5 µL, 10 µM) the mixture was incubated at 60 °C for 15 min. PEX-2: Followed by

addition of a mixture of natural rNTPs (2.0 µL, 16 mM), TGK polymerase (2.0 µL, 10 µM) and further incubation at 60 °C for 2 h. Finally, sample (**FAM-Cy5-Cy3-riboswitch**) was treated by rSAP (1 U) and TurboDNase (2 U) at 37 °C for 30 min, followed by addition of Proteinase K (1.6 U) and further incubation at 37 °C for 30 min. Crude reaction mixtures containing either the reaction intermediates (after SNI-1, PEX-1, SNI-2) or the final product (after PEX-2, **FAM-Cy5-Cy3-riboswitch**) were after TurboDNase and Proteinase K treatment purified using QIAquick nucleotide removal kit according to standard supplier's protocol, prior to dPAGE analysis.

**Semi-preparative scale PEX reaction for preparation of FAM-Cy5-Cy3-riboswitch**

Reaction mixture containing ssDNA template – **templ_ribosw71_A** (0.96 µL, 100 µM) and 5'-(6-FAM)-labelled RNA primer – **FAM-RNA-prim_23nt** (0.8 µL, 100 µM) in presence of ThermoPol buffer (2 µL, 10X) was heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C $s^{-1}$) in a thermal cycler with heated lid (100 °C). SNI-1: After additon of **rU$^{Cy5}$TP** (2 µL, 1 mM) and TGK polymerase (0.8 µL, 50 µM), the final reaction mixture (14 µL) was incubated at 45 °C for 15 min. Mixture was combined with rSAP (1 U) and incubated at 37 °C for 30 min, followed by heat denaturation at 65 °C for 10 min. PEX-1: Reaction was combined with a mixture of natural rNTPs (4.0 µL, 4 mM), additional amount of TGK polymerase (2.0 µL, 10 µM) and incubated at 45 °C for 15 min, followed by incubation at 60 °C for 15 min. Then the mixture was combined with rSAP (2 U), TurboDNase (4 U) and incubated at 37 °C for 30 min, followed by heat denaturation at 75 °C for 15 min. After addition of second ssDNA template – **templ_ribosw71_B** (1.0 µL, 100 µM), the mixture was heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C $s^{-1}$). SNI-2: After addition of **rU$^{Cy3}$TP (**2 µL, 1 mM) and TGK polymerase (3 µL, 10 µM) the mixture was incubated at 60 °C for 15 min. PEX-2: Followed by addition of a mixture of natural rNTPs (4.0 µL, 16 mM), TGK polymerase (4.0 µL, 10 µM) and further incubation at 60 °C for 2 h. The final product (**FAM-Cy5-Cy3-riboswitch**) was treated with TurboDNase (6 U) at 37 °C for 30 min, followed by addition of Proteinase K (2.4 U) and further incubation at 37 °C for 30 min. Crude reaction mixtures containing either reaction intermediates (after SNI-1, PEX-1, SNI-2) or the final product (after PEX-2, **FAM-Cy5-Cy3-riboswitch**) were after TurboDNase and Proteinase K treatment purified using QIAquick nucleotide removal kit according to standard supplier's protocol. Samples were two times eluted with 15 µL of $H_2O$, lyophilised to dryness and again diluted in 10 µL of $H_2O$ prior to analysis.

**Preparative scale PEX reaction for preparation of Cy5-Cy3-riboswitch**

Reaction was performed following the procedure in section 0, just 10X scaled up and with usage of non-fluorescently labelled RNA primer – **RNA-prim_23nt**. Crude reaction mixture was purified

using QIAquick nucleotide removal kit according to standard supplier's protocol and evaporated to dryness. Aliquot of the sample was again diluted in 10 µL of $H_2O$ prior to mass analysis and the remaining sample was used for FRET measurements.

**FRET measurements of the Cy5-Cy3-riboswitch**

The sample (**Cy5-Cy3-riboswitch**) was diluted with TRIS buffer (50 mM, pH 7.5), NaCl (50 mM) and $MgSO_4$ (4 mM) to final 1 µM concentration. Reaction mixture (100 µL) was heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C $s^{-1}$). Afterwards, the reaction mix was combined with urea (4 M) and increasing concentration of adenine ligand (5 µM, 10 µM, 100 µM, 1 mM, 1.5 mM, 2 mM, 2.5 mM, 3 mM, 4 mM, 8 mM). The mixture was transferred to a 100 µL quartz cuvette and fluorescent spectra were measured at 25 °C. Fluorescence spectrum was recorded with the excitation wavelength λex = 530 nm and the range of the emission spectra was 540-800 nm. The negative control FRET measurements were performed under same conditions with usage of a mixture of Cy5-labelled (**Cy5-DNA-oligo**) and Cy3-labelled (**Cy3-DNA-oligo**) ssDNA oligonucleotides. All experiments were performed in triplicates.

### 5.2.23 Enzymatic synthesis of mRNA

**Preparation of ssDNA template – templ_IRES**

Reaction mixture (50 µL) containing Q5 reaction buffer (1X), mixture of dNTPs (200 µM), plasmid (5 ng), labelled forward DNA primer – **5'-(dual-Bio)-DNA-FOR-prim_29nt** (0.5 µM), reverse DNA primer – **DNA-REV-prim_26nt** (0.5 µM) and Q5 HF DNA polymerase (1 U) was incubated following the standard protocol. The reaction mixture was purified according to standard supplier's protocol using QIAquick PCR purification kit, followed by magnetoseparation. For generation of larger quantities of dsDNA template, the PCR reaction was 20X scaled-up, purified using the NucleoSpin gel and PCR clean-up midi kit, followed by magnetoseparation for preparation of ssDNA templates. Aliquots of either dsDNA or ssDNA samples were combined with TriTrack DNA loading dye (6X) and analysed on 1% agarose gel containing TAE (1X) and SybrGold (1X) in TAE running buffer (1X) at 120 V.

**Preparation of ssDNA template – templ_IRES-prolonged**

Reaction mixture (50 µL) containing Q5 reaction buffer (1X), mixture of dNTPs (200 µM), plasmid (5 ng), labelled forward DNA primer – **5'-(dual-Bio)-DNA-FOR-prim_29nt** (0.5 µM), reverse DNA primer – **DNA-REV-prim_24nt** (0.5 µM) and Q5 HF DNA polymerase (1 U) was incubated following the standard protocol. The reaction mixture was purified according to standard supplier's

protocol using QIAquick PCR purification kit, followed by magnetoseparation. For generation of larger quantities of dsDNA template, the PCR reaction was 20X scaled-up, purified using the NucleoSpin gel and PCR clean-up midi kit, followed by magnetoseparation for preparation of ssDNA templates. Aliquots of either dsDNA or ssDNA samples were combined with TriTrack DNA loading dye (6X) and analysed on 1% agarose gel containing TAE (1X) and SybrGold (1X) in TAE running buffer (1X) at 120 V.

**Preparation of ssDNA template – templ_IRES-nLuc**

Reaction mixture (50 μL) containing Q5 reaction buffer (1X), mixture of dNTPs (200 μM), plasmid (5 ng), labelled forward DNA primer – **5'-(dual-Bio)-DNA-FOR-prim_29nt** (0.5 μM), reverse DNA primer – **DNA-REV-prim_48nt** (0.5 μM) and Q5 HF DNA polymerase (1 U) was incubated following the standard protocol. The reaction mixture was purified according to standard supplier's protocol using QIAquick PCR purification kit, followed by magnetoseparation. For generation of larger quantities of dsDNA template, the PCR reaction was 20X scaled-up, purified using the NucleoSpin gel and PCR clean-up midi kit, followed by magnetoseparation for preparation of ssDNA templates. Aliquots of either dsDNA or ssDNA samples were combined with TriTrack DNA loading dye (6X) and analysed on 1% agarose gel containing TAE (1X) and SybrGold (1X) in TAE running buffer (1X) at 120 V.

**Preparation of IRES-RNA**

Reaction was performed in total volume of 10 μL in ThermoPol buffer (1X). The ssDNA template – **templ_IRES** (0.24 μM) and 5'-(Cy5)-labelled RNA primer – **Cy5-mRNA-prim** (0.2 μM) in presence of ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C $s^{-1}$). After addition of a mixture of rNTPs (0.4 mM) and TGK polymerase (0.1 μM) the reaction was incubated at 65 °C for 4 h, followed by final work up using the standard protocol.

**Synthesis of IRES-RNA_prolonged**

Reaction was performed in total volume of 10 μL in ThermoPol buffer (1X). The ssDNA template – **templ_IRES-prolonged** (0.24 μM) and 5'-(Cy5)-labelled RNA primer – **Cy5-mRNA-prim** (0.2 μM) in presence of ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C $s^{-1}$). After addition of a mixture of rNTPs (0.4 mM) and TGK polymerase (1.0 μM) the reaction was incubated at 65 °C for 4 h, followed by final work up using the standard protocol.

**Synthesis of natural mRNA (mRNA-nat)**

Reaction was performed in total volume of 10 μL in ThermoPol buffer (1X). The ssDNA template – **templ_IRES-nLuc** (0.24 μM) and 5'-(Cy5)-labelled RNA primer – **Cy5-mRNA-prim** (0.2 μM) in presence of ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C $s^{-1}$). After addition of a mixture of rNTPs (0.4 mM) and TGK polymerase (0.75 μM) the reaction was incubated at 65 °C for 4 h, followed by final work up using the standard protocol.

**Synthesis of fully modified mRNA (mRNA-full)**

Reaction was performed in total volume of 10 μL in ThermoPol buffer (1X). The ssDNA template – **templ_IRES-nLuc** (0.24 μM) and 5'-(Cy5)-labelled RNA primer – **Cy5-mRNA-prim** (0.2 μM) in presence of ThermoPol buffer were heated up to 95 °C for 30 sec and then cooled down to 3 °C (0.1 °C $s^{-1}$). After addition of a mixture of rATP, rUTP, **rC$^{Me}$TP**, rGTP (0.4 mM) and TGK polymerase (0.75 μM) the reaction was incubated at 65 °C for 4 h, followed by final work up using the standard protocol.

**Synthesis of gene modified mRNA (mRNA-gene)**

Reaction mixture (10 μL) containing ThermoPol buffer (1X), mixture of rATP, rUTP, **rC$^{Me}$TP**, rGTP (0.2 mM), ssDNA template – **templ_IRES-nLuc** (0.24 μM), 5'-(Cy5)-labelled RNA primer – **Cy5-IRES-RNA** (0.2 μM) and TGK polymerase (0.75 μM) was heated up to 95 °C for 30 sec followed by incubation at 65 °C for 4 h. Final work up was performed according to standard protocol.

**Synthesis of single-site modified mRNA (mRNA-init)**

Reaction mixture (10 μL) containing ThermoPol buffer (1X), **rC$^{Me}$TP** (0.2 mM), ssDNA template – **templ_IRES-nLuc** (0.24 μM), 5'-(Cy5)-labelled RNA primer **– Cy5-IRES-RNA** (0.2 μM) and TGK polymerase (0.75 μM) was heated up to 95 °C for 30 sec followed by incubation at 65 °C for 20 min. After this, the mixture was combined with rSAP (1 U) and incubated at 37 °C for 30 min, followed by enzyme denaturation at 65 °C for 5 min. Then a mixture of rNTPs (4 mM, 1 μL) was added and the reaction was further incubated at 65 °C for 4 h, followed by final work up using the standard protocol.

**Synthesis of single-site modified mRNA (mRNA-mid)**

Reaction mixture (10 μL) containing ThermoPol buffer (1X), **rC$^{Me}$TP** (0.2 mM), ssDNA template – **templ_IRES-nLuc** (0.24 μM), 5'-(Cy5)-labelled RNA primer – **Cy5-IRES-RNA_prolonged** (0.2 μM) and TGK polymerase (0.75 μM) was heated up to 95 °C for 30 sec followed by incubation at 65

°C for 20 min. After this, the mixture was combined with rSAP (1 U) and incubated at 37 °C for 30 min, followed by enzyme denaturation at 65 °C for 5 min. Then a mixture of rNTPs (4 mM, 1 µL) was added and the reaction was further incubated at 65 °C for 4 h, followed by final work up using the standard protocol.

**Bisulfite conversion of mRNA with single modification**

Approximately 20 ng of single-site modified mRNA were treated with EZ RNA methylation kit according to supplier's protocol, followed by elution with 15 µL of water.

**One step RT-PCR reaction of bisulfite-treated mRNA**

Reaction mixture (12.5 µL) containing bisulfite-converted mRNA (1 µL), forward DNA primer – **DNA-FOR-prim_26nt** (0.5 µM), reverse DNA primer – **DNA-REV-prim_21nt** (0.5 µM), LunaScript Multiplex one-step RT-PCR reaction mix (1X) and LunaScript Multiplex one-step RT-PCR enzyme mix (1X) was incubated according to standard protocol. Crude reaction mix was combined with TriTrack DNA loading dye (1X) and analysed on 1% agarose gel containing TAE (1X) and SybrSafe (1X) in TAE running buffer (1X) at 120 V. For generation of larger quantities of the desired product, the reaction was 10X scaled-up. The desired PCR products were visualised by UV shadowing, excised, and purified according to standard supplier's protocol using QIAquick gel extraction kit, followed by elution with $H_2O$.

**Sample preparation for Sanger sequencing**

Mixture (10 µL) containing 45 ng of bisulfite-treated mRNA and DNA primer – **DNA-REV-prim_21nt** (5 pmol) was subjected to Sanger sequencing analysis.

### 5.2.24 mRNA translation studies

**Protocol for *in vitro* translation in rabbit reticulocyte lysate system**

Reaction mixture (10 µL) containing 50 ng of mRNA, Ribolock RNase inhibitor (40 U), complete amino acids mixture (50 µM) and nuclease-treated rabbit reticulocyte lysate was incubated at 30 °C for 1.5 h in a thermal cycler with heated lid (65 °C). Translation reactions were stopped by addition of cycloheximide solution (100 µM, in DMSO). 1 µL aliquot of the quenched reaction mixture was combined with 20 µL of reconstituted Nano-Glo luciferase assay reagent prepared by combining Nano-Glo luciferase assay substrate and Nano-Glo luciferase assay buffer (1:50) according to manufacturer's protocol. The mixture was incubated at room temperature for 3 min and 20 µL aliquots were transferred to white opaque 384-well microplate prior to luminescence measurements with 1000

ms integration time on Tecan Spark Multimode Reader. The data are reported as counts/sec. For negative control, mRNA was omitted.

**Protocol for *in cellulo* translation in HEK293T cell system**

$3\times10^4$ HEK293T cells were seeded in 96-well plate (black wells, clear bottom) 18 h prior to transfection in 40 µL Dulbecco's Modified Eagle Medium containing 10% fetal bovine serum. Lipofectamine MessengerMAX (0.125 µL) was diluted in 5 µL OptiMEM medium. At the same time mRNA (25 ng) was mixed with 5 µL OptiMEM and incubated for 10 min. Both mixes were combined and incubated for 5 min at ambient temperature. Afterwards, the combined transfection mix was added to cells. After 4 h at 37 °C, 5% $CO_2$ in humidified incubator, cells were equilibrated at 25 °C for 10 min. Cells were combined with 50 uL of reconstituted Nano-Glo luciferase assay reagent prepared by combining Nano-Glo luciferase assay substrate and Nano-Glo luciferase assay buffer (1:50) according to manufacturer's protocol. Then, the plate was mixed for 3 min at 1 mm circular diameter on Tecan Spark Multimode Reader, followed by luminescence measurement with 1000 ms integration time. The data are reported as counts/sec. For negative controls, mRNA and/or Lipofectamine was omitted.

# 6 Appendix

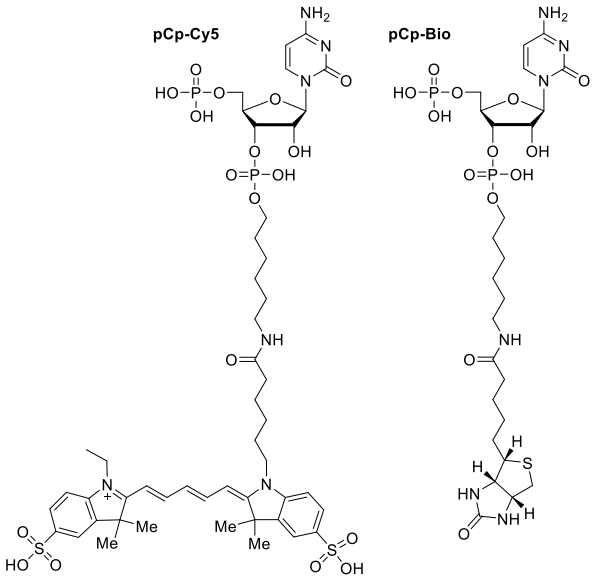



**Figure 103. Chemical structures of pCp-Cy5 and pCp-Bio used in ligation reactions.**

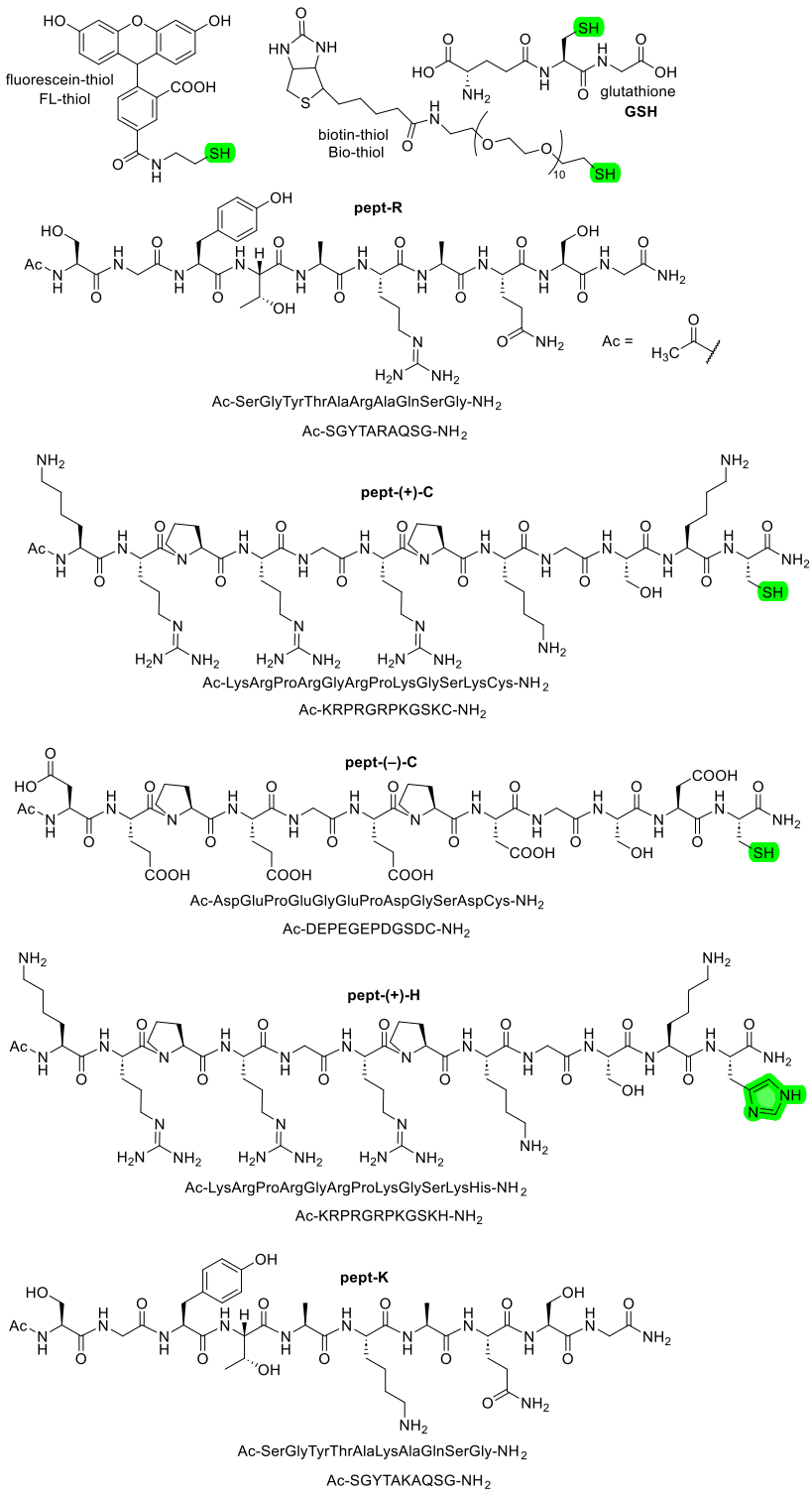



**Figure 104. Chemical structures of (bio)molecules with thiol functionality and various peptides used in bioconjugation reactions. Targeted functional group is highlighted in green.**

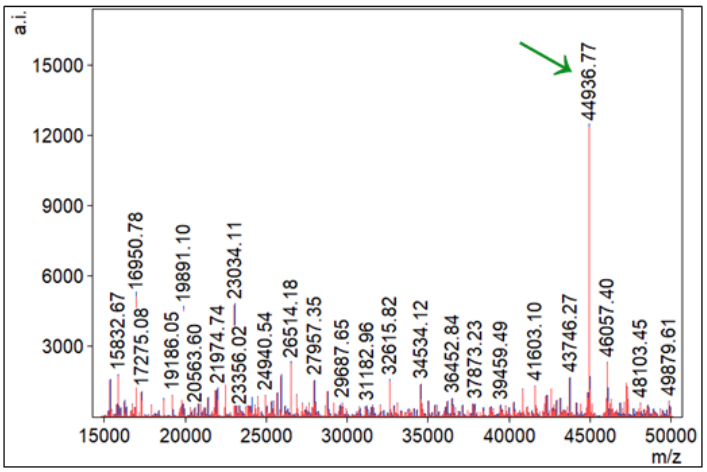



**Figure 105. Deconvoluted mass spectrum of 20RNA_1A$^{CA\text{-}HuR}$ conjugate. calculated mass: 44937.89 Da (with loss of N-terminal methionine); found mass: 44936.77 Da (with loss of N-terminal methionine); Δ = 1.12 Da.**

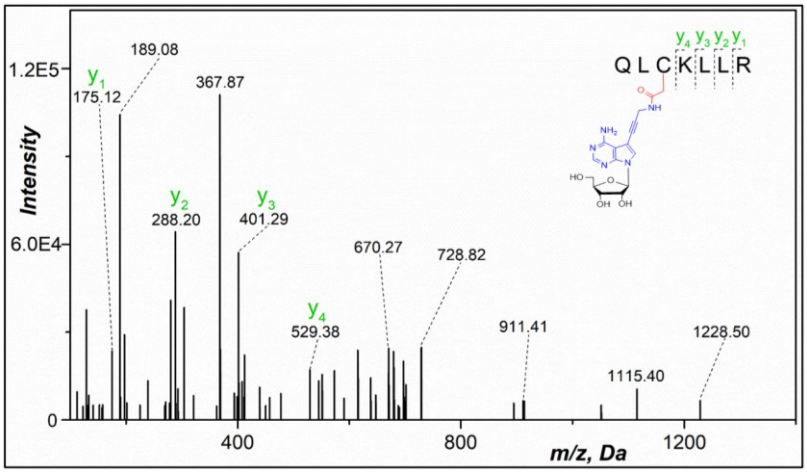



**Figure 106. Nano-ESI$^{+}$-MS/MS spectrum of peptide from 20RNA_1A$^{CA\text{-}HIV\text{-}RT}$ conjugate digest. Modification cross-linked to cysteine (position C280). *m/z* acquired: 411.8861 Da, z = 3+.**

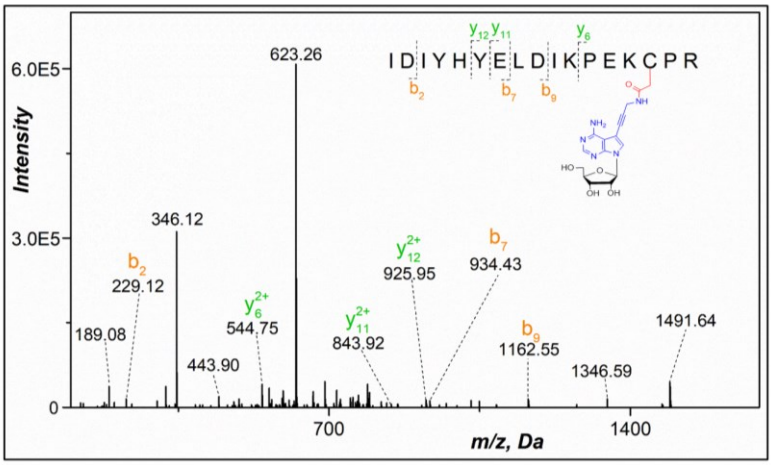



**Figure 107. Nano-ESI$^{+}$-MS/MS spectrum of peptide from 20RNA_1A$^{CA\text{-}hAgo2}$ conjugate digest. Modification cross-linked to cysteine (position C84). *m*/*z* acquired: 623.5595 Da, z = 4+.**

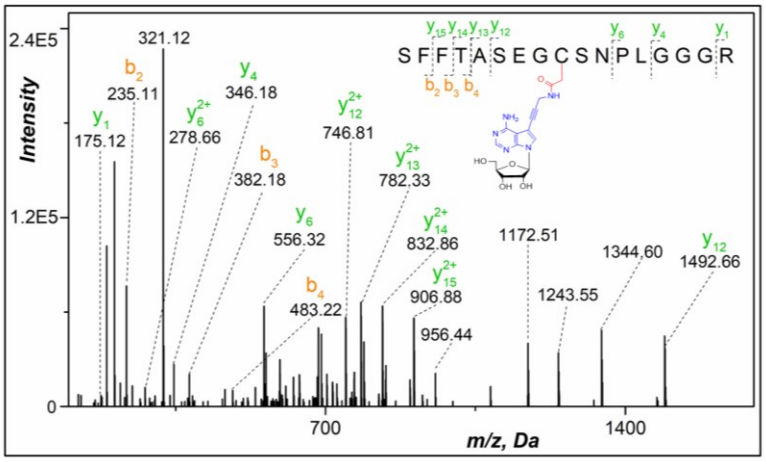



**Figure 108. Nano-ESI$^{+}$-MS/MS spectrum of peptide from 20RNA_1A$^{CA\text{-}hAgo2}$ conjugate digest. Modification cross-linked to cysteine (position C206). *m*/*z* acquired: 682.6306 Da, z = 3+.**

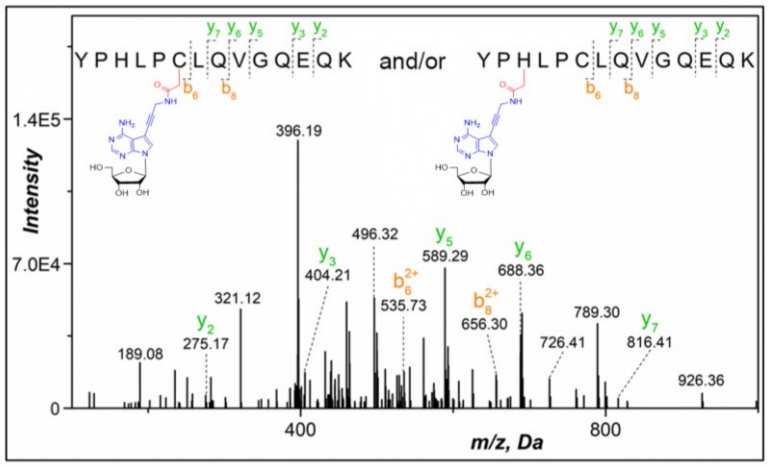



**Figure 109. Nano-ESI$^{+}$-MS/MS spectrum of peptide from 20RNA_1A$^{CA\text{-}hAgo2}$ conjugate digest. Modification cross-linked to cysteine (position C345) and/or to histidine (position H342). *m*/*z* acquired: 500.4941 Da, z = 4+.**

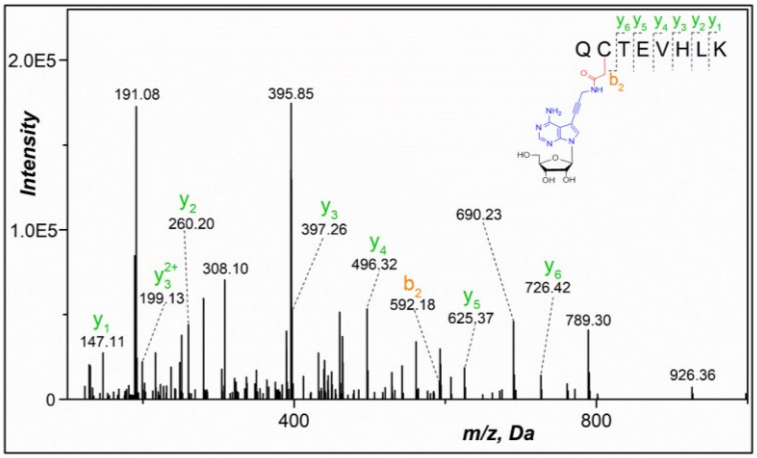



**Figure 110. Nano-ESI$^{+}$-MS/MS spectrum of peptide from 20RNA_1A$^{CA\text{-}hAgo2}$ conjugate digest. Modification cross-linked to cysteine (position C480). *m*/*z* acquired: 439.8680 Da, z = 3+.**

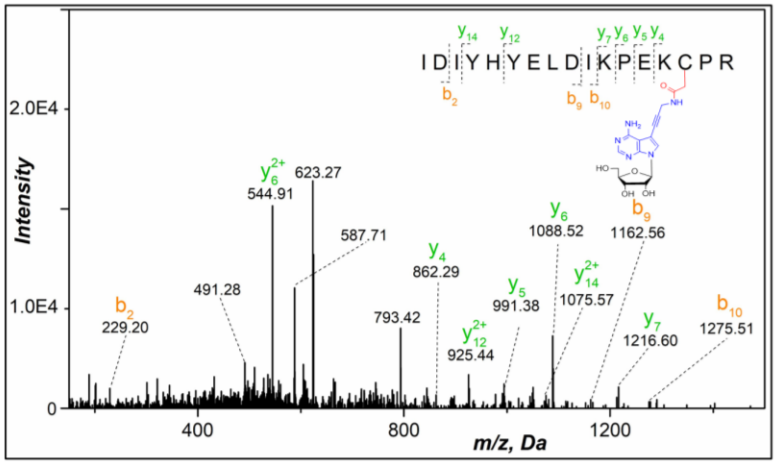



**Figure 111. Nano-ESI$^{+}$-MS/MS spectrum of peptide from 20RNA_1A$^{CA\text{-}hAgo2}$ conjugate digest. Modification cross-linked to cysteine (position C84). *m*/*z* acquired: 623.5577 Da, z = 4+.**

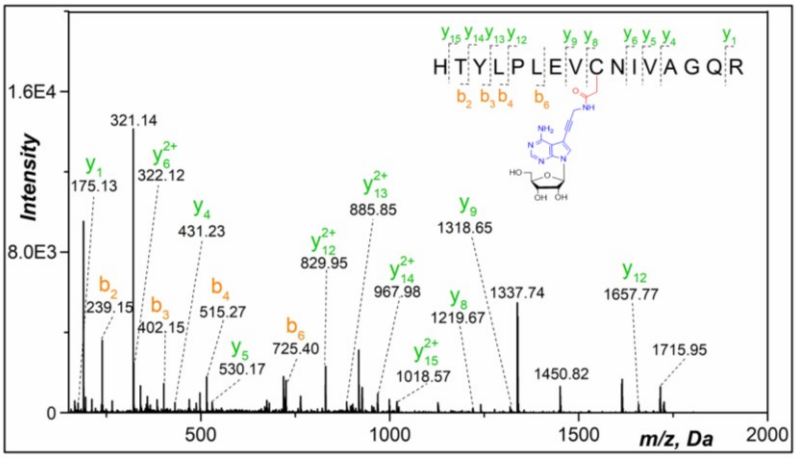



**Figure 112. Nano-ESI$^{+}$-MS/MS spectrum of peptide from 20RNA_1A$^{CA\text{-}hAgo2}$ conjugate digest. Modification cross-linked to cysteine (position C362). *m*/*z* acquired: 724.6937 Da, z = 3+.**

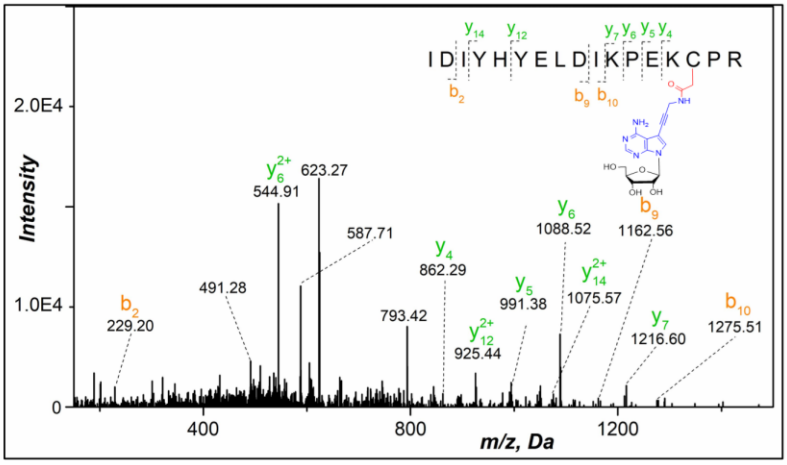



**Figure 113. Nano-ESI$^{+}$-MS/MS spectrum of peptide from 20RNA_1A$^{CA\text{-}hAgo2}$ conjugate digest. Modification cross-linked to cysteine (position C84). *m*/*z* acquired: 623.5577 Da, z = 4+.**

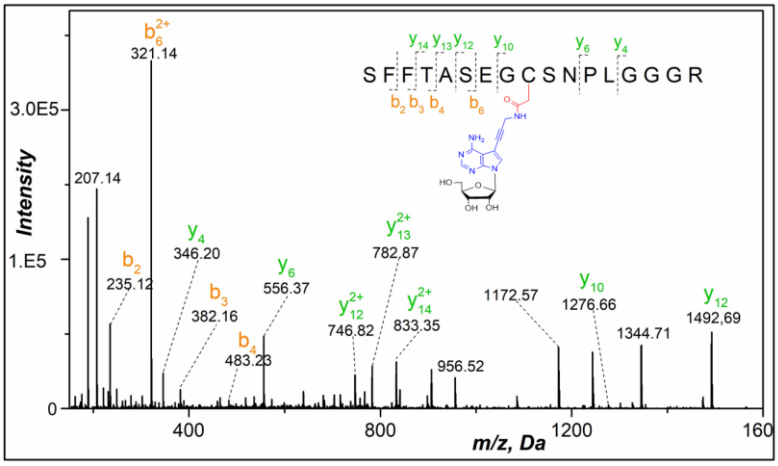



**Figure 114. Nano-ESI$^{+}$-MS/MS spectrum of peptide from 20RNA_1A$^{CA\text{-}hAgo2}$ conjugate digest. Modification cross-linked to cysteine (position C206). *m*/*z* acquired: 682.6306 Da, z = 3+.**

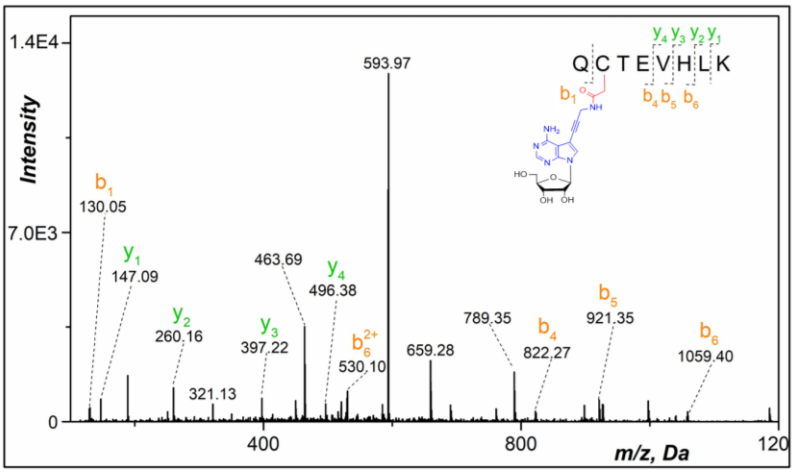



**Figure 115. Nano-ESI$^{+}$-MS/MS spectrum of peptide from 20RNA_1A$^{CA\text{-}hAgo2}$ conjugate digest. Modification cross-linked to cysteine (position C480). *m*/*z* acquired: 659.2984 Da, z = 2+.**

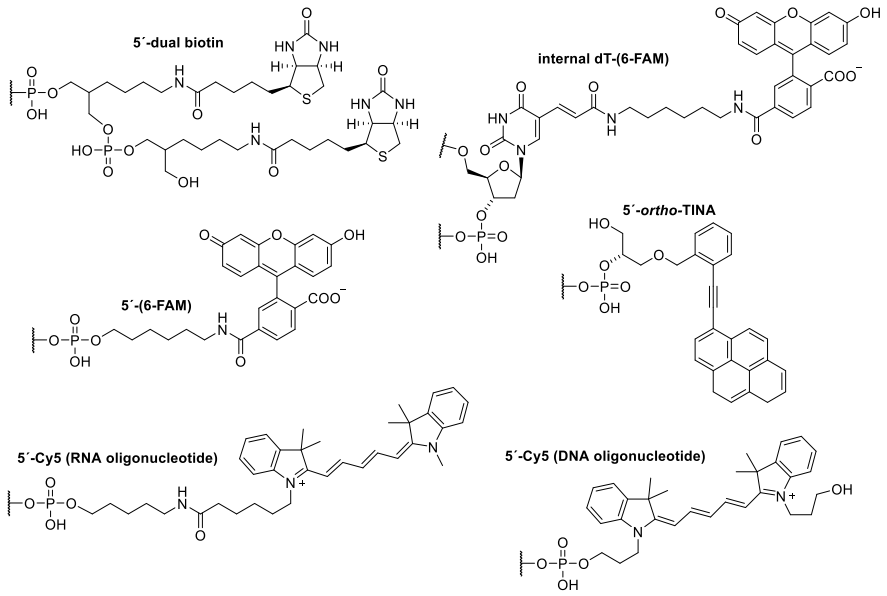



**Figure 116. Structures of modifications of synthetic oligonucleotides.**

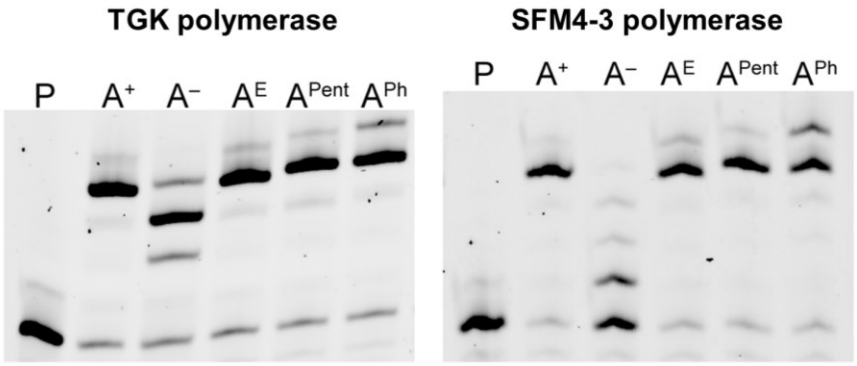



**Figure 117. 22.5% dPAGE analysis of PEX reaction with templ_19nt_A. (P) RNA primer; ($A^+$) positive control, mixture of rATP, rGTP; ($A^-$) negative control, mixture of rGTP and $H_2O$; ($A^E$) modification, mixture of $rA^ETP$, rGTP; ($A^{Pent}$) modification, mixture of $rA^{Pent}TP$, rGTP; ($A^{Ph}$) modification, mixture of $rA^{Ph}TP$, rGTP. FAM scan. Left figure – PEX reaction performed with TGK polymerase, right figure – PEX reaction performed with SFM4-3 polymerase.**

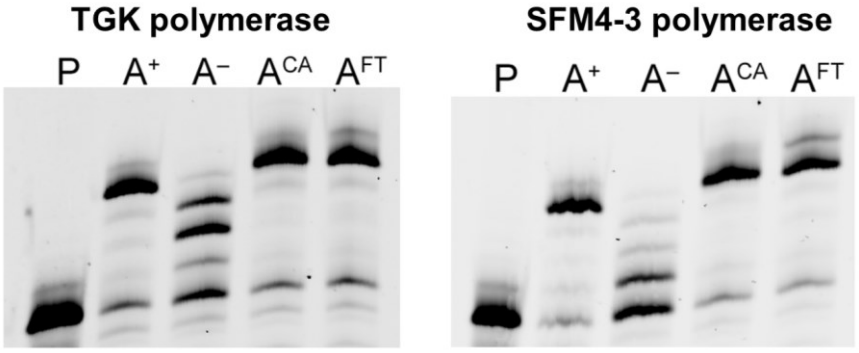



**Figure 118. 22.5% dPAGE analysis of PEX reaction with templ_19nt_A. (P) RNA primer; ($A^+$) positive control, mixture of rATP, rGTP; ($A^-$) negative control, mixture of rGTP and $H_2O$; ($A^{CA}$) modification, mixture of $rA^{CA}TP$, rGTP; ($A^{FT}$) modification, mixture of $rA^{FT}TP$, rGTP. FAM scan. Left figure – PEX reaction performed with TGK polymerase, right figure – PEX reaction performed with SFM4-3 polymerase.**

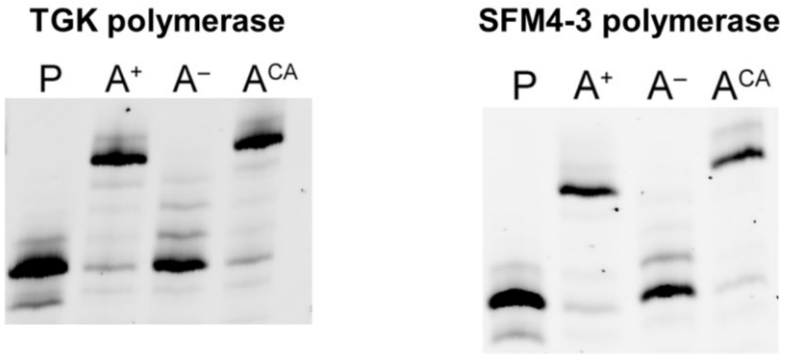



**Figure 119. 22.5% dPAGE analysis of PEX reaction with templ_19nt_A under reduced temperature. (P) RNA primer; ($A^+$) positive control, mixture of rATP, rGTP; ($A^-$) negative control, mixture of rGTP and $H_2O$; ($A^{CA}$) modification, mixture of $rA^{CA}TP$, rGTP. FAM scan. Left figure – PEX reaction performed with TGK polymerase, right figure – PEX reaction performed with SFM4-3 polymerase.**

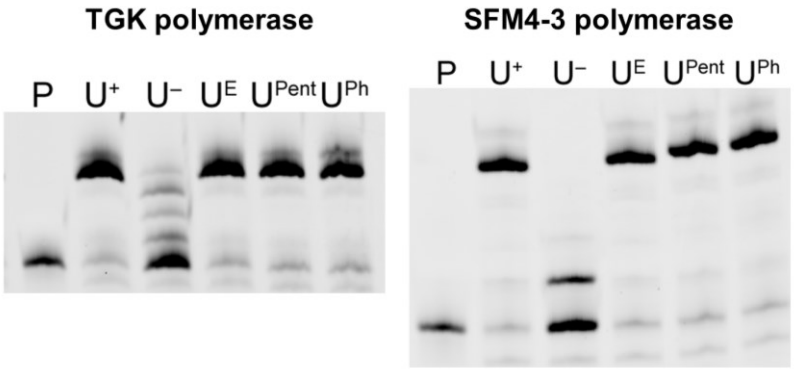



**Figure 120. 22.5% dPAGE analysis of PEX reaction with templ_19nt_U. (P) RNA primer; ($U^{+}$) positive control, mixture of rUTP, rGTP; ($U^{-}$) negative control, mixture of rGTP and $H_2O$; ($U^{E}$) modification, mixture of $rU^{E}TP$, rGTP; ($U^{Pent}$) modification, mixture of $rU^{Pent}TP$, rGTP; ($U^{Ph}$) modification, mixture of $rU^{Ph}TP$, rGTP. FAM scan. Left figure – PEX reaction performed with TGK polymerase, right figure – PEX reaction performed with SFM4-3 polymerase.**

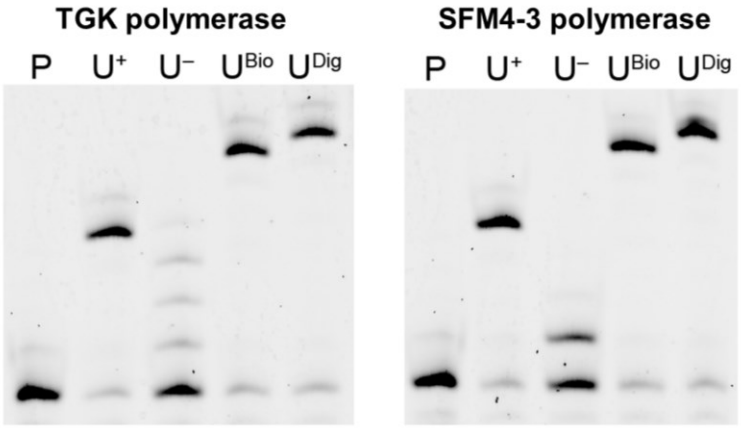



**Figure 121. 22.5% dPAGE analysis of PEX reaction with templ_19nt_U. (P) RNA primer; ($U^{+}$) positive control, mixture of rUTP, rGTP; ($U^{-}$) negative control, mixture of rGTP and $H_2O$; ($U^{Bio}$) modification, mixture of $rU^{Bio}TP$, rGTP; ($U^{Dig}$) modification, mixture of $rU^{Dig}TP$, rGTP. FAM scan. Left figure – PEX reaction performed with TGK polymerase, right figure – PEX reaction performed with SFM4-3 polymerase.**

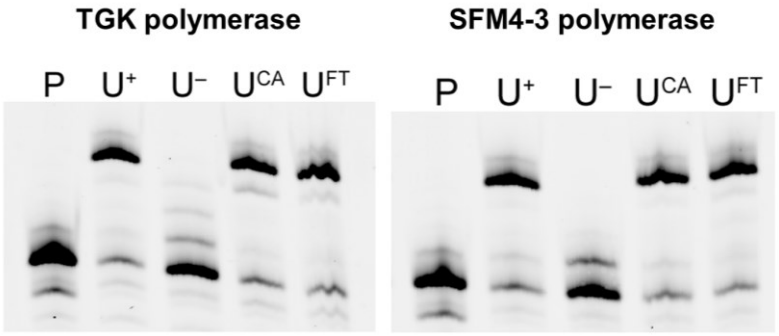



**Figure 122. 22.5% dPAGE analysis of PEX reaction with templ_19nt_U. (P) RNA primer; ($U^{+}$) positive control, mixture of rUTP, rGTP; ($U^{-}$) negative control, mixture of rGTP and $H_2O$; ($U^{CA}$) modification, mixture of $rU^{CA}TP$, rGTP; ($U^{FT}$) modification, mixture of $rU^{FT}TP$, rGTP. FAM scan. Left figure – PEX reaction performed with TGK polymerase, right figure – PEX reaction performed with SFM4-3 polymerase.**

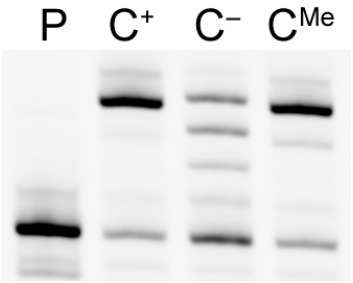



**Figure 123. 22.5% dPAGE analysis of PEX reaction with templ_19nt_C. (P) RNA primer; ($C^{+}$) positive control, mixture of rCTP, rGTP; ($C^{-}$) negative control, mixture of rGTP and $H_2O$; ($C^{Me}$) modification, mixture of $rC^{Me}TP$, rGTP. FAM scan. PEX reaction performed with TGK polymerase.**

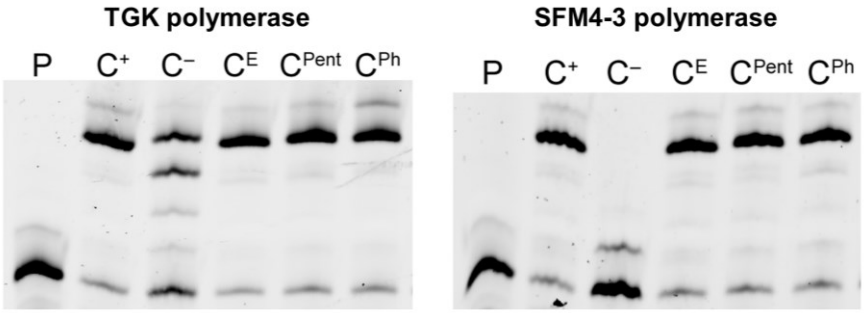



**Figure 124. 22.5% dPAGE analysis of PEX reaction with templ_19nt_C. (P) RNA primer; ($C^{+}$) positive control, mixture of rCTP, rGTP; ($C^{-}$) negative control, mixture of rGTP and $H_2O$; ($C^{E}$) modification, mixture of $rC^{E}TP$, rGTP; ($C^{Pent}$) modification, mixture of $rC^{Pent}TP$, rGTP; ($C^{Ph}$) modification, mixture of $rC^{Ph}TP$, rGTP. FAM scan. Left figure – PEX reaction performed with TGK polymerase, right figure – PEX reaction performed with SFM4-3 polymerase.**

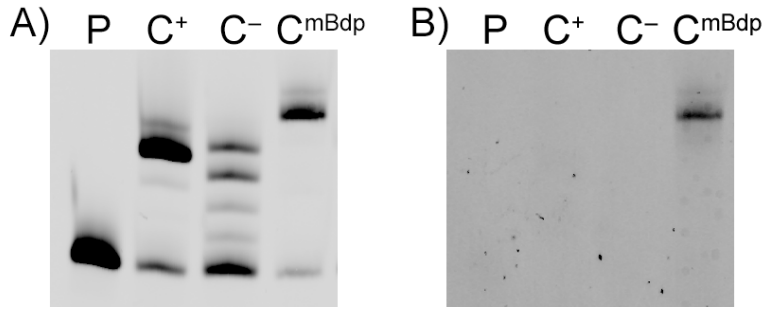



**Figure 125. 22.5% dPAGE analysis of PEX reaction with templ_19nt_C. (P) RNA primer; ($C^{+}$) positive control, mixture of rCTP, rGTP; ($C^{-}$) negative control, mixture of rGTP and $H_2O$; ($C^{mBdp}$) modification, mixture of $rC^{mBdp}TP$, rGTP. A) Cy5 scan; B) FAM scan (visualisation of mBdp-modification). PEX reaction performed with TGK polymerase.**

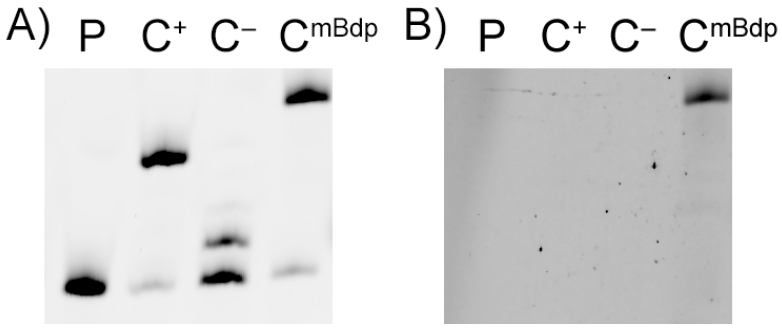



**Figure 126. 22.5% dPAGE analysis of PEX reaction with templ_19nt_C. (P) RNA primer; ($C^{+}$) positive control, mixture of rCTP, rGTP; ($C^{-}$) negative control, mixture of rGTP and $H_2O$; ($C^{mBdp}$) modification, mixture of $rC^{mBdp}TP$, rGTP. A) Cy5 scan; B) FAM scan (visualisation of mBdp-modification). PEX reaction performed with SFM4-3 polymerase.**

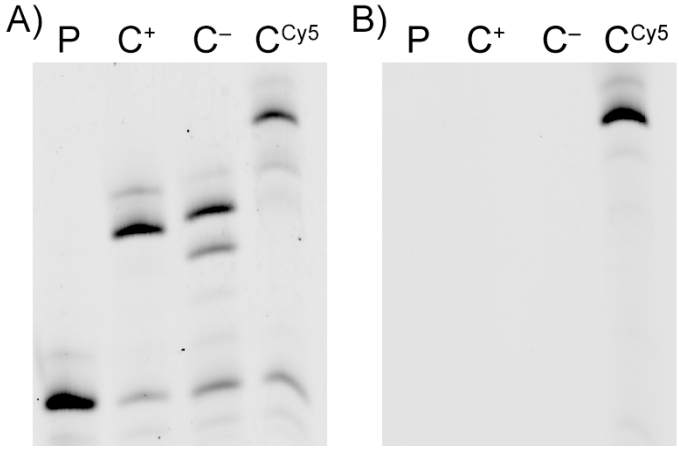



**Figure 127. 22.5% dPAGE analysis of PEX reaction with templ_19nt_C. (P) RNA primer; ($C^{+}$) positive control, mixture of rCTP, rGTP; ($C^{-}$) negative control, mixture of rGTP and $H_2O$; ($C^{Cy5}$) modification, mixture of $rC^{Cy5}TP$, rGTP. A) FAM scan; B) Cy5 scan (visualisation of Cy5-modification). PEX reaction performed with TGK polymerase.**

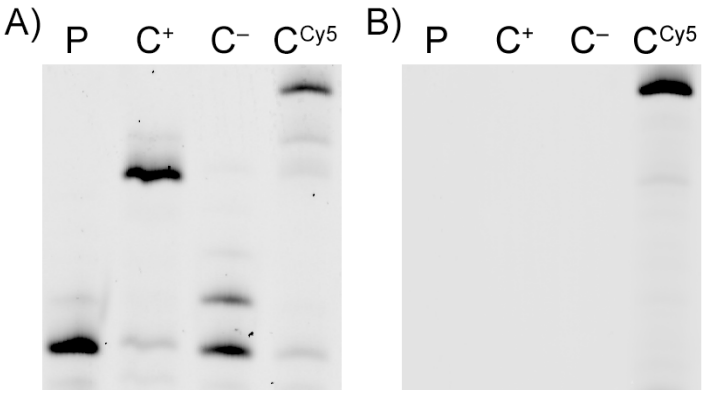



**Figure 128. 22.5% dPAGE analysis of PEX reaction with templ_19nt_C. (P) RNA primer; ($C^+$) positive control, mixture of rCTP, rGTP; ($C^-$) negative control, mixture of rGTP and $H_2O$; ($C^{Cy5}$) modification, mixture of $rC^{Cy5}TP$, rGTP. A) FAM scan; B) Cy5 scan (visualisation of Cy5-modification). PEX reaction performed with SFM4-3 polymerase.**

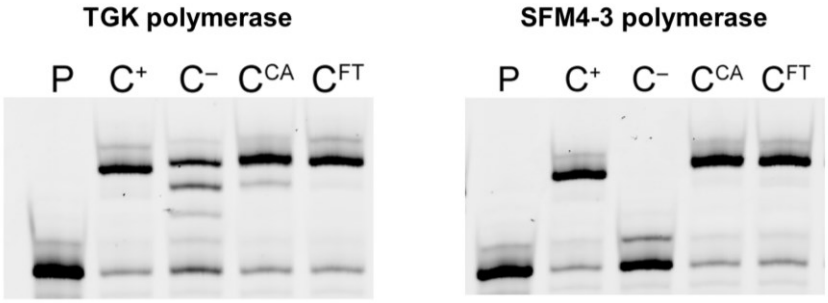



**Figure 129. 22.5% dPAGE analysis of PEX reaction with templ_19nt_C. (P) RNA primer; ($C^+$) positive control, mixture of rCTP, rGTP; ($C^-$) negative control, mixture of rGTP and $H_2O$; ($C^{CA}$) modification, mixture of $rC^{CA}TP$, rGTP; ($C^{FT}$) modification, mixture of $rC^{FT}TP$, rGTP. FAM scan. Left figure – PEX reaction performed with TGK polymerase, right figure – PEX reaction performed with SFM4-3 polymerase.**

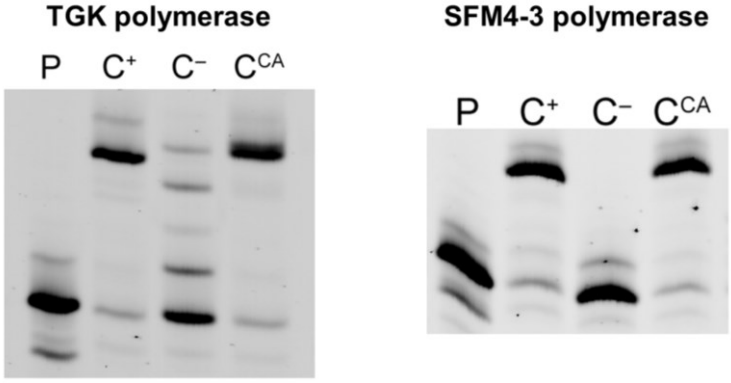



**Figure 130. 22.5% dPAGE analysis of PEX reaction with templ_19nt_C under reduced temperature. (P) RNA primer; ($C^+$) positive control, mixture of rCTP, rGTP; ($C^-$) negative control, mixture of rGTP and $H_2O$; ($C^{CA}$) modification, mixture of $rC^{CA}TP$, rGTP. FAM scan. Left figure – PEX reaction performed with TGK polymerase, right figure – PEX reaction performed with SFM4-3 polymerase.**

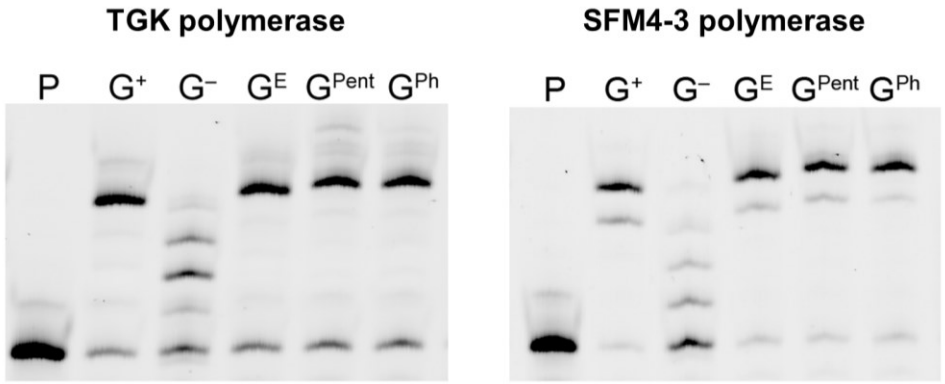



**Figure 131. 22.5% dPAGE analysis of PEX reaction with templ_19nt_G. (P) RNA primer; ($G^+$) positive control, mixture of rUTP, rGTP; ($G^-$) negative control, mixture of rUTP and $H_2O$; ($G^E$) modification, mixture of rUTP, $rG^ETP$; ($G^{Pent}$) modification, mixture of rUTP, $rG^{Pent}TP$; ($G^{Ph}$) modification, mixture of rUTP, $rG^{Ph}TP$. FAM scan. Left figure – PEX reaction performed with TGK polymerase, right figure – PEX reaction performed with SFM4-3 polymerase.**

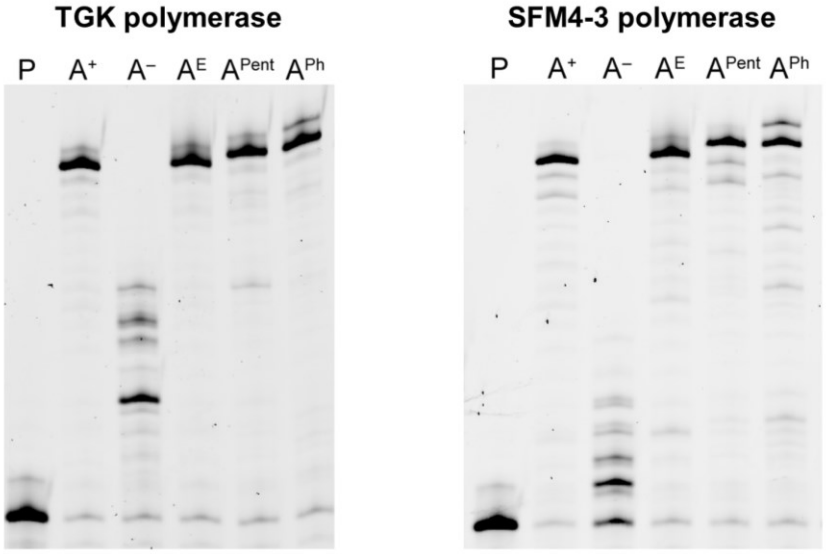



**Figure 132. 22.5% dPAGE analysis of PEX reaction with 5'-(TINA)-templ_31nt. (P) RNA primer; ($A^+$) positive control, all natural rNTPs; ($A^-$) negative control, mixture of rUTP, rCTP, rGTP and $H_2O$; ($A^E$) modification, mixture of $rA^ETP$, rUTP, rCTP, rGTP; ($A^{Pent}$) modification, mixture of $rA^{Pent}TP$, rUTP, rCTP, rGTP; ($A^{Ph}$) modification, mixture of $rA^{Ph}TP$, rUTP, rCTP, rGTP. FAM scan. Left figure – PEX reaction performed with TGK polymerase, right figure – PEX reaction performed with SFM4-3 polymerase.**

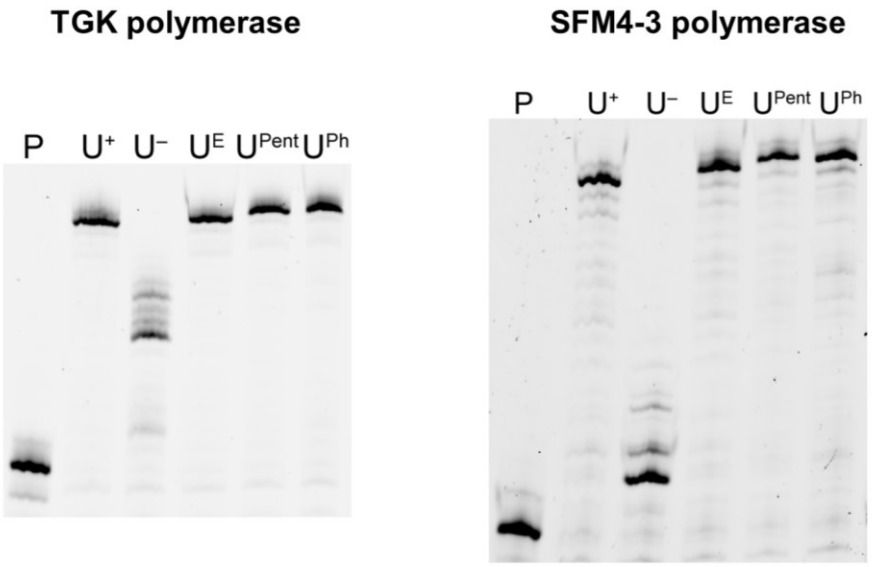



**Figure 133. 22.5% dPAGE analysis of PEX reaction with 5'-(TINA)-templ_31nt. (P) RNA primer; ($U^{+}$) positive control, all natural rNTPs; ($U^{-}$) negative control, mixture of rATP, rCTP, rGTP and $H_2O$; ($U^{E}$) modification, mixture of rATP, $rU^{E}TP$, rCTP, rGTP; ($U^{Pent}$) modification, mixture of rATP, $rU^{Pent}TP$, rCTP, rGTP; ($U^{Ph}$) modification, mixture of rATP, $rU^{Ph}TP$, rCTP, rGTP. FAM scan. Left figure – PEX reaction performed with TGK polymerase, right figure – PEX reaction performed with SFM4-3 polymerase.**

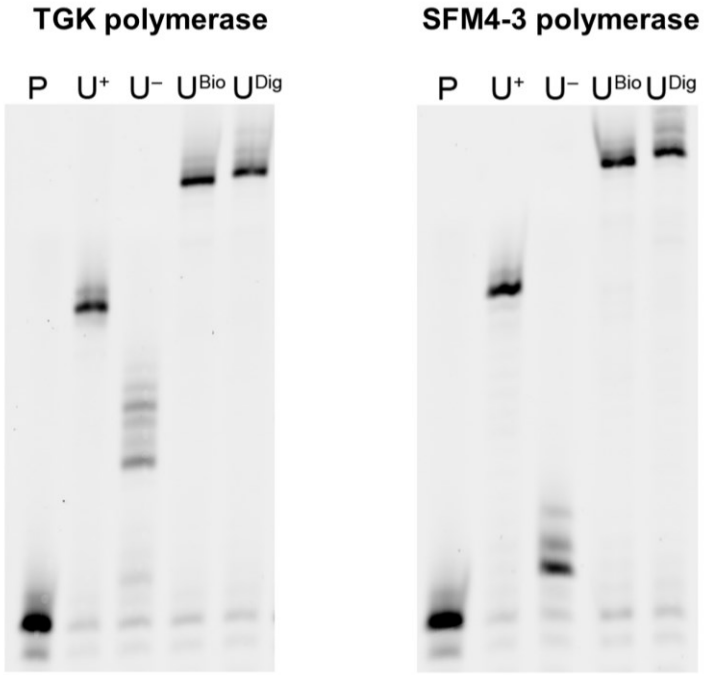



**Figure 134. 22.5% dPAGE analysis of PEX reaction with 5'-(TINA)-templ_31nt. (P) RNA primer; ($U^{+}$) positive control, all natural rNTPs; ($U^{-}$) negative control, mixture of rATP, rCTP, rGTP and $H_2O$; ($U^{Bio}$) modification, mixture of rATP, $rU^{Bio}TP$, rCTP, rGTP; ($U^{Dig}$) modification, mixture of rATP, $rU^{Dig}TP$, rCTP, rGTP. FAM scan. Left figure – PEX reaction performed with TGK polymerase, right figure – PEX reaction performed with SFM4-3 polymerase.**

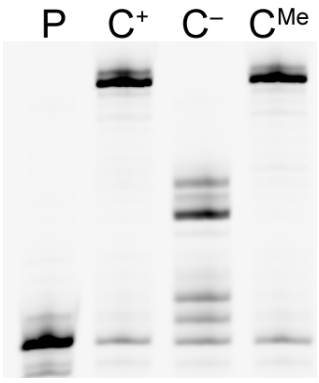



**Figure 135. 22.5% dPAGE analysis of PEX reaction with 5'-(TINA)-templ_31nt. (P) RNA primer; ($C^+$) positive control, all natural rNTPs; ($C^-$) negative control, mixture of rATP, rUTP, rGTP and $H_2O$; ($C^{Me}$) modification, mixture of rATP, rUTP, $rC^{Me}TP$, rGTP. FAM scan. PEX reaction performed with TGK polymerase.**

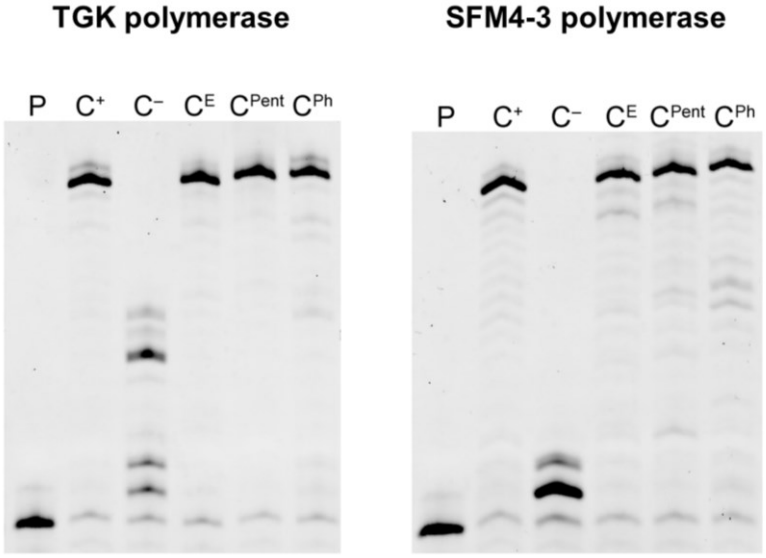



**Figure 136. 22.5% dPAGE analysis of PEX reaction with 5'-(TINA)-templ_31nt. (P) RNA primer; ($C^+$) positive control, all natural rNTPs; ($C^-$) negative control, mixture of rATP, rUTP, rGTP and $H_2O$; ($C^E$) modification, mixture of rATP, rUTP, $rC^ETP$, rGTP; ($C^{Pent}$) modification, mixture of rATP, rUTP, $rC^{Pent}TP$, rGTP; ($C^{Ph}$) modification, mixture of rATP, rUTP, $rC^{Ph}TP$, rGTP. FAM scan. Left figure – PEX reaction performed with TGK polymerase, right figure – PEX reaction performed with SFM4-3 polymerase.**

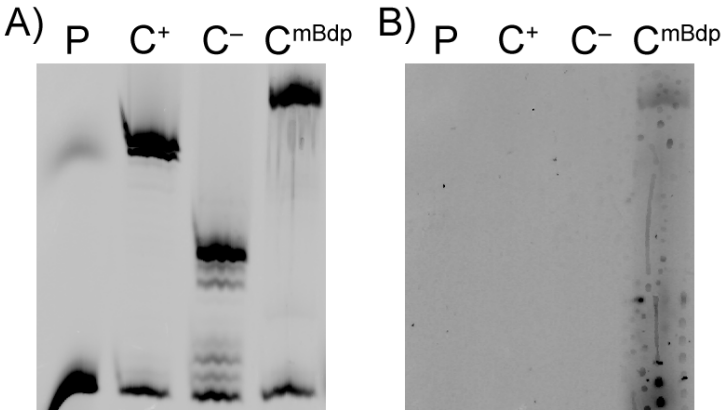



**Figure 137. 22.5% dPAGE analysis of PEX reaction with 5'-(TINA)-templ_31nt. (P) RNA primer; ($C^+$) positive control, all natural rNTPs; ($C^-$) negative control, mixture of rATP, rUTP, rGTP and $H_2O$; ($C^{mBdp}$) modification, mixture of rATP, rUTP, $rC^{mBdp}TP$, rGTP. A) Cy5 scan; B) FAM scan (visualisation of mBdp-modification). PEX reaction performed with TGK polymerase.**

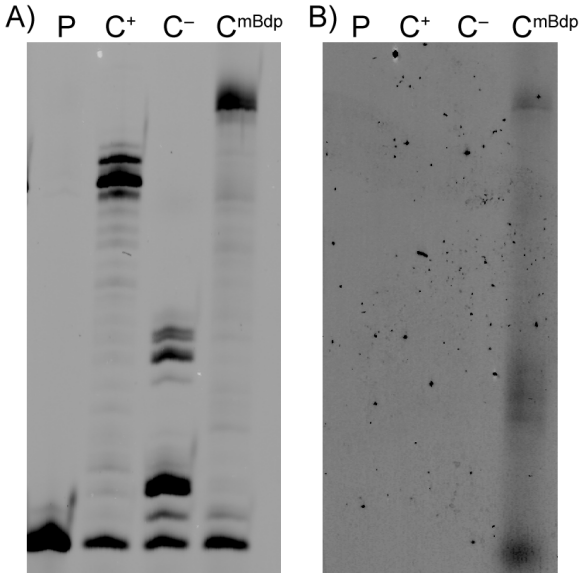



**Figure 138. 22.5% dPAGE analysis of PEX reaction with 5'-(TINA)-templ_31nt. (P) RNA primer; ($C^+$) positive control, all natural rNTPs; ($C^-$) negative control, mixture of rATP, rUTP, rGTP and $H_2O$; ($C^{mBdp}$) modification, mixture of rATP, rUTP, $rC^{mBdp}TP$, rGTP. A) Cy5 scan; B) FAM scan (visualisation of mBdp-modification). PEX reaction performed with SFM4-3 polymerase.**

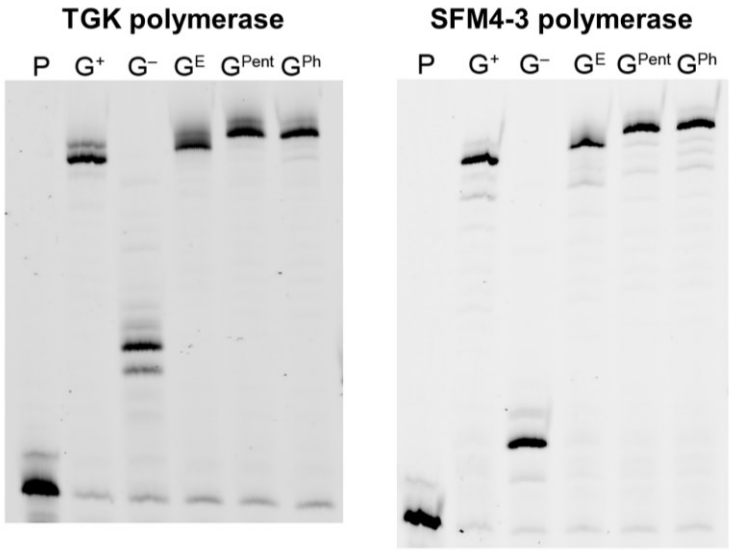



**Figure 139. 22.5% dPAGE analysis of PEX reaction with 5'-(TINA)-templ_31nt. (P) RNA primer; ($G^{+}$) positive control, all natural rNTPs; ($G^{-}$) negative control, mixture of rATP, rUTP, rCTP and $H_2O$; ($G^{E}$) modification, mixture of rATP, rUTP, rCTP, $rG^{E}TP$; ($G^{Pent}$) modification, mixture of rATP, rUTP, rCTP, $rG^{Pent}TP$; ($G^{Ph}$) modification, mixture of rATP, rUTP, rCTP, $rG^{Ph}TP$. FAM scan. Left figure – PEX reaction performed with TGK polymerase, right figure – PEX reaction performed with SFM4-3 polymerase.**

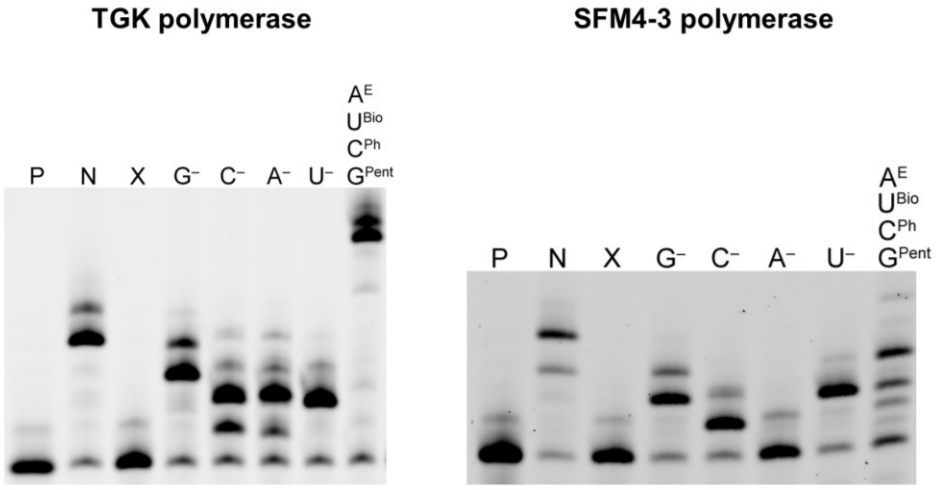



**Figure 140. 12.5% dPAGE analysis of PEX reaction with templ_19nt_mix. (P) RNA primer; (N) positive control, all natural rNTPs; (X) negative control, no rNTPs; ($G^{-}$) negative control, mixture of rATP, rUTP, rCTP and $H_2O$; ($C^{-}$) negative control, mixture of rATP, rUTP, rGTP and $H_2O$; ($A^{-}$) negative control, mixture of rUTP, rCTP, rGTP and $H_2O$; ($U^{-}$) negative control, mixture of rATP, rCTP, rGTP and $H_2O$; ($A^{E}$, $U^{Bio}$, $C^{Ph}$, $G^{Pent}$) modification, mixture of $rA^{E}TP$, $rU^{Bio}TP$, $rC^{Ph}TP$, $rG^{Ph}TP$. FAM scan. Left figure – PEX reaction performed with TGK polymerase, right figure – PEX reaction performed with SFM4-3 polymerase.**

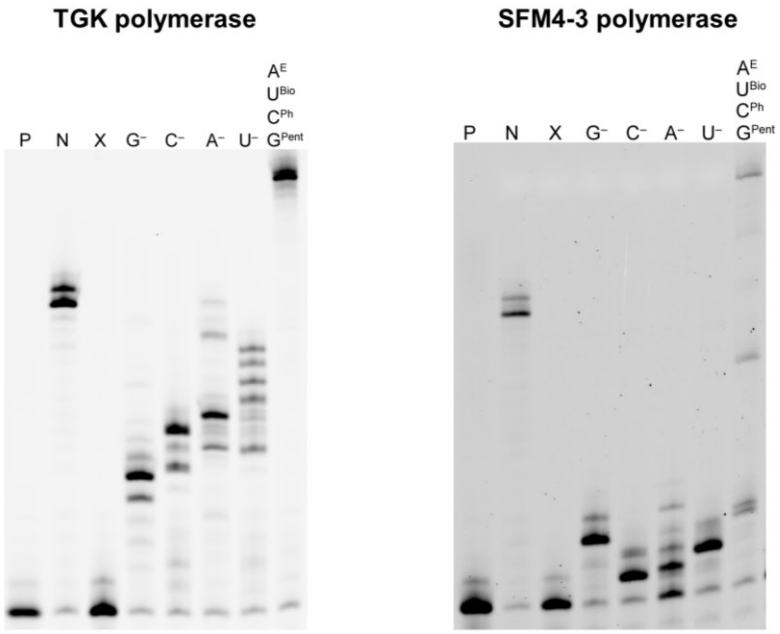



**Figure 141. 12.5% dPAGE analysis of PEX reaction with 5'-(TINA)-templ_31nt. (P) RNA primer; (N) positive control, all natural rNTPs; (X) negative control, no rNTPs; ($G^-$) negative control, mixture of rATP, rUTP, rCTP and $H_2O$; ($C^-$) negative control, mixture of rATP, rUTP, rGTP and $H_2O$; ($A^-$) negative control, mixture of rUTP, rCTP, rGTP and $H_2O$; ($U^-$) negative control, mixture of rATP, rCTP, rGTP and $H_2O$; ($A^E$, $U^{Bio}$, $C^{Ph}$, $G^{Pent}$) modification, mixture of $rA^ETP$, $rU^{Bio}TP$, $rC^{Ph}TP$, $rG^{Ph}TP$. FAM scan. Left figure – PEX reaction performed with TGK polymerase, right figure – PEX reaction performed with SFM4-3 polymerase.**

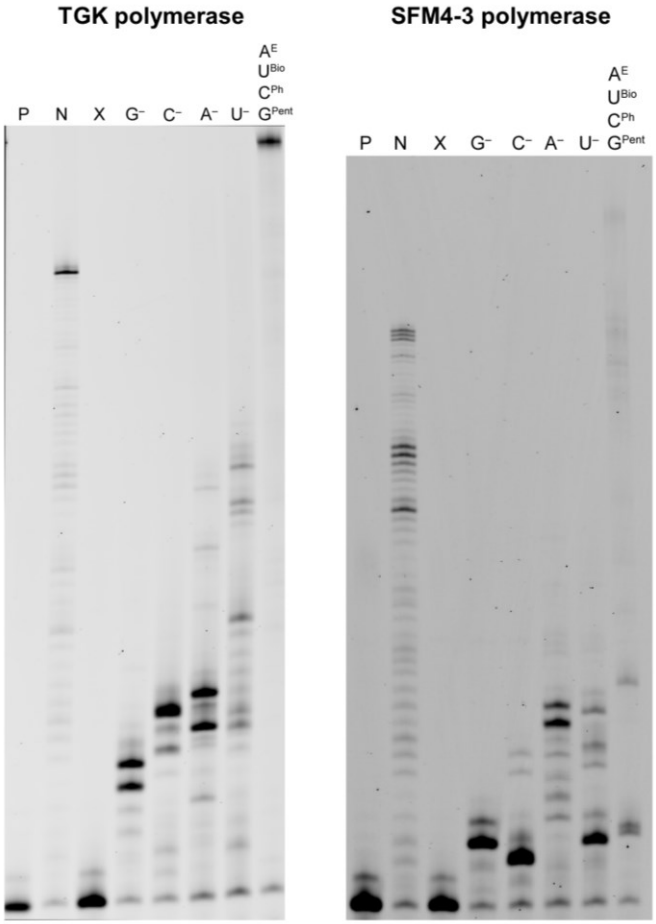



**Figure 142. 12.5% dPAGE analysis of PEX reaction with templ_65nt. (P) RNA primer; (N) positive control, all natural rNTPs; (X) negative control, no rNTPs; ($G^-$) negative control, mixture of rATP, rUTP, rCTP and $H_2O$; ($C^-$) negative control, mixture of rATP, rUTP, rGTP and $H_2O$; ($A^-$) negative control, mixture of rUTP, rCTP, rGTP and $H_2O$; ($U^-$) negative control, mixture of rATP, rCTP, rGTP and $H_2O$; ($A^E$, $U^{Bio}$, $C^{Ph}$, $G^{Pent}$) modification, mixture of $rA^{E}TP$, $rU^{Bio}TP$, $rC^{Ph}TP$, $rG^{Ph}TP$. FAM scan. Left figure – PEX reaction performed with TGK polymerase, right figure – PEX reaction performed with SFM4-3 polymerase.**

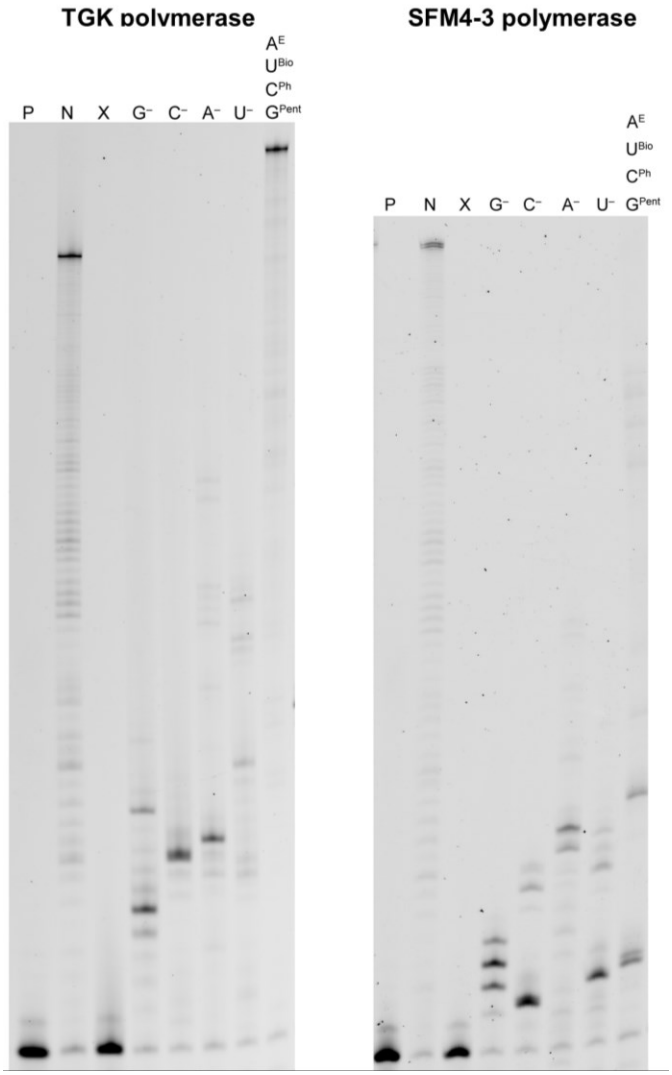



**Figure 143. 12.5% dPAGE analysis of PEX reaction with templ_98nt. (P) RNA primer; (N) positive control, all natural rNTPs; (X) negative control, no rNTPs; ($G^-$) negative control, mixture of rATP, rUTP, rCTP and $H_2O$; ($C^-$) negative control, mixture of rATP, rUTP, rGTP and $H_2O$; ($A^-$) negative control, mixture of rUTP, rCTP, rGTP and $H_2O$; ($U^-$) negative control, mixture of rATP, rCTP, rGTP and $H_2O$; ($A^E$, $U^{Bio}$, $C^{Ph}$, $G^{Pent}$) modification, mixture of $rA^ETP$, $rU^{Bio}TP$, $rC^{Ph}TP$, $rG^{Ph}TP$. FAM-scan. Left figure – PEX reaction performed with TGK polymerase, right figure – PEX reaction performed with SFM4-3 polymerase.**

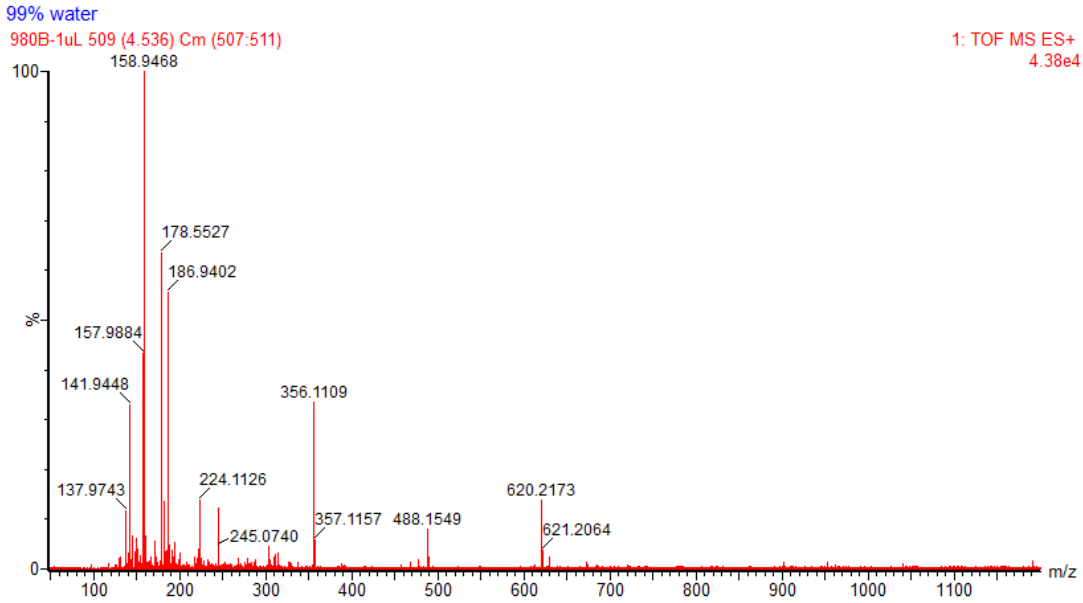



**Figure 144. 19RNA_C$^{CA}$ digest, raw spectrum, calculated mass: 620.2065 Da, found mass: 620.2173 Da (product).**

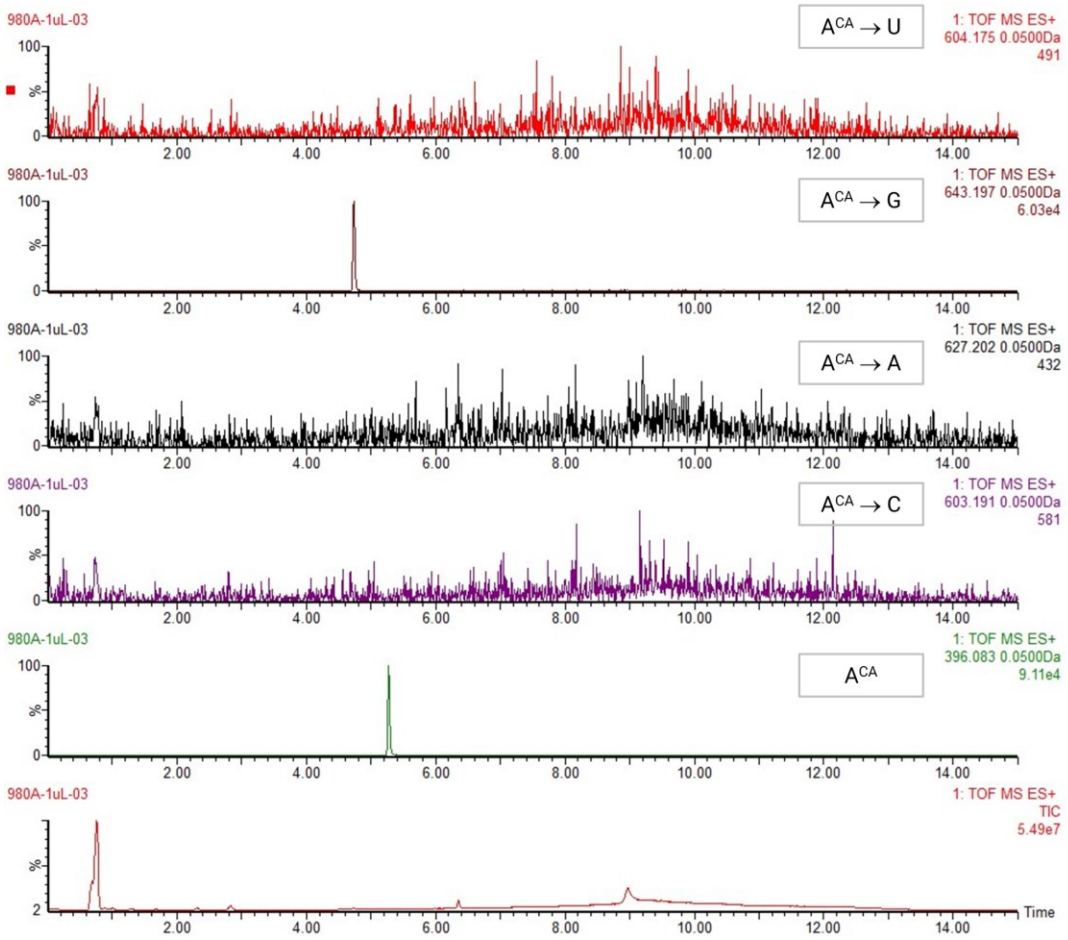



**Figure 145. Chromatogram from 19RNA_A$^{CA}$ digest. Extracted mass chromatograms for dinucleoside conjugates of A$^{CA}$ to putative indicated nucleosides generated after RNA digestion. Lowest panel represents TIC chromatogram.**

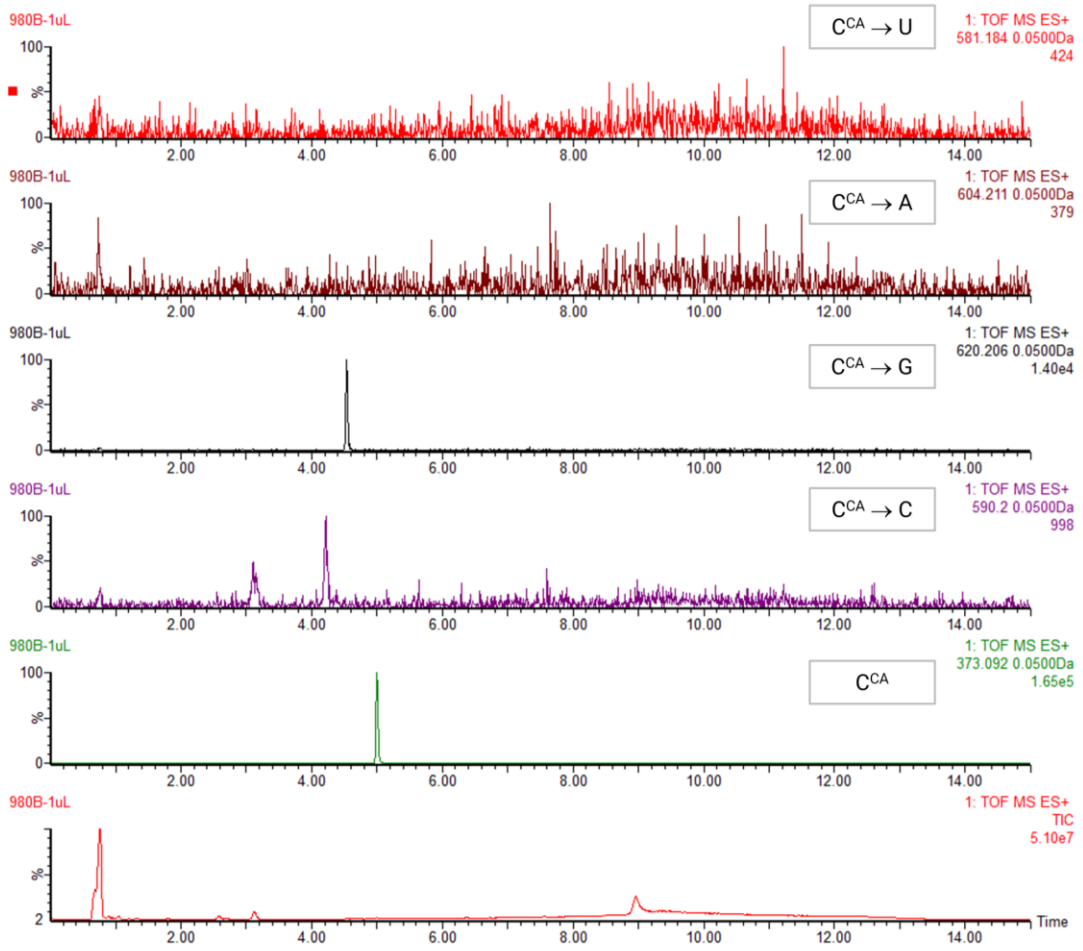



**Figure 146. Chromatogram from 19RNA_C$^{CA}$ digest. Extracted mass chromatograms for dinucleoside conjugates of C$^{CA}$ to putative indicated nucleosides generated after RNA digestion. Lowest panel represents TIC chromatogram.**

**Table 21. List of mRNA and mRNA intermediates prepared in this study.**

**Cy5-IRES-RNA**

GACUCACUAUAGGGCCCCUCUCCCUCCCCCCCCCCCUAACGUUACUGGCCGAAGCCGCUUGGAAUAAGGCCGGUGUGCGUUU
GUCUAUAUGUUAUUUUCCACCAUAUUGCCGUCUUUUGGCAAUGUGAGGGCCCGGAAACCUGGCCCUGUCUUCUUGACGAGC
AUUCCUAGGGGUCUUUCCCCUCUCGCCAAAGGAAUGCAAGGUCUGUUGAAUGUCGUGAAGGAAGCAGUUCCUCUGGAAGCU
UCUUGAAGACAAACAACGUCUGUAGCGACCCUUUGCAGGCAGCGGAACCCCCCACCUGGCGACAGGUGCCUCUGCGGCCAA
AAGCCACGUGUAUAAGAUACACCUGCAAAGGCGGCACAACCCCAGUGCCACGUUGUGAGUUGGAUAGUUGUGGAAAGAGUC
AAAUGGCUCUCCUCAAGCGUAUUCAACAAGGGGCUGAAGGAUGCCCAGAAGGUACCCCAUUGUAUGGGAUCUGAUCUGGGG
CCUCGGUGCACAUGCUUUACAUGUGUUUAGUCGAGGUUAAAAAAACGUCUAGGCCCCCCGAACCACGGGGACGUGGUUUUC
CUUUGAAAAACACGAUGAUAAUAUGGCCACAACCAUGGU

**Cy5-IRES-RNA_prolonged**

GACUCACUAUAGGGCCCCUCUCCCUCCCCCCCCCCCUAACGUUACUGGCCGAAGCCGCUUGGAAUAAGGCCGGUGUGCGUUU
GUCUAUAUGUUAUUUUCCACCAUAUUGCCGUCUUUUGGCAAUGUGAGGGCCCGGAAACCUGGCCCUGUCUUCUUGACGAGC
AUUCCUAGGGGUCUUUCCCCUCUCGCCAAAGGAAUGCAAGGUCUGUUGAAUGUCGUGAAGGAAGCAGUUCCUCUGGAAGCU
UCUUGAAGACAAACAACGUCUGUAGCGACCCUUUGCAGGCAGCGGAACCCCCCACCUGGCGACAGGUGCCUCUGCGGCCAA
AAGCCACGUGUAUAAGAUACACCUGCAAAGGCGGCACAACCCCAGUGCCACGUUGUGAGUUGGAUAGUUGUGGAAAGAGUC
AAAUGGCUCUCCUCAAGCGUAUUCAACAAGGGGCUGAAGGAUGCCCAGAAGGUACCCCAUUGUAUGGGAUCUGAUCUGGGG
CCUCGGUGCACAUGCUUUACAUGUGUUUAGUCGAGGUUAAAAAAACGUCUAGGCCCCCCGAACCACGGGGACGUGGUUUUC
CUUUGAAAAACACGAUGAUAAUAUGGCCACAACCAUGGUCUUCACACUCGAAGAUUUCGUUGGGGACUGGCGACAGACAGC

CGGCUACAACCUGGACCAAGUCCUUGAACAGGGAGGUGUGUCCAGUUUGUUUCAGAAUCUCGGGGUGUCCGUAACUCCGAU
CCAAAGGAUUGUCCUGAGCGGUGAAAAUGGGCUGAAGAUCGACAUCCAUGUCAUCAUCCCGUAUGAAGGUCUGAG

**Cy5-mRNA-nat**

GACUCACUAUAGGGCCCCUCUCCCUCCCCCCCCCCUAACGUUACUGGCCGAAGCCGCUUGGAAUAAGGCCGGUGUGCGUUU
GUCUAUAUGUUAUUUUCCACCAUAUUGCCGUCUUUUGGCAAUGUGAGGGCCCGGAAACCUGGCCCUGUCUUCUUGACGAGC
AUUCCUAGGGGUCUUUCCCCUCUCGCCAAAGGAAUGCAAGGUCUGUUGAAUGUCGUGAAGGAAGCAGUUCCUCUGGAAGCU
UCUUGAAGACAAACAACGUCUGUAGCGACCCUUUGCAGGCAGCGGAACCCCCCACCUGGCGACAGGUGCCUCUGCGGCCAA
AAGCCACGUGUAUAAGAUACACCUGCAAAGGCGGCACAACCCCAGUGCCACGUUGUGAGUUGGAUAGUUGUGGAAAGAGUC
AAAUGGCUCUCCUCAAGCGUAUUCAACAAGGGGCUGAAGGAUGCCCAGAAGGUACCCCAUUGUAUGGGAUCUGAUCUGGGG
CCUCGGUGCACAUGCUUUACAUGUGUUUAGUCGAGGUUAAAAAAACGUCUAGGCCCCCCGAACCACGGGGACGUGGUUUUC
CUUUGAAAAACACGAUGAUAAUAUGGCCACAACCAUGGUCUUCACACUCGAAGAUUUCGUUGGGGACUGGCGACAGACAGC
CGGCUACAACCUGGACCAAGUCCUUGAACAGGGAGGUGUGUCCAGUUUGUUUCAGAAUCUCGGGGUGUCCGUAACUCCGAU
CCAAAGGAUUGUCCUGAGCGGUGAAAAUGGGCUGAAGAUCGACAUCCAUGUCAUCAUCCCGUAUGAAGGUCUGAGCGGCGA
CCAAAUGGGCCAGAUCGAAAAAAUUUUUAAGGUGGUGUACCCUGUGGAUGAUCAUCACUUUAAGGUGAUCCUGCACUAUGG
CACACUGGUAAUCGACGGGGUUACGCCGAACAUGAUCGACUAUUUCGGACGGCCGUAUGAAGGCAUCGCCGUGUUCGACGG
CAAAAAGAUCACUGUAACAGGGACCCUGUGGAACGGCAACAAAAUUAUCGACGAGCGCCUGAUCAACCCCGACGGCUCCCU
GCUGUUCCGAGUAACCAUCAACGGAGUGACCGGCUGGCGGCUGUGCGAACGCAUUCUGGCGUAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAA

**Cy5-mRNA-init**

GACUCACUAUAGGGCCCCUCUCCCUCCCCCCCCCCUAACGUUACUGGCCGAAGCCGCUUGGAAUAAGGCCGGUGUGCGUUU
GUCUAUAUGUUAUUUUCCACCAUAUUGCCGUCUUUUGGCAAUGUGAGGGCCCGGAAACCUGGCCCUGUCUUCUUGACGAGC
AUUCCUAGGGGUCUUUCCCCUCUCGCCAAAGGAAUGCAAGGUCUGUUGAAUGUCGUGAAGGAAGCAGUUCCUCUGGAAGCU
UCUUGAAGACAAACAACGUCUGUAGCGACCCUUUGCAGGCAGCGGAACCCCCCACCUGGCGACAGGUGCCUCUGCGGCCAA
AAGCCACGUGUAUAAGAUACACCUGCAAAGGCGGCACAACCCCAGUGCCACGUUGUGAGUUGGAUAGUUGUGGAAAGAGUC
AAAUGGCUCUCCUCAAGCGUAUUCAACAAGGGGCUGAAGGAUGCCCAGAAGGUACCCCAUUGUAUGGGAUCUGAUCUGGGG
CCUCGGUGCACAUGCUUUACAUGUGUUUAGUCGAGGUUAAAAAAACGUCUAGGCCCCCCGAACCACGGGGACGUGGUUUUC
CUUUGAAAAACACGAUGAUAAUAUGGCCACAACCAUGGU**C**UUCACACUCGAAGAUUUCGUUGGGGACUGGCGACAGACAGC
CGGCUACAACCUGGACCAAGUCCUUGAACAGGGAGGUGUGUCCAGUUUGUUUCAGAAUCUCGGGGUGUCCGUAACUCCGAU
CCAAAGGAUUGUCCUGAGCGGUGAAAAUGGGCUGAAGAUCGACAUCCAUGUCAUCAUCCCGUAUGAAGGUCUGAGCGGCGA
CCAAAUGGGCCAGAUCGAAAAAAUUUUUAAGGUGGUGUACCCUGUGGAUGAUCAUCACUUUAAGGUGAUCCUGCACUAUGG
CACACUGGUAAUCGACGGGGUUACGCCGAACAUGAUCGACUAUUUCGGACGGCCGUAUGAAGGCAUCGCCGUGUUCGACGG
CAAAAAGAUCACUGUAACAGGGACCCUGUGGAACGGCAACAAAAUUAUCGACGAGCGCCUGAUCAACCCCGACGGCUCCCU
GCUGUUCCGAGUAACCAUCAACGGAGUGACCGGCUGGCGGCUGUGCGAACGCAUUCUGGCGUAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAA

**Cy5-mRNA-mid**

GACUCACUAUAGGGCCCCUCUCCCUCCCCCCCCCCUAACGUUACUGGCCGAAGCCGCUUGGAAUAAGGCCGGUGUGCGUUU
GUCUAUAUGUUAUUUUCCACCAUAUUGCCGUCUUUUGGCAAUGUGAGGGCCCGGAAACCUGGCCCUGUCUUCUUGACGAGC
AUUCCUAGGGGUCUUUCCCCUCUCGCCAAAGGAAUGCAAGGUCUGUUGAAUGUCGUGAAGGAAGCAGUUCCUCUGGAAGCU
UCUUGAAGACAAACAACGUCUGUAGCGACCCUUUGCAGGCAGCGGAACCCCCCACCUGGCGACAGGUGCCUCUGCGGCCAA
AAGCCACGUGUAUAAGAUACACCUGCAAAGGCGGCACAACCCCAGUGCCACGUUGUGAGUUGGAUAGUUGUGGAAAGAGUC
AAAUGGCUCUCCUCAAGCGUAUUCAACAAGGGGCUGAAGGAUGCCCAGAAGGUACCCCAUUGUAUGGGAUCUGAUCUGGGG
CCUCGGUGCACAUGCUUUACAUGUGUUUAGUCGAGGUUAAAAAAACGUCUAGGCCCCCCGAACCACGGGGACGUGGUUUUC
CUUUGAAAAACACGAUGAUAAUAUGGCCACAACCAUGGUCUUCACACUCGAAGAUUUCGUUGGGGACUGGCGACAGACAGC
CGGCUACAACCUGGACCAAGUCCUUGAACAGGGAGGUGUGUCCAGUUUGUUUCAGAAUCUCGGGGUGUCCGUAACUCCGAU
CCAAAGGAUUGUCCUGAGCGGUGAAAAUGGGCUGAAGAUCGACAUCCAUGUCAUCAUCCCGUAUGAAGGUCUGAG**C**GGCGA
CCAAAUGGGCCAGAUCGAAAAAAUUUUUAAGGUGGUGUACCCUGUGGAUGAUCAUCACUUUAAGGUGAUCCUGCACUAUGG
CACACUGGUAAUCGACGGGGUUACGCCGAACAUGAUCGACUAUUUCGGACGGCCGUAUGAAGGCAUCGCCGUGUUCGACGG
CAAAAAGAUCACUGUAACAGGGACCCUGUGGAACGGCAACAAAAUUAUCGACGAGCGCCUGAUCAACCCCGACGGCUCCCU
GCUGUUCCGAGUAACCAUCAACGGAGUGACCGGCUGGCGGCUGUGCGAACGCAUUCUGGCGUAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAA

**Cy5-mRNA-gene**

GACUCACUAUAGGGCCCCUCUCCCUCCCCCCCCCCUAACGUUACUGGCCGAAGCCGCUUGGAAUAAGGCCGGUGUGCGUUU
GUCUAUAUGUUAUUUUCCACCAUAUUGCCGUCUUUUGGCAAUGUGAGGGCCCGGAAACCUGGCCCUGUCUUCUUGACGAGC
AUUCCUAGGGGUCUUUCCCCUCUCGCCAAAGGAAUGCAAGGUCUGUUGAAUGUCGUGAAGGAAGCAGUUCCUCUGGAAGCU
UCUUGAAGACAAACAACGUCUGUAGCGACCCUUUGCAGGCAGCGGAACCCCCCACCUGGCGACAGGUGCCUCUGCGGCCAA
AAGCCACGUGUAUAAGAUACACCUGCAAAGGCGGCACAACCCCAGUGCCACGUUGUGAGUUGGAUAGUUGUGGAAAGAGUC

AAAUGGCUCUCCUCAAGCGUAUUCAACAAGGGGCUGAAGGAUGCCCAGAAGGUACCCCAUUGUAUGGGAUCUGAUCUGGGG
CCUCGGUGCACAUGCUUUACAUGUGUUUAGUCGAGGUUAAAAAAACGUCUAGGCCCCCCGAACCACGGGGACGUGGUUUUC
CUUUGAAAAACACGAUGAUAAUAUGGCCACAACCAUGGUCUUCACACUCGAAGAUUUCGUUGGGGACUGGCGACAGACAGC
CGGCUACAACCUGGACCAAGUCCUUGAACAGGGAGGUGUGUCCAGUUUGUUUCAGAAUCUCGGGGUGUCCGUAACUCCGAU
CCAAAGGAUUGUCCUGAGCGGUGAAAAUGGGCUGAAGAUCGACAUCCAUGUCAUCAUCCCGUAUGAAGGUCUGAGCGGCGA
CCAAAUGGGCCAGAUCGAAAAAAUUUUUAAGGUGGUGUACCCUGUGGAUGAUCAUCACUUUAAGGUGAUCCUGCACUAUGG
CACACUGGUAAUCGACGGGGUUACGCCGAACAUGAUCGACUAUUUCGGACGGCCGUAUGAAGGCAUCGCCGUGUUCGACGG
CAAAAAGAUCACUGUAACAGGGACCCUGUGGAACGGCAACACAAAAUUAUCGACGAGCGCCUGAUCAACCCCGACGGCUCCCU
GCUGUUCCGAGUAACCAUCAACGGAGUGACCGGCUGGCGGCUGUGCGAACGCAUUCUGGCGUAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAA

**Cy5-mRNA-full**

GACUCACUAUAGGGCCCCUCUCCCUCCCCCCCCCCUAACGUUACUGGCCGAAGCCGCUUGGAAUAAGGCCGGUGUGCGUUU
GUCUAUAUGUUAUUUUCCACCAUAUUGCCGUCUUUUGGCAAUGUGAGGGCCCGGAAACCUGGCCCUGUCUUCUUGACGAGC
AUUCCUAGGGGUCUUUCCCCUCUCGCCAAAGGAAUGCAAGGUCUGUUGAAUGUCGUGAAGGAAGCAGUUCCUCUGGAAGCU
UCUUGAAGACAAACAACGUCUGUAGCGACCCUUUGCAGGCAGCGGAACCCCCCACCUGGCGACAGGUGCCUCUGCGGCCAA
AAGCCACGUGUAUAAGAUACACCUGCAAAGGCGGCACAACCCCAGUGCCACGUUGUGAGUUGGAUAGUUGUGGAAAGAGUC
AAAUGGCUCUCCUCAAGCGUAUUCAACAAGGGGCUGAAGGAUGCCCAGAAGGUACCCCAUUGUAUGGGAUCUGAUCUGGGG
CCUCGGUGCACAUGCUUUACAUGUGUUUAGUCGAGGUUAAAAAAACGUCUAGGCCCCCCGAACCACGGGGACGUGGUUUUC
CUUUGAAAAACACGAUGAUAAUAUGGCCACAACCAUGGUCUUCACACUCGAAGAUUUCGUUGGGGACUGGCGACAGACAGC
CGGCUACAACCUGGACCAAGUCCUUGAACAGGGAGGUGUGUCCAGUUUGUUUCAGAAUCUCGGGGUGUCCGUAACUCCGAU
CCAAAGGAUUGUCCUGAGCGGUGAAAAUGGGCUGAAGAUCGACAUCCAUGUCAUCAUCCCGUAUGAAGGUCUGAGCGGCGA
CCAAAUGGGCCAGAUCGAAAAAAUUUUUAAGGUGGUGUACCCUGUGGAUGAUCAUCACUUUAAGGUGAUCCUGCACUAUGG
CACACUGGUAAUCGACGGGGUUACGCCGAACAUGAUCGACUAUUUCGGACGGCCGUAUGAAGGCAUCGCCGUGUUCGACGG
CAAAAAGAUCACUGUAACAGGGACCCUGUGGAACGGCAACACAAAAUUAUCGACGAGCGCCUGAUCAACCCCGACGGCUCCCU
GCUGUUCCGAGUAACCAUCAACGGAGUGACCGGCUGGCGGCUGUGCGAACGCAUUCUGGCGUAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAA

**Methyl-C modification is highlighted in bold.**

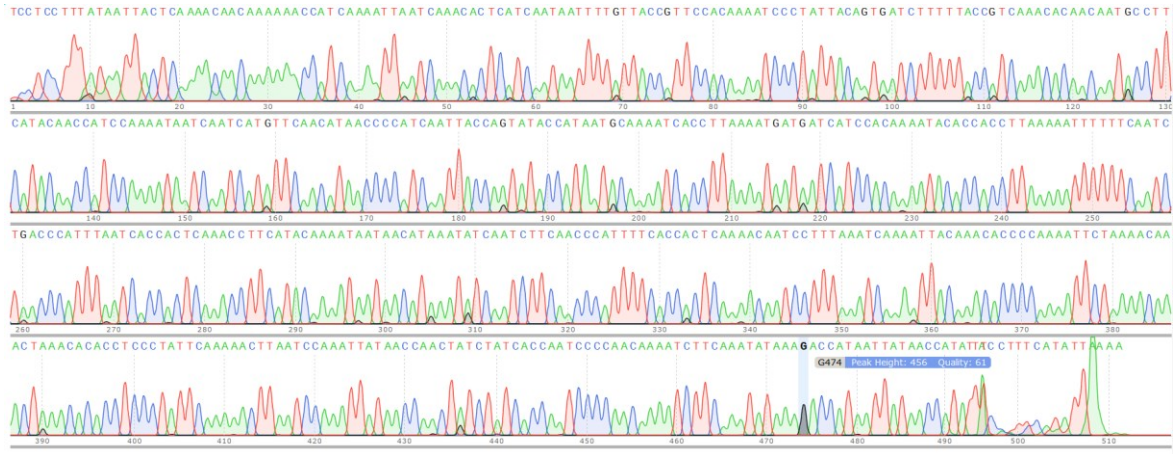



**Figure 147. Full sequencing data from mRNA-init.**

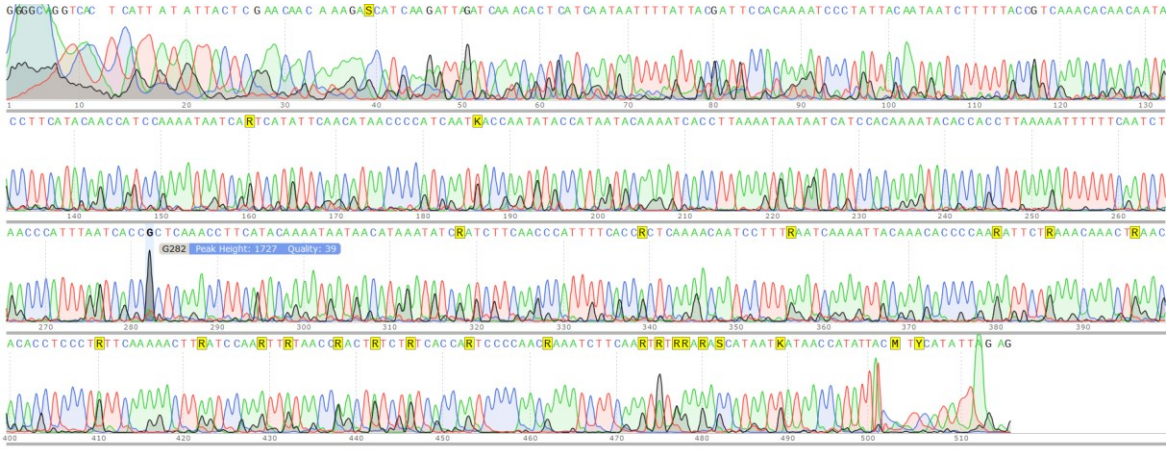


**Figure 148. Full sequencing data from mRNA-mid.**

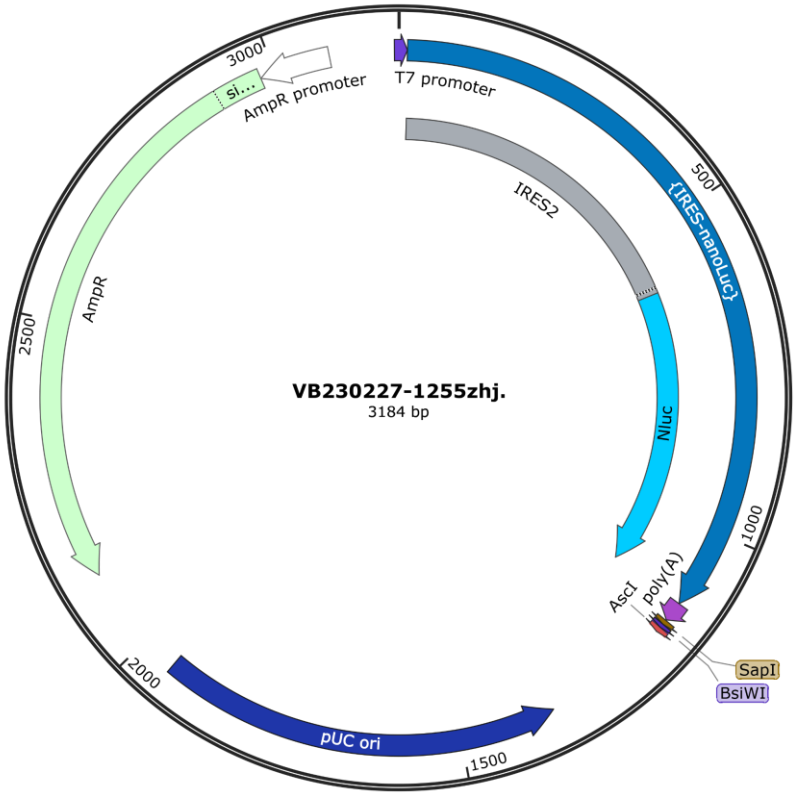



**Figure 149. Plasmid construct for generation of DNA templates for mRNA synthesis.**

# 7 References


[1] R. Dahm, *Hum. Genet.* **2008**, *122*, 565–581.

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

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

[4] F. Gros, H. Hiatt, W. Gilbert, C. G. Kurland, R. W. Risebrough, J. D. Watson, *Nature* **1961**, *190*, 581–585.

[5] A. Rich, D. R. Davies, *J. Am. Chem. Soc.* **1956**, *78*, 3548–3549.

[6] B. Van Beek, M. A. Van Bochove, T. A. Hamlin, F. Matthias Bickelhaupt, *Electron. Struct.* **2019**, *1*, 024001–024012.

[7] C. Cheong, G. Varani, I. Tinoco, *Nature* **1990**, *346*, 680–682.

[8] P. W. Davis, R. W. Adamiak, I. Tinoco, *Biopolymers* **1990**, *29*, 109–122.

[9] S. Arnott, D. W. L. Hukins, S. D. Dover, W. Fuller, A. R. Hodgson, *J. Mol. Biol.* **1973**, *81*, 107–122.

[10] R. Sinha, G. S. Kumar, *J. Phys. Chem. B* **2009**, *113*, 13410–13420.

[11] S. Kumari, A. Bugaut, J. L. Huppert, S. Balasubramanian, *Nat. Chem. Biol. 2007 34* **2007**, *3*, 218–221.

[12] P. J. Hsu, Y. Zhu, H. Ma, Y. Guo, X. Shi, Y. Liu, M. Qi, Z. Lu, H. Shi, J. Wang, Y. Cheng, G. Luo, Q. Dai, M. Liu, X. Guo, J. Sha, B. Shen, C. He, *Cell Res.* **2017**, *27*, 1115–1127.

[13] A. Serganov, D. J. Patel, *Nat. Rev. Genet. 2007 810* **2007**, *8*, 776–790.

[14] S. Van Lint, C. Heirman, K. Thielemans, K. Breckpot, *Hum. Vaccin. Immunother.* **2013**, *9*, 265–274.

[15] P. Odonoghue, J. Ling, D. Söll, *RNA Biol.* **2018**, *15*, 423–426.

[16] Y. Hori, C. Engel, T. Kobayashi, *Nat. Rev. Mol. Cell Biol.* **2023**, *24*, 414–429.

[17] L. Statello, C. J. Guo, L. L. Chen, M. Huarte, *Nat. Rev. Mol. Cell Biol.* **2020**, *22*, 96–118.

[18] A. M. Ardekani, M. M. Naeini, *Avicenna J. Med. Biotechnol.* **2010**, *2*, 161–179.

[19] R. A. Padgett, *eLS* **2015**, 1–8.

[20] Z. hao Huang, Y. ping Du, J. tao Wen, B. feng Lu, Y. Zhao, *Cell Death Discov.* **2022**, *8*, 1–10.

[21] E. Janzen, E. Janzen, C. Blanco, H. Peng, J. Kenchel, J. Kenchel, I. A. Chen, I. A. Chen, I. A. Chen, *Chem. Rev.* **2020**, *120*, 4879–4897.

[22] L. Verduci, E. Tarcitano, S. Strano, Y. Yarden, G. Blandino, *Cell Death Dis.* **2021**, *12*, 1–12.

[23] I. A. Roundtree, M. E. Evans, T. Pan, C. He, *Cell* **2017**, *169*, 1187–1200.

[24] S. H. Boo, Y. K. Kim, *Exp. Mol. Med.* **2020**, *52*, 400–408.

[25] P. Boccaletto, M. A. MacHnicka, E. Purta, P. Pitkowski, B. Baginski, T. K. Wirecki, V. De Crécy-Lagard, R. Ross, P. A. Limbach, A. Kotter, M. Helm, J. M. Bujnicki, *Nucleic Acids Res.* **2018**, *46*, D303–D307.

[26] S. Zaccara, R. J. Ries, S. R. Jaffrey, *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 608–624.

[27] I. A. Roundtree, M. E. Evans, T. Pan, C. He, *Cell* **2017**, *169*, 1187–1200.

[28] F. F. Davis, W. Allen, *J. Biol. Chem.* **1957**, *227*, 907–915.

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

[30] M. Ryczek, M. Pluta, L. Błaszczyk, A. Kiliszek, *Appl. Sci.* **2022**, *12*, 1543–1560.

[31] S. L. Beaucage, C. B. Reese, *Curr. Protoc. Nucleic Acid Chem.* **2009**, *38*, 1–31.

[32] S. L. Beaucage, R. P. Iyer, *Tetrahedron* **1993**, *49*, 6123–6194.

[33] S. L. Beaucage, M. H. Caruthers, *Tetrahedron Lett.* **1981**, *22*, 1859–1862.

[34] S. L. Beaucage, R. P. Iyer, *Tetrahedron* **1992**, *48*, 2223–2311.

[35] M. A. Behlke, *Oligonucleotides* **2008**, *18*, 305–319.

[36] D. Siolas, C. Lerner, J. Burchard, W. Ge, P. S. Linsley, P. J. Paddison, G. J. Hannon, M. A. Cleary, *Nat. Biotechnol.* **2005**, *23*, 227–231.

[37] C. Rinaldi, M. J. A. Wood, *Nat. Rev. Neurol.* **2017**, *14*, 9–21.

[38] M. Flamme, L. K. McKenzie, I. Sarac, M. Hollenstein, *Methods* **2019**, *161*, 64–82.

[39] S. SomozaÀlvaro, *Chem. Soc. Rev.* **2008**, *37*, 2668–2675.

[40] X. Wei, *Tetrahedron* **2013**, *69*, 3615–3637.

[41] S. Ni, H. Yao, L. Wang, J. Lu, F. Jiang, A. Lu, G. Zhang, *Int. J. Mol. Sci.* **2017**, *18*, 1683–1704.

[42] Y. Shiba, H. Masuda, N. Watanabe, T. Ego, K. Takagaki, K. Ishiyama, T. Ohgi, J. Yano, *Nucleic Acids Res.* **2007**, *35*, 3287–3296.

[43] N. Milisavljevič, P. Perlíková, R. Pohl, M. Hocek, *Org. Biomol. Chem.* **2018**, *16*, 5800–5807.

[44] M. Yoshikawa, T. Kato, T. Takenishi, *Bull. Chem. Soc. Jpn.* **1969**, *42*, 3505–3508.

[45] M. Yoshikawa, T. Kato, T. Takenishi, *Tetrahedron Lett.* **1967**, *8*, 5065–5068.

[46] A. R. Kore, M. Shanmugasundaram, A. Senthilvelan, B. Srinivasan, *Nucleosides, Nucleotides and Nucleic Acids* **2012**, *31*, 423–431.

[47] V. Borsenberger, M. Kukwikila, S. Howorka, *Org. Biomol. Chem.* **2009**, *7*, 3826–3835.

[48] J. Ludwig, F. Eckstein, *J. Org. Chem.* **1989**, *54*, 631–635.

[49] J. Caton-Williams, L. Lin, M. Smith, Z. Huang, *Chem. Commun.* **2011**, *47*, 8142–8144.

[50] S. Warnecke, C. Meier, *J. Org. Chem.* **2009**, *74*, 3024–3030.

[51] J. Caton-Williams, L. Lin, M. Smith, Z. Huang, *Chem. Commun.* **2011**, *47*, 8142–8144.

[52] W. Wu, C. L. F. Meyers, R. F. Borch, *Org. Lett.* **2004**, *6*, 2257–2260.

[53] Q. Sun, J. P. Edathil, R. Wu, E. D. Smidansky, C. E. Cameron, B. R. Peterson, *Org. Lett.* **2008**, *10*, 1703–1706.

[54] M. Hocek, *J. Org. Chem.* **2014**, *79*, 9914–9921.

[55] K. H. Shaughnessy, *European J. Org. Chem.* **2006**, *2006*, 1827–1835.

[56] K. H. Shaughnessy, *Molecules* **2015**, *20*, 9419–9454.

[57] R. B. DeVasher, L. R. Moore, K. H. Shaughnessy, *J. Org. Chem.* **2004**, *69*, 7919–7927.

[58] A. L. Casalnuovo, J. C. Calabrese, *J. Am. Chem. Soc.* **1990**, *112*, 4324–4330.

[59] M. Srinivasa Rao, N. Esho, C. Sergeant, R. Dembinski, *J. Org. Chem.* **2003**, *68*, 6788–6790.

[60] F. Mohajer, M. M. Heravi, V. Zadsirjan, N. Poormohammad, *RSC Adv.* **2021**, *11*, 6885–6925.

[61] J. Dadová, P. Vidláková, R. Pohl, L. Havran, M. Fojta, M. Hocek, *J. Org. Chem.* **2013**, *78*, 9627–9637.

[62] D. L. Leone, M. Hubálek, R. Pohl, V. Sýkorová, M. Hocek, *Angew. Chemie Int. Ed.* **2021**, *60*, 17383–17387.

[63] M. Kuba, T. Kraus, R. Pohl, M. Hocek, *Chem. – A Eur. J.* **2020**, *26*, 11950–11954.

[64] D. L. Leone, R. Pohl, M. Hubálek, M. Kadeřábková, M. Krömer, V. Sýkorová, M. Hocek, *Chem. – A Eur. J.* **2022**, *28*, e202104208.

[65] D. Kodr, C. P. Yenice, A. Simonova, D. P. Saftić, R. Pohl, V. Sýkorová, M. Ortiz, L. Havran, M. Fojta, Z. J. Lesnikowski, C. K. O'Sullivan, M. Hocek, *J. Am. Chem. Soc.* **2021**, *143*, 7124–7134.

[66] M. Kuba, P. Khoroshyy, M. Lepšík, E. Kužmová, D. Kodr, T. Kraus, M. Hocek, *Angew. Chemie Int. Ed.* **2023**, *62*, e202307548.

[67] M. Li, N. Lin, Z. Huang, L. Du, C. Altier, H. Fang, B. Wang, *J. Am. Chem. Soc.* **2008**, *130*, 12636–12638.

[68] S. Serdjukow, F. Kink, B. Steigenberger, M. Tomás-Gamasa, T. Carell, *Chem. Commun.* **2014**, *50*, 1861–1863.

[69] K. Bergen, A. L. Steck, S. Strütt, A. Baccaro, W. Welte, K. Diederichs, A. Marx, *J. Am. Chem. Soc.* **2012**, *134*, 11840–11843.

[70] A. Hottin, A. Marx, *Acc. Chem. Res.* **2016**, *49*, 418–427.

[71] A. Hottin, K. Betz, K. Diederichs, A. Marx, *Chem. – A Eur. J.* **2017**, *23*, 2109–2118.

[72] H. Cahová, A. Panattoni, P. Kielkowski, J. Fanfrlík, M. Hocek, *ACS Chem. Biol.* **2016**, *11*, 3165–3171.

[73] P. Kielkowski, J. Fanfrlík, M. Hocek, *Angew. Chemie Int. Ed.* **2014**, *53*, 7552–7555.

[74] S. Borkotoky, C. K. Meena, G. M. Bhalerao, A. Murali, *Sci. Reports 2017 71* **2017**, *7*, 1–12.

[75] M. Chamberlin, J. Mcgrath, L. Waskell, *Nature* **1970**, *228*, 227–231.

[76] J. Yu, G. Oster, *Biophys. J.* **2012**, *102*, 532–541.

[77] B. Beckert, B. Masquida, *Methods Mol. Biol.* **2011**, *703*, 29–41.

[78] T. Conrad, I. Plumbom, M. Alcobendas, R. Vidal, S. Sauer, *Commun. Biol. 2020 31* **2020**, *3*, 1–8.

[79] D. L. Lyakhov, B. He, X. Zhang, F. W. Studier, J. J. Dunn, W. T. McAllister, *J. Mol. Biol.* **1998**, *280*, 201–213.

[80] J. A. Brown, Z. Suo, *Biochemistry* **2011**, *50*, 1135–1142.

[81] J. L. Ong, D. Loakes, S. Jaroslawski, K. Too, P. Holliger, *J. Mol. Biol.* **2006**, *361*, 537–550.

[82] N. Staiger, A. Marx, *ChemBioChem* **2010**, *11*, 1963–1966.

[83] E. O. McCullum, J. C. Chaput, *Chem. Commun.* **2009**, 2938–2940.

[84] G. Xia, L. Chen, T. Sera, M. Fa, P. G. Schultz, F. E. Romesberg, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 6597–6602.

[85] P. H. Patel, L. A. Loeb, *J. Biol. Chem.* **2000**, *275*, 40266–40272.

[86] R. C. Cadwell, G. F. Joyce, *Genome Res.* **1992**, *2*, 28–33.

[87] N. Staiger, A. Marx, *ChemBioChem* **2010**, *11*, 1963–1966.

[88] C. Cozens, V. B. Pinheiro, A. Vaisman, R. Woodgate, P. Holliger, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 8067–8072.

[89] M. R. Dunn, C. Otto, K. E. Fenton, J. C. Chaput, *ACS Chem. Biol.* **2016**, *11*, 1210–1219.

[90] G. Xia, L. Chen, T. Sera, M. Fa, P. G. Schultz, F. E. Romesberg, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 6597–6602.

[91] T. Chen, N. Hongdilokkul, Z. Liu, R. Adhikary, S. S. Tsuen, F. E. Romesberg, *Nat. Chem. 2016 86* **2016**, *8*, 556–562.

[92] T. Chen, F. E. Romesberg, *J. Am. Chem. Soc.* **2017**, *139*, 9949–9954.

[93] T. Chen, F. E. Romesberg, *Biochemistry* **2017**, *56*, 5227–5228.

[94] Q. Shao, T. Chen, K. Sheng, Z. Liu, Z. Zhang, F. E. Romesberg, *J. Am. Chem. Soc.* **2020**, *142*, 2125–2128.

[95] L. Baronti, H. Karlsson, M. Marušič, K. Petzold, *Anal. Bioanal. Chem. 2018 41014* **2018**, *410*, 3239–3252.

[96] J. F. Milligan, D. R. Groebe, G. W. Witherell, O. C. Uhlenbeck, *Nucleic Acids Res.* **1987**, *15*, 8783–8798.

[97] E. Zaychikov, L. Denissova, T. Meier, M. Götte, H. Heumann, *J. Biol. Chem.* **1997**, *272*, 2259–2267.

[98] Z. J. Kartje, H. I. Janis, S. Mukhopadhyay, K. T. Gagnon, *J. Biol. Chem.* **2021**, *296*, 100175–100176.

[99] C. T. Martin, J. E. Coleman, *Biochemistry* **1987**, *26*, 2690–2696.

[100] M. Maslak, C. T. Martin, *Biochemistry* **1994**, *33*, 6918–6924.

[101] J. D. Vaught, T. Dewey, B. E. Eaton, *J. Am. Chem. Soc.* **2004**, *126*, 11231–11237.

[102] C. Kao, M. Zheng, S. Rüdisser, *RNA* **1999**, *5*, 1268–1272.

[103] C. Helmling, S. Keyhani, F. Sochor, B. Fürtig, M. Hengesbach, H. Schwalbe, *J. Biomol. NMR* **2015**, *63*, 67–76.

[104] H. Karlsson, L. Baronti, K. Petzold, *RNA* **2020**, *26*, 1023–1037.

[105] E. A. Doherty, J. A. Doudna, *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *69*, 597–615.

[106] C. E. Weinberg, Z. Weinberg, C. Hammann, *Nucleic Acids Res.* **2019**, *47*, 9480–9494.

[107] S. R. Price, N. Ito, C. Oubridge, J. M. Avis, K. Nagai, *J. Mol. Biol.* **1995**, *249*, 398–408.

[108] Y. L. Ahmed, R. Ficner, *RNA Biol.* **2014**, *11*, 427–432.

[109] P. R. Langer, A. A. Waldrop, D. C. Ward, *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 6633–6637.

[110] J. T. George, S. G. Srivatsan, *Methods* **2017**, *120*, 28–38.

[111] J. T. George, S. G. Srivatsan, *Bioconjug. Chem.* **2017**, *28*, 1529–1536.

[112] M. B. Walunj, A. A. Tanpure, S. G. Srivatsan, *Nucleic Acids Res.* **2018**, *46*, e65.

[113] Y. Zheng, P. A. Beal, *Bioorg. Med. Chem. Lett.* **2016**, *26*, 1799–1802.

[114] A. A. Sawant, P. P. Mukherjee, R. K. Jangid, S. Galande, S. G. Srivatsan, *Org. Biomol. Chem.* **2016**, *14*, 5832–5842.

[115] H. Rao, A. A. Tanpure, A. A. Sawant, S. G. Srivatsan, *Nat. Protoc. 2012 76* **2012**, *7*, 1097–1112.

[116] P. A. Asare-Okai, E. Agustin, D. Fabris, M. Royzen, *Chem. Commun.* **2014**, *50*, 7844–7847.

[117] C. C. Smith, M. Hollenstein, C. J. Leumann, *RSC Adv.* **2014**, *4*, 48228–48235.

[118] M. Liu, H. Jinmei, H. Abe, Y. Ito, *Bioorg. Med. Chem. Lett.* **2010**, *20*, 2964–2967.

[119] N. K. Vaish, A. W. Fraley, J. W. Szostak, L. W. McLaughlin, *Nucleic Acids Res.* **2000**, *28*, 3316–3322.

[120] L. S. McCoy, D. Shin, Y. Tor, *J. Am. Chem. Soc.* **2014**, *136*, 15176–15184.

[121] S. G. Srivatsan, Y. Tor, *J. Am. Chem. Soc.* **2007**, *129*, 2044–2053.

[122] A. A. Tanpure, S. G. Srivatsan, *ChemBioChem* **2014**, *15*, 1309–1316.

[123] C. Domnick, F. Eggert, S. Kath-Schorr, *Chem. Commun.* **2015**, *51*, 8253–8256.

[124] Y. Liu, E. Holmstrom, J. Zhang, P. Yu, J. Wang, M. A. Dyba, D. 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.

[125] T. H. Tahlrov, D. Temiakov, M. Anikin, V. Patlan, W. T. McAllister, D. G. Vassylyev, S.

Yokoyama, *Nature* **2002**, *420*, 43–50.

[126] Y. Sohn, H. Shen, C. Kang, *Methods Enzymol.* **2003**, *371*, 170–179.

[127] M. Ondruš, V. Sýkorová, L. Bednárová, R. Pohl, M. Hocek, *Nucleic Acids Res.* **2020**, *48*, 11982–11993.

[128] D. Kodr, C. P. Yenice, A. Simonova, D. P. Saftić, R. Pohl, V. Sýkorová, M. Ortiz, L. Havran, M. Fojta, Z. J. Lesnikowski, C. K. O'Sullivan, M. Hocek, *J. Am. Chem. Soc.* **2021**, *143*, 7124–7134.

[129] N. Kupriková, M. Ondruš, L. Bednárová, M. Riopedre-Fernandez, L. Poštová Slavětínská, V. Sýkorová, M. Hocek, *Nucleic Acids Res.* **2013**, *1*, 13–14.

[130] T. Hultman, S. Stahl, E. Homes, M. Uhlén, *Nucleic Acids Res.* **1989**, *17*, 4937–4946.

[131] Y. Wakimoto, J. Jiang, H. Wakimoto, *Curr. Protoc. Mol. Biol.* **2014**, *107*, 2.15.1-2.15.9.

[132] M. Yuce, H. Kurt, H. Budak, *Anal. Methods* **2013**, *6*, 548–557.

[133] M. Citartan, T. H. Tang, S. C. Tan, S. C. B. Gopinath, *World J. Microbiol. Biotechnol.* **2011**, *27*, 1167–1173.

[134] L. Civit, A. Fragoso, C. K. O'Sullivan, *Anal. Biochem.* **2012**, *431*, 132–138.

[135] C. Liang, D. Li, G. Zhang, H. Li, N. Shao, Z. Liang, L. Zhang, A. Lu, G. Zhang, *Analyst* **2015**, *140*, 3439–3444.

[136] P. Güixens-Gallardo, M. Hocek, P. Perlíková, *Bioorg. Med. Chem. Lett.* **2016**, *26*, 288–291.

[137] P. Ménová, H. Cahová, M. Plucnara, L. Havran, M. Fojta, M. Hocek, *Chem. Commun.* **2013**, *49*, 4652–4654.

[138] R. K. Saiki, D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, H. A. Erlich, *Science* **1988**, *239*, 487–491.

[139] E. Paredes, M. Evans, S. R. Das, *Methods* **2011**, *54*, 251–259.

[140] K. Krell, D. Harijan, D. Ganz, L. Doll, H. A. Wagenknecht, *Bioconjug. Chem.* **2020**, *31*, 990–1011.

[141] L. Wicke, J. W. Engels, *Bioconjug. Chem.* **2012**, *23*, 627–642.

[142] J. Schoch, S. Ameta, A. Jäschke, *Chem. Commun.* **2011**, *47*, 12536–12537.

[143] A. A. Sawant, A. A. Tanpure, P. P. Mukherjee, S. Athavale, A. Kelkar, S. Galande, S. G. Srivatsan, *Nucleic Acids Res.* **2016**, *44*, e16.

[144] M. L. Winz, A. Samanta, D. Benzinger, A. Jäschke, *Nucleic Acids Res.* **2012**, *40*, e78.

[145] J. Mattay, M. Dittmar, A. Rentmeister, *Curr. Opin. Chem. Biol.* **2021**, *63*, 46–56.

[146] J. M. Holstein, D. Stummer, A. Rentmeister, *Chem. Sci.* **2015**, *6*, 1362–1369.

[147] J. M. Holstein, D. Schulz, A. Rentmeister, *Chem. Commun.* **2014**, *50*, 4478–4481.

[148] D. Schulz, J. M. Holstein, A. Rentmeister, *Angew. Chemie Int. Ed.* **2013**, *52*, 7874–7878.

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

[150] M. R. Stark, J. A. Pleiss, M. Deras, S. A. Scaringe, S. D. Rader, *RNA* **2006**, *12*, 2014–2019.

[151] W. C. Kurschat, J. Müller, R. Wombacher, M. Helm, *RNA* **2005**, *11*, 1909–1914.

[152] J. Hertler, K. Slama, B. Schober, Z. Özrendeci, V. Marchand, Y. Motorin, M. Helm, *Nucleic Acids Res.* **2022**, *50*, e115.

[153] A. Sudakov, B. Knezic, M. Hengesbach, B. Fürtig, E. Stirnal, H. Schwalbe, *Chem. – A Eur. J.* **2023**, *29*, e202203368.

[154] K. Lang, R. Micura, *Nat. Protoc.* **2008**, *3*, 1457–1466.

[155] A. Mamot, P. J. Sikorski, A. Siekierska, P. De Witte, J. Kowalska, J. Jemielity, *Nucleic Acids Res.* **2021**, *50*, e3.

[156] S. Gerstberger, M. Hafner, T. Tuschl, *Nat. Rev. Genet.* **2014**, *15*, 829–845.

[157] A. Castello, B. Fischer, K. Eichelbaum, R. Horos, B. M. Beckmann, C. Strein, N. E. Davey, D. T. Humphreys, T. Preiss, L. M. Steinmetz, J. Krijgsveld, M. W. Hentze, *Cell* **2012**, *149*, 1393–1406.

[158] M. Nechay, R. E. Kleiner, *Curr. Opin. Chem. Biol.* **2020**, *54*, 37–44.

[159] C. A. McHugh, P. Russell, M. Guttman, *Genome Biol.* **2014**, *15*, 1–10.

[160] M. Ramanathan, D. F. Porter, P. A. Khavari, *Nat. Methods 2019 163* **2019**, *16*, 225–234.

[161] J. Ule, K. Jensen, A. Mele, R. B. Darnell, *Methods* **2005**, *37*, 376–386.

[162] M. Hafner, M. Katsantoni, T. Köster, J. Marks, J. Mukherjee, D. Staiger, J. Ule, M. Zavolan, *Nat. Rev. Methods Prim.* **2021**, *1*, 1–23.

[163] D. D. Licatalosi, A. Mele, J. J. Fak, J. Ule, M. Kayikci, S. W. Chi, T. A. Clark, A. C. Schweitzer, J. E. Blume, X. Wang, J. C. Darnell, R. B. Darnell, *Nature* **2008**, *456*, 464–469.

[164] E. L. Van Nostrand, G. A. Pratt, A. A. Shishkin, C. Gelboin-Burkhart, M. Y. Fang, B. Sundararaman, S. M. Blue, T. B. Nguyen, C. Surka, K. Elkins, R. Stanton, F. Rigo, M. Guttman, G. W. Yeo, *Nat. Methods* **2016**, *13*, 508–514.

[165] J. König, K. Zarnack, G. Rot, T. Curk, M. Kayikci, B. Zupan, D. J. Turner, N. M. Luscombe, J. Ule, *Nat. Struct. Mol. Biol. 2010 177* **2010**, *17*, 909–915.

[166] M. Hafner, M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano, A. C. Jungkamp, M. Munschauer, A. Ulrich, G. S. Wardle, S. Dewell, M. Zavolan, T. Tuschl, *Cell* **2010**, *141*, 129–141.

[167] C. Gräwe, S. Stelloo, F. A. H. van Hout, M. Vermeulen, *Trends Biotechnol.* **2021**, *39*, 890–900.

[168] J. I. Perez-Perri, B. Rogell, T. Schwarzl, F. Stein, Y. Zhou, M. Rettel, A. Brosig, M. W.

Hentze, *Nat. Commun. 2018 91* **2018**, *9*, 1–13.

[169] A. G. Baltz, M. Munschauer, B. Schwanhäusser, A. Vasile, Y. Murakawa, M. Schueler, N. Youngs, D. Penfold-Brown, K. Drew, M. Milek, E. Wyler, R. Bonneau, M. Selbach, C. Dieterich, M. Landthaler, *Mol. Cell* **2012**, *46*, 674–690.

[170] R. M. L. Queiroz, T. Smith, E. Villanueva, M. Marti-Solano, M. Monti, M. Pizzinga, D. M. Mirea, M. Ramakrishna, R. F. Harvey, V. Dezi, G. H. Thomas, A. E. Willis, K. S. Lilley, *Nat. Biotechnol.* **2019**, *37*, 169–178.

[171] E. C. Urdaneta, C. H. Vieira-Vieira, T. Hick, H. H. Wessels, D. Figini, R. Moschall, J. Medenbach, U. Ohler, S. Granneman, M. Selbach, B. M. Beckmann, *Nat. Commun.* **2019**, *10*, 1–17.

[172] V. Shchepachev, S. Bresson, C. Spanos, E. Petfalski, L. Fischer, J. Rappsilber, D. Tollervey, *Mol. Syst. Biol.* **2019**, *15*, e8689.

[173] J. Trendel, T. Schwarzl, R. Horos, A. Prakash, A. Bateman, M. W. Hentze, J. Krijgsveld, *Cell* **2019**, *176*, 391–403.

[174] X. Zheng, S. Cho, H. Moon, T. J. Loh, H. N. Jang, H. Shen, *Methods Mol. Biol.* **2016**, *1421*, 35–44.

[175] K. Leppek, G. Stoecklin, *Nucleic Acids Res.* **2014**, *42*, e13.

[176] J. R. Hogg, K. Collins, *RNA* **2007**, *13*, 868–880.

[177] W. A. Velema, Z. Lu, *JACS Au* **2023**, *3*, 316–332.

[178] J. E. Hearst, *Chem. Res. Toxicol.* **1989**, *2*, 69–75.

[179] T. W. Nilsen, *Cold Spring Harb. Protoc.* **2014**, *9*, 996–1000.

[180] U. Zaman, F. M. Richter, R. Hofele, K. Kramer, T. Sachsenberg, O. Kohlbacher, C. Lenz, H. Urlaub, *Mol. Cell. Proteomics* **2015**, *14*, 3196–3210.

[181] E. A. Hoffman, B. L. Frey, L. M. Smith, D. T. Auble, *J. Biol. Chem.* **2015**, *290*, 26404–26411.

[182] K. Möller, J. Rinke, A. Ross, G. Buddle, R. Brimacombe, *Eur. J. Biochem.* **1977**, *76*, 175–187.

[183] A. Expert-Bezançon, C. Chiaruttini, *Methods Enzymol.* **1988**, *164*, 310–318.

[184] M. Buisson, A. M. Reboud, *FEBS Lett.* **1982**, *148*, 247–250.

[185] M. Ascano, M. Hafner, P. Cekan, S. Gerstberger, T. Tuschl, *Wiley Interdiscip. Rev. RNA* **2012**, *3*, 159–177.

[186] K. M. Meisenheimer, P. L. Meisenheimer, M. C. Willis, T. H. Koch, *Nucleic Acids Res.* **1996**, *24*, 981–982.

[187] M. C. Willis, B. J. Hicke, O. C. Uhlenbeck, T. R. Cech, T. H. Koch, *Science (80-. ).* **1993**, *262*, 1255–1257.

[188] K. M. Meisenheimer, P. L. Meisenheimer, T. H. Koch, *Methods Enzymol.* **2000**, *318*, 88–104.

[189] J. M. Gott, M. C. Willis, T. H. Koch, O. C. Uhlenbeck, *Biochemistry* **1991**, *30*, 6290–6295.

[190] A. Favre, C. Saintomé, J. L. Fourrey, P. Clivio, P. Laugâa, *J. Photochem. Photobiol. B Biol.* **1998**, *42*, 109–124.

[191] M. E. Harris, E. L. Christian, *Methods Enzymol.* **2009**, *468*, 127–146.

[192] G. W. Preston, A. J. Wilson, *Chem. Soc. Rev.* **2013**, *42*, 3289–3301.

[193] J. Sumranjit, S. J. Chung, *Molecules* **2013**, *18*, 10425–10451.

[194] F. Muttach, F. Mäsing, A. Studer, A. Rentmeister, *Chem. – A Eur. J.* **2017**, *23*, 5988–5993.

[195] K. L. Buchmueller, B. T. Hill, M. S. Platz, K. M. Weeks, **2003**, 10850–10861.

[196] C. Costas, E. Yuriev, K. L. Meyer, T. S. Guion, M. M. Hanna, *Nucleic Acids Res.* **2000**, *28*, 1849–1858.

[197] Z. Zhang, J. Lin, Z. Liu, G. Tian, X. M. Li, Y. Jing, X. Li, X. D. Li, *J. Am. Chem. Soc.* **2022**, *144*, 20979–20997.

[198] A. E. Arguello, A. N. Deliberto, R. E. Kleiner, *J. Am. Chem. Soc.* **2017**, *139*, 17249–17252.

[199] H. S. Jeong, G. Hayashi, A. Okamoto, *ACS Chem. Biol.* **2015**, *10*, 1450–1455.

[200] K. Nakamoto, Y. Ueno, *J. Org. Chem.* **2014**, *79*, 2463–2472.

[201] C. D. Spicer, B. G. Davis, *Nat. Commun. 2014 51* **2014**, *5*, 1–14.

[202] D. L. Leone, R. Pohl, M. Hubálek, M. Kadeřábková, M. Krömer, V. Sýkorová, M. Hocek, *Chem. – A Eur. J.* **2022**, *28*, e202104208.

[203] D. L. Leone, M. Hubálek, R. Pohl, V. Sýkorová, M. Hocek, *Angew. Chemie Int. Ed.* **2021**, *60*, 17383–17387.

[204] M. Krömer, M. Brunderová, I. Ivancová, L. Poštová Slavětínská, M. Hocek, *Chempluschem* **2020**, *85*, 1164–1170.

[205] I. Ivancová, R. Pohl, M. Hubálek, M. Hocek, *Angew. Chemie Int. Ed.* **2019**, *58*, 13345–13348.

[206] A. Olszewska, R. Pohl, M. Brázdová, M. Fojta, M. Hocek, *Bioconjug. Chem.* **2016**, *27*, 2089–2094.

[207] J. Dadová, P. Orság, R. Pohl, M. Brázdová, M. Fojta, M. Hocek, *Angew. Chemie Int. Ed.* **2013**, *52*, 10515–10518.

[208] J. Dadová, M. Vrábel, M. Adámik, M. Brázdová, R. Pohl, M. Fojta, M. Hocek, *Chem. – A Eur. J.* **2015**, *21*, 16091–16102.

[209] M. Hocek, *Acc. Chem. Res.* **2019**, *52*, 1730–1737.

[210] I. Ivancová, D. L. Leone, M. Hocek, *Curr. Opin. Chem. Biol.* **2019**, *52*, 136–144.

[211] S. Ji, I. Fu, S. Naldiga, H. Shao, A. K. Basu, S. Broyde, N. Y. Tretyakova, *Nucleic Acids Res.* **2018**, *46*, 6455–6469.

[212] M. Fonvielle, N. Sakkas, L. Iannazzo, C. Le Fournis, D. Patin, D. Mengin-Lecreulx, A. El-Sagheer, E. Braud, S. Cardon, T. Brown, M. Arthur, M. Etheve-Quelquejeu, *Angew. Chemie Int. Ed.* **2016**, *55*, 13553–13557.

[213] W. Dai, A. Li, N. J. Yu, T. Nguyen, R. W. Leach, M. Wühr, R. E. Kleiner, *Nat. Chem. Biol.* **2021**, *17*, 1178–1187.

[214] V. Khoddami, B. R. Cairns, *Nat. Biotechnol.* **2013**, *31*, 458–464.

[215] T. C. Roberts, R. Langer, M. J. A. Wood, *Nat. Rev. Drug Discov.* **2020**, *19*, 673–694.

[216] Y. Kawamoto, Y. Wu, Y. Takahashi, Y. Takakura, *Adv. Drug Deliv. Rev.* **2023**, *199*, 114872–114921.

[217] A. M. Quemener, L. Bachelot, A. Forestier, E. Donnou-Fournet, D. Gilot, M. D. Galibert, *Wiley Interdiscip. Rev. RNA* **2020**, *11*, e1594.

[218] X. Shen, D. R. Corey, *Nucleic Acids Res.* **2018**, *46*, 1584–1600.

[219] T. Yoshida, K. Morihiro, Y. Naito, A. Mikami, Y. Kasahara, T. Inoue, S. Obika, *Nucleic Acids Res.* **2022**, *50*, 7224–7234.

[220] B. Hu, L. Zhong, Y. Weng, L. Peng, Y. Huang, Y. Zhao, X. J. Liang, *Signal Transduct. Target. Ther. 2020 51* **2020**, *5*, 1–25.

[221] H. Peacock, A. Kannan, P. A. Beal, C. J. Burrows, *J. Org. Chem.* **2011**, *76*, 7295–7300.

[222] S. R. Suter, J. Sheu-Gruttadauria, N. T. Schirle, R. Valenzuela, A. A. Ball-Jones, K. Onizuka, I. J. Macrae, P. A. Beal, *J. Am. Chem. Soc.* **2016**, *138*, 8667–8669.

[223] W. Zhang, Q. X. Liu, Z. H. Guo, J. S. Lin, *Molecules* **2018**, *23*, 344–369.

[224] G. Mayer, G. Mayer, *Angew. Chemie Int. Ed.* **2009**, *48*, 2672–2689.

[225] P. Bayat, R. Nosrati, M. Alibolandi, H. Rafatpanah, K. Abnous, M. Khedri, M. Ramezani, *Biochimie* **2018**, *154*, 132–155.

[226] N. Razmi, B. Baradaran, M. Hejazi, M. Hasanzadeh, J. Mosafer, A. Mokhtarzadeh, M. de la Guardia, *Biosens. Bioelectron.* **2018**, *113*, 58–71.

[227] Y. Seok Kim, N. H. Ahmad Raston, M. Bock Gu, *Biosens. Bioelectron.* **2016**, *76*, 2–19.

[228] E. Bagheri, K. Abnous, M. Alibolandi, M. Ramezani, S. M. Taghdisi, *Biosens. Bioelectron.* **2018**, *111*, 1–9.

[229] C. Tuerk, L. Gold, *Science (80-. ).* **1990**, *249*, 505–510.

[230] A. D. Ellington, J. W. Szostak, *Nature* **1990**, *346*, 818–822.

[231] Z. Zhuo, Y. Yu, M. Wang, J. Li, Z. Zhang, J. Liu, X. Wu, A. Lu, G. Zhang, B. Zhang, *Int. J. Mol. Sci.* **2017**, *18*, 2142–2161.

[232] P. E. Burmeister, S. D. Lewis, R. F. Silva, J. R. Preiss, L. R. Horwitz, P. S. Pendergrast, T. G. McCauley, J. C. Kurz, D. M. Epstein, C. Wilson, A. D. Keefe, *Chem. Biol.* **2005**, *12*, 25–33.

[233] J. Ruckman, L. S. Green, J. Beeson, S. Waugh, W. L. Gillette, D. D. Henninger, L. Claesson-Welsh, N. Janjić, *J. Biol. Chem.* **1998**, *273*, 20556–20567.

[234] L. S. Green, D. Jellinek, C. Bell, L. A. Beebe, B. D. Feistner, S. C. Gill, F. M. Jucker, N. Janjić, *Chem. Biol.* **1995**, *2*, 683–695.

[235] E. W. M. Ng, D. T. Shima, P. Calias, E. T. Cunningham, D. R. Guyer, A. P. Adamis, *Nat. Rev. Drug Discov.* **2006**, *5*, 123–132.

[236] J. P. Vavalle, M. G. Cohen, *Future Cardiol.* **2012**, *8*, 371–382.

[237] A. Amundarain, F. Pastor, F. Prósper, X. Agirre, *Cancers (Basel).* **2022**, *14*, 5471–5484.

[238] C. Maasch, K. Buchner, D. Eulberg, S. Vonhoff, S. Klussmann, *Nucleic Acids Symp. Ser.* **2008**, *52*, 61–62.

[239] M. A. Williams, G. J. Mckay, U. Chakravarthy, *Cochrane Database Syst. Rev.* **2014**, *1*, CD009300.

[240] F. Schwoebel, L. T. Van Eijk, D. Zboralski, S. Sell, K. Buchner, C. Maasch, W. G. Purschke, M. Humphrey, S. Zöllner, D. Eulberg, F. Morich, P. Pickkers, S. Klussmann, *Blood* **2013**, *121*, 2311–2315.

[241] E. K. Waters, R. M. Genga, M. C. Schwartz, J. A. Nelson, R. G. Schaub, K. A. Olson, J. C. Kurz, K. E. McGinness, *Blood* **2011**, *117*, 5514–5522.

[242] A. D. Keefe, S. Pai, A. Ellington, *Nat. Rev. Drug Discov.* **2010**, *9*, 537–550.

[243] X. Ni, M. Castanares, A. Mukherjee, S. E. Lupold, *Curr. Med. Chem.* **2011**, *18*, 4206–4214.

[244] J. Zhou, J. Rossi, *Nat. Rev. Drug Discov.* **2016**, *16*, 181–202.

[245] A. B. Vogel, I. Kanevsky, Y. Che, K. A. Swanson, A. Muik, M. Vormehr, L. M. Kranz, K. C. Walzer, S. Hein, A. Güler, J. Loschko, M. S. Maddur, A. Ota-Setlik, K. Tompkins, J. Cole, B. G. Lui, T. Ziegenhals, A. Plaschke, D. Eisel, S. C. Dany, S. Fesser, S. Erbar, F. Bates, D. Schneider, B. Jesionek, B. Sänger, A. K. Wallisch, Y. Feuchter, H. Junginger, S. A. Krumm, A. P. Heinen, P. Adams-Quack, J. Schlereth, S. Schille, C. Kröner, R. de la Caridad Güimil Garcia, T. Hiller, L. Fischer, R. S. Sellers, S. Choudhary, O. Gonzalez, F. Vascotto, M. R. Gutman, J. A. Fontenot, S. Hall-Ursone, K. Brasky, M. C. Griffor, S. Han, A. A. H. Su, J. A. Lees, N. L. Nedoma, E. H. Mashalidis, P. V. Sahasrabudhe, C. Y. Tan, D. Pavliakova, G. Singh, C. Fontes-Garfias, M. Pride, I. L. Scully, T. Ciolino, J. Obregon, M. Gazi, R. Carrion, K. J. Alfson, W. V. Kalina, D. Kaushal, P. Y. Shi, T. Klamp, C. Rosenbaum, A. N. Kuhn, Ö. Türeci, P. R. Dormitzer, K. U. Jansen, U. Sahin, *Nature* **2021**, *592*, 283–289.

[246] K. Karikó, H. Muramatsu, F. A. Welsh, J. Ludwig, H. Kato, S. Akira, D. Weissman, *Mol. Ther.* **2008**, *16*, 1833–1840.

[247] X. Hou, T. Zaks, R. Langer, Y. Dong, *Nat. Rev. Mater.* **2021**, *6*, 1078–1094.

[248] E. Rohner, R. Yang, K. S. Foo, A. Goedel, K. R. Chien, *Nat. Biotechnol.* **2022**, *40*, 1586–1600.

[249] D. D. Kang, H. Li, Y. Dong, *Adv. Drug Deliv. Rev.* **2023**, *199*, 114961–114972.

[250] K. Karikó, H. Muramatsu, J. Ludwig, D. Weissman, *Nucleic Acids Res.* **2011**, *39*, e142.

[251] R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, P. Horvath, *Science* **2007**, *315*, 1709–1712.

[252] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, *Science* **2012**, *337*, 816–821.

[253] P. Mali, K. M. Esvelt, G. M. Church, *Nat. Methods 2013 1010* **2013**, *10*, 957–963.

[254] A. Latorre, A. Latorre, Á. Somoza, *Angew. Chem. Int. Ed. Engl.* **2016**, *55*, 3548–3550.

[255] A. Hendel, R. O. Bak, J. T. Clark, A. B. Kennedy, D. E. Ryan, S. Roy, I. Steinfeld, B. D. Lunstad, R. J. Kaiser, A. B. Wilkens, R. Bacchetta, A. Tsalenko, D. Dellinger, L. Bruhn, M. H. Porteus, *Nat. Biotechnol.* **2015**, *33*, 985–989.

[256] D. O'Reilly, Z. J. Kartje, E. A. Ageely, E. Malek-Adamian, M. Habibian, A. Schofield, C. L. Barkau, K. J. Rohilla, L. B. Derossett, A. T. Weigle, M. J. Damha, K. T. Gagnon, *Nucleic Acids Res.* **2019**, *47*, 546–558.

[257] Y. K. Kim, *Exp. Mol. Med.* **2022**, *54*, 455–465.

[258] M. W. Hentze, A. Castello, T. Schwarzl, T. Preiss, *Nat. Rev. Mol. Cell Biol. 2018 195* **2018**, *19*, 327–341.

[259] J. W. Bae, S. C. Kwon, Y. Na, V. N. Kim, J. S. Kim, *Nat. Struct. Mol. Biol.* **2020**, *27*, 678–682.

[260] A. Bourderioux, P. Nauš, P. Perlíková, R. Pohl, I. Pichová, I. Votruba, P. Džubák, P. Konečný, M. Hajdúch, K. M. Stray, T. Wang, A. S. Ray, J. Y. Feng, G. Birkus, T. Cihlar, M. Hocek, *J. Med. Chem.* **2011**, *54*, 5498–5507.

[261] H. Tian, Y. Xu, S. Liu, D. Jin, J. Zhang, L. Duan, W. Tan, *Molecules* **2017**, *22*, 694–704.

[262] R. Sousa, Y. J. Chung, J. P. Rose, B. C. Wang, *Nature* **1993**, *364*, 593–599.

[263] N. T. Schirle, I. J. MacRae, *Science (80-. ).* **2012**, *336*, 1037–1040.

[264] W. J. Ma, S. Cheng, C. Campbell, A. Wright, H. Furneaux, *J. Biol. Chem.* **1996**, *271*, 8144–8151.

[265] S. G. Sarafianos, B. Marchand, K. Das, D. M. Himmel, M. A. Parniak, S. H. Hughes, E. Arnold, *J. Mol. Biol.* **2009**, *385*, 693–713.

[266] F. Liu, J. Z. H. Zhang, Y. Mei, *Sci. Rep.* **2016**, *6*, 1–11.

[267] M. Majumder, P. Chakraborty, S. Mohan, S. Mehrotra, V. Palanisamy, *Adv. Drug Deliv. Rev.* **2022**, *188*, 114442–114454.

[268] C. Barnes, A. Kanhere, *Methods Mol. Biol.* **2016**, *1480*, 99–113.

[269] E. J. Fialcowitz-White, B. Y. Brewer, J. D. Ballin, C. D. Willis, E. A. Toth, G. M. Wilson, *J. Biol. Chem.* **2007**, *282*, 20948–20959.

[270] R. R. Breaker, *Nature* **2004**, *432*, 838–845.

[271] F. Wachowius, C. Höbartner, *ChemBioChem* **2010**, *11*, 469–480.

[272] A. Ovcharenko, F. P. Weissenboeck, A. Rentmeister, *Angew. Chemie Int. Ed.* **2021**, *60*, 4098–4103.

[273] Y. Liu, P. Yu, M. Dyba, R. Sousa, J. R. Stagno, Y. X. Wang, *Methods* **2016**, *103*, 4–10.

[274] S. A. N. McElhinny, D. Kumar, A. B. Clark, D. L. Watt, B. E. Watts, E. B. Lundström, E. Johansson, A. Chabes, T. A. Kunkel, *Nat. Chem. Biol.* **2010**, *6*, 774–781.

[275] P. Güixens-Gallardo, Z. Zawada, J. Matyašovský, D. Dziuba, R. Pohl, T. Kraus, M. Hocek, *Bioconjug. Chem.* **2018**, *29*, 3906–3912.

[276] T. Pesnot, L. M. Tedaldi, P. G. Jambrina, E. Rosta, G. K. Wagner, *Org. Biomol. Chem.* **2013**, *11*, 6357–6371.

[277] K. Hölz, A. Pavlic, J. Lietard, M. M. Somoza, *Sci. Reports 2019 91* **2019**, *9*, 1–12.

[278] P. J. Mchugh, J. Knowland, *Nucleic Acids Res.* **1995**, *23*, 1664–1670.

[279] J. F. Lemay, J. C. Penedo, R. Tremblay, D. M. J. Lilley, D. A. A. Lafontaine, *Chem. Biol.* **2006**, *13*, 857–868.

[280] P. A. Dalgarno, J. Bordello, R. Morris, P. St-Pierre, A. Dubé, I. D. W. Samuel, D. A. Lafontaine, J. Carlos Penedo, *Nucleic Acids Res.* **2013**, *41*, 4253–4265.

[281] R. Marques, R. Lacerda, L. Romão, *Biomedicines* **2022**, *10*, 1865–1895.

[282] Y. Yang, Z. Wang, *J. Mol. Cell Biol.* **2019**, *11*, 911–919.

[283] J. S. Kieft, *Trends Biochem. Sci.* **2008**, *33*, 274–283.

[284] C. G. England, E. B. Ehlerding, W. Cai, *Bioconjug. Chem.* **2016**, *27*, 1175–1187.

[285] Y. Li, T. O. Tollefsbol, *Methods Mol. Biol.* **2011**, *791*, 11–21.

[286] M. Schaefer, T. Pollex, K. Hanna, F. Lyko, *Nucleic Acids Res.* **2009**, *37*, e12.

[287] A. Krause, A. Hertl, F. Muttach, A. Jäschke, *Chem. – A Eur. J.* **2014**, *20*, 16613–16619.