

# **Artificial nucleic acids with enhanced functionalities**

## **Inaugural-Dissertation**

zur Erlangung des Doktorgrades (Dr. rer. nat.)  
der Mathematisch-Naturwissenschaftlichen Fakultät  
der Universität zu Köln

vorgelegt von  
Hannah Depmeier  
aus Velbert

Köln  
2024



Berichtersteller:

*Prof. in Dr. Stephanie Kath-Schorr*

*Prof. Dr. Ulrich Baumann*

Tag der mündlichen Prüfung:

9. September 2024





Parts of this thesis have been published in the *Journal of American Chemical Society*:

Depmeier, H.; Kath-Schorr, S., Expanding the Horizon of the Xeno Nucleic Acid Space: Threose Nucleic Acids with Increased Information Storage. *J. Am. Chem. Soc.* **2024**, *146* (11), 7743–7751. DOI: 10.1021/jacs.3c14626.



*Everything is possible.*

*The impossible just takes longer.*

Dan Brown (Digital Fortress)



## Abstracts

This thesis examines a diverse range of chemically modified nucleic acids for various different applications. In total, four research projects were pursued, all of which share a common theme in the functionalization of nucleic acids to enhance their functionalities.

To date, a plethora of artificial nucleic acid analogs with enhanced functional properties have been described in the literature. However, most of the currently described artificial nucleic acids comprise a modification in only one structural element of the tripartite structure of nucleic acids, which is comprised of a phosphate backbone, a sugar moiety, and a nucleobase. This work explores the enzymatic synthesis of a novel artificial nucleic acid variant with modifications in the sugar and the nucleobase moieties. Therefore, the nuclease-resistant  $\alpha$ -L-threofuranosyl nucleic acid (TNA) was combined with the unnatural hydrophobic base pair **TPT3:NaM**. In this context, the  $\alpha$ -L-threofuranosyl **NaM** triphosphate and the TNA nucleotides bearing the canonical nucleobases were synthesized. The enzymatic incorporation of the highly modified nucleotide building blocks bearing the unnatural **TPT3** and **NaM** nucleobases into TNA was investigated using different TNA-compatible polymerases. The enzymatic synthesis of **TPT3**-modified TNA was successfully achieved, whereas the preparation of **NaM**-modified TNA presented greater challenges. This study marks the first enzymatic synthesis of TNA with an expanded genetic alphabet (exTNA), opening promising opportunities in nucleic acid therapeutics, particularly for the selection of nuclease-resistant, high-affinity aptamers.

The most prevalent obstacle that compromises the fidelity of enzymatic TNA synthesis is the occurrence of Hoogsteen base pairs between two guanine nucleobases within the enzyme's active site. The current solution is to utilize 7-deaza-7-modified threoguanosine triphosphates during enzymatic TNA synthesis, which have the Hoogsteen base pairing site sterically blocked. However, this approach results in modified TNA polymers, which are not always desired. In this study, a photocleavable, traceless approach based on a photolabile *N*<sup>7</sup>-functionalization of guanine was devised to reduce the propensity of Hoogsteen base pairing. This approach allows for the removal of the functionalization post-enzymatic preparation through irradiation with light, thereby restoring the native TNA. In this context, the synthesis of a *N*<sup>7</sup>-benzophenone-photocaged threoguanosine was achieved, and thus significant synthetic progress towards the desired benzophenone-photocaged threoguanosine triphosphate was made.

Click-SELEX (systematic evolution of ligands by exponential enrichment) is a valuable method based on the copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC) for the selection of therapeutic and diagnostic DNA aptamers with enhanced chemical diversity. However, a

recently discovered obstacle during click-SELEX is the three-dimensional folding of the selected clickmers in dependence of residual copper ions. This results in a non-native folding of the clickmer, which restricts its applicability. The development of a click-SELEX approach based on the inverse electron-demand Diels-Alder (IEDDA) reaction would provide copper-free reaction conditions, circumventing the aforementioned problem. Synthetic efforts to enable an IEDDA-based click-SELEX were performed. A modified deoxyadenosine triphosphate was synthesized, comprising a reactive cyclopropene moiety that is attached to the nucleobase via a carbamate linker. The incorporation of the cyclopropene-modified nucleotide into DNA and the downstream functionalization via IEDDA using a tetrazine derivative were successfully demonstrated. However, the current approach is constrained by the instability of the modification, as evidenced by the decomposition of the cyclopropene-carbamate linker during storage. Overcoming the observed instabilities of the cyclopropene-carbamate linker will potentially enable a copper-free click-SELEX method.

The Z-conformation of RNA is a rare and transient structure that occurs predominantly in guanine (**G**)- and cytosine (**C**)-rich sequences. Z-RNA is a ligand for the Z-nucleic acid binding protein 1 (ZBP1), which contains two tandem Z-nucleic acid binding domains, Z $\alpha$ 1 and Z $\alpha$ 2. Binding of Z-RNA to ZBP1 initiates the ZBP1-mediated necroptosis pathway. *In vivo* knock-out studies indicate that the Z $\alpha$ 2 domain is the primary driver for triggering necroptosis. Consequently, the functional significance of two Z $\alpha$  domains in tandem orientation remains elusive. Therefore, the functional significance of the two tandem Z $\alpha$  domains for ligand binding was investigated in a simplified *in vitro* system. Initially, the hypothesis was tested whether two tandem Z $\alpha$  domains have the function to also bind double-stranded A-RNA and convert it into the Z-conformation. Therefore, the wild-type murine ZBP1 was successfully expressed and purified. A **GC**-rich hairpin RNA with 8-methylguanosine modifications was utilized to stabilize the Z-conformation under close-to-native conditions *in vitro*. An electrophoretic mobility shift assay demonstrated that wild-type ZBP1 binds specifically to Z-RNA, but not to A-RNA, indicating that the tandem Z $\alpha$  domains in ZBP1 do not induce A-to-Z transition of RNA. To further elucidate the function of the individual Z $\alpha$  domains for ligand binding, the binding affinity of each Z $\alpha$  domain to Z-RNA was planned to be investigated in an isolated manner. This was aspired to determine whether two tandem Z $\alpha$  domains bind more strongly to Z-RNA than one domain and whether mutation of one Z $\alpha$  domain impedes binding of the Z-RNA ligand. Therefore, ZBP1 mutants with loss-of-function mutations in either the Z $\alpha$ 1 domain, the Z $\alpha$ 2 domain, or both Z $\alpha$  domains were successfully expressed and purified. A comparative *in vitro* interaction study between Z-RNA and these ZBP1 mutants is a subject for future investigation.

## Zusammenfassung

Funktionalisierte Nukleinsäureanaloga können für eine Vielzahl von Zwecken und Anwendungen entwickelt und eingesetzt werden. Diese Arbeit umfasst vier Forschungsprojekte, die alle auf die chemische Funktionalisierung von Nukleinsäuren zurückgreifen.

In der Literatur wurde bereits eine Vielzahl von künstlichen Nukleinsäureanaloga mit verbesserten funktionellen Eigenschaften beschrieben. Die meisten dieser künstlichen Nukleinsäuren weisen jedoch nur eine Modifikation in der Nukleinsäurestruktur auf, also entweder im Phosphat-Rückgrat, dem Zucker oder der Nukleobase. In dieser Arbeit wurde die enzymatische Synthese einer neuen künstlichen Nukleinsäure mit Modifikationen am Zucker und der Nukleobase untersucht. Dazu wurde die nuklease-resistente  $\alpha$ -L-Threonukleinsäure (TNA) mit dem unnatürlichen hydrophoben Basenpaar **TPT3:NaM** kombiniert. In diesem Zusammenhang wurden das  $\alpha$ -L-Threofuranosyl-**NaM**-Triphosphat und die TNA-Nukleotide mit den natürlichen Nukleobasen synthetisiert. Der enzymatische Einbau der stark modifizierten TNA Nukleotidbausteine mit den unnatürlichen Nukleobasen **TPT3** und **NaM** wurde mit verschiedenen TNA-kompatiblen Polymerasen untersucht. Die enzymatische Synthese von **TPT3**-modifizierter TNA war erfolgreich, während der Einbau der **NaM** Nukleobase in die TNA frühzeitig zum Abbruch führte. Diese Studie präsentiert die erste enzymatische Synthese von TNA mit einem erweiterten genetischen Alphabet (exTNA), was vielversprechende Möglichkeiten für Nukleinsäure-Therapeutika eröffnet, insbesondere für die Selektion von nuklease-resistenten, hochaffinen Aptameren.

Der häufigste Grund, der die Genauigkeit der enzymatischen TNA-Synthese beeinflusst ist das Auftreten von Hoogsteen-Basenpaaren zwischen zwei Guanin-Nukleobasen im aktiven Zentrum der Polymerase. Die derzeitige Lösung besteht in der Verwendung von 7-Deaza-7-modifizierten Threoguanosintriphosphaten, bei denen die Hoogsteen-Basenpaarungsstelle sterisch blockiert ist. Allerdings werden dadurch modifizierte TNA-Polymere hergestellt, was nicht in jedem experimentellen Kontext erwünscht ist. In dieser Studie wurde daher ein Ansatz entwickelt, der auf einer photolabilen  $N^7$ -Funktionalisierung von Guanin basiert, um die Neigung zur Hoogsteen-Basenpaarung zu reduzieren. Dieser Ansatz ermöglicht es, die Funktionalisierung nach der enzymatischen Herstellung mittels Licht zu entfernen, wodurch native TNA ohne Modifikationen am Guanin erhalten wird. Im Rahmen dieser Studie wurde ein  $N^7$ -Benzophenon-modifiziertes Threoguanosin synthetisiert, was einen bedeutenden synthetischen Fortschritt in Richtung des gewünschten Benzophenon-modifizierten Threoguanosintriphosphats darstellt.

Click-SELEX (engl.: Systematic Evolution of Ligands by Exponential Enrichment) ist eine Methode die mittels Kupfer(I)-katalysierter Azid-Alkin-Cycloaddition (CuAAC) die Selektion von DNA-Aptameren mit erhöhter chemischer Diversität ermöglicht. Ein kürzlich identifiziertes Problem in der Click-SELEX ist jedoch die dreidimensionale Faltung der selektierten Clickmere in Abhängigkeit von verbleibenden Kupferionen. Dies führt zu einer nicht-nativen Faltung des Clickmers, was seine Anwendbarkeit erheblich einschränkt. Die Entwicklung eines Click-SELEX-Ansatzes basierend auf der Diels-Alder-Reaktion mit inversem Elektronenbedarf (iEDDA) würde kupferfreie Reaktionsbedingungen ermöglichen und das oben genannte Problem umgehen. In dieser Arbeit wurden erste synthetische Schritte gemacht, um eine iEDDA-basierte Click-SELEX zu ermöglichen. Es wurde ein modifiziertes Desoxyadenosintriphosphat synthetisiert, das eine reaktive Cyclopropeneinheit enthält, die über einen Carbamat-Linker an die Nukleobase gebunden ist. Der Einbau des Cyclopropen-modifizierten Nukleotids in die DNA und die anschließende Funktionalisierung mittels iEDDA unter Verwendung eines Tetrazinderivats waren erfolgreich. Der derzeitige Ansatz wird jedoch durch die Instabilität des Cyclopropen-Carbamat-Linkers während der Lagerung beeinträchtigt. Daher ist eine Verbesserung des Linkersystems erforderlich, um in Zukunft eine kupferfreie Click-SELEX-Methode zu etablieren.

Die Z-Konformation der RNA ist eine seltene und eher transiente Struktur, die hauptsächlich in Guanin (**G**)- und Cytosin (**C**)-reichen Sequenzen vorkommt. Z-RNA ist ein Ligand für das Z-Nukleinsäure-bindende Protein 1 (ZBP1), das zwei aufeinanderfolgende Z-Nukleinsäure-Bindungsdomänen,  $Z\alpha 1$  und  $Z\alpha 2$ , enthält. Die Bindung von Z-RNA an ZBP1 initiiert die ZBP1-vermittelte Nekroptose. *In vivo* Studien deuten darauf hin, dass die  $Z\alpha 2$ -Domäne der primäre Auslöser der Nekroptose ist. Die funktionelle Bedeutung von zwei aufeinanderfolgenden  $Z\alpha$ -Domänen ist deswegen noch unklar. Daher wurde ein vereinfachtes *in vitro* System verwendet, um die molekulare Interaktion zwischen ZBP1 und RNA genauer zu untersuchen. Zunächst wurde die Hypothese getestet, ob zwei Tandem- $Z\alpha$ -Domänen doppelsträngige A-RNA binden können und in die Z-Konformation überführen können. Dazu wurde Wildtyp-ZBP1 erfolgreich exprimiert und aufgereinigt. Eine **GC**-reiche Hairpin-RNA mit 8-Methylguanosin-Modifikationen wurde verwendet, um die Z-Konformation unter nahezu nativen Bedingungen *in vitro* zu stabilisieren. Eine Affinitätselektrophorese zeigte, dass Wildtyp-ZBP1 spezifisch an Z-RNA, aber nicht an A-RNA bindet, was darauf hindeutet, dass Tandem- $Z\alpha$ -Domänen keine A-zu-Z-Konformationsänderung bei RNA herbeiführen. Um die funktionelle Bedeutung der einzelnen  $Z\alpha$ -Domänen weiter aufzuklären, sollte die Bindungsaffinität jeder  $Z\alpha$ -Domäne isoliert untersucht werden. Dabei sollte festgestellt werden, ob die Mutation einer  $Z\alpha$ -Domäne die Bindung des Z-RNA-Liganden behindert und ob zwei aufeinanderfolgende  $Z\alpha$ -Domänen stärker an Z-RNA binden als eine. Daher wurden



ZBP1-Varianten mit Funktionsverlustmutationen entweder in der Z $\alpha$ 1-Domäne, in der Z $\alpha$ 2-Domäne oder in beiden Z $\alpha$ -Domänen erfolgreich exprimiert und aufgereinigt. Eine vergleichende molekulare *in vitro* Interaktionsstudie zwischen Z-RNA und diesen ZBP1-Mutanten muss zukünftig noch durchgeführt werden.



## Vorwort

Die vorliegende Arbeit wurde im Zeitraum von Oktober 2019 bis Juli 2024 am Institut für Organische Chemie der Universität zu Köln und am Institut für Life & Medical Sciences (LIMES) der Universität Bonn unter der Betreuung von Frau *Prof. 'in Dr. Stephanie Kath-Schorr* angefertigt.

An dieser Stelle möchte ich mich bei allen bedanken, die mich während meiner Promotionszeit unterstützt und motiviert haben.

An erster Stelle gilt mein ausdrücklicher Dank Frau *Prof. 'in Dr. Stephanie Kath-Schorr*, die es mir ermöglicht hat, diese spannenden Forschungsthemen zu bearbeiten. Ich danke ihr für die fachliche Unterstützung, die engagierte Betreuung und die wissenschaftliche Freiheit, die sie mir gewährt hat.

Mein Dank gilt auch den Mitgliedern meiner Prüfungskommission. Insbesondere danke ich Herrn *Prof. Dr. Ulrich Baumann* für die freundliche Übernahme des Zweitgutachtens. Herrn *Prof. Dr. Mathias Wickleder* danke ich für die Übernahme des Prüfungsvorsitzes. Bei Herrn *Dr. Jörg-Martin Neudörfel* bedanke ich mich für die Schriftführung.

Weiterhin danke ich *Prof. 'in Dr. Ines Neundorf* für die Zweitbetreuung meiner Promotion an der Universität zu Köln.

Bedanken möchte ich mich auch bei meinen aktuellen und ehemaligen *Kolleg\*innen* und *Freund\*innen* des *Arbeitskreises Kath-Schorr*. Ich danke euch für eure stetige Unterstützung und Hilfsbereitschaft, die angenehme Arbeitsatmosphäre und die lustige Zeit, die wir auch außerhalb des Labors zusammen verbracht haben. An dieser Stelle möchte ich einige Personen besonders hervorheben.

Mein besonderer Dank gilt *Eva Hoffmann*, mit der ich den Weg vom Master bis hierher gegangen bin. Danke für die vielen hilfreichen Tipps, die zahlreichen Diskussionen und die emotionale Unterstützung während dieser Zeit. Für das Korrekturlesen dieser Arbeit danke ich *Dr. Christof Domnick*, *Niclas Zips* und *Axel Miltz*. Außerdem möchte ich mich bei denjenigen bedanken, mit denen ich während dieser Zeit ein Büro oder Labor geteilt habe: *Dr. Lisa Bornewasser*, *Dr. Christof Domnick*, *Dr. Frank Eggert*, *Eva Hoffmann*, *Ekaterina Kulko* und *Matea Rabar*. Vielen Dank für die kooperative Zusammenarbeit, die angenehme Arbeitsatmosphäre, die vielen hilfreichen Tipps und Ratschläge und die schöne und lustige gemeinsame Zeit. Ebenso möchte ich mich bei *Finn Dicke* und *Lukas Neu* für die gute Zusammenarbeit im Rahmen des Z-RNA Projektes bedanken. Des Weiteren möchte ich mich bei *Tobias Behn* für seine unendliche Hilfsbereitschaft bei synthetischen Fragen bedanken. *Sarwar Aziz* möchte ich für seine Unterstützung bei der Synthese danken. Mein Dank gilt auch allen *Studierenden*, die meine Arbeit im Rahmen von Praktika und Abschlussarbeiten begleitet

haben. Außerdem möchte ich mich bei *Axel Mitz* bedanken, der mir besonders in der Endphase meiner Promotion eine große Stütze war und mich immer wieder ermutigt und bestärkt hat. Des Weiteren danke ich *Robert Dörrenhaus*, *Eva Hoffmann* und *Philip Wagner* für unsere kleinen Konzert Trips, die auch außerhalb des Labors für lustige und unvergessliche Erinnerungen gesorgt haben. Außerdem möchte ich mich bei meiner DnD-Runde, bestehend aus *Tobias Behn*, *Axel Miltz*, *Andre Zenz* und *Niclas Zips*, für die spannenden und unterhaltsamen Abende als Ausgleich nach einem langen Labortag bedanken.

Mein Dank geht auch an *Prof. Dr. Günter Mayer* und *Prof. Dr. Michael Famulok* vom LIMES Institut der Universität Bonn. Ich bedanke mich für die Großzügigkeit bei der Mitnutzung von Geräten und Arbeitsplätzen. Außerdem möchte ich mich bei allen *Mitarbeiter\*innen* der *Arbeitsgruppen Mayer* und *Famulok* für den wissenschaftlichen Austausch und die angenehme Arbeitsatmosphäre während meiner Zeit an der Universität Bonn bedanken.

Des Weiteren möchte ich mich bei allen *Kooperationspartner\*innen* der *Forschungsgruppen Pasparakis* (CECAD Köln) und *Mayer* (LIMES Institut der Universität Bonn) für eine gute Zusammenarbeit und den spannenden wissenschaftlichen Austausch bedanken. Ebenfalls möchte ich mich bei *Dr. Jan Gebauer* aus der *Arbeitsgruppe Baumann* (Universität zu Köln) für die Unterstützung bei den MST Messungen bedanken.

Mein Dank richtet sich auch an alle *Mitarbeiterinnen* und *Mitarbeiter* in den zentralen Einrichtungen der Universität zu Köln und der Universität Bonn. Mein besonderer Dank gilt den Abteilungen der *NMR* und *Massenspektrometrie* für die zuverlässige und schnelle Messung meiner Proben. Herrn *Dr. Jörg-Martin Neudörfl* danke ich für die Röntgenstrukturanalyse.

Nicht zuletzt möchte ich mich bei meinen *Freund\*innen* und meiner *Familie* bedanken. Ihr habt mich stets unterstützt, motiviert und mir emotionalen Rückhalt gegeben.

Ein großes Dankeschön geht an *Anna Strauß* für ihre emotionale Unterstützung und ihre motivierenden Worte, die mir unglaublich viel Kraft gegeben haben. Ich danke *Tamara Krampe*, die immer an mich geglaubt hat und mich immer wieder daran erinnert hat, wie weit ich auf meinem persönlichen Lebensweg schon gekommen bin.

Ebenso danke ich *Daniel Jaschky* für sein großes Verständnis, seine Geduld und seine ermutigende Unterstützung. Danke, dass du meine Stütze außerhalb der akademischen Welt bist.

Mein ganz besonderer Dank gilt aber vor allem meinen Eltern, *Andrea* und *Klaus-Walter Depmeier*, und meiner Schwester, *Lisa Depmeier*, die mich stets auf meinem Weg bestärkt haben. Vielen Dank für eure uneingeschränkte Unterstützung und euer Vertrauen in mich - ohne euch wäre das alles nicht möglich gewesen!

# Contents

<b>1</b>	<b>Theoretical background.....</b>	<b>1</b>
1.1	Artificial nucleic acids .....	3
1.1.1	Phosphate backbone-modified nucleic acids .....	4
1.1.2	Xeno nucleic acids.....	6
1.1.3	Nucleobase-modified nucleic acids.....	14
1.1.4	State-of-the-art: artificial nucleic acids with a dual modification pattern.....	20
1.1.5	Bioorthogonal click reactions for post-synthetic functionalization of nucleic acids.....	24
1.1.6	Artificial nucleic acids as therapeutic tools.....	27
1.2	Chemical modifications to alter nucleic acid conformations .....	35
1.2.1	Naturally occurring conformations of nucleic acids .....	35
1.2.2	Chemical modifications to stabilize the Z-conformation of nucleic acids .....	38
1.2.3	Z-DNA-binding proteins in disease .....	40
<b>2</b>	<b>Research objectives.....</b>	<b>45</b>
<b>3</b>	<b>Results and discussion .....</b>	<b>49</b>
3.1	Towards TNA with an expanded genetic alphabet.....	50
3.1.1	Synthesis of TNA nucleotides.....	50
3.1.2	Synthesis of the deoxyribofuranosyl TPT3 phosphoramidite.....	59
3.1.3	TNA primer extension assays for the preparation of TNA with an expanded genetic alphabet .....	60
3.2	A traceless approach to suppress Hoogsteen base pair formation in enzymatic TNA preparation.....	91
3.2.1	Synthesis towards a threoguanosine building block containing a photolabile <i>N</i> <sup>7</sup> -modification .....	92
3.3	Towards cyclopropene-modified nucleotides for an iEDDA click-SELEX approach ...	102
3.3.1	Synthesis of cyclopropene-modified deoxyribonucleotides .....	104

---

3.3.2	Enzymatic preparation of cyclopropene-modified DNA and downstream iEDDA click functionalization .....	112
3.3.3	Stability studies of the cyclopropene-carbamate modification .....	115
3.4	Investigation of the interaction between mZBP1 and Z-RNA .....	121
3.4.1	Expression and purification of mZBP1 variants.....	121
3.4.2	<i>In vitro</i> interaction studies of Z-RNA and mZBP1 .....	123
<b>4</b>	<b>Conclusion and outlook.....</b>	<b>127</b>
<b>5</b>	<b>Materials and methods.....</b>	<b>133</b>
5.1	Software and data analysis .....	133
5.2	Chemical methods.....	134
5.2.1	General chemical methods and equipment.....	134
5.2.2	NMR spectroscopy .....	135
5.2.3	HPLC analysis and purification .....	135
5.2.4	HPLC-MS spectrometry.....	136
5.2.5	GC-MS spectrometry.....	136
5.2.6	HR MS spectrometry .....	136
5.2.7	Chemical syntheses .....	137
5.3	Biochemical methods .....	182
5.3.1	Equipment .....	182
5.3.2	Consumables .....	184
5.3.3	Reagents and buffers .....	185
5.3.4	Enzymes .....	189
5.3.5	Bacterial cells, vectors, and plasmids .....	189
5.3.6	Kits .....	190
5.3.7	General biochemical methods .....	191
5.3.8	General bacterial cell culture methods.....	196
5.3.9	PCR-based cloning of Kod-RI TNA polymerase .....	201

---

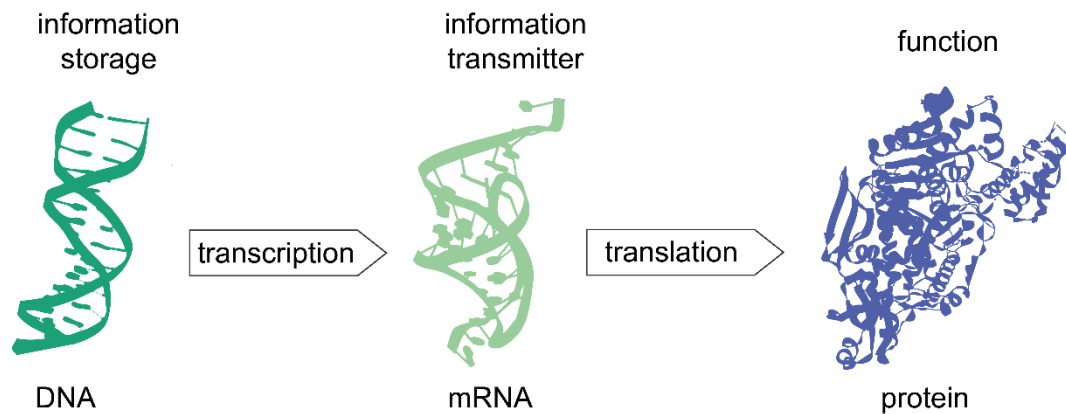
5.3.10	Site-directed mutagenesis of the Kod-RI coding sequence to the Kod-RSGA coding sequence .....	214
5.3.11	Expression and purification of TNA polymerases.....	217
5.3.12	TNA primer extension assays .....	220
5.3.13	MD simulations of tTPT3 and tNaM elongation in a nascent TNA strand .....	228
5.3.14	Preparation of cyclopropene-modified DNA .....	247
5.3.15	Expression and purification of mZBP1 variants.....	249
5.3.16	Interaction studies between mZBP1 variants and fluorescently-labeled RNA in A- and Z-conformation.....	251
<b>6</b>	<b>References.....</b>	<b>255</b>
<b>7</b>	<b>Appendix.....</b>	<b>283</b>
7.1	Supplementary figures and data.....	283
7.2	NMR spectra .....	313
7.3	Crystallography data .....	341
7.4	List of synthesized compounds.....	342
7.5	List of sequences .....	346
7.5.1	DNA primers.....	346
7.5.2	List of RNA, DNA and, TNA sequences.....	349
7.5.3	Vector sequences.....	352
7.5.4	Gene and plasmid sequences .....	358
7.5.5	Amino acid sequences .....	388
7.6	List of abbreviations .....	390
7.7	List of tables.....	394
7.8	List of figures.....	398
7.9	List of schemes .....	405
<b>8</b>	<b>Eidesstattliche Erklärung .....</b>	<b>409</b>





# 1 Theoretical background

The central dogma of molecular biology was postulated by Francis Crick in 1958 and illustrates the directional flow of genetic information within a biological system.<sup>1</sup> Specifically, he described the transfer of sequence information from nucleic acids (including deoxyribonucleic acid - DNA and ribonucleic acid - RNA) to proteins as an unidirectional process: "Once information has got into a protein it can't get out again."<sup>2</sup> In essence, to further embellish this dogma, DNA serves as the repository of genetic information, while RNA is best known as an intermediary structure represented by messenger RNA (mRNA) that facilitates the propagation of the encoded genetic information from DNA for the synthesis of proteins (Figure 1).<sup>3</sup> Proteins, in turn, perform a variety of functions in organisms, including DNA replication and transcription, gene expression regulation, structural support, catalysis of cellular biochemical reactions, molecular transport, and signalling.<sup>4</sup>

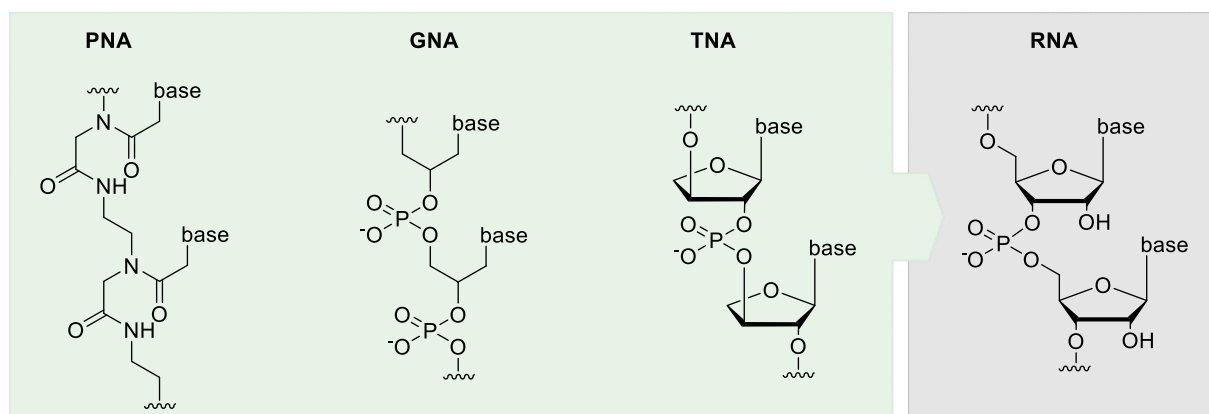


**Figure 1: Schematic representation of the unidirectional flow of biological information according to the central dogma of molecular biology.**<sup>1</sup> PDB files: DNA: 5F9I<sup>5</sup>, mRNA: 2TPK<sup>6</sup>, protein: 4K8Z<sup>7</sup>.

However, it must be noted that this strict interpretation of the central dogma is now considered obsolete and incorrect.<sup>8-10</sup> With the discovery of prion proteins capable of self-replication without the contribution of DNA and RNA<sup>11,12</sup> and the discovery that reverse transcriptases (RTs) are present not only in retroviruses but also in prokaryotes and eukaryotes,<sup>13-16</sup> the central dogma was challenged. Most interestingly, however, research on RNA has revealed that RNA is not only a transient copy for genetic information transfer, as assumed in the central dogma, but that a large part of the transcriptome possesses multiple functions besides the coding capacity.<sup>17</sup> These RNAs are called non-coding RNAs (ncRNAs) and their functions range from signaling to catalytic activity to gene expression regulation.<sup>18-20</sup> The unraveling of these diverse functions of ncRNA led to the postulation of the RNA world hypothesis, which suggests that RNA, with its ability to store genetic information and its catalytic activity, plays a key role in the origin of life.<sup>21</sup> The RNA world hypothesis is one out of several hypotheses<sup>22-25</sup> that seek to explain the emergence of life on earth. The RNA world hypothesis postulates that life as we know it was preceded by a world whose life relied on ribonucleic acids as universal

building blocks for storing genetic information and catalyzing chemical reactions. This hypothesis is supported by several factors, including the presence of self-replicating RNA molecules<sup>26,27</sup> and the composition of the ribosome<sup>28–31</sup>. Ribosomes consist largely of ribosomal RNA (rRNA) and thus utilize RNA catalysis for the fundamental process of protein biosynthesis.<sup>28–31</sup> In addition to insights derived from current biological understanding, researchers have also supported the RNA world hypothesis by experimentally substantiating plausible chemical scenarios for *ab initio* formation of RNA that might have happened under simulated prebiotic conditions.<sup>32–39</sup> Nonetheless, many aspects of the RNA world hypothesis remain inconsistent and elusive.<sup>40</sup> Major concerns include the inherent instability of RNA, its distinct chirality as well as the non-enzymatic production of RNA despite its length and complexity.<sup>40</sup>

Thus, researchers started to investigate the chemical etiology of nucleic acids, proposing simpler molecular scaffolds as the basis for an information-carrying biopolymer that preceded RNA.<sup>41–43</sup> Considering a so-called pre-RNA world, many potential alternatives to RNA were generated from RNA's close structural neighborhood, which might have been more likely to be formed *de novo* under prebiotic conditions.<sup>44</sup> Among these simpler RNA precursors, acyclic nucleic acids such as uncharged peptide nucleic acid (PNA) and glycol nucleic acid (GNA) as well as cyclic RNA analogs such as threose nucleic acid (TNA) have been proposed (Figure 2).<sup>41,43,45,46</sup>

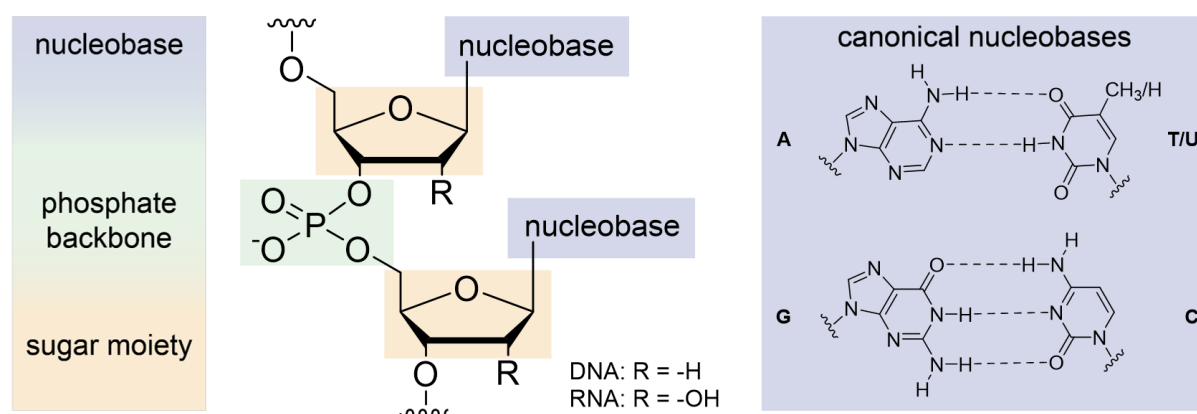


**Figure 2: Possible RNA progenitors within a pre-RNA world scenario.** Examples of acyclic and cyclic RNA analogs with potential prebiotic relevance. From left to right: peptide nucleic acid (PNA), glycol nucleic acid (GNA), threose nucleic acid (TNA), and for comparison, ribonucleic acid (RNA).<sup>43</sup>

However, not only research on the etiology of nucleic acid structure has contributed to the synthesis of RNA analogs, but also biomedical research has initiated the synthesis of non-canonical nucleic acids due to their highly interesting properties.<sup>47</sup> In this context, various modified and artificial nucleic acids have been developed that encompass modifications in the connecting phosphate linkages, the sugar moiety or the nucleobase.<sup>48–50</sup>

## 1.1 Artificial nucleic acids

Artificial nucleic acids are a group of synthetic nucleic acid analogs that carry distinct non-natural modifications in the tripartite structure of nucleic acid polymers. In recent years, the development of modified and artificial nucleic acid analogs has increased significantly paving the way for novel nucleic acid therapies and diagnostics. Several modified nucleic acids as well as artificial nucleic acid polymers entirely composed of unnatural building blocks have been reported in the literature.<sup>49,50</sup> These modifications include changes to the connecting phosphate backbone<sup>51,52</sup>, the sugar moiety<sup>48</sup>, or the nucleobase<sup>53</sup> (Figure 3).



**Figure 3: Possible modification sites in the tripartite structure of nucleic acids.** Modifications to nucleic acids can be implemented in either the phosphate backbone (green), the sugar moiety (orange), or the nucleobase (blue). The canonical nucleobases adenine (A), thymine (T), cytosine (C), and guanine (G) are depicted on the right. Sugar residues are omitted for clarity.

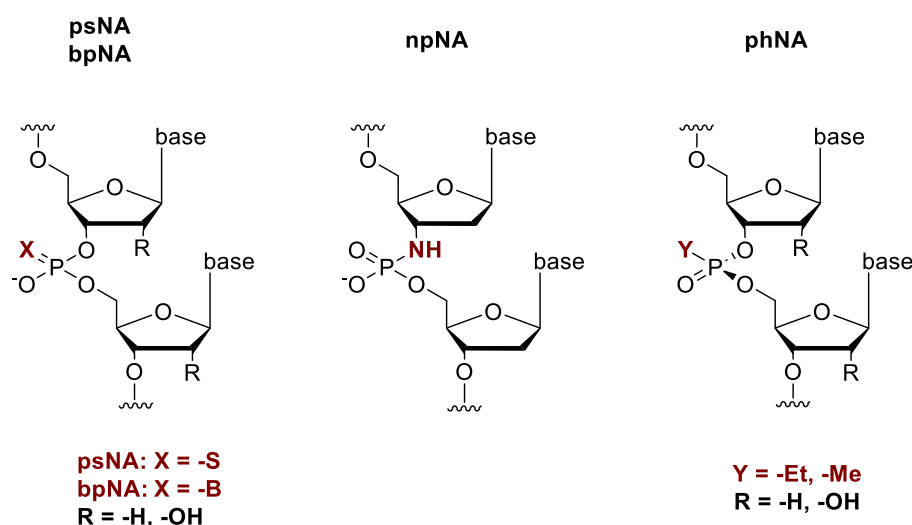
In general, tailor-made modifications differently affect the stability and dynamics of antiparallel Watson-Crick duplexes, the stability towards nucleases, and the chemical diversity of the nucleic acid.<sup>49,54</sup> In more detail, nuclease stability can be enhanced by modifying the sugar or phosphate moiety. Moreover, modifications to the sugar and phosphate backbones have a significant impact on the hybridization efficiency of the oligonucleotide. Conversely, alterations in the nucleobase moiety primarily increase the chemical diversity of nucleic acids.<sup>55</sup> However, it should be noted that replication and transcription processes are also affected by modifications to the nucleosidic scaffold.<sup>49,54</sup> While chemical solid-phase synthesis is a viable option for implementing numerous modifications to nucleic acids up to a length of approximately 100 to 200 nucleotides (nt),<sup>50,56,57</sup> polymerase-mediated preparation of artificial nucleic acids without length limitation frequently proves to be a considerably challenging process.<sup>50</sup> For enzymatic synthesis of artificial nucleic acids, it is of particular importance that the modifications do not impede the cross-hybridization with canonical nucleic acids and that the highly substrate-specific polymerases can process the non-canonical nucleotides.<sup>50,58</sup> In many cases, the use of specifically engineered polymerases is necessary to process highly modified nucleotide substrates.<sup>50,58</sup> Although progress has been made in polymerase engineering over the past few years, the enzymatic synthesis of artificial nucleic acids,

particularly those with more drastic modifications, remains challenging and has not been achieved for all artificial genetic polymers.<sup>58,59</sup> However, the processing of modified nucleotides by polymerases is not only of great importance for circumventing the length limitations of solid-phase synthesis but also for the development of aptamer-based therapies.<sup>58</sup> Aptamers are short, single-stranded, and highly structured nucleic acids that can bind specific targets by shape complementarity. Such aptamers are evolved *in vitro* using polymerase-mediated Darwinian evolution experiments (for further details on aptamers and their directed *in vitro* evolution, please refer to Chapter 1.1.6).<sup>60</sup>

Due to the large number of artificial nucleic acids reported in the literature, only a selection of the most common examples of artificial nucleic acids will be presented in the following chapters. Interesting properties of the presented artificial nucleic acids with respect to the development of therapeutic and diagnostic tools will be highlighted in more detail. This includes the aforementioned aspects pertaining to nuclease stability, chemical diversity, hybridization to canonical nucleobases, and enzymatic preparation.

### 1.1.1 Phosphate backbone-modified nucleic acids

The phosphate linkages represent one part in nucleic acids that can be chemically modified. Alterations to the phosphodiester backbone affect the nuclease stability as well as their pharmacokinetic properties.<sup>61</sup> So far, various modifications of the phosphate backbone have been reported in the literature, ranging from minor modifications to complete replacement of the phosphate group, resulting in several anionic, cationic, or neutral phosphodiester backbone-modified nucleic acids. Figure 4 depicts some examples of artificial nucleic acids with minor modifications of the phosphate linkage that will be described hereafter.



**Figure 4: Examples of repeating units of phosphate backbone-modified artificial nucleic acids.** From left to right: phosphorothioate nucleic acid (psNA), boranophosphate nucleic acid (bpNA), phosphoramidate nucleic acid (npNA), and (*S*)<sub>p</sub>-*P*-alkyl-phosphonate nucleic acid (phNA). The modifications are highlighted in red.

Simple modifications, such as in phosphorothioate (psNA)<sup>52</sup> and boranophosphate (bpNA)<sup>62</sup> nucleic acids, represent one way to modify the phosphate linkages of DNA and RNA (Figure 4). In psNA and bpNA, one of the non-bridging alpha oxygens of the phosphate group is replaced by either a sulfur ( $\alpha$ S) or borano ( $\alpha$ B) group, respectively.<sup>52,62</sup> It has been demonstrated that both phosphorothioate and boranophosphate modifications confer higher nuclease resistance to the nucleic acid analog compared to natural phosphodiester-linked ONs.<sup>63–65</sup> In addition, both nucleic acid analogs are capable of hybridizing to DNA and RNA,<sup>63,64,66</sup> but these duplexes are reported to be destabilized compared to those between two natural phosphodiester-linked nucleic acid strands.<sup>63,67</sup> Regarding the enzymatic preparation of psNA and bpNA,  $\alpha$ B and  $\alpha$ S nucleotides have been shown to be good substrates for naturally occurring polymerases, and a partial substitution of the canonical nucleotides is feasible.<sup>51,68,69</sup> Additionally, a fully modified bpNA was prepared with moderate fidelity using a 3'-5' exonuclease-deficient variant of the Klenow fragment of *Escherichia coli* (*E.coli*) DNA polymerase I.<sup>70</sup> In the case of psNA, enzymatic production with a fully substituted phosphorothioate backbone has been reported using genetically engineered mutants of *Thermus aquaticus* (Taq) polymerase.<sup>71</sup>

Phosphoramidate nucleic acids (npNA) are another family of backbone-modified nucleic acid analogs (Figure 4). In phosphoramidate oligonucleotides, the 3'-bridging oxygen of the phosphate moiety is replaced by an amino group, thereby forming N3'-P5' phosphoramidate bonds as the internucleotide linkages.<sup>72,73</sup> npNA is characterized by a very strong affinity for single-stranded RNA (ssRNA) and single-stranded DNA (ssDNA).<sup>73,74</sup> Furthermore, npNA is resistant to nuclease degradation.<sup>72,74</sup> Lelyveld *et al.* have evolved a DNA-dependent npNA polymerase derived from DNA polymerase I from *geobacillus stearothermophilus* that is capable of moderate npNA synthesis.<sup>75</sup> Interestingly, not only has DNA template-directed synthesis of npNA been achieved, but npNA also serves as a template for reverse transcription back into DNA using wildtype and engineered reverse transcriptases.<sup>76</sup>

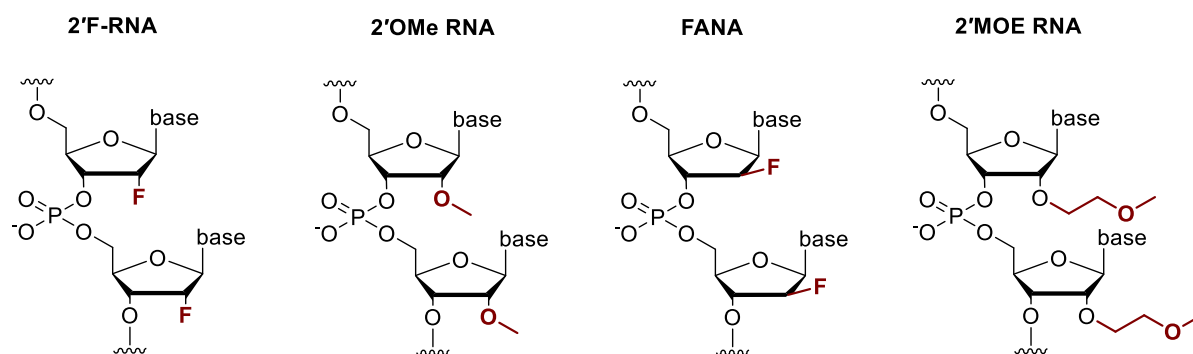
One example of a nucleic acid mimic with an electroneutral small backbone modification is ( $S$ )<sub>p</sub>-*P*-alkyl-phosphonate nucleic acid (phNA; Figure 4). In phNA, alkyl groups, such as methyl and ethyl chains, are attached to the central phosphorus of the phosphate moiety, replacing one of the non-bridging oxygen atoms.<sup>77,78</sup> This results in an uncharged *P*-alkyl-phosphonodiester backbone.<sup>78</sup> phNA is capable of forming duplexes with complementary canonical nucleic acids and possesses high nuclease stability.<sup>79,80</sup> However, the enzymatic synthesis of phNA by naturally occurring enzymes is challenging, and only sporadic incorporation of a phNA thymidine (phT) triphosphate (TP) has been demonstrated.<sup>77,81</sup> More recently, Holliger and coworkers reported the discovery of an engineered polymerase and reverse transcriptase capable of DNA template-directed synthesis of phNA and its reverse transcription back into DNA.<sup>78</sup>

In addition to the implementation of minor modifications in the phosphate backbone of nucleic acids, complete replacement of the phosphate group with new chemical moieties has also been performed.<sup>82–86</sup> This confers superior nuclease stability to such artificial nucleic acids.<sup>84,87,88</sup> However, these extreme modifications to the natural backbone structure certainly result in impaired enzymatic compatibility. Therefore, these alternative backbone replacements will not be discussed here in further detail.

### 1.1.2 Xeno nucleic acids

The second possibility to create artificial nucleic acids is to chemically modify the sugar moiety of the nucleosidic scaffold. Such nucleic acid modifications range from small chemical changes in the (deoxy)ribose sugar, to backbones constructed from sugars other than the naturally occurring (deoxy)riboses, to structurally completely different chemical entities. Alterations implemented in the sugar moiety of nucleic acids typically increase biological stability and affinity to complementary nucleic acids.<sup>61</sup>

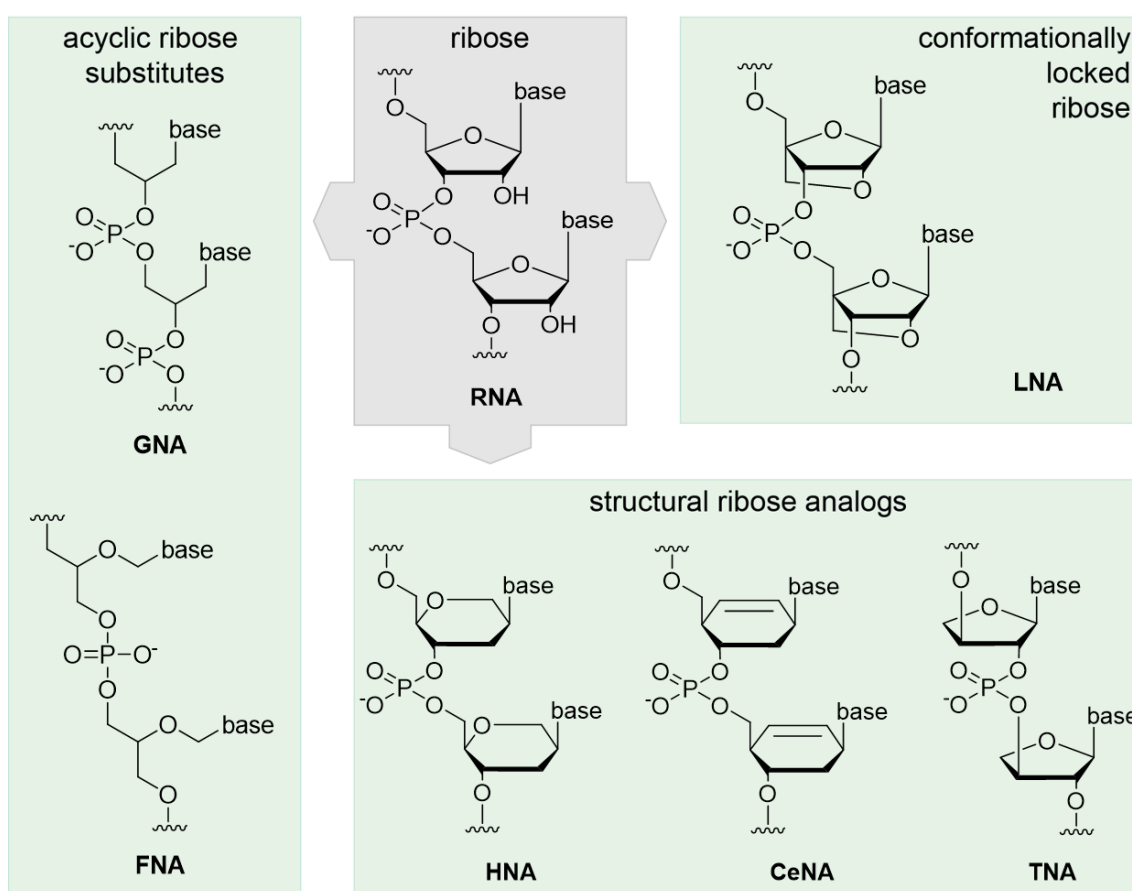
Examples of artificial nucleic acids with minor modifications of the sugar moiety include modifications of the 2'-position of the sugar as found in 2'-fluoro RNA (2'-F-RNA), 2'-O-methyl RNA (2'-OMe RNA), 2'-fluoro-arabinonucleic acid (FANA), and 2'-O-methoxyethyl RNA (2'-MOE RNA; Figure 5).<sup>54</sup>



**Figure 5: Examples of repeating units of artificial nucleic acids with minor modifications at the 2'-position of the sugar moiety.** From left to right: 2'-fluoro RNA (2'-F-RNA), 2'-O-methyl RNA (2'-OMe RNA), 2'-fluoro-arabinonucleic acid (FANA), and 2'-O-methoxyethyl RNA (2'-MOE RNA). The modifications are highlighted in red.

These modifications at the 2'-position of the sugar moiety proved to be compatible with enzymatic synthesis by naturally occurring or engineered polymerases, to increase duplex stability, and to confer enhanced biological stability to the nucleic acids.<sup>50,54,89,90</sup> However, this chapter will focus on more drastic sugar modifications in which the chemical composition and structure of the sugar moiety is completely altered. Such artificial nucleic acids with sugar analogs and even more divergent unnatural backbone structures are called xeno nucleic acids (XNA).<sup>91,92</sup>

One class of xeno nucleic acids are those containing conformationally locked sugar moieties, so-called bridged nucleic acids (BNA). In BNAs, the 2' oxygen and the 4' carbon of the ribose sugar are connected by different bridging moieties.<sup>93</sup> The most prominent example of this class is the  $\beta$ -D-ribo isomer of locked nucleic acid (LNA; Figure 7). LNA contains a 2'-4'-methylene bridge, which confines the torsion angles in the furanose sugar and locks it into the 3'-*endo* conformation.<sup>94-96</sup> Interestingly, the rigid modification confers improved thermodynamic properties to LNA, as this pre-structuring of the molecule provides an entropic advantage in hybridization with nucleic acids.<sup>97</sup> Furthermore, LNA offers improved biological stability in blood serum and cell extracts compared to natural DNA and RNA.<sup>98,99</sup>



**Figure 6: Examples of repeating units of xeno nucleic acids compared to RNA.** XNAs with acyclic ribose substitutes: glycol nucleic acid (GNA) and flexible nucleic acid (FNA). XNA with conformationally locked ribose:  $\beta$ -D-locked nucleic acid (LNA). XNAs with structural ribose analogs: hexitol nucleic acid (HNA), cyclohexene nucleic acid (CeNA), and threose nucleic acid (TNA).

The enzymatic synthesis of LNA by naturally occurring polymerases is limited and only partially LNA-modified DNA was achieved.<sup>100,101</sup> Eventually, fully modified LNA synthesis was accomplished using genetically engineered polymerases. Holliger and coworkers demonstrated LNA synthesis with Pol7C as well as reverse transcription of LNA with RT521K. Both enzymes are derived from TgoT polymerase, a variant of the replicative polymerase from *Thermococcus gorgonarius* (Tgo). However, the fidelity of LNA synthesis and reverse transcription using these polymerases was relatively low.<sup>102</sup> Recently, Obika and colleagues

have developed a genetically engineered LNA polymerase as well as a RT that are characterized by greatly improved efficiency and fidelity.<sup>103</sup> Both enzymes are variants of Kod DNA polymerase, a family B DNA polymerase derived from *Thermococcus kodakarensis* (Kod). These variants, namely Kod DGLNK and Kod DLK, exhibit distinct capabilities: Kod DGLNK is proficient in synthesizing LNA from a DNA template, and Kod DLK is capable of synthesizing DNA from an LNA template.<sup>103</sup>

Besides the most prominent  $\beta$ -D-LNA, there are additional seven configurational isomers as well as diverse structural analogs that are conformationally locked by different bridging moieties, e.g., amido-bridged nucleic acid (AmNA), *N*-MeO-amino bridged nucleic acid (*N*-MeO-BNA), and carbocyclic LNA (cLNA).<sup>104,105</sup>

Apart from conformationally locked sugar moieties, the sugar in xeno nucleic acids can be replaced by acyclic groups that may be less structurally related to the natural (deoxy)ribose sugar. One example of such a xeno nucleic acid is glycol nucleic acid (GNA; Figure 6), an acyclic artificial nucleic acid in which the (deoxy)ribose sugar is substituted by propylene glycol.<sup>106</sup> The propylene glycol units consist of only three carbon atoms, which are linked by phosphodiester bonds.<sup>106</sup> GNA can form stable antiparallel Watson-Crick homoduplexes in a homochiral manner, which exhibit even greater thermodynamic and thermal stability than their cognate counterparts.<sup>106,107</sup> However, GNA cannot cross-hybridize with DNA and its hybridization to RNA is rather poor.<sup>107</sup> In addition, GNA possesses an increased stability toward 3'-exonuclease degradation. It has been shown that DNA modified with 3'-terminal GNA residues is protected from degradation by snake venom phosphodiesterase (SVPDE).<sup>108</sup> To date, enzymatic template-dependent GNA synthesis remains a challenge, most likely due to the incompatibility of GNA to hybridize with DNA.<sup>109,110</sup> Interestingly, despite the incompatibility of DNA and GNA to form heteroduplexes, Bst polymerase, a family A polymerase from *Bacillus stearothermophilus*, is able to perform moderate reverse transcription of GNA into DNA.<sup>111</sup>

Another example of an acyclic nucleic acid with the ribose sugar substituted by a different chemical group is flexible nucleic acid (FNA; Figure 6). In FNA, the sugar moiety is substituted by the mixed acetal aminal of formyl glycerol, which is then linked by phosphodiester groups. More specifically, FNA is a conservative simplification of RNA and DNA in which the 2'-carbon of the sugar moiety is formally deleted.<sup>41,112</sup> This structural simplification reduces the number of stereocenters from four in RNA to only one (pro)stereocenter in FNA. However, this structural simplification results in a more flexible backbone,<sup>113</sup> and thus FNA does not cross-hybridize with DNA.<sup>114,115</sup> Nevertheless, the enzymatic oligomerization of up to seven FNA nucleotides has been enzymatically catalyzed by polymerases.<sup>113</sup> In addition, FNA can serve as a template for template-directed non-enzymatic oligomerization of RNA.<sup>116</sup> Similar to GNA, FNA is only slowly degraded by SVPDE and thus has increased stability against 3'-exonucleases compared to the parent polymers.<sup>115,117</sup>



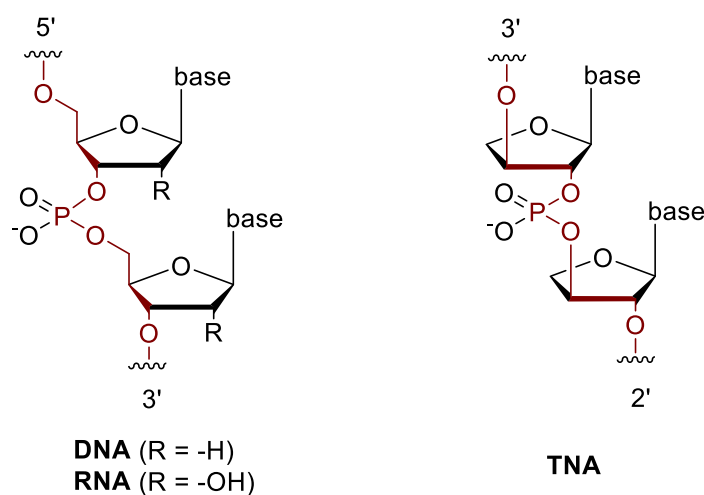
Lastly, xeno nucleic acids can also comprise structural analogs of the natural (deoxy)ribose sugar, which are either derived from their close structural neighborhood or only slightly resemble their natural counterparts. A ring-expanded analog of DNA is hexitol nucleic acid (HNA; Figure 6). In HNA, an additional methylene group is inserted between the 4' oxygen and the 1' carbon atom resulting in an 1,5-anhydrohexitol scaffold.<sup>118</sup> The internucleotide linkages in HNA are phosphodiester bonds connected via the 4'- and 6'-hydroxyl groups of the 1,5-anhydrohexitol. Furthermore, the nucleobase is shifted from the 1' to the 2' position.<sup>118,119</sup> Interestingly, the stability of an RNA-HNA duplex is higher compared to an RNA-RNA duplex.<sup>120</sup> In addition, HNA has enhanced stability towards the 3'-exonuclease SVPDE compared to RNA and DNA.<sup>118</sup> Although HNA synthesis has been tested using several naturally occurring polymerases, the strong substrate specificity of these enzymes hindered its efficiency.<sup>121,122</sup> Efficient HNA synthesis was only achieved when Holliger and coworkers evolved a DNA-dependent HNA polymerase, Pol6G12. Additionally, they developed a HNA-dependent DNA polymerase, RT521. Both polymerases are mutants of the aforementioned TgoT polymerase and are capable of reliable HNA transcription and reverse transcription with a reasonable length.<sup>102</sup>

An analog of HNA is the so-called cyclohexene nucleic acid (CeNA) in which the OCH<sub>2</sub> moiety of HNA is substituted by an isosteric unsaturated moiety, thus eliminating the ring oxygen (Figure 6).<sup>118</sup> Similar to HNA, CeNA can form stable homoduplexes with itself as well as heteroduplexes with RNA and DNA.<sup>123,124</sup> Interestingly, CeNA hybridizes more strongly with RNA than with DNA.<sup>124</sup> Furthermore, CeNA has been shown to have improved serum stability compared to DNA.<sup>123</sup> Although, naturally occurring polymerases from different evolutionary families accept CeNA nucleotides as substrates, DNA-templated CeNA synthesis and CeNA-templated DNA synthesis by canonical enzymes remains rather inefficient.<sup>125</sup> Holliger and coworkers engineered an efficient HNA polymerase as well as a reverse transcriptase derived from the aforementioned TgoT polymerase.<sup>102</sup> Pol7C, a polymerase that can also accept LNA nucleotides as substrates, was also found to be capable of CeNA synthesis. In addition, RT521K was identified to possess CeNA-RT activity besides its ability to reverse transcribe LNA back into DNA.<sup>102</sup>

Finally, another analog from the close structural neighborhood of the (deoxy)ribose sugar is the threose sugar (Figure 6). Nucleic acids based on this threose sugar bring along interesting characteristics and properties, which will be discussed in more detail in the next chapter.

### 1.1.2.1 $\alpha$ -L-Threofuranosyl nucleic acid

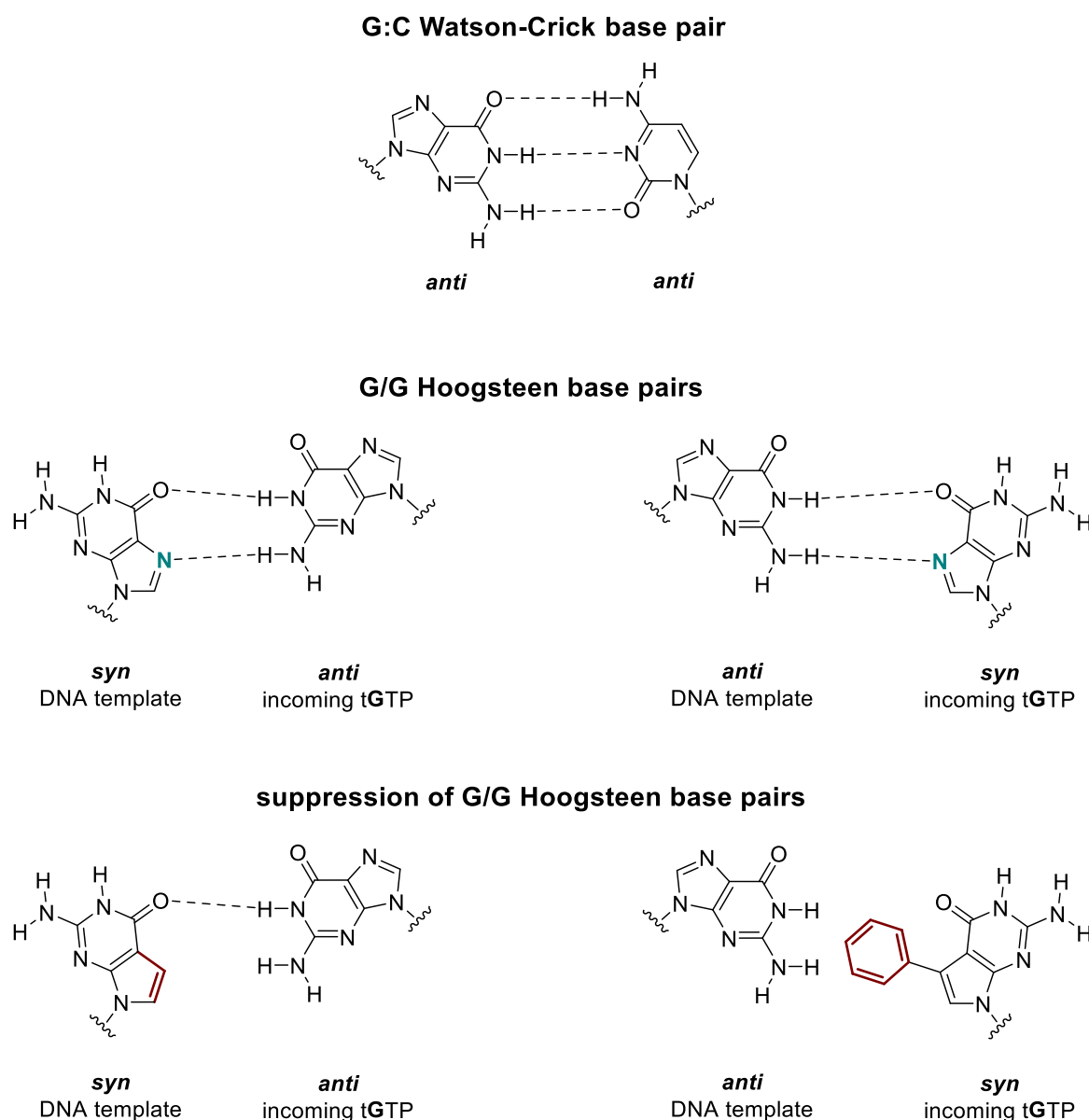
$\alpha$ -L-threofuranosyl nucleic acid (TNA) is a xeno nucleic acid notable for its special properties and distinct features. Two decades ago, TNA was first proposed by Albert Eschenmoser as a possible ancestor of RNA to elucidate the etiology of the nucleic acid structure as we know it today.<sup>44,126,127</sup> TNA differs from DNA and RNA in its backbone structure. The natural (deoxy)ribose sugar is replaced by a congener, the tetrose sugar  $\alpha$ -L-threose. This substitution of the (deoxy)ribose sugar with  $\alpha$ -L-threose results in an artificial genetic polymer with a five-atom repeat unit, one atom shorter than the six-atom periodicity prevalent in DNA and RNA (Figure 7).<sup>126,128</sup> The backbone of TNA further features phosphodiester linkages that are vicinally connected via the 3'- and 2'-hydroxyl groups of the threose sugar, contrasting with the naturally occurring 5'-3' phosphodiester linkages found in DNA and RNA.<sup>126</sup>



**Figure 7: Schematic comparison between the canonical nucleic acids (RNA/DNA) and TNA.** The differences in the backbone periodicity between TNA and RNA/DNA are highlighted in red.

Despite these structural differences from natural nucleic acids, TNA not only forms homoduplexes with itself,<sup>129</sup> but also cross-hybridizes via antiparallel Watson-Crick base pairing with complementary DNA, and even more strongly with RNA.<sup>130,131</sup> The chemical simplicity of TNA combined with its ability to cross-pair with complementary RNA, has led to the investigation of TNA in studies concerning the origin of life. TNA has been proposed as a potential progenitor of RNA in the hypothetical pre-RNA world scenario, with considerations of prebiotic conditions on primitive earth favoring TNA formation over *de novo* RNA formation.<sup>113,132</sup> Notably, Heuberger and Switzer demonstrated the non-enzymatic oligomerization of RNA from a cytosine-rich TNA template with activated ribofuranosyl guanosine monophosphates, suggesting a potential “genetic takeover” by RNA.<sup>113</sup> Furthermore, similar template-copying experiments with polymerases unveiled that certain polymerases are able to copy short DNA segments using a TNA template<sup>133</sup> as well as vice versa.<sup>121,134</sup> This finding motivated researchers to identify DNA-dependent TNA polymerases capable of extensive TNA synthesis along with efficient TNA-dependent DNA polymerases.

Such enzymes would enable information to be transferred back and forth between TNA and DNA, facilitating *in vitro* selection and Darwinian evolution experiments with TNA.<sup>134,135</sup> This idea gained increasing attention as TNA was shown to form secondary structures similar to DNA and RNA, such as hairpins and G-quadruplexes,<sup>126,136</sup> and to have improved biological stability compared canonical nucleic acids.<sup>137,138</sup> TNA has been demonstrated to remain stable in the presence of nucleases, including RQ1 DNase, RNase A, and the highly active 3'-exonuclease SVPDE.<sup>137,138</sup> Furthermore, TNA is not digested after 7 days of incubation in 50 % human serum or human liver microsomes and even shows superior stability relative to FANA and 2'F RNA.<sup>137,139</sup> Obviously, the interest in TNA for Darwinian evolution to develop therapeutic and diagnostic tools is very plausible, and thus several approaches to find more efficient TNA polymerases and reverse transcriptases have been reported in the literature.<sup>134,138,140–142</sup> In this context, Therminator DNA polymerase was identified as the only commercially available polymerase capable of synthesizing TNA oligonucleotides of significant length from a DNA template.<sup>138,141,143,144</sup> Therminator DNA polymerase is a site-specifically engineered (D141A / E143A / A485L) exonuclease-deficient form of the native DNA polymerase from *Thermococcus* species 9°N-7 with an enhanced ability to incorporate modified substrates.<sup>141,143,145,146</sup> Nevertheless, Therminator polymerase prefers natural dNTPs as substrates and has reduced activity when using  $\alpha$ -L-threofuranosyl nucleoside triphosphates (tNTPs) as substrates. Interestingly, this difference between natural and unnatural substrates becomes negligible upon addition of manganese ions during enzymatic TNA preparation.<sup>143</sup> However, Therminator polymerase still has shortcomings during enzymatic TNA synthesis with a strong sequence bias resulting in reduced fidelity and truncation during TNA synthesis.<sup>135,138,141,147</sup> Specifically, **G**-rich templates cause Therminator polymerase to stall and truncate TNA synthesis.<sup>135</sup> To further elucidate the fidelity of Therminator polymerase, Chaput and coworkers performed sequencing reactions of complementary DNA (cDNA) products obtained after reverse transcription of enzymatically prepared TNA oligonucleotides.<sup>138</sup> The detailed analysis of the mutations showed that deoxyguanosine (d**G**) to deoxycytidine (d**C**) transversions after one replication and reverse transcription cycle are the main reason for the reduced fidelity. This indicated a misincorporation of  $\alpha$ -L-threofuranosyl guanosine triphosphate (t**G**TP) opposite d**G** in the template. Although the fidelity of Therminator polymerase is remarkable for the enzymatic preparation of an XNA polymer, the observed sequence bias and polymerase stalling are particularly troublesome for Darwinian evolution experiments that require an accurate generation of large, unbiased pools of random and diverse TNA sequences.<sup>135,147</sup> A later study showed that the termination of enzymatic TNA synthesis on DNA templates containing consecutive d**G** residues is a result of the formation of Hoogsteen d**G**/t**G** base pairs in the enzyme active site (Figure 8).<sup>147</sup>



**Figure 8: C:G Watson-Crick base pair and G/G Hoogsteen base pair.** Top: Watson-Crick base pair between guanine and cytosine, both adopting the *anti*-conformation relative to the sugar. Middle: **G/G** Hoogsteen base pair with the electronegative  $N^7$  nitrogen shown in green, which allows the alternative base pairing mode. In the Hoogsteen base pair, the guanosine in the template adopts the *syn*-conformation, while the incoming tGTP adopts the *anti*-conformation or vice versa. Bottom: Suppression of **G/G** Hoogsteen base pair formation. The  $C^7$  (red) of 7-deaza-2'-deoxy-guanosine in the template suppresses **G/G** Hoogsteen base pairing when the guanosine in the template adopts the *syn*-conformation.<sup>147</sup> The phenyl group (red) of 7-deaza-7-phenyl tGTP suppresses **G/G** Hoogsteen base pairing when the incoming tGTP adopts the *syn*-conformation.<sup>148</sup> Threose sugar residues are omitted for clarity.

In this alternative base pairing interaction, the two nucleobases are accommodated in opposite *N*-glycosidic bond conformations, with one nucleotide adopting the *anti*-conformation and the other the *syn*-conformation.<sup>147</sup> This results in two hydrogen bonds between the Watson-Crick face of one guanine base and the major groove face of the other guanine residue (Figure 8, middle).<sup>149</sup> Hoogsteen **G/G** base pairs have been shown to be structurally similar to a regular Watson-Crick **G:C** base pair in a DNA duplex.<sup>150</sup> This structural similarity is thought to be the

reason why several natural and engineered polymerases encounter problems on **G**-rich templates.<sup>147</sup> The formation of Hoogsteen base pairs during TNA synthesis by Terminator polymerase was suppressed by complete substitution of **dG** in the DNA template with 7-deaza-2'-deoxy-guanosine (**7dG**). **7dG** is an analog of guanosine in which the  $N^7$  nitrogen of guanine is replaced by a carbon atom. This modification deletes one of the essential hydrogen bond acceptors in the **G/G** Hoogsteen base pair and restricts its stable formation (Figure 8, bottom).<sup>147</sup>

The development of a new polymerase is another attempt to avoid Hoogsteen base pairing during TNA synthesis, which could also simultaneously improve the efficiency of the polymerase-mediated TNA synthesis. To address the limitations of the Terminator DNA polymerase, the Chaput laboratory engineered the Kod-RI polymerase, a B-family-derived polymerase from archaeal hyperthermophilic species *Thermococcus kodakarensis*. Kod-RI polymerase carries 3'-5' exonuclease silencing mutations (D141A / E143A) as well as mutations that result in an improved TNA synthesis (A485R / E664I).<sup>140</sup> It was reported that Kod-RI polymerase outperformed Terminator polymerase with 5-fold faster extension rates and 20-fold higher fidelity.<sup>151</sup> However, a subsequent fidelity analysis, analogous to the previous fidelity measurements with Terminator polymerase, again revealed a strong sequence bias. Kod-RI achieved a fidelity of 99.7 % when using DNA templates containing only **dA**, **dT**, and **dC**. However, the fidelity dropped to 93 % when using templates containing all four genetic letters (**dA**, **dT**, **dC**, and **dG**). Again, the drop in fidelity was attributed to **dG** to **dC** transversions after one cycle of replication and reverse transcription, indicating a propensity for **dG/tG** mispairing during TNA synthesis.<sup>148</sup> In a first approach to overcome **dG/tG** mispairing, the unnatural **7dG** analog was again used instead of the natural **dG** residue in the DNA template. Surprisingly, TNA synthesis with the **7dG**-containing template still resulted in **dG** to **dC** transversions after one replication and reverse transcription cycle. Consequently, the incoming **tGTP** adopts the *syn*-conformation and the **dG** in the templating strand is in the *anti*-conformation (Figure 8, middle). Thus, in this case, modification of the  $N^7$  nitrogen of the **tGTP** would obstruct the formation of the **dG/tG** Hoogsteen base pair. In fact, 7-deaza-7-phenyl **tGTP** drastically reduced the **dG** to **dC** transversion, yielding an overall fidelity of 99.1 % sufficient for Darwinian evolution experiments (Figure 8, bottom).<sup>148</sup> However, this solution leads to the generation of TNA molecules bearing modified guanine residues, which is not always desired. Therefore, further polymerase evolution was necessary to circumvent the use of modified **tGTP** substrates.

In an effort to improve the performance of the Kod-RI polymerase, Chaput and coworkers developed the Kod-RSGA polymerase.<sup>152</sup> Kod-RSGA carries the same 3'-5' exonuclease silencing mutations (D141A / E143A) as Kod-RI polymerase. In addition, two mutations (A485R / N491S) have been introduced in the finger domain that increase catalytic activity.

Furthermore, two additional modifications in the thumb subdomain (R606G / T723A) have been implemented, which increase the efficiency of TNA synthesis and the amount of full-length TNA product.<sup>152</sup> It is notable that Kod-RSGA polymerase demonstrates robust performance even with challenging templates. Kod-RSGA polymerase exhibits high fidelity in copying **GG** repeats in templates, showing superiority over Terminator and Kod-RI polymerases. Nonetheless, the largest source of error, albeit to a much lesser extent, is still the **dG/tG** mismatch during TNA synthesis.<sup>152</sup> In summary, Kod-RSGA polymerase was shown to be the first efficient manganese-independent TNA polymerase with 20-fold increased activity over Kod-RI polymerase, high fidelity, and low template sequence bias for TNA synthesis.<sup>153</sup> In addition to the necessity for DNA-templated TNA polymerases, TNA-templated DNA polymerases are also required for *in vitro* selection of TNA aptamers. To date, SuperScript II (SSII) and *Bacillus stearothermophilus* (Bst) I polymerases have been identified as the most efficient polymerases for that purpose.<sup>138</sup> Strikingly, Bst I polymerase is more efficient in catalyzing the DNA-dependent TNA synthesis and achieves superior fidelity compared to SSII.<sup>142</sup> In combination, the genetically engineered Kod-RSGA TNA polymerase and the naturally occurring TNA reverse transcriptase Bst I polymerase allow faithful information transfer back and forth between DNA and TNA with an overall fidelity of 99.1 %, enabling Darwinian evolution experiments with TNA.<sup>154</sup> Interesting therapeutic and diagnostic applications of TNA with a focus on *in vitro* selection of functional TNA molecules will be presented in Chapter 1.1.6.1.

### 1.1.3 Nucleobase-modified nucleic acids

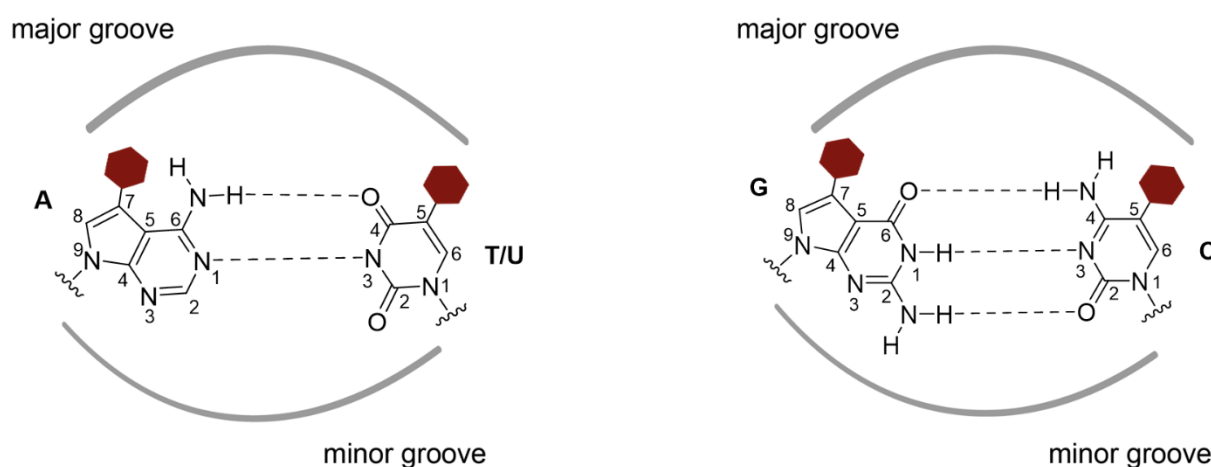
Besides introducing modifications in the sugar and phosphate backbone, the third way to generate artificial nucleic acids is to modify the nucleobase moiety. This is of interest because the chemical diversity of nucleic acids is limited to the four naturally occurring nucleobases adenine, thymine, cytosine, and guanine. By introducing chemical modifications to the nucleobase moiety of nucleic acids, their chemical diversity is increased, their functional density is enhanced, and new binding modalities are created.<sup>155</sup>

Modifications of the nucleobase moiety range from small modifications of the canonical nucleobases and reconfiguration of their hydrogen bonding pattern<sup>156–158</sup> to completely new non-canonical nucleobases.<sup>159–162</sup>

#### 1.1.3.1 Functionalization of canonical nucleobases

The simplest and most straightforward method to create artificial nucleobases with increased chemical diversity is to attach chemical groups to the natural purine and pyrimidine

nucleobases. However, the scope for the modification of canonical nucleobases is limited because they provide the recognition site for Watson-Crick base pairing. The most attractive modification sites are those exposed to the major groove of the nucleic acid duplex, which is the case for the C<sup>5</sup> position of pyrimidine bases and the N<sup>7</sup> position of purine bases (Figure 9).<sup>163</sup> Of note, modifications to the purine bases are usually made using their 7-deaza derivatives.<sup>163–167</sup> Modifications at these positions are usually well tolerated and, if not too bulky, do not significantly affect base pairing behavior, duplex geometry, or processing by cognate polymerases.<sup>168</sup> Examples of chemical residues used to increase the nucleobase diversity while maintaining polymerase capability include *e.g.* halogens, amides, acrylates, as well as aromatic and guanidinium groups, and amino acid-like residues.<sup>69,164,166,167,169</sup>



**Figure 9: Attractive sites for modification of the canonical pyrimidine and purine nucleobases.** Modifications (red hexagons) exposed to the major groove are usually well accommodated within the duplex and do not interfere with polymerase processing. The pyrimidine nucleobases thymine and cytosine are preferentially modified at the C<sup>5</sup> position. The purine nucleobases adenine and guanine are preferentially modified at the C<sup>7</sup> position of their 7-deaza derivatives.<sup>163</sup> The modifications are attached to the nucleobases by different linkers. Modifications include *e.g.* halogens, amides, acrylates, as well as aromatic and guanidinium groups, and amino acid-like residues. Sugar residues are omitted for clarity.

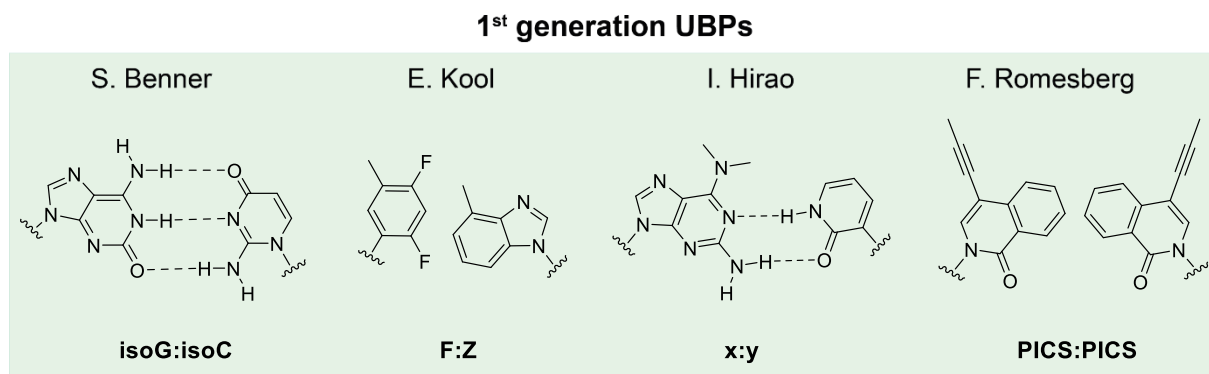
In addition to these modifications, the attachment of small reactive handles to the nucleobase moiety is also very common. Such small reactive handles are well accepted by polymerases and allow the attachment of a variety of chemical groups, including bulky residues, after enzymatic synthesis of the nucleic acid.<sup>170</sup> Probably the most commonly used artificial nucleoside bearing a reactive handle in the nucleobase moiety is C<sup>5</sup>-alkyne-modified 5-ethynyl-2'-deoxyuridine (EdU) because of its facile synthesis.<sup>170</sup> Other reactive groups used as reactive handles on nucleobases include *e.g.* vinyl and azide groups, primary amines, strained alkenes and alkynes, as well as iodo and formyl groups.<sup>171–177</sup> The further functionalization of such reactive handles with various chemical moieties requires specialized chemistries, which will be discussed in Chapter 1.1.5.<sup>170</sup>

Functionalized nucleobases as presented in this chapter still rely on the recognition of the four canonical nucleobases by Watson-Crick base pairing and result in a global modification of

nucleic acids. Thus, the selective introduction of independent novel modifications going beyond the four canonical nucleobases cannot be achieved with this strategy.

### 1.1.3.2 Unnatural base pairs: a historical review

An approach to truly augment the chemical diversity of nucleic acids is to expand the genetic alphabet, rather than attaching functional moieties to canonical nucleobases. The idea of an alternative or expanded genetic alphabet was first raised in the context of earlier life forms more than half a century ago,<sup>178</sup> and has since received increasing attention. Pioneering work on artificially expanding the genetic code was published by the Benner laboratory in 1989.<sup>156,157</sup> In an attempt to explore the potential for alternative information storage on primitive earth, they made the first attempt towards a man-made base pair. Using the most apparent strategy, they created a new unnatural base pair (UBP), isoguanine (**isoG**) and isocytosine (**isoC**), which resembles the natural Watson-Crick base pairing system by forming an orthogonal hydrogen bonding pattern (Figure 10).<sup>156</sup>



**Figure 10: Overview of the first generation UBPs to expand the genetic alphabet.** Sugar residues are omitted for clarity.

Several polymerases, such as the Klenow fragment of *E.coli* DNA polymerase I, avian myeloblastosis virus (AMV)-RT, and T7 RNA polymerase, accepted the new isomeric base pair as a substrate for replication and transcription processes.<sup>156,179</sup> In principle, this would allow the use of the **isoG:isoC** base pair as an independent third base pair. However, the mismatch of **T** and **U** opposite to a tautomeric form of **isoG** that forms under aqueous conditions at physiological pH presented a major drawback. Additionally, substantial deamination of **isoC** in an alkaline environment indicated reduced stability of the unnatural base.<sup>179</sup>

About a decade later, Morales and Kool reported the first UBP relying on hydrophobic interactions. They synthesized the nucleobases **F** and **Z** (Figure 10), which are nonpolar shape analogs of the pyrimidine nucleobase **T** and the purine nucleobase **A**, respectively.<sup>159,180</sup> The **F:Z** base pair demonstrated that hydrogen bonds, characteristic for natural Watson-Crick base



pairing, are not essential for effective base pairing, but that shape and size also play an important role.

Although the **F:Z** base pair isosterically mimics the canonical **A:T** Watson-Crick base pair nearly perfectly, its incorporation during replication with the exonuclease-deficient (exo-) Klenow fragment exhibited a lower efficiency. Additionally, significant self-pairing of **F** and misincorporation of deoxythymidine triphosphate (dTTP) opposite **Z** in the template was reported to be an issue.<sup>159</sup>

Shortly thereafter, the Romesberg group reported the development of a self-pairing base pair, **PICS:PICS** (Figure 10), which creates an orthogonal UBP by utilizing hydrophobic interactions and van der Waals forces.<sup>181</sup> The **PICS** nucleobase is based on an isocarbostyryl scaffold and thus, unlike the **F** and **Z** nucleobases, is not a shape analog of the canonical nucleobases. Interestingly, the Klenow fragment (exo-) incorporated d**PICSTP** opposite to d**PICS** in the template, with a strong preference for self-pairing over non-cognate pairing with one of the canonical nucleotides. Although processing of **PICS** by the Klenow fragment occurs with reasonable efficiency, extension of the oligonucleotide after **PICS** incorporation proved inefficient due to nucleobase stacking.<sup>181,182</sup>

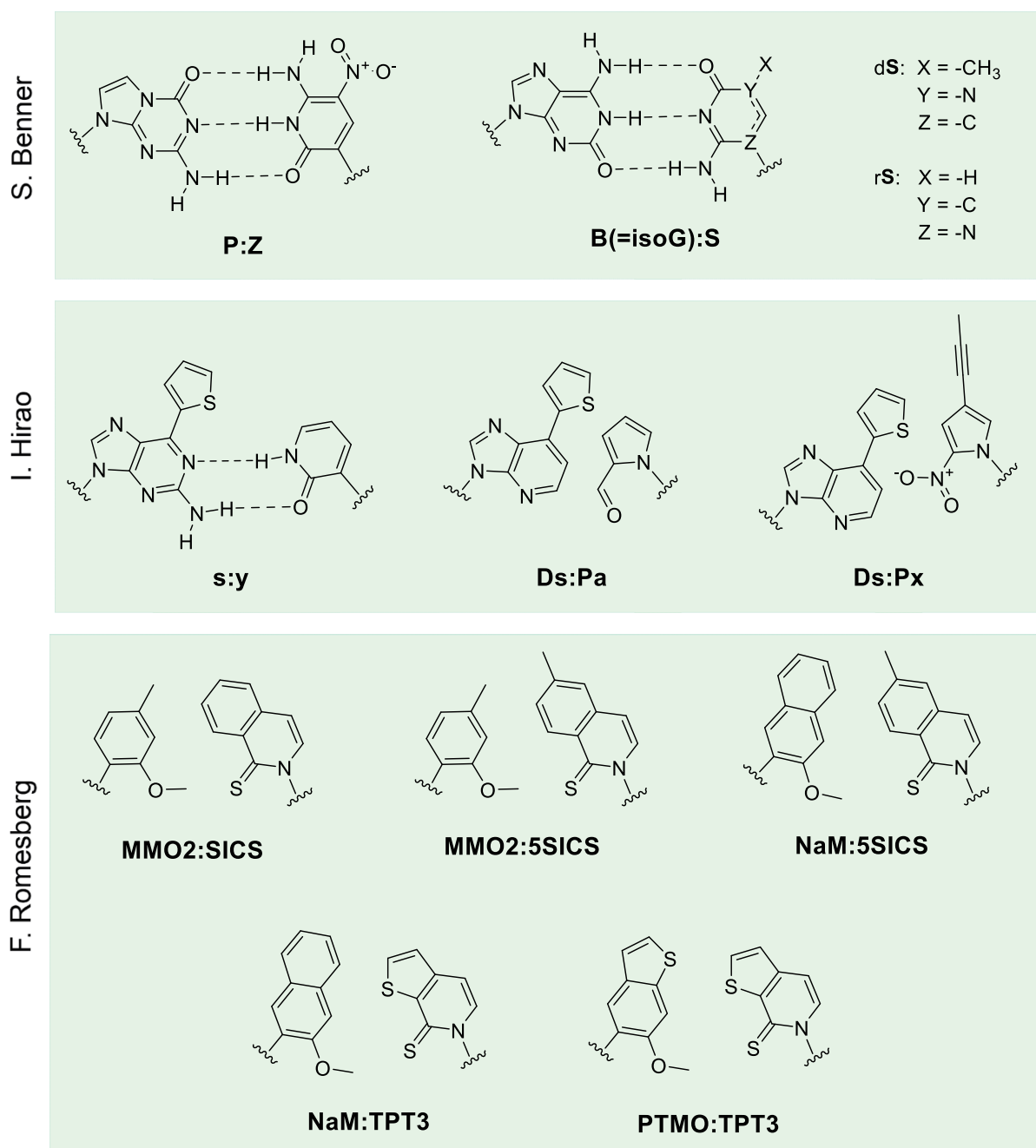
During the same time, Hirao's laboratory also focused on the generation of UBPs. Using purine and pyridine scaffolds, the UBP **x:y** (Figure 10) was developed based on the strategies of shape complementarity and hydrogen bonding for base pair formation.<sup>183</sup> Selective incorporation of d**yTP** into DNA and **yTP** into RNA opposite **x** in the template was achieved by the exonuclease-proficient (exo+) Klenow fragment and T7 RNA polymerase, respectively.<sup>183,184</sup> However, the dimethylamino group of **x** disturbs the stacking interactions between the unnatural and the adjacent base pair in a duplex, particularly limiting the efficient replication of the UBP.<sup>185</sup>

This impressive work on UBPs has motivated researchers to continue studies in this emerging field to address major drawbacks of the first generation of UBPs, such as lack of specificity and efficiency during enzymatic incorporation.

For example, the Benner group reported improved variants of the iso**G** nucleobase to reduce mispairing events during the enzymatic synthesis of the iso**G**:iso**C** UBP.<sup>186,187</sup> Extending the idea of creating novel base pairs by shuffling hydrogen bonding patterns, the Benner group developed several new nucleobases that each pair via three hydrogen bonds to form UBPs.<sup>188</sup> In 2019, an eight-letter nucleic acid, called the Hachimoji nucleic acid (*hachi* – eight; *moji* – letters), was reported that contains the **P:Z** and **B:S** UBPs (Figure 11) in addition to the four natural nucleobases.<sup>158</sup> Interestingly, iso**G** reoccurs in this system, but has been renamed to **B**. To date, efficient enzymatic synthesis and replication of Hachimoji DNA has not been demonstrated, but the **P:Z** UBP has been shown to be compatible with enzymatic replication

and transcription processes.<sup>189–191</sup> Notably, transcription of Hachimoji DNA into Hachimoji RNA was achieved by an engineered variant of T7 polymerase.<sup>158</sup>

## 2<sup>nd</sup> generation UBPs



**Figure 11: Overview of the second generation UBPs to expand the genetic alphabet.** Sugar residues are omitted for clarity.

Not only the Benner group, but also the Hirao group continued their research on artificial nucleobases. Efforts to further improve their first generation **x:y** UBP resulted in the **s:y** UBP (Figure 11). In the **s** unnatural base, the dimethylamino group of **x** was exchanged by a thiophene to promote stacking interactions with neighboring nucleobases. Compared to the **x:y** UBP, this modified version of the UBP was more efficiently incorporated into DNA by enzymes.<sup>185</sup> Nevertheless, the incorporation specificity of the **s:y** base pair is limited due to the

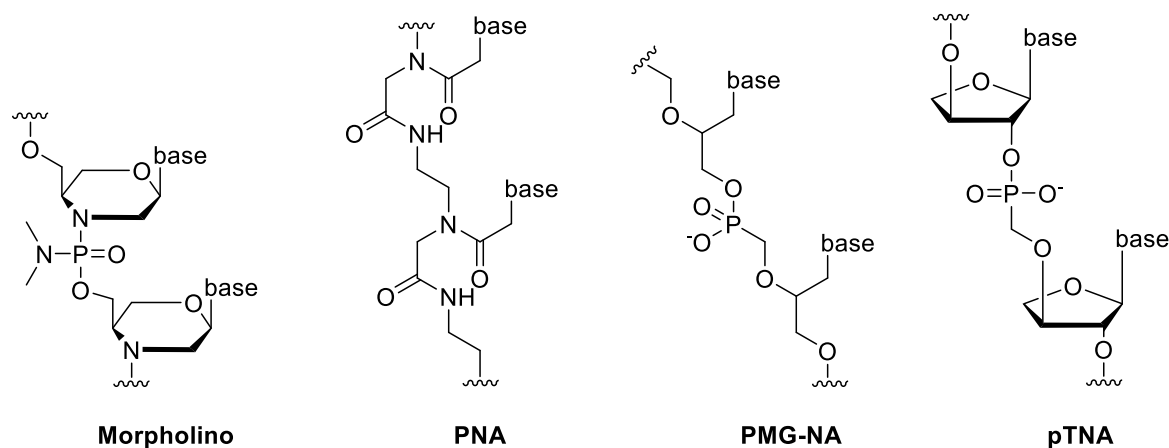
tendency of the **y** UB to form a mismatch with adenine.<sup>53,185</sup> Further steps to improve the shape and electrostatic complementarity of the UBP were made and combined with the concept of hydrophobicity, ultimately resulting in the **Ds:Pa** UBP (Figure 11).<sup>192</sup> While the **Ds** base is a modified version of **s** with the 2-amino group removed, the **Pa** UB is based on a pyrrole scaffold with an aldehyde group, which is smaller than the previous pyridine scaffold of **y**. The **Ds:Pa** UBP was found to be efficiently replicated and transcribed by Vent DNA and T7 RNA polymerases, respectively.<sup>192</sup> Notably, it was shown that the **Pa** UB can be further modified with functional residues and used for the enzymatic preparation of site-specifically functionalized DNA and RNA.<sup>192–194</sup> Although the replication efficiency of this UBP was high, major drawbacks included **Ds/Ds** self-pairing and **A/Pa** mispairing.<sup>192</sup> Rational revision of this base pair to eliminate the observed mismatch events eventually led to the **Ds:Px** UBP (Figure 11).<sup>195–197</sup> In **Px**, the aldehyde functionality of **Pa** is exchanged with a nitro group and the UB is additionally decorated with a propynyl group. On the one hand, the nitro group electrostatically prevents mispairing with non-cognate **As**. On the other hand, the propynyl group increases the incorporation efficiency of **Px** opposite **Ds** compared to that of **Ds** opposite **Ds**, making the formation of **Ds/Ds** self-pairs unfavorable.<sup>197</sup> This highly improved UBP represents a breakthrough in the field of UBPs as it is completely orthogonal to the natural base pairs and replicates efficiently with more than 99.9 % fidelity using Deep Vent (exo+) polymerase.<sup>198</sup>

The Romesberg group has also invested a tremendous amount of work to improve their first artificial **PICS:PICS** UBP to generate novel hydrophobic base pairs with improved properties (Figure 11).<sup>199–206</sup> In 2008, the Romesberg group performed two complementary high throughput screenings of 3600 potential hydrophobic UBP candidates to identify novel UBPs that are well accommodated by DNA polymerases. Both screens identified the UBP **dSICS:dMMO2** as the most promising lead candidate for further optimization (Figure 11). Using this UBP and related derivatives, detailed structure activity relationship analysis identified essential functional groups necessary for successful polymerase processing.<sup>207</sup> Further optimization of the **dSICS:dMMO2** UBP resulted in the **d5SICS:dMMO2** base pair, with **d5SICS** having superior properties compared to **dSICS** due to reduced self-pairing of **d5SICS** (Figure 11). The hydrophobic **d5SICS:MMO2** UBP is efficiently and selectively replicated by the Klenow fragment and several other naturally occurring polymerases.<sup>207,208</sup> Further work was invested in the improvement of the **MMO2** base, culminating in the **d5SICS:dNaM** UBP, where **NaM** has an increased aromatic surface area compared to **MMO2** (Figure 11).<sup>160</sup> This was the first UBP created by the Romesberg laboratory with replication efficiencies and fidelities nearly approaching those of natural DNA.<sup>209</sup> Crystallization of the **d5SICS:dNaM** UBP in complex with KlenTaq DNA polymerase revealed that the polymerase active site induces a Watson-Crick-like structure during the processing of the **dNaM:d5SICS** base pair.<sup>210</sup> However,

in the post-insertion position of the enzyme, the UBP resembles a cross-strand intercalated structure, thereby reducing the affinity with which the next correct nucleotide binds.<sup>211</sup> Therefore, efforts were made to further improve the structural characteristics of this UBP, which eventually yielded **dTPT3** as an even more suitable counterpart to **dNaM** (Figure 11). The **dTPT3:dNaM** system is characterized by natural-like replication efficiency and fidelity using OneTaq polymerase. Remarkably, the efficiency is only four times lower than amplification of all natural DNA, and the fidelity is greater than 99.98%.<sup>161</sup> An alternative counterpart for **TPT3** was later developed by the Romesberg group and culminated in the **PTMO** unnatural nucleobase (Figure 11).<sup>212</sup> The **PTMO** nucleobase has a reduced ring size in comparison to **NaM**, resulting in increased incorporations of 41% in single-nucleotide extension experiments with T4 polynucleotide kinase, in contrast to 35% for **NaM:TPT3**. Nevertheless, in polymerase chain reaction (PCR) with OneTaq polymerase, the incorporation of **dNaMTP** opposite **dTPT3** was more efficient than that of **dPTMOTP**.<sup>212</sup> Notably, the **TPT3** scaffold was found to have a unique tolerance to modifications, as replication and transcription containing functionalized **TPT3** were shown to proceed with only slightly reduced fidelity.<sup>161,213–215</sup> This observation allows for the efficient introduction of functionalized nucleotides and thus the site-specific modification of nucleic acids even further enhancing their chemical diversity.

#### 1.1.4 State-of-the-art: artificial nucleic acids with a dual modification pattern

The artificial nucleic acids discussed thus far have in common that the chemical modifications implemented are typically limited to one segment of the nucleic acid's tripartite structure, which consists of a phosphodiester backbone, a sugar moiety, and a nucleobase. Although advances have been made in the synthesis of unnatural nucleotide building blocks<sup>216</sup> and in polymerase engineering<sup>59</sup> to facilitate the enzymatic generation of artificial nucleic acids, doubly modified nucleic acids are still relatively rare.<sup>217,218</sup> However, especially highly modified nucleic acids with enhanced properties regarding chemical diversity, secondary structure formation, and nuclease resistance that can be prepared by polymerases hold great promise for the Darwinian evolution of nucleic acids for the development of novel, innovative therapeutic and diagnostic tools.



**Figure 12: Examples of repeating units of artificial nucleic acids with a dual modification pattern.** From left to right: morpholino nucleic acid, peptide nucleic acid (PNA), phosphonomethylglycerol nucleic acid (PMG-NA) and phosphonomethyl threose nucleic acid (pTNA). These doubly modified artificial nucleic acids comprise modifications in the sugar and the phosphate moieties.

A very prominent example of a doubly modified nucleic acid is the so-called morpholino nucleic acid devised by Summerton in 1985 (Figure 12).<sup>219</sup> In morpholino oligonucleotides, the (deoxy)ribose sugar is substituted by methylenemorpholine rings carrying the nucleobases. The methylenemorpholine rings are connected through phosphorodiamidate groups instead of phosphate groups.<sup>220</sup> Morpholino oligonucleotides are completely resistant to enzymatic degradation by *e.g.* DNases, RNases, proteases, and phosphodiesterases and are stable in plasma and serum.<sup>221</sup> In addition, morpholinos have a high affinity for binding complementary RNA, with a lower binding affinity compared to an RNA:RNA duplex, but a higher binding affinity than DNA to RNA.<sup>219</sup> In particular, highly modified morpholinos with modifications at the C<sup>5</sup> positions of the pyrimidine nucleobases have been reported in the literature to exhibit even higher binding affinities.<sup>222</sup> However, to date, morpholinos can only be prepared synthetically, *e.g.* employing chlorophosphoramidate chemistries and polymerase processing has not been demonstrated yet.<sup>223</sup>

Another prominent example of a distinctive artificial genetic polymer with a dual modification pattern is peptide nucleic acid (PNA).<sup>224</sup> In PNA, the sugar-phosphate backbone is substituted by a linear, achiral, uncharged *N*-(2-aminoethyl)glycin polymer in which the repeating *N*-(2-aminoethyl)glycin units are linked by peptide bonds (Figure 12).<sup>224</sup> PNA hybridizes to complementary DNA and RNA in parallel or antiparallel orientation with high sequence specificity and a remarkable thermal stability.<sup>225</sup> This remarkable thermal stability is attributed to the uncharged PNA backbone, which lacks the electrostatic repulsion of natural nucleic acid duplexes.<sup>225</sup> Additionally, PNA is resistant to protease and nuclease degradation.<sup>226</sup> However, efficient enzymatic synthesis of PNA by polymerases is unlikely to be achieved in the near future. Currently, PNA synthesis relies on either chemical solid-phase methods<sup>227</sup> or non-enzymatic template-dependent polymerization<sup>228</sup>.

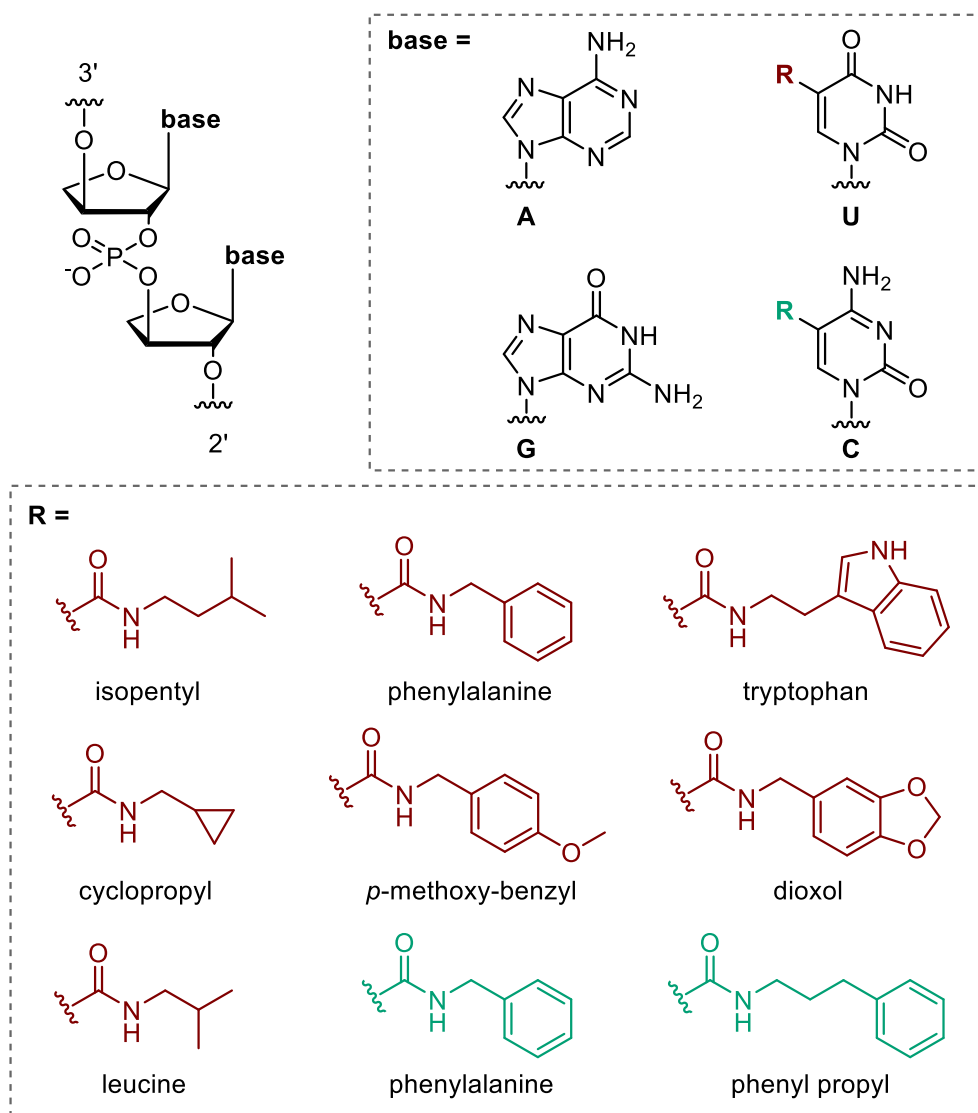
Another acyclic nucleic acid with a dual modification pattern is phosphonomethylglycerol (PMG) nucleic acid, which consists of repeating chiral phosphonomethylglycerol (PMG) units (Figure 12).<sup>229</sup> The cross-hybridization of an all PMG-NA to RNA and DNA is reported to be extremely weak. Nevertheless, PMG-based diphosphates were found to be partially accepted as substrates by both Klenow (exo-) DNA polymerase and Vent (exo-) polymerase. However, only up to three PMG-based diphosphates could be incorporated sequentially using a DNA template, likely due to inefficient hybridization of PMG-NA and DNA. Hence, chemical solid-phase synthesis is the only reliable way to synthesize PMG-NA by now.<sup>229</sup>

The three artificial nucleic acids mentioned above are characterized by a dual modification pattern and therefore have interesting properties in terms of binding affinity to natural nucleic acids and nuclease stability. However, due to incompatibility with enzymatic synthesis, these artificial nucleic acids are not suitable candidates for Darwinian evolution experiments.

Another dual-modified artificial nucleic acid with a phosphonomethyl-based backbone that is certainly compatible with enzymatic synthesis is phosphonomethyl-threosyl nucleic acid (pTNA). pTNA was developed by Herdewijn, Pinheiro, and colleagues and involves modifications to both the sugar and the phosphodiester linkages.<sup>230</sup> pTNA features a threofuranosyl sugar linked by phosphonomethyl groups between the 3'- and 2'-hydroxyl groups of the threose sugar, which confers superior nuclease resistance (Figure 12).<sup>134,137,231</sup> However, similar to PMG-NA, pTNA does not form stable duplexes with complementary DNA and RNA strands. Nevertheless, enzymatic synthesis of pTNA is possible using the engineered TgoT\_EPFLH polymerase derived from the previously mentioned TgoT polymerase.<sup>230</sup> Although, pTNA has a dual modification pattern and can be prepared enzymatically, it is not clear whether pTNA can form secondary structures to enable Darwinian evolution experiments.<sup>153</sup>

All of the aforementioned artificial nucleic acids with a dual modification pattern combine modifications in the sugar and the phosphate backbone. Nevertheless, artificial nucleic acids that combine a nuclease-resistant sugar or phosphate modification with a nucleobase modification would be of particular interest for the selection of aptamers. This approach would not only result in a more nuclease-resistant oligonucleotide, but would also enhance its chemical diversity beyond the canonical nucleobases.

Recent advances in increasing the chemical diversity of nuclease-resistant sugar-modified xeno nucleic acids were made by Chaput and coworkers.<sup>232-234</sup> They reported the enzymatic synthesis of functionally enhanced TNA by combining the nuclease-resistant threose backbone with canonical nucleobases modified with hydrophobic residues at the C<sup>5</sup> position of pyrimidine nucleobases (Figure 13).<sup>232-234</sup> These modifications were designed to resemble the nonpolar amino acid side chains prevalent in antibody paratopes,<sup>235</sup> thereby bringing novel hydrophobic binding modalities into play.



**Figure 13: TNA with an increased chemical density developed by the Chaput research group.**<sup>232–234</sup> The C<sup>5</sup>-position of pyrimidines has been modified with hydrophobic and amino acid-like residues. Some examples of installed modifications are shown here.

Nonetheless, this functionalized TNA still relies on the recognition of the four natural nucleobases through Watson-Crick base pairing, resulting in a global modification of the nucleic acid. Therefore, this approach does not allow for the targeted incorporation of additional, independent modifications beyond the four standard nucleobases. However, as it was shown that the TNA nucleotides with functionalized canonical nucleobases were well accommodated by the Kod-RSGA TNA polymerase, the potential for further advances in this area is indicated.<sup>232–234</sup>

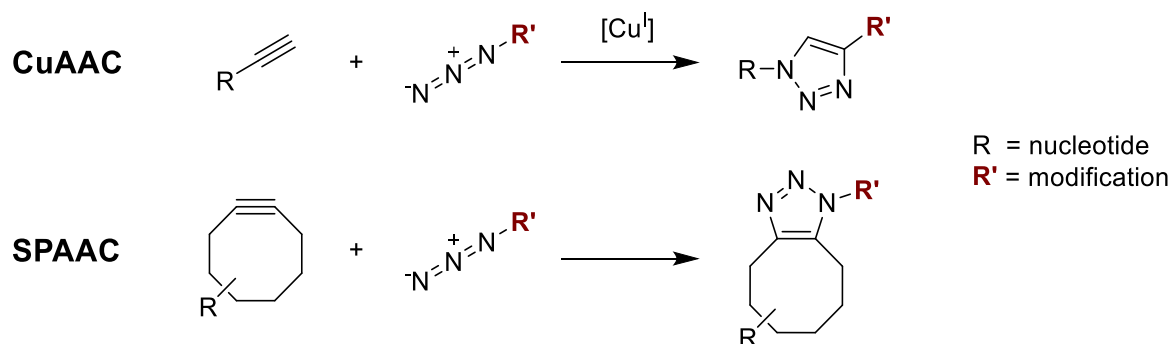
### 1.1.5 Bioorthogonal click reactions for post-synthetic functionalization of nucleic acids

The introduction of modifications to the nucleobase moiety of nucleic acids represents a valuable technique for increasing the functional density of nucleic acid molecules (see also Chapter 1.1.3.1 and 1.1.4).<sup>236–238</sup> This is of particular interest in the context of advancing nucleic acid therapeutics through the principle of Darwinian evolution, as such modifications have the ability to expand the range of addressable targets.<sup>237,238</sup> However, the direct integration of complex modifications during the enzymatic synthesis of an oligonucleotide can prove to be very laborious. The synthesis of modified nucleotides requires multiple intricate steps, and the introduction of each new modification requires the synthesis of the entire modified nucleotide. Moreover, the presence of bulky groups can impede the enzymatic preparation of nucleobase-modified nucleic acids.<sup>237</sup>

Modern methodologies have addressed these challenges by employing chemically modified nucleotides during the enzymatic synthesis of nucleic acids, which are equipped with small reactive handles in the nucleobase moiety instead of large and bulky groups.<sup>239</sup> These handles facilitate the subsequent attachment of various chemical moieties after the enzymatic synthesis of the oligonucleotide. The small reactive handles are specifically designed to minimize interference with polymerase activity and to ensure separation of the attachment step of large functional groups from the incorporation phase during enzymatic synthesis.<sup>239</sup> Such a strategy implements the attachment of a plethora of modifications that were previously inaccessible via an enzymatic approach.<sup>240</sup> In particular, a single modification initially incorporated into the nucleic acid can subsequently serve as a platform for the generation of multiple nucleic acids bearing different modifications with minimal effort.<sup>241,242</sup>

However, such reactions used to functionalize nucleic acids with integrated reactive groups must meet certain criteria. These criteria include the bioorthogonality and physiological reaction conditions of the post-synthetic attachment reaction.<sup>240,243,244</sup> In this context, click reactions are particularly well-suited, due to their modularity, high yield, stereospecificity, minimal byproduct formation, and simple reaction conditions.<sup>245</sup> Examples of click chemistry reactions include nucleophilic substitutions, Michael additions, epoxidations, dihydroxylations, and cycloadditions.<sup>170</sup> To date, many of these reactions have been used to bioorthogonally modify biomolecules of interest.<sup>243</sup>





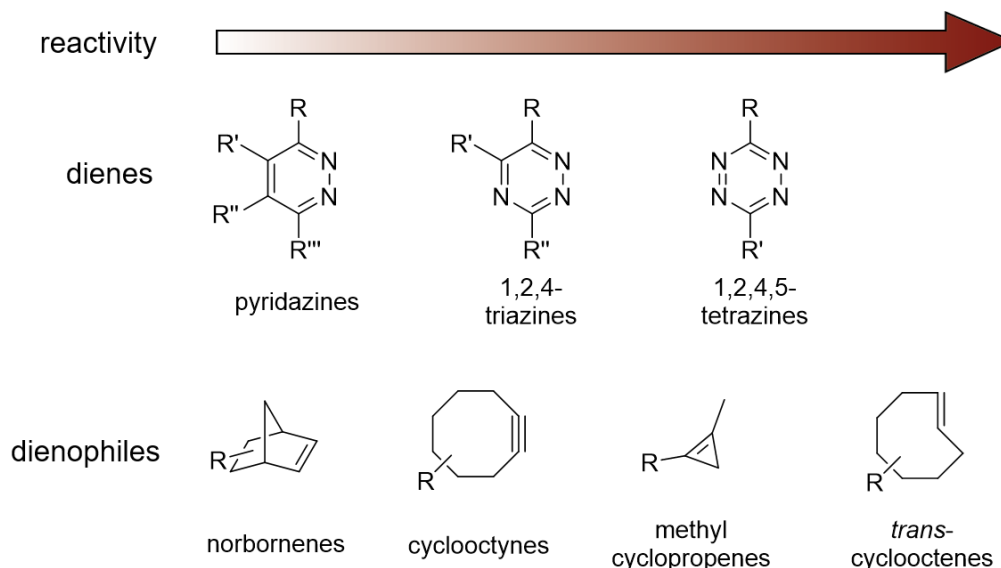
**Figure 14: Click reactions used in bioorthogonal applications:** Probably the most common click reaction, the copper-catalyzed azide-alkyne cycloaddition (CuAAC), and its copper-free alternative, the strain-promoted azide-alkyne cycloaddition (SPAAC). In this context, the alkyne moiety is attached to a nucleotide (R) and the azide is connected to the desired modification (R', red).

Probably the most commonly used reaction for downstream attachment of diverse chemical moieties is the classical copper-catalyzed azide-alkyne cycloaddition (CuAAC), which has been widely used to modify nucleic acids to increase their chemical diversity.<sup>170</sup> In CuAAC, an azide moiety readily reacts with a terminal alkyne moiety to form a covalent 1,2,3-triazole linkage at high reaction rates due to Cu<sup>I</sup> catalysis (Figure 14).<sup>246,247</sup> The use of alkyne-modified nucleotides in the enzymatic synthesis of nucleic acids, such as the aforementioned EdU, allows for CuAAC-mediated downstream attachment of virtually any functional group that can be conjugated to an azide. However, Cu<sup>I</sup> is susceptible to oxidation under aqueous reaction conditions.<sup>170</sup> To enable the reaction to proceed under aqueous conditions, Cu<sup>II</sup> salts are used that are reduced *in situ* by water-soluble, biocompatible reducing agents such as ascorbate.<sup>170,248</sup> However, under these reaction conditions, reactive oxygen species (ROS) are formed,<sup>248</sup> which cause oxidative degradation and strand excision of nucleic acids.<sup>170,249</sup> Therefore, Cu<sup>I</sup> stabilizing ligands and small amounts of acetonitrile as a co-solvent are used in bioorthogonal CuAAC reactions to protect the biomolecule from oxidation and accelerate the reaction.<sup>250,251</sup> Another common obstacle to the modification of biomolecules employing CuAAC is the complete removal of toxic copper ions after the click reaction.<sup>252</sup> Although CuAAC click chemistry has been widely used to modify nucleic acids, including the application in Darwinian evolution experiments,<sup>241,253–255</sup> these drawbacks limit the implementation of the CuAAC reaction for nucleic acid functionalization.

A copper-free alternative to CuAAC is the strain-promoted azide-alkyne cycloaddition (SPAAC). Instead of terminal alkynes, cyclic strained alkynes such as cyclooctyne derivatives are used (Figure 14). The ring strain promotes the reaction and avoids the need for a toxic metal catalyst.<sup>256</sup> However, the reaction rate of SPAAC is rather low compared to CuAAC, and the reactive cyclooctyne moieties are more bulky and more likely to interfere with polymerase processing.<sup>257</sup>

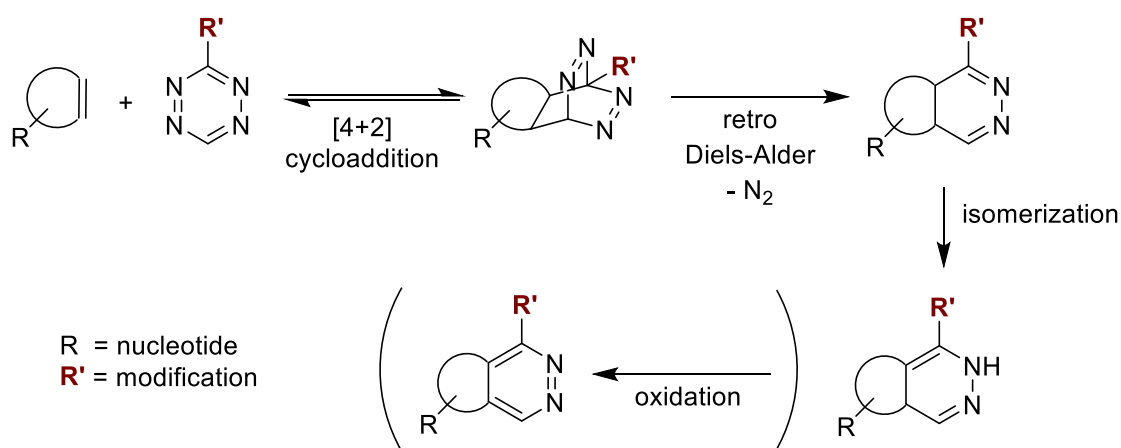
A very attractive bioorthogonal reaction is the inverse electron-demand Diels-Alder (iEDDA) reaction. This reaction is gaining increasing attention due to fast reaction kinetics, high

selectivity, and high yields under aqueous conditions without the need for a catalyst.<sup>258,259</sup> In the iEDDA reaction, electron-poor dienes react rapidly with electron-rich dienophiles.<sup>260</sup> Commonly used dienes include tetrazine and triazine derivatives and to a lesser extent pyridazines, while dienophiles include strained alkenes and alkynes such as norbornene, methylcyclopropene, *trans*-cyclooctenes, and cyclooctyne derivatives (Figure 15).<sup>261,262</sup>



**Figure 15: Commonly used dienes and dienophiles in inverse electron-demand Diels-Alder reactions in order of reactivity.**<sup>258,261</sup>

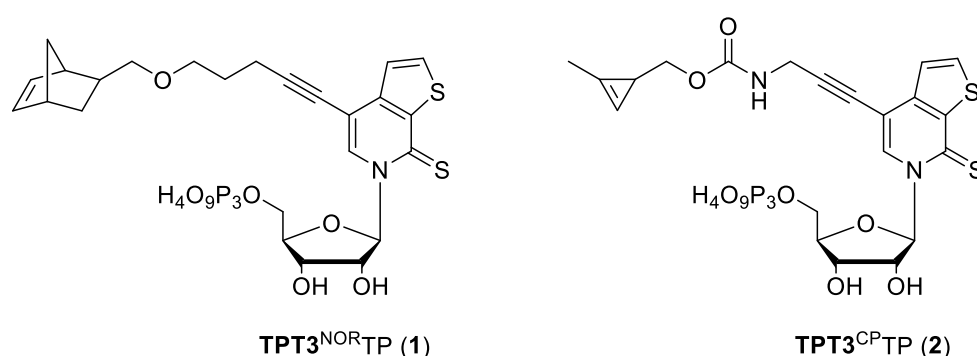
The reaction between tetrazines and strained alkenes proceeds via an initial reversible cycloaddition to give a strained bicyclic intermediate. Upon nitrogen expulsion, which renders the reaction irreversible, the adduct undergoes a retro Diels-Alder to ultimately yield a 4,5-dihydropyridazine, which is either isomerized to the 1,4-dihydro isomer or oxidized to the pyridazine end product (Figure 16).<sup>261</sup>



**Figure 16: Reaction cascade of the bioorthogonal inverse electron-demand Diels-Alder reaction between strained alkenes and 1,2,4,5-tetrazines.** In this context, the strained alkene moiety is attached to a nucleotide (R) and the tetrazine is connected to the desired modification (R', red).

Because this reaction does not require a copper catalyst, non-toxic as well as non-denaturing reaction conditions are provided, making the iEDDA reaction suitable for nucleic acid

modification. To facilitate the functionalization of nucleic acids via iEDDA, nucleotides with strained alkenes attached to the nucleobase via a linker are used during the enzymatic preparation of the nucleic acid.<sup>214,215,263,264</sup> This allows virtually any chemical moiety that can be conjugated to a tetrazine to be attached to the nucleic acid after its enzymatic preparation. In the Kath-Schorr group, norbornene-modified and cyclopropene-modified nucleotides have been used in the preparation of nucleic acids to enable the downstream functionalization of nucleic acids with fluorophores and spin labels. More in detail, the norbornene moiety is attached to the nucleobase via an ether linkage, while the cyclopropene moiety is attached via a carbamate linker.<sup>213–215,264–266</sup> Two representative examples of norbornene- and methylcyclopropene-modified **TPT3** nucleoside triphosphates, **TPT3<sup>NOR</sup>TP (1)** and **TPT3<sup>CP</sup>TP (2)**, developed by the Kath-Schorr group are shown in Figure 17.<sup>213,214</sup>



**Figure 17: TPT3 nucleoside triphosphates with reactive handles used in the Kath-Schorr research group for downstream nucleic acid functionalization via the iEDDA click reaction.** Left: a norbornene reactive handle attached to the unnatural **TPT3** nucleobase via an ether linker.<sup>213</sup> Right: a cyclopropene mini-tag attached to the unnatural **TPT3** nucleobase via a carbamate linker.<sup>214</sup>

To date, the iEDDA reaction has been extensively used by our research group and other laboratories to attach reporter groups to nucleic acids to study their structures, dynamics, folding, and functions at a molecular level, as well as their localization in a cellular context.<sup>213,215,264,267</sup> However, the iEDDA reaction has so far not been used to increase the chemical diversity of nucleic acids for *in vitro* selection experiments as has been done with CuAAC.

### 1.1.6 Artificial nucleic acids as therapeutic tools

Nucleic acid therapeutics have emerged as promising and versatile tools for the treatment of various diseases. The number of nucleic acid-based therapeutics that have been approved to date demonstrate their potential.<sup>268</sup> In particular, their high specificity, broad functionalities, tunable immunogenicity, and their reproducible batch-to-batch production make nucleic acids extremely interesting for drug discovery.<sup>269</sup> In contrast to other pharmacological modalities such as small molecules and antibodies, which primarily target proteins, nucleic acid-based

therapeutics have more diverse mechanisms of action.<sup>269</sup> Nucleic acid therapies offer a broad platform of technologies that either target the underlying cause of a disease at the genomic or transcriptomic level or act at a proteomic level. These diverse technologies include e.g. antisense oligonucleotides (ASOs), RNA interference (RNAi), mRNA vaccines, as well as aptamer-based therapies.<sup>270</sup> Several oligonucleotide-based therapeutics using these technologies have been approved and are already in use for the treatment of diseases.<sup>268</sup>

Although nucleic acid-based therapeutics provide a broad toolbox for curative applications, it quickly became clear that major drawbacks of nucleic acid-based therapeutics include their low stability in body fluids and their narrow chemical diversity. This is because canonical nucleic acids are susceptible to degradation by endogenous nucleases, and their chemical diversity is limited because the genetic code contains only four different nucleobases, all of which rely on hydrogen bonding. Therefore, artificial nucleic acids are of particular interest for therapeutic and diagnostic applications to overcome these drawbacks.<sup>270,271</sup>

To date, several nucleic acid-based therapeutics, as well as potential candidates, have been developed using artificial building blocks that enhance the applicability of nucleic acids in drug discovery. The nucleic acid therapeutics approved to date on the market also made use of modifications to improve the drugs in terms of e.g. binding affinity, nuclease resistance, and chemical stability.<sup>268</sup>

Modifications to ASOs and RNAi-based therapeutics are straightforward because modifications can be introduced during solid-phase synthesis of the small oligonucleotides.<sup>272</sup> However, introduction of modifications is mostly limited to the phosphate backbone and sugar moiety, as their mechanism of action is based on the binding of complementary target sequences.<sup>273</sup> Reported modifications to the phosphate backbone include, for example, phosphorothioate modifications, while 2'OMe and 2'MOE were extensively used as sugar modifications.<sup>270</sup> In addition, more drastic modifications to the sugar moiety have been implemented. In the case of ASOs, this includes the use of LNA, TNA, PNA and morpholino backbones.<sup>270,274,275</sup> In addition, small interfering RNAs (siRNAs) containing LNA and TNA modifications have been reported.<sup>276–278</sup>

The modification of mRNA vaccines is similarly constrained, as modifications must be incorporated by enzymes during production and must not interfere with translation after administration.<sup>279</sup> Nevertheless, modifications have shown to be crucial for the stability, immunogenicity and translational efficiency of mRNA vaccines. The two naturally occurring nucleobase modifications, pseudouridine ( $\Psi$ ) and *N*<sup>1</sup>-methylpseudouridine ( $^1\Psi$ ), are probably the most widely used modifications within the coding sequence of mRNA vaccines.<sup>279,280</sup>

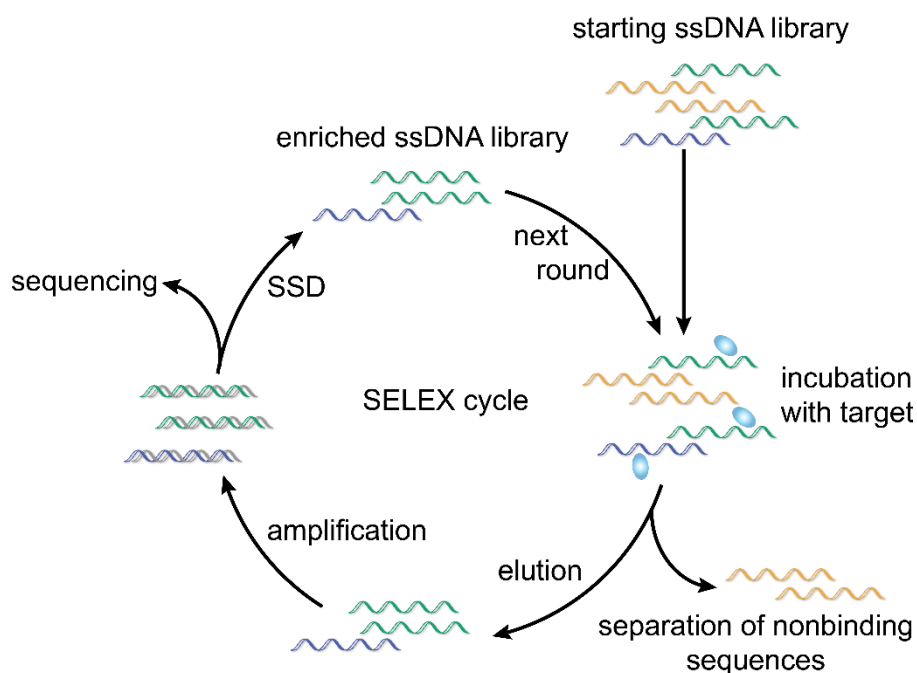
In contrast, aptamers can be modified to a greater extent than other nucleic acid-based drugs, as their mechanism of action is not based on complementary nucleic acid recognition or dependent on translational processing. Aptamers are single-stranded nucleic acids that fold

into distinct three-dimensional structures, enabling them to bind targets with high specificity and affinity through shape complementarity.<sup>60</sup> This enables not only the introduction of phosphate and sugar modifications as well as naturally occurring nucleobase modifications, but also the introduction of artificial modifications to the nucleobase, which significantly expands their chemical diversity and, consequently, the addressable target spectrum.<sup>253</sup> In the following chapter, the *in vitro* selection of aptamers will be presented, including approaches that allow the selection of modified aptamers with enhanced functionalities.

### 1.1.6.1 *In vitro* selection and evolution of artificial aptamers

As already indicated above, aptamers are relatively short single-stranded nucleic acids that form unique tertiary structures. Aptamers can selectively and tightly bind to a specific target, but unlike ASOs and RNAi-based therapeutics, binding is determined not only by the primary sequence and complementary binding, but rather by the three-dimensional structure that the aptamer folds into. Target recognition and aptamer binding are strongly driven by shape-dependent interactions as well as electrostatic, hydrophobic, and hydrogen-bonding interactions.<sup>281</sup> Thus, aptamers are considered the nucleic acid pendants of antibodies, but have several advantages over antibodies. Aptamers are smaller, cheaper to produce, stable during long-term storage, and no batch-to-batch variabilities occur. In addition, aptamers have low immunogenicity, better pH and thermal stability, as well as a higher degree of modifiability, allowing for a broader range of targets.<sup>282,283</sup> Moreover, aptamers can bind to a wide variety of targets, including small molecules, proteins, viruses, toxins, and even cells.<sup>283</sup>

In order to develop an aptamer that binds with high affinity to a specific target among the wide range of possible targets, *in vitro* Darwinian evolution experiments are performed. More specifically, aptamers are screened against a specific target by Systematic Evolution of Ligands by Exponential Enrichment (SELEX; Figure 18).



**Figure 18: Schematic representation of Systematic Evolution of Ligands by Exponential Enrichment (SELEX).** The SELEX cycle starts with a ssDNA library, which is incubated with the target. Unbound sequences are removed and bound sequences are eluted. The eluted sequences are PCR amplified and the dsDNA library is converted back into a ssDNA library by single-strand displacement (SSD). The enriched library is then submitted to the next SELEX cycle. The sequence information of the final identified aptamer is determined by sequencing.<sup>284</sup>

In brief, the SELEX process to select DNA aptamers uses a randomized ssDNA oligonucleotide library as a starting point. The random regions are flanked by consistent primer binding sites at the 3'- and 5'-ends to allow easy and simultaneous amplification of all oligonucleotides within the library. In the first step, the oligonucleotides are incubated with the immobilized target and unbound sequences are removed. Second, bound sequences are eluted and amplified by PCR, resulting in an enriched dsDNA library. The enriched ssDNA library must then be restored by single strand displacement (SSD) to allow secondary structure formation of the DNA strands in the next SELEX round. Several iterative cycles of increasing evolutionary pressure are performed to identify the oligonucleotide sequence with the highest target affinity. The selected aptamer sequence is then identified by sequencing. In the case of RNA aptamer selection using an RNA library, a reverse transcription step must be included. This allows PCR amplification of the enriched cDNA library, followed by a transcription step to reconstitute the RNA library.<sup>284</sup>

Since the invention of SELEX,<sup>285,286</sup> a plethora of aptamers have been discovered, the process has been optimized, and innovative ideas have been implemented to extend the original method.<sup>287</sup> Although aptamers have several advantages over *e.g.* antibodies, there are still some drawbacks. The biological stability under physiological conditions due to rapid renal filtration and nuclease degradation, as well as their narrow chemical diversity pose common obstacles.<sup>288</sup> To overcome these major drawbacks, there has been a particular interest in the development of tailor-made modified aptamers, as different modifications presented in

Chapters 1.1.1 to 1.1.4 can lead to aptamers with increased stability and enhanced chemical diversity. Typically, these modifications are introduced in a stepwise manner. Initially, unmodified aptamers are selected *in vitro*, and then individual positions of the selected aptamer are systematically modified post-selection by chemical solid-phase synthesis in a laborious process with the objective of fine-tuning the aptamer's properties. Subsequently, the impact of these modifications on the aptamer's binding affinity to the target must be evaluated to ensure that the modification does not interfere with target binding.<sup>288,289</sup> A more straightforward approach to circumvent laborious post-selection modifications is the direct implementation of modifications during the *in vitro* aptamer selection process. However, in order to implement modifications in aptamers during SELEX, the modifications must be compatible with enzymes, and the SELEX process must be adapted to the required conditions.<sup>288</sup> Some selected examples of aptamers that are modified in their phosphate moiety, sugar, or nucleobase are presented below, including a short description of the adapted SELEX methods.

Reported backbone modifications for aptamers mainly include thio-aptamers, which have been designed to improve nuclease stability and binding affinity. In thio-aptamers, the phosphodiester linkage is partially substituted by phosphorothioate and phosphorodithioate groups by using  $\alpha$ S-ATP in SELEX.<sup>290</sup> For example, thio-aptamers have been successfully selected against E-selectin, a cell adhesion molecule expressed in response to inflammatory stimuli,<sup>291</sup> the hyaluronic acid-binding domain of the glycoprotein CD44, which is responsible for cancer metastasis and tumor growth,<sup>68</sup> and the RNase H domain of human immunodeficiency virus (HIV)-1 RT, which is involved in viral proliferation.<sup>292</sup>

A greater variety exists for sugar modified aptamers reported in the literature. Aptamers modified at the 2' position of sugars include 2'F, 2'FANA, and 2'OMe modifications.<sup>293–295</sup> These modifications have been shown to effectively increase aptamer stability and binding affinity.<sup>296</sup> The first two aptamer drugs, Macugen and Izervay, approved by the U.S. Food and Drug Administration (FDA) in 2004 and 2023, respectively, also contain 2'F and 2'OMe modifications in the sugar moiety.<sup>297,298</sup> Macugen targets the vascular endothelial growth factor isoform VEGF165 for the treatment of neovascular age-related macular degeneration.<sup>299</sup> Izervay targets the complementary protein C5 and is used for the treatment of geographic atrophy secondary to age-related macular degeneration.<sup>298</sup> Certainly, 2'-OMe or 2'-F modifications of the sugar backbone have been systematically introduced post-selection, which is costly and time-consuming compared to an exclusively enzymatic selection strategy.<sup>139,300</sup> Therefore, many researchers are focussing on the evolution of polymerases capable of directly incorporating non-canonical substrates during SELEX.<sup>48</sup>

Polymerase evolution is particularly important to enable SELEX with sugar modifications that are more drastic than the 2'-modifications as is the case with XNA. Aptamers based on XNA scaffolds, such as LNA, HNA and TNA aptamers, are of particular interest because of their

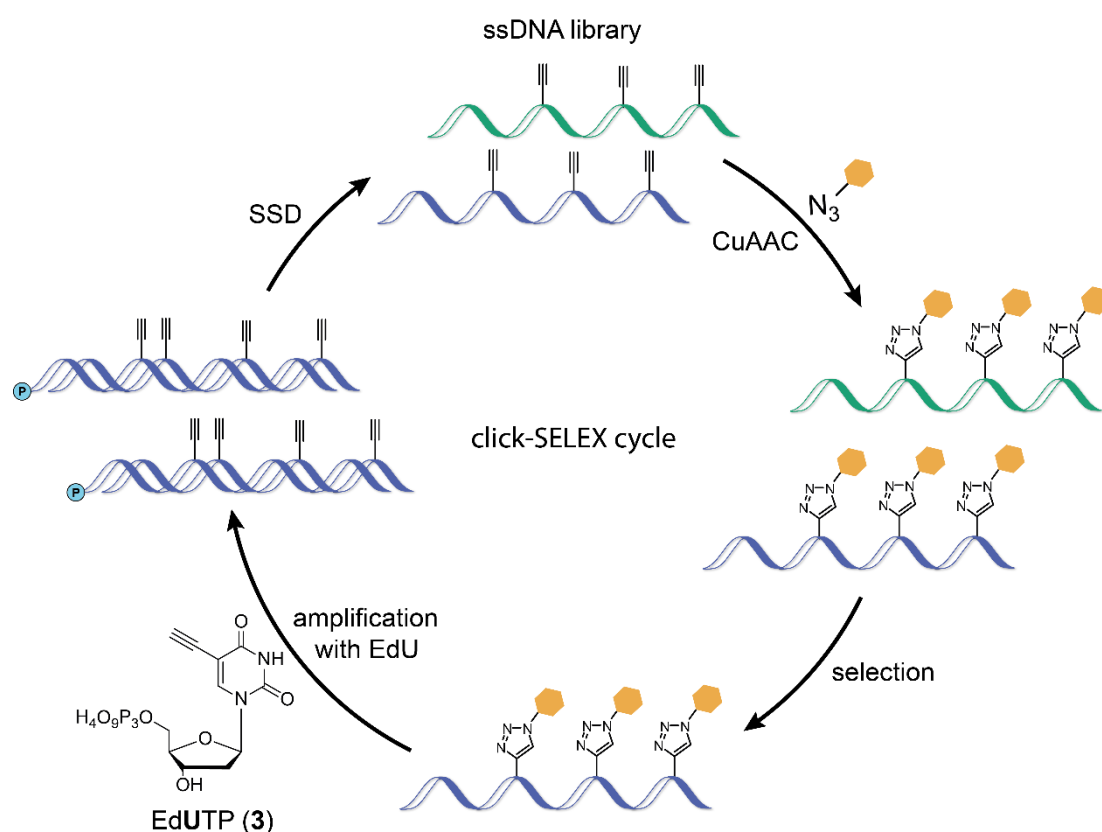
extremely reduced susceptibility to nucleases while maintaining excellent binding affinities.<sup>50</sup> Therefore, enzymes have been developed that allow XNA synthesis from a DNA template and reverse transcription back into DNA (XNA → DNA → XNA) to enable SELEX. Some of these enzymes with an extended substrate scope have already been presented in Chapter 1.1.2. An XNA aptamer partially modified with LNA monomers was selected as a blood coagulant against human thrombin using a SELEX approach. Unfortunately, this approach was rather inefficient due to low efficiency of the replication system.<sup>301</sup> Noteworthy, this approach did not use the engineered LNA polymerase Kod DGLNK and reverse transcriptase Kod DLK to select an LNA aptamer, which could have led to better results. Direct aptamer selection from HNA libraries using SELEX provided HNA aptamers against several targets, including hen-egg lysozyme as a model system and HIV trans-activating response element, a relevant target for anti-HIV therapies.<sup>102</sup> Aptamers have also been selected from TNA libraries against HIV-RT, an enzyme essential for HIV replication, and ochratoxin A (OTA), demonstrating the ability of TNA aptamers to recognize small molecules.<sup>139,148</sup> In addition, a TNA aptamer has been developed to target the cell surface protein programmed cell death ligand (PDL)-1, which is prevalent on cancer cells. The aptamer blocks immune evasion by cancer cells and activates the anti-tumor immune response, resulting in inhibited tumor growth.<sup>302</sup>

An interesting modification of the SELEX strategy to at least avoid the need for XNA reverse transcriptases are SELEX approaches coupled to DNA display methods.<sup>144,303</sup> DNA display SELEX has so far been successfully performed with TNA. In this approach, the DNA template strand contains a short hairpin structure that serves as a self-priming region. Thus, the TNA sequence that is enzymatically appended from this primer is covalently linked to its coding DNA sequence, forming a chimeric DNA/TNA hairpin. The TNA portion of the duplex is then displaced by a DNA sequence by extending a DNA primer that hybridizes in the loop region of the hairpin structure. This produces single-stranded TNA molecules that can fold into three-dimensional structures and can be incubated with the target. By physically linking the genotype (DNA template) to the phenotype (TNA), the DNA template sequence encoding the enriched TNA sequences can be used directly in PCR amplification, eliminating the need for a reverse transcription step.<sup>144,303,304</sup>

Moreover, nucleobase modifications have been extensively used to increase the chemical diversity of aptamers, thereby generating novel binding modalities that allow targeting of previously inaccessible molecules.<sup>289</sup> In the field of functionalized canonical nucleobases, slow off-rate aptamers (SOMAmers) have gained increasing attention.<sup>238</sup> SOMAmers are aptamers bearing protein-like side-chain modifications in the nucleobase moiety, usually attached to the C<sup>5</sup> of uridine triphosphate. Hydrophobic modifications in particular proved to have a positive effect on dissociation rate constants, resulting in high affinity SOMAmers.<sup>238</sup> The development of click-SELEX for the selection of so-called clickmers has taken this principle to the next level:



By incorporating one reactive alkyne group into the aptamer, various post-synthetic modifications can be attached (Figure 19).<sup>305</sup> Click-SELEX uses EdUTP (**3**) instead of dTTP in PCRs, which in turn allows the post-PCR functionalization of the alkyne-modified nucleic acids via CuAAC with any functional group that can be connected to an azide. Click-SELEX differs from regular SELEX in that a functionalization step using CuAAC click chemistry is performed prior to incubation of the ssDNA library with the target. The enriched library is then PCR amplified again using EdUTP (**3**) instead of dTTP, allowing for another round of functionalization and enrichment. With this technique, even large functional groups can be attached to the aptamer without interfering with the PCR amplification.<sup>253</sup>



**Figure 19: Schematic representation of one click-SELEX cycle.** In click-SELEX, an alkyne-modified ssDNA library is used. The library is then functionalized with an azide-conjugated modification using click chemistry (CuAAC). Subsequently, the functionalized oligonucleotides are incubated with the target. The unbound sequences are removed and the bound sequences are recovered (selection). The recovered sequences are amplified by PCR using EdUTP (**3**) instead of dTTP and one 5'-phosphorylated primer (amplification). In the next step, the dsDNA library is converted into an ssDNA library by λ-exonuclease digestion of the 5'-phosphorylated DNA strand (SSD). The resulting ssDNA library can then be submitted into the next click-SELEX cycle.<sup>253</sup>

It is noteworthy that by using the split-and-combine variant of the click-SELEX cycle, screening for clickmers with different chemical entities can be performed simultaneously. This involves dividing the EdU-modified library into aliquots, functionalizing each aliquot with different chemical residues via CuAAC, and then pooling the aliquots for enrichment.<sup>241</sup> Using this technique, a clickmer was selected that recognizes the transplant rejection biomarker CXCL9.<sup>254</sup>

More extensive nucleobase modifications of aptamers to fully augment the chemical diversity can be achieved by genetic alphabet expansion. The introduction of UBPs into aptamers allows the incorporation of novel chemical entities in addition to the canonical nucleobases. This results in even greater chemical diversity compared to modifications attached to canonical nucleobases, as the number of different building blocks is increased to more than four.<sup>306</sup> Due to the increased chemical diversity and novel binding modalities provided by the UBP, aptamers with enhanced affinities can be selected for targets that are challenging for DNA libraries containing only natural nucleobases.<sup>307</sup> Komito *et al.* first reported the successful selection of DNA aptamers bearing five letters employing the well-studied **Ds:Px** UBP, which is amplified with high fidelity and efficiency.<sup>307</sup> Aptamers bearing the **Ds** UB were selected to bind the aforementioned VEGF165 and the cytokine interferon- $\gamma$  (IF- $\gamma$ ), a relevant target for the treatment of hyperinflammation, as well as the von Willebrand Factor (VWF) A1-Domain, which is targeted for the treatment of hematological disorders.<sup>307,308</sup> For the selection of aptamers with an expanded genetic code, so-called genetic alphabet expansion SELEX (exSELEX) is applied.<sup>309</sup> exSELEX is similar to the regular SELEX procedure, but since only **Ds**-containing sequences should be included in the aptamer sequence, the complementary **Px**-containing sequences must be removed after the PCR. Therefore, one primer with a 15 nt overhang is used during PCR, which allows later separation of the **Ds**- and **Px**-containing strands by gel electrophoresis. Another difference to the regular SELEX procedure is the analysis of the sequence information of the selected aptamer. Since the sequencing of unnatural nucleobases is hardly feasible, the sequences from the oligonucleotide library are attached to unique identification tags. These small tags encode the relevant sequence information and the positions of the UBs.<sup>309</sup> To date, aptamers bearing the **TPT3:NaM** base pair have not been prepared using exSELEX. However, Wang *et al.* sought to replace the **Ds** UB in the **Ds**-modified IF- $\gamma$  targeting aptamer originally selected by Hirao's team with the **TPT3** UB. Interestingly, three **TPT3** modifications in the aptamer resulted in increased target binding affinity, highlighting the potential for improved molecular recognition.<sup>310</sup>

Finally, one of the most recent advances in aptamer selection will be presented: the selection of xeno nucleic acid-based aptamers with functionalized nucleobases. This is an impressive example of aptamers with two modification sites, namely the sugar and the nucleobase. These aptamers have been termed threomers and consist of an all TNA backbone with all pyrimidine nucleobases having hydrophobic amino acid-like residues at the  $C^5$  position.<sup>232–234</sup> This approach somewhat combines the concept of XNA aptamers and SOMAmers to generate highly stable and diverse aptamers with remarkable binding affinities. Two threomers were selected using the previously presented DNA display method for SELEX. One threomer was selected against the spike protein subunit 1 of SARS-CoV-2, which is critical for the viral life cycle and thus a potential target for drug therapies. The other threomer was selected against

the Tumor Necrosis Factor alpha (TNF $\alpha$ ), which is relevant for the treatment of inflammatory diseases.<sup>232</sup>

Although only two aptamer-based drugs have been approved to date,<sup>297,298</sup> the field of aptamer research has great potential for drug development that has not yet been exhausted. This potential is represented not only by the active research in progress described above, but also by the numerous aptamers targeting severe diseases such as ocular diseases, cancer, inflammation, and infections that have reached various stages of clinical trials.<sup>311</sup> Besides the therapeutic use of aptamers, targeted drug delivery with aptamer-drug conjugates as well as diagnostic applications for the detection of cognate disease markers are also possible.<sup>312–314</sup> Obviously, the field of aptamer research still leaves many possibilities for further exploration.

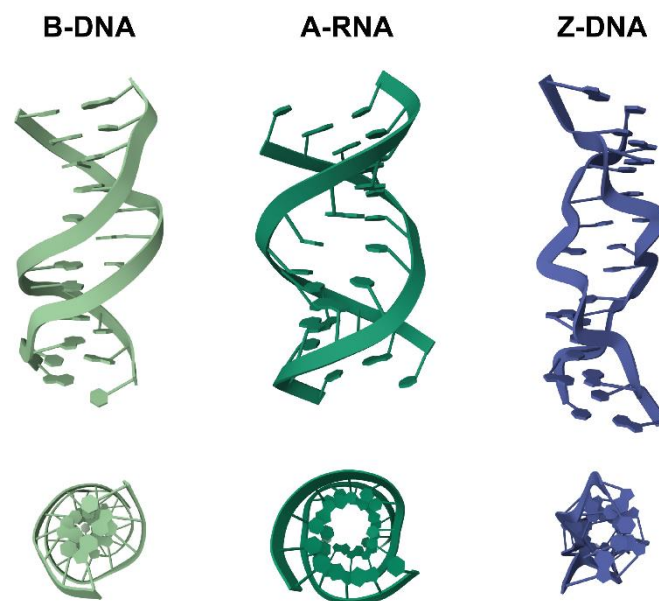
## 1.2 Chemical modifications to alter nucleic acid conformations

In addition to modifications that alter the nuclease stability, duplex stability, and chemical diversity of nucleic acids, modifications that induce conformational changes of nucleic acids can also be installed. In total, there are three naturally occurring nucleic acid conformations, namely the A-, B-, and Z-form, all of which assemble into antiparallel duplexes (Figure 20). Structural features that are crucial for nucleic acid conformations include the conformation of the sugar moiety, *i.e.*, C3'-endo and C2'-endo sugar puckers, and the possible *syn*- and *anti*-conformations of the nucleobases relative to the sugar.<sup>315</sup>

### 1.2.1 Naturally occurring conformations of nucleic acids

Under physiological conditions, dsDNA typically adopts a right-handed B-type helix, with the negatively charged backbone facing outwards and the stacked base pairs oriented inwards (Figure 20, left). The nucleobases are nearly perpendicular to the helical axis, with approximately 10.5 base pairs (bp) per helical turn and a nucleobase distance (axial rise) of 3.4 Å, resulting in a helical pitch of 36 Å. In B-DNA, the sugar conformation is predominantly C2'-endo.<sup>315</sup> Remarkably, DNA can also adopt an A-type helical structure under certain conditions, such as in environments with relative humidity below 75 %, high salt concentrations, or in solutions containing organic solvents.<sup>316,317</sup> The stability of A-DNA and B-DNA in distinct environments can be explained by different binding patterns of cations and water molecules for the two conformations.<sup>318,319</sup> From a structural point of view, the base pairs in A-DNA are shifted to the periphery of the helix, creating a cavity at the center of the helix. Furthermore, each helical turn encompasses 11 base pairs that are tilted approximately 20°

relative to the helical axis, with an axial rise of 2.55 Å, resulting in a helical pitch of 28 Å. The sugar conformation in A-DNA predominantly adopts the C3'-endo conformation. These structural differences result in an A-type helix that is wider in diameter (23 Å) and more compact compared to the narrower (20 Å), more elongated B-DNA helix. In addition, the major groove of A-DNA is deeper and the minor groove shallower than that of B-DNA. In both the A-type and B-type helices, the nucleobases are oriented in the *anti*-conformation relative to the sugar.<sup>315,320</sup>



**Figure 20: Different conformations adopted by nucleic acids.** From left to right: B-DNA, A-RNA and Z-DNA duplexes. Bottom: Top view of the helical center. PDB files: B-DNA: 6CQ3<sup>321</sup>, A-RNA: 413D<sup>322</sup>, Z-DNA: 4OCB<sup>323</sup>.

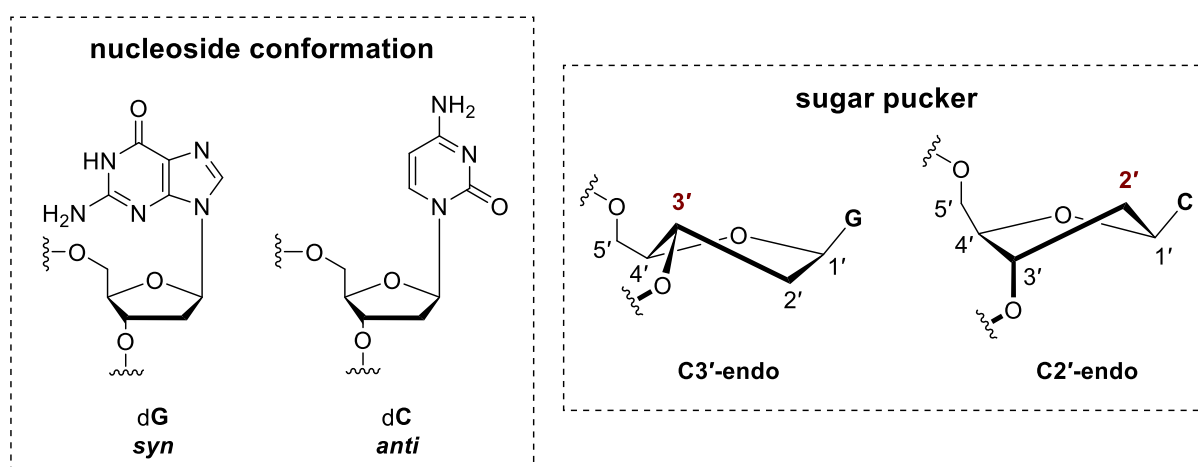
A very similar A-type helix is naturally formed by dsRNA (Figure 20). Similar to A-DNA, A-RNA has a diameter of 28 Å and forms a right-handed helix with a C3'-endo sugar pucker and 11 base pairs per turn. Only the axial rise between two nucleobases is somewhat between the values of B-DNA and A-DNA and amounts to 2.8 Å, resulting in a helical pitch of approximately 31 Å. Thus, A-RNA forms a slightly more elongated duplex than A-DNA.<sup>315</sup>

Z-DNA represents the least common conformation of DNA (Figure 20). The first crystal structure of a double-helical Z-DNA fragment was solved by Rich in 1979.<sup>324</sup> Z-DNA forms a left-handed helix that is even thinner (18 Å) and longer than that of B-DNA. A helical turn consists of approximately 11.6 base pairs with the nucleobases shifted off-axis and an approximate axial rise of 3.7 Å, resulting in a helical pitch of 45 Å. The minor groove of Z-DNA is extremely narrow and deep, while the major groove is almost completely flattened out.<sup>315,325</sup> Interestingly, not only DNA but also RNA can adopt the Z-conformation. Z-DNA and Z-RNA are structurally equivalent, although their naturally occurring B- and A-conformations have significant structural differences. The helical parameters of Z-DNA and Z-RNA are very similar, the only difference being that Z-RNA has a slightly smaller minor groove than Z-DNA.<sup>325</sup> The helical parameters of the different nucleic acid conformations are summarized in Table 1.

**Table 1: Comparison of the helical parameters of the different naturally occurring nucleic acid conformations.**

Helical parameters	B-DNA	A-DNA	A-RNA	Z-DNA	Z-RNA
Diameter	20 Å	23 Å	28 Å	18 Å	18 Å
Helical pitch	36 Å	28 Å	31 Å	45 Å	45 Å
Helical turn	10.5 bp	11 bp	11 bp	11.6 bp	11.6 bp
Axial rise	3.4 Å	2.55 Å	2.8 Å	3.7 Å	3.7 Å
Sugar pucker	C2'-endo	C3'-endo	C3'-endo	C: 2'-endo G: C3'-endo	C: 2'-endo G: C3'-endo
Nucleoside conformation	<i>anti</i>	<i>anti</i>	<i>anti</i>	C: <i>anti</i> G: <i>syn</i>	C: <i>anti</i> G: <i>syn</i>
Handedness	right-handed	right-handed	right-handed	left-handed	left-handed

The Z-conformation of nucleic acids is found primarily in alternating purine-pyrimidine sequences, especially those with high **GC**-content. Its backbone has a characteristic zig-zag pattern caused by alternating *syn*- and *anti*-conformations of the successive purine and pyrimidine nucleobases relative to the sugar.<sup>315,325</sup> Of note, purines are more likely to adopt the *syn*-conformation than pyrimidines due to lower energy barriers. Therefore, alternating purine-pyrimidine sequences are particularly well suited to provide the necessary switching between the *syn*- and *anti*-conformations of the consecutive purine and pyrimidine nucleosides, respectively (Figure 21). Within the purine nucleosides, guanosine has a higher tendency to adopt the *syn*-conformation than adenosine, explaining the preferred adoption of the Z-conformation for **GC**-rich sequences.<sup>325</sup>



**Figure 21: Nucleoside conformations and sugar pucker found in Z-nucleic acids.**<sup>325</sup> Here, the conformational differences are shown for dG and dC in Z-DNA as representative examples for the purine and pyrimidine nucleosides in Z-nucleic acids. Left: In Z-DNA, guanine adopts a *syn*-conformation relative to the sugar and cytosine adopts an *anti*-conformation. Right: The deoxyribose adopts a 3'-endo sugar pucker with guanine as the nucleobase and a C2'-endo sugar pucker with cytosine as the nucleobase.

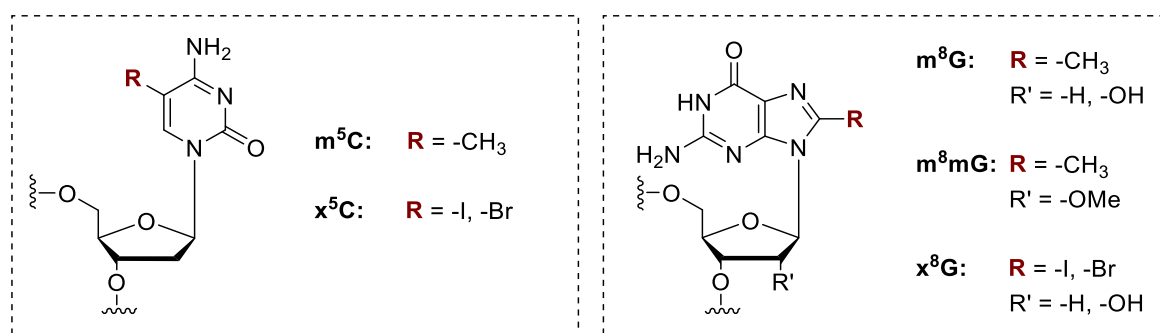
In addition, not only does the conformation of the *N*-glycosidic bond change, but the sugar also adopts a different conformation. For **(GC)<sub>n</sub>** sequences, this means that cytidine adopts an *anti*-conformation and the sugar adopts a C2'-endo conformation, while the complementary guanosine adopts a *syn*-conformation and the sugar adopts a C3'-endo pucker (Figure 21).<sup>325</sup> Due to the *syn*-conformation of **G** relative to the sugar, the *N*<sup>7</sup> and *C*<sup>8</sup> atoms of guanine are facing to the outer side of the molecule, whereas these sites are buried inside the helix in A-RNA, A-DNA and B-DNA.<sup>315,325</sup>

The zig-zag pattern of the Z-nucleic acid backbone certainly causes some phosphate groups to be in closer proximity than in B-DNA, resulting in greater electrostatic repulsion. Thus, *in vitro* Z-nucleic acid formation is promoted by high salt concentrations or even more by polyvalent cations that shield the repulsion of adjacent phosphate groups.<sup>315,325</sup> And indeed, the transition of B-DNA to Z-DNA was first observed by changes in the circular dichroism spectrum of d**(GC)<sub>n</sub>** sequences in high salt solutions.<sup>326</sup> Noteworthy, the formation of Z-RNA requires even more drastic conditions than the formation of Z-DNA.<sup>327</sup>

Nucleic acids in the Z-conformation are relatively difficult to study because they do not form a stable Z-type double helix *in vivo*, but rather a transient one.<sup>328,329</sup> Although high salt concentrations can promote the transition to the Z-conformation, these conditions do not represent those found in physiological environments. Therefore, strategies to chemically manipulate the nucleobases and lock the conformation of nucleic acids in the Z-geometry have been developed.

### 1.2.2 Chemical modifications to stabilize the Z-conformation of nucleic acids

The transition of DNA and RNA conformations from B-to-Z and A-to-Z can be promoted not only by high salt concentrations, but also by covalently attached chemical modifications to the nucleobases, including methylation and halogenation.<sup>325</sup> Different modifications can be applied to induce the transition to the Z-conformation in alternating purine-pyrimidine **(GC)<sub>n</sub>** sequences, which will be presented hereafter.



**Figure 22: Nucleobase modifications at the C<sup>5</sup> position of cytosine and the C<sup>8</sup> position of guanine stabilize the Z-conformation of nucleic acids.** Common modifications that stabilize the Z-conformation of nucleic acids include methylation, iodination, and bromination. Left: 5-Methylcytosine (m<sup>5</sup>C) and C<sup>5</sup>-halogenated cytosine (x<sup>5</sup>C) stabilize DNA in its Z-conformation.<sup>330–332</sup> Right: 8-Methylguanosine (m<sup>8</sup>G), C<sup>8</sup>-halogenated guanosine (x<sup>8</sup>G), and 2'-O-methyl-8-methylguanosine (m<sup>8</sup>mG) stabilize both DNA and RNA in its Z-conformation.<sup>325</sup>

Methylation at the C<sup>5</sup> position of cytosines (m<sup>5</sup>C) can destabilize B-DNA and promote the transition to the Z-conformation of DNA (Figure 22). The introduction of a methyl group increases the hydrophobicity present in the major groove of the nucleic acid, thus destabilizing B-DNA while promoting Z-DNA geometry, which is less perturbed by this modification. Regardless of salt concentration, d(m<sup>5</sup>CG)<sub>n</sub> sequences undergo B-to-Z transitions.<sup>330,331</sup> A similar effect can be observed when halogenation of the C<sup>5</sup> position of cytosine (x<sup>5</sup>C) is used instead of methylation.<sup>332</sup> In contrast, the same modifications<sup>332</sup> in RNA cannot induce the transition of r(m<sup>5</sup>CG)<sub>n</sub> and r(x<sup>5</sup>CG)<sub>n</sub> sequences from A-RNA to Z-RNA and still require high salt concentrations to do so.<sup>331</sup>

In contrast, modifications at the C<sup>8</sup> position of guanine, stabilize the formation of Z-DNA as well as Z-RNA (Figure 22).<sup>325</sup> Similar to modifications at the C<sup>5</sup> position of cytosine, methylation (m<sup>8</sup>G) or halogenation (x<sup>8</sup>G) can be applied to the C<sup>8</sup> position of guanine.<sup>331</sup> In B-DNA and A-RNA, which contain guanosine in the *anti*-conformation, modifications at the C<sup>8</sup> position of guanine are not tolerated by the nucleic acid geometry due to steric considerations. In contrast, the C<sup>8</sup> atom of guanine in the *syn*-conformation, as found in Z-conformation nucleic acids, faces the outside of the helix, minimizing the steric effect of modifications at this position. To accommodate modifications at this position in the duplex structure, the guanosine adopts the *syn*-conformation, thereby promoting A-to-Z and B-to-Z transitions.<sup>315,325</sup> Using this technique, the steric bulk at the C<sup>8</sup> positions virtually locks the guanosine in the *syn*-conformation, stabilizing the Z-conformation to a large extent without the need for high ionic strength.<sup>325</sup>

Methyl groups and halogens can be introduced at the desired positions of guanine and cytosine using two distinct approaches: either by chemical synthesis of modified phosphoramidite building blocks for solid-phase synthesis, or after the oligonucleotide synthesis. The chemical synthesis of the modified phosphoramidite building blocks involves multiple synthetic steps, but allows for the site-specific introduction of modifications during solid-phase synthesis.<sup>333,334</sup> In the post-synthetic modification approach, the native nucleic acid is brominated by incubation

in aqueous saturated bromine solution. However, the major drawback of the latter approach is that it results in random bromination to varying degrees at both the C<sup>5</sup> position of cytosine and the C<sup>8</sup> position of guanine.<sup>335</sup>

A number of studies on Z-conformation-stabilizing modifications are reported in literature. Interestingly, Z-DNA is more stabilized by 8-methylguanosine (m<sup>8</sup>G) than by its deoxynucleoside counterpart, 8-methyl-2'-deoxyguanosine (m<sup>8</sup>dG), due to the introduction of hydrophilic groups in solvent-exposed regions.<sup>334</sup> Moreover, it is not necessary to modify every nucleobase in the oligonucleotide: In the case of m<sup>8</sup>G and 2'-O-methyl-8-methylguanosine (m<sup>8</sup>mG), it has been shown that the selective introduction of one or two modifications already promotes the transition of DNA or RNA to the Z-conformation under physiological conditions.<sup>333,334</sup> Notably, the m<sup>8</sup>mG modification also facilitates the formation of Z-RNA with a sequence containing one AU base pair, which is usually disfavored.<sup>333</sup>

Chemical modifications represent a versatile and powerful tool for stabilizing the Z-conformation of RNA and DNA. The targeted modification during chemical solid-phase synthesis is particularly advantageous, as it is not random as the post-synthetic modification approach. Such Z-conformation stabilizers permit the investigation of Z-RNA and Z-DNA under physiological conditions, which is not possible when using high salt concentrations to stabilize the Z-conformation.<sup>325</sup>

### 1.2.3 Z-DNA-binding proteins in disease

The presence of Z-DNA-binding proteins was first evidenced in *Drosophila nuclei* in 1982.<sup>336</sup> About a decade later, adenosine deaminase acting on RNA 1 (ADAR1) was the first protein identified to contain a Z-DNA-binding domain (Z $\alpha$ ).<sup>337,338</sup> Thereafter, other Z-DNA-binding proteins were identified by alignment of conserved Z $\alpha$  domains.<sup>339</sup> Among these, one Z-DNA binding protein was identified that has a remarkably high sequence similarity to the Z $\alpha$  domain of ADAR1, namely Z-DNA-binding protein 1 (ZBP1).<sup>340</sup> ZBP1 was first found in the stromal tissue of tumor-bearing mice.<sup>341</sup> To date, ADAR1 and ZBP1 are the only Z-DNA-binding proteins identified in mammals.<sup>342</sup>

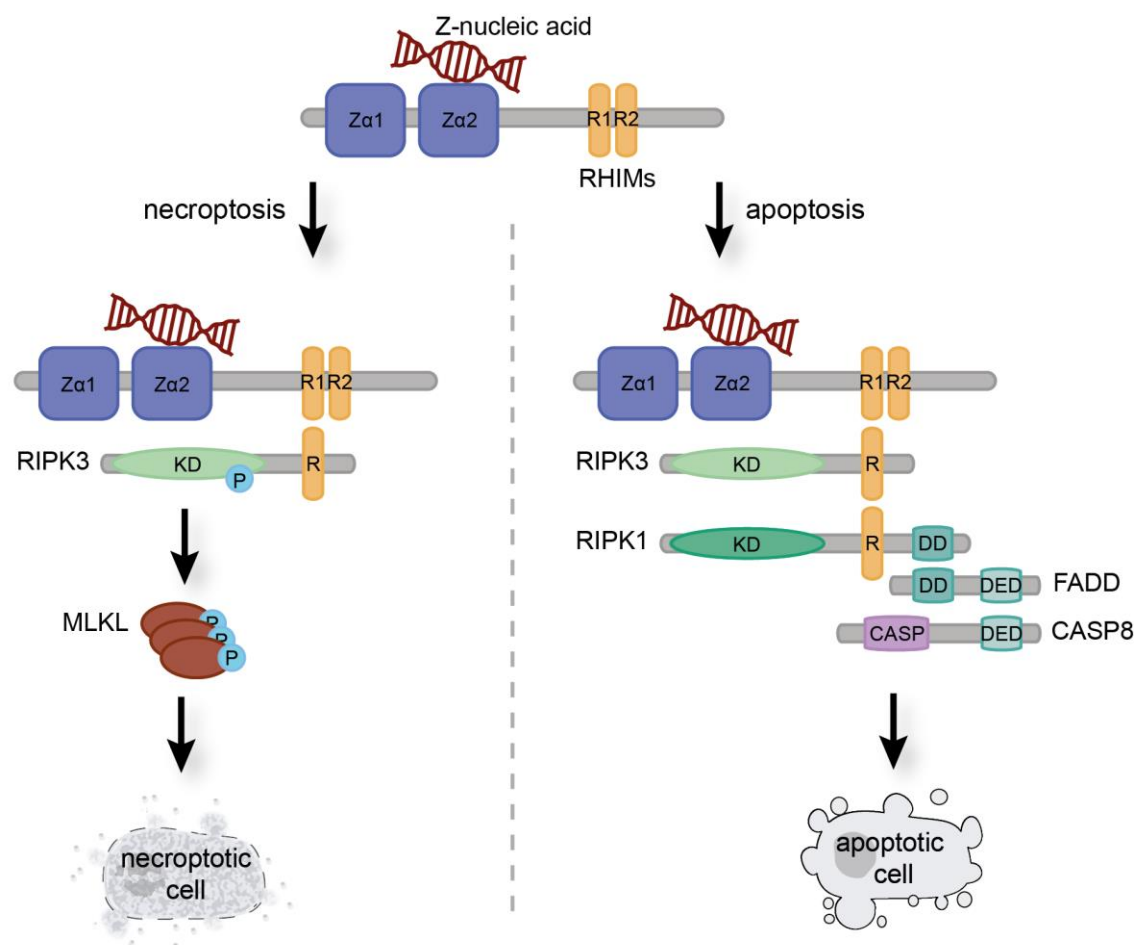
ZBP1 contains two functional Z $\alpha$  domains, termed Z $\alpha$ 1 and Z $\alpha$ 2 (Figure 23). The Z $\alpha$  domains of ZBP1 have been shown to bind both Z-DNA and Z-RNA.<sup>343</sup> In addition to the N-terminal Z $\alpha$  domains, ZBP1 contains two intermediate receptor-interacting protein homotypic interaction motif (RHIM) domains that are responsible for interactions with other RHIM domain-containing proteins.<sup>339</sup> Moreover, ZBP1 carries a C-terminal signaling domain for induction of an interferon-I (IFN-I) response. ZBP1 can shuttle between the nucleus and cytoplasm, and its location within the cell depends on the presence of its Z-nucleic acid (Z-NA) ligand.<sup>344</sup>





**Figure 23: Three dimensional structure of the Z-DNA binding protein 1 from *mus musculus*.** Left: AlphaFold<sup>345,346</sup> predicted structure of the murine ZBP1 with the two Z $\alpha$  domains highlighted in blue and the RHIM domains R1 and R2 highlighted in yellow (AFDB: AF-Q9QY24-F1). Right: CocrySTALLIZATION of the Z $\alpha$ 1 domain (blue) with a six basepair Z-DNA duplex highlighted in red (PDB: 1J75<sup>347</sup>).<sup>340</sup>

Initially, ZBP1 was recognized as an innate immune sensor that triggers the activation of both the IFN-I response and the NF- $\kappa$ B signaling pathway, serving as a defense mechanism against invading pathogens such as viruses.<sup>348,349</sup> Later, ZBP1 was shown to also play a pivotal role in the host's defense against invasive pathogens and tumor cells by programmed cell death.<sup>344</sup>



**Figure 24: ZBP1-mediated cell death pathways upon Z-nucleic acid binding: necroptosis and apoptosis.** In the necroptosis pathway, receptor-interacting serine/threonine-protein kinase (RIPK) 3 is recruited by the RIP homotypic interaction motif (RHIM) domains of ZBP1. RIPK3 is autophosphorylated and then in turn activates mixed

lineage kinase domain-like pseudokinase (MLKL) by phosphorylation with the kinase domain (KD). Activated MLKL oligomerizes and induces necroptosis.<sup>350</sup> In the apoptosis pathway, RIPK3 recruits RIPK1 through the RHIM domain. RIPK1 then interacts with Fas-associated protein with death domain (FADD) through the death domain (DD). FADD in turn recruits caspase-8 (CASP8) through its death effector domain (DED) to induce apoptosis.<sup>344</sup>

Upon recognition of exogenous virus-derived Z-NA by the Z $\alpha$  domains of the host's endogenous ZBP1, ZBP1 mediates the induction of a cell death pathway (Figure 24). Initially, RIPK3 is recruited by ZBP1 through the RHIM domains, resulting in the autophosphorylation of RIPK3. The RIPK3 recruitment then simultaneously activates two forms of cell death: necroptosis and apoptosis.<sup>351</sup> These cell death pathways promote the clearance of pathogens such as viruses and eliminate infected cells. In the necroptosis pathway, RHIM-mediated recruitment of mixed lineage kinase domain-like pseudokinase (MLKL) by RIPK3 occurs, and RIPK3 in turn activates MLKL by phosphorylation with its kinase domain (KD). In the nucleus, activated MLKL oligomerizes and translocates to the nuclear envelope, triggering nuclear envelope disruption and DNA leakage into the cytosol. In the cytosol, activated MLKL oligomerizes and translocates to the plasma membrane to damage its integrity and induce necroptosis, a highly inflammatory form of cell death.<sup>350</sup> In the apoptotic pathway, RIPK3 recruits RIPK1 through their RHIM domains. RIPK1 in turn recruits Fas-associated protein with death domain (FADD) through the common death domain (DD). The conserved death effector domains (DED) in FADD and caspase-8 (CASP8) facilitate their interaction, leading to activation of CASP8. CASP8 ultimately induces apoptosis, which does not elicit an immune response compared to necroptosis.<sup>344</sup> This apoptotic mechanism counteracts the ZBP1-mediated necroptosis pathway, thereby inhibiting inflammation.<sup>352</sup>

Interestingly, advanced *in vivo* knock-out studies of ZBP1-mediated signaling have revealed that the Z $\alpha$ 2 domain is crucial for the initiation of the ZBP1 pathway. While the Z $\alpha$ 2 domain alone has been shown to trigger downstream signaling in several contexts, the Z $\alpha$ 1 domain alone was unable to induce a similar response.<sup>353–357</sup>

Although ZBP1-mediated cell death pathways limit pathogen replication and stimulate the immune response, excessive or aberrant ZBP1-mediated cell death can also lead to undesirable and even lethal inflammation.<sup>344</sup> To name one well-known example: ZBP1-mediated cell death restricts the viral load during SARS-CoV-2 infection, but at the same time causes severe inflammation and lung dysfunction in critical patients.<sup>344,358</sup> In addition, endogenous activation of the ZBP1 signaling pathway can lead to severe chronic and auto-inflammatory diseases.<sup>344,359</sup> On the other hand, endogenous activation of the ZBP1-mediated cell death pathway can also promote tumor death, thereby limiting tumor size and metastasis.<sup>360,361</sup>

Both pathogens and hosts have evolved strategies that antagonize ZBP1 signaling to promote viral infection or sustain immune homeostasis, respectively, by inhibiting the ligand-sensing function of ZBP1 or the ZBP1-RIPK3 interaction.<sup>344,353,360</sup> In the host, this function is performed

by the aforementioned ADAR1 by binding of the Z $\alpha$  domain of ADAR1 to the Z $\alpha$  domain of ZBP1, thereby blocking the downstream interaction with RIPK3.<sup>360</sup>

The ZBP1 signaling pathway is extremely complex and involves many other aspects that are not discussed here. Moreover, much remains to be learned, and further research that supports a better understanding of the ZBP1 pathway may provide concepts for the development of new ZBP1-related drugs for the treatment of infections, inflammatory diseases, and tumor immunity.<sup>339,344</sup>



## 2 Research objectives

This thesis investigates a variety of chemical modifications of nucleic acids for different applications. In this context, four research projects were conceived, all of which share a theme in the functionalization of nucleic acids. The projects either aimed at the development of these functionalized nucleic acids themselves for different applications or at their utilization in advanced studies. The motivations and objectives of the different projects, including the planned experimental implementation, are presented concisely in the order of their appearance in this thesis.

The main project of this thesis aimed to develop an XNA with an expanded genetic alphabet (exXNA) comprising an alternative sugar backbone and three base pairs, one unnatural and two canonical. Such an exXNA has not yet been reported in the literature. The combination of these two modifications has the potential to overcome current drawbacks of natural nucleic acid aptamers, which suffer from narrow chemical diversity and poor stability under physiological conditions.<sup>288</sup> In order to enable exXNA aptamer selection, enzymatic preparation of the novel exXNA is an inevitable prerequisite. Therefore, the objective was to fuse the nuclease-resistant TNA backbone<sup>137</sup> with the unnatural hydrophobic base pair **TPT3:NaM**, developed by the Romesberg group.<sup>209</sup> This would result in TNA with an expanded genetic alphabet (exTNA). The combination of the two modifications was anticipated to provide a greater chemical diversity and enhanced stability of the resulting exTNA compared to canonical nucleic acids. Moreover, this combination of the threose sugar with the **TPT3:NaM** UBP was anticipated to be promising with regard to polymerase-mediated synthesis, as both modifications on their own are reported to be good substrates for various polymerases.<sup>140,141,152,209</sup>

Initially, the objective was to enable the experimental implementation of enzymatic exTNA preparation. To this end, the synthesis of the TNA nucleoside triphosphates required for the polymerase-mediated synthesis of exTNA was aspired. As part of previous research conducted during my master's thesis, the highly modified  $\alpha$ -L-threofuranosyl-based **TPT3** nucleotide and the  $\alpha$ -L-threofuranosyl-based thymine nucleotide were successfully synthesized.<sup>362</sup> Building on this foundation, the synthesis of the highly modified  $\alpha$ -L-threofuranosyl-based **NaM** nucleotide and the three TNA nucleotides bearing the canonical nucleobases cytosine, adenine, and guanine was planned. Following the synthesis of the TNA nucleotides, the investigation of the enzymatic synthesis of exTNA using primer extension assays employing different polymerases was aspired. This necessitates the use of DNA templates containing the corresponding unnatural base, which directs the incorporation of **TPT3-** and **NaM**-bearing TNA nucleotides. Consequently, the synthesis of a deoxyribofuranosyl **TPT3** phosphoramidite was envisaged, as this building block is not

commercially available and is required for **TPT3**-modified DNA template solid-phase synthesis. In order to assess the capability of polymerases to incorporate the **TPT3**- and **NaM**-bearing TNA nucleotides, primer extension assays were aspired to be analyzed by denaturing polyacrylamide gel electrophoresis (dPAGE) and liquid chromatography mass spectrometry (LC-MS).

The second project sought to enhance the fidelity of enzymatic TNA preparation with canonical nucleobases, which is susceptible to the occurrence of **dG/tG** Hoogsteen base pair formation catalyzed by TNA polymerases.<sup>147,148</sup> Currently described solutions to prevent **dG/tG** mismatches involve modifications that sterically block the Hoogsteen base pairing site of the guanine scaffold utilizing 7-deaza-7-substituted guanosine residues.<sup>148</sup> However, this approach yields a TNA sequence containing modified guanine residues, which is not always desired.

Therefore, a cleavable and traceless modification strategy to block the Hoogsteen base pair formation during enzymatic TNA synthesis was envisaged. It was intended to synthesize an  $\alpha$ -L-threofuranosyl guanine nucleotide with a photolabile benzophenone group attached to the  $N^7$  position of the guanine nucleobase. It was planned to investigate whether this modification would block the Hoogsteen base pairing site during enzymatic TNA synthesis, yet allow its subsequent removal by irradiation with light to release the unmodified, native TNA sequence. Moreover, it was contemplated to determine whether the proposed benzophenone-modified  $\alpha$ -L-threofuranosyl guanine nucleotide enhances the fidelity of TNA synthesis by reducing **dG/tG** mismatch formation, thus enabling high-fidelity polymerase-mediated TNA synthesis without permanent base modifications.

The third project aimed to expand the functional density of DNA through the use of click chemistry. More in detail, a novel click-SELEX method based on the iEDDA reaction instead of the CuAAC was conceived. This project was initiated in response to a recent observation that aptamers selected by CuAAC-based click-SELEX fold into their three-dimensional structures in dependence of copper ions, which are not sufficiently removed after the CuAAC click functionalization step of the DNA. The presence of copper ions is likely to result in non-native folding of the aptamer, restricting its applicability. Consequently, the iEDDA reaction was considered a promising alternative, as it is a bio-orthogonal click reaction characterized by fast reaction rates without the need for a copper catalyst.<sup>261</sup> It has been demonstrated in the literature that DNA polymerases are capable of incorporating cyclopropene-modified nucleotides with high efficiency and fidelity.<sup>214,363</sup> Therefore, a small cyclopropene modification was considered a suitable reactive handle that was sought to be attached to the nucleobase via a linker system. This cyclopropene moiety in turn allows the attachment of

tetrazine-conjugated modifications post-amplification to enhance the chemical diversity of DNA.

During the course of this project, it was contemplated to synthesize cyclopropene-modified deoxyribonucleotides bearing adenine and uracil as nucleobases. Furthermore, it was intended to investigate whether the cyclopropene-modified nucleotides indeed serve as substrates for cognate DNA polymerases through the use of primer extension assays, followed by dPAGE and LC-MS analysis. In addition, the utilization of a tetrazine-fluorophore conjugate was envisaged to functionalize cyclopropene-modified DNA and examine the click functionalization step via dPAGE. Moreover, it was planned to evaluate the thermal stability of the reactive cyclopropene modification by nuclear magnetic resonance (NMR) and LC-MS, as it is repeatedly exposed to elevated temperatures over extended time periods during the consecutive SELEX cycles. This was considered an important aspect, given that the literature indicates limited stability of differently substituted cyclopropenes.<sup>364,365</sup> Ultimately, it was aspired to employ the cyclopropene-modified nucleotides in the click-SELEX method with the intention of eventually selecting modified DNA clickmers.

The fourth project aimed to investigate the interaction between ZBP1 and Z-RNA in *in vitro*. This project was motivated by published data indicating that the Z $\alpha$ 2 domain is the primary driver of the ZBP1-mediated necroptosis pathway *in vivo*, while the Z $\alpha$ 1 domain appears to be redundant.<sup>350,353–355,357</sup> Consequently, the functional significance of the two Z $\alpha$  domains in tandem orientation remains elusive and requires further investigations. The following research questions are to be elucidated: (1) Can two tandem Z $\alpha$  domains also bind double-stranded A-RNA and convert it into the Z-conformation? (2) Do both Z $\alpha$  domains isolated from each other have the same affinity to Z-RNA? (3) Can two tandem domains bind Z-RNA with higher affinity? (4) Does mutation of one Z $\alpha$  domain impede binding of the Z-RNA ligand? In order to address the aforementioned questions, it was intended to perform interaction studies utilizing electrophoretic mobility shift assays (EMSA). Therefore, it was conceived to express and purify the wild-type (WT) murine (m) ZBP1 (mZBP1<sup>WT</sup>) and three different variants with loss-of-function mutations in either the Z $\alpha$ 1 domain (mZBP1<sup>mZ $\alpha$ 1</sup>), the Z $\alpha$ 2 domain (mZBP1<sup>mZ $\alpha$ 2</sup>), or both Z $\alpha$  domains (mZBP1<sup>mZ $\alpha$ 1-2</sup>). Moreover, in order to stabilize the Z-conformation of RNA for *in vitro* studies, the use of a chemical modification was contemplated to allow interaction studies under close-to-native conditions. As previously proposed in the literature, the substitution of guanosine with 8-methylguanosine in a **GC**-rich RNA hairpin was considered to stabilize the Z-RNA conformation.<sup>333,334</sup> Initially, it was aspired to comparatively investigate the binding of mZBP1<sup>WT</sup> to Z-RNA and A-RNA to elucidate whether a conversion from A-to-Z conformation can be induced by two tandem Z $\alpha$  domains. In addition, it was intended to investigate the interaction of each Z $\alpha$  domain with Z-RNA in an

isolated manner to elucidate the binding affinity of the individual Z $\alpha$  domains, their interdependence, and their contribution to ligand binding.

The use of this simplified *in vitro* system to study the interaction between ZBP1 and either A- or Z-RNA was expected to provide a more comprehensive understanding of the functional significance of the tandem Z $\alpha$  domains.



### 3 Results and discussion

This chapter presents the findings of four comprehensive research projects, including a detailed discussion of the key results.

Chapter 3.1 outlines the process toward the development of TNA with an expanded genetic alphabet. In this context, the synthesis of various nucleotide building blocks was achieved. Firstly, the synthetic route of the novel  $\alpha$ -L-threofuranosyl-based **NaM** nucleotide is outlined. Secondly, the synthesis of TNA nucleotides bearing the canonical nucleobases cytosine, adenine, and guanine is presented. Thirdly, the synthesis of a novel **dTPT3** phosphoramidite for DNA solid-phase synthesis is reported. Subsequently, investigations were conducted to examine the enzymatic processing of the **TPT3**- and **NaM**-bearing TNA nucleotides by Terminator, Kod-RI, and Kod-RSGA polymerases. In order to achieve this, the cloning, expression, and purification of the latter two polymerases was performed.

Chapter 3.2 presents a traceless approach to prevent **dG/tG** Hoogsteen base pair formation during enzymatic TNA synthesis. The synthetic approach toward a threoguanosine (**tG**) bearing a photolabile benzophenone modification at the  $N^7$  position of guanine, with the objective of blocking the Hoogsteen base pairing site, is outlined in detail.

Chapter 3.3 presents the synthesis of cyclopropene-modified deoxyadenosine and deoxyuridine triphosphates for an iEDDA-based click-SELEX approach. In this context, the processing of the cyclopropene-modified deoxyadenosine triphosphate by Taq DNA polymerase, as well as the subsequent functionalization of the cyclopropene-modified DNA, was demonstrated. Moreover, the thermal stability of the cyclopropene moiety is discussed in detail.

Chapter 3.4 presents the expression and purification of wild-type mZBP1 and three mZBP1 mutants, which were prepared for the use in interaction studies with Z-RNA. Furthermore, the initial findings from preliminary *in vitro* interaction studies between mZBP1<sup>WT</sup> and 8-methylguanosine-modified Z-RNA, in comparison to A-RNA, are outlined.

The overall progress of this work offers significant insights and advancements across the different topics that will likely contribute to future investigations in these research fields.

### 3.1 Towards TNA with an expanded genetic alphabet

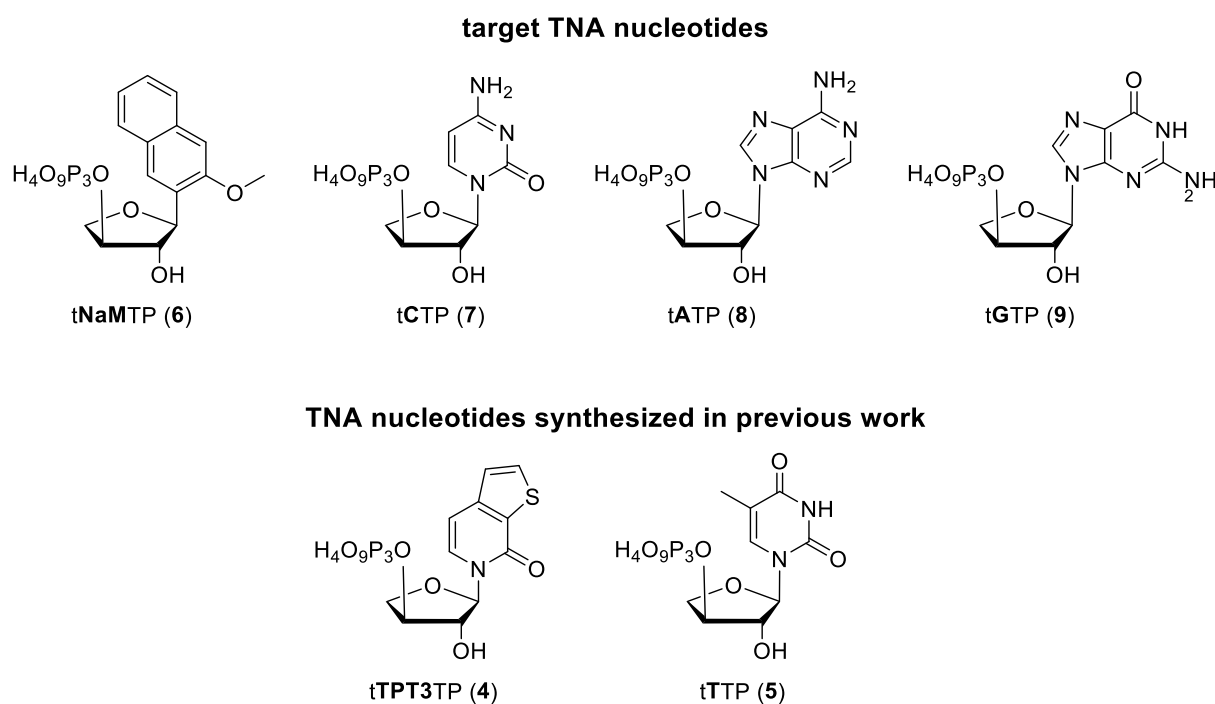
To date, there are only a few examples of artificial nucleic acids with a dual modification pattern, and those that can be prepared enzymatically are particularly rare.<sup>217</sup> Nevertheless, nucleic acids with artificial sugar analogs and modified nucleobases have great potential for the selection of novel nuclease-resistant, high-affinity aptamers. This is due to the excellent stability to nucleases provided by the artificial sugar and the greater chemical diversity owing to the modified nucleobase.<sup>217,288</sup> This thesis thus investigates the possibility of developing a novel and unprecedented artificial nucleic acid based on an XNA scaffold in combination with an extended genetic alphabet that can be produced by enzymatic approaches. To the best of my knowledge, an XNA with a third UBP, namely an expanded genetic alphabet XNA (exXNA), has not yet been described in the literature. Consequently, the nuclease-resistant TNA backbone was fused with the hydrophobic unnatural base pair **TPT3:NaM** to develop an expanded genetic alphabet TNA (exTNA).

This chapter presents the required steps for the development of an exTNA. This includes the synthesis of the highly modified **NaM**-containing TNA nucleotide (Chapter 3.1.1.1), as well as the synthesis of the three TNA nucleotides bearing the canonical nucleobases cytosine, adenine, and guanine (Chapter 3.1.1.2). Furthermore, the synthesis of a **dTPT3** phosphoramidite was performed (Chapter 3.1.2). This is essential for the solid-phase DNA synthesis of **dTPT3**-modified DNA templates for TNA primer extension assays employing the **NaM**-containing TNA nucleotide. Furthermore, the results of TNA primer extension assays towards the preparation of **tTPT3**- and **tNaM**-modified exTNA using different TNA-compatible polymerases are presented (Chapter 3.1.3).

#### 3.1.1 Synthesis of TNA nucleotides

The enzymatic preparation of TNA with an expanded genetic alphabet necessitates the synthesis of TNA nucleotides (tNTPs) bearing the canonical nucleobases, as well as tNTPs bearing the unnatural nucleobases **TPT3** and **NaM** (Figure 25).

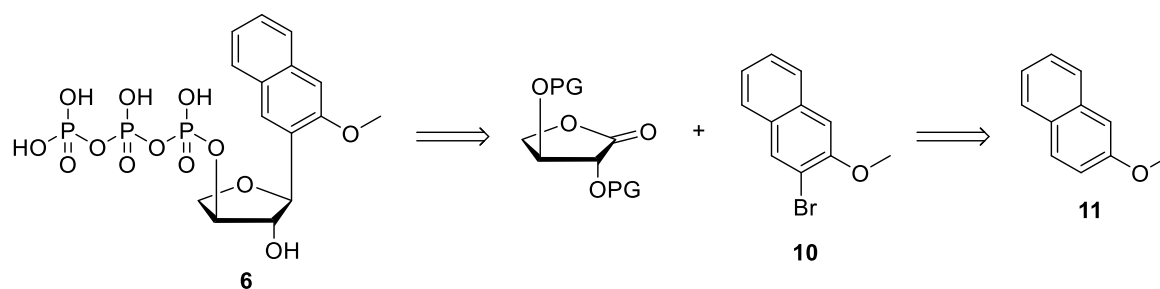
The highly modified  $\alpha$ -L-threofuranosyl **TPT3** nucleoside triphosphate (**tTPT3TP**, **4**) and the  $\alpha$ -L-threofuranosyl thymine nucleoside triphosphate (**tTTP**, **5**) were prepared during my master's thesis and can be readily utilized in TNA primer extension assays.<sup>362</sup> Consequently, the syntheses of the highly modified  $\alpha$ -L-threofuranosyl **NaM** nucleoside triphosphate (**tNaMTP**, **6**), as well as the tNTPs bearing the canonical nucleobases cytosine (**tCTP**, **7**), adenine (**tATP**, **8**), and guanine (**tGTP**, **9**) were performed and are presented hereafter.



**Figure 25: Required TNA nucleotides for the enzymatic preparation of TNA with an expanded genetic alphabet.** Top: Target TNA nucleotides that were synthesized in the framework of this thesis. Bottom: TNA nucleotides that were synthesized during previous work.<sup>362</sup>

### 3.1.1.1 Synthesis of the $\alpha$ -L-threofuranosyl NaM nucleotide

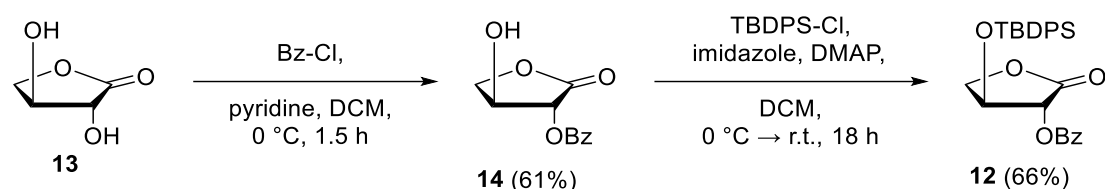
A retrosynthetic analysis was initially conducted with the objective of elucidating the synthesis of **tNaMTP (6)**. The retrosynthetic pathway is depicted in Scheme 1.



**Scheme 1: Retrosynthetic considerations providing access to tNaMTP (6).**

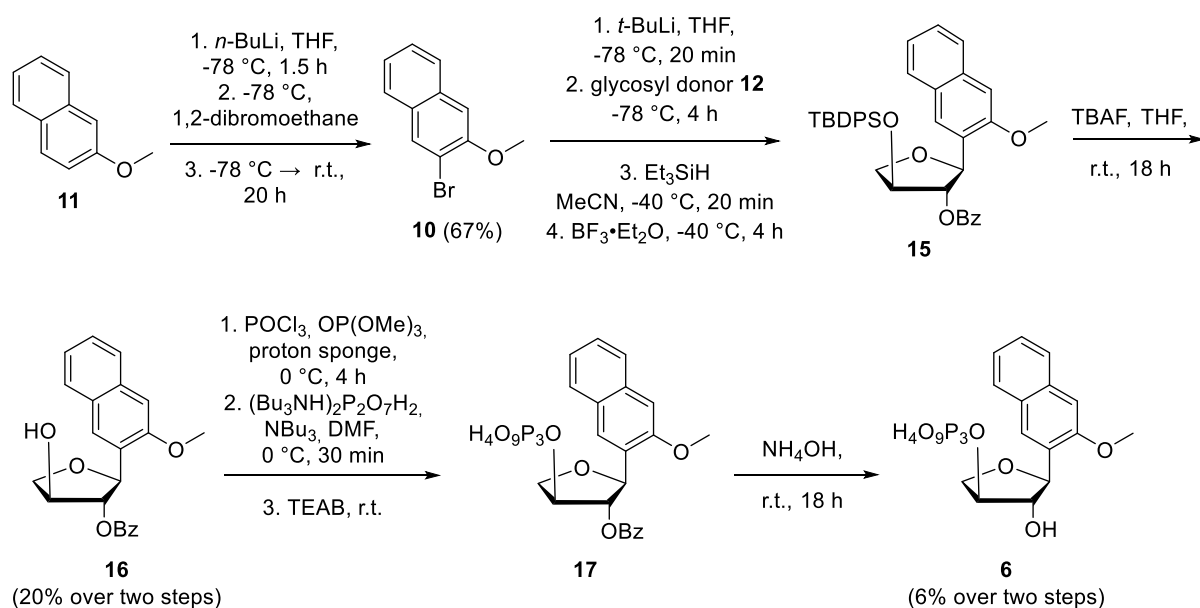
In brief, retrosynthetic considerations entail the disassembly of nucleotide **6** into a glycosyl donor and 2-bromo-3-methoxynaphthalene (**10**). The 2-bromo-3-methoxynaphthalene (**10**) can be further dismantled to give commercially available 2-methoxynaphthalene (**11**) as the starting material. In the forward direction, the bromination of 2-methoxynaphthalene (**11**) in the *ortho*-position was planned to be performed according to a procedure described in the literature by Niimi *et al.* to obtain the nucleobase precursor **10**.<sup>366</sup> With regard to the glycosyl donor, it is important to use orthogonal protecting groups for the 2'- and 3'-hydroxyl functionalities. This enables selective deprotection of the 3'-OH, thereby facilitating the desired

3'-triphosphorylation in the final step. The synthesis of a suitably protected glycosyl donor is literature-known, and its preparation was envisioned according to published procedures.<sup>146,367,368</sup> The establishment of the C-glycosidic bond between the lactone functionality of the glycosyl donor and the 2-bromo-3-methoxynaphthalene (**10**) was envisaged to be performed in a manner analogous to protocols for the preparation of the literature-known ribofuranosyl **NaM** nucleoside.<sup>369</sup> Finally, the 3'-triphosphate was intended to be reintroduced utilizing standard triphosphorylation protocols for nucleosides.<sup>370</sup>



**Scheme 2: Synthesis of glycosyl donor 12.**

To facilitate the synthesis of **tNaMTP** (**6**), the suitably protected glycosyl donor **12** was first prepared according to a previously published procedure by the Chaput group (Scheme 2).<sup>367</sup> This was achieved starting from commercially available  $\alpha$ -L-threonolactone (**13**), using 1.1 equivalents of benzoyl chloride (Bz-Cl) in 1:10 pyridine-dichloromethane (DCM) at 0 °C. Using this procedure, 2-O-benzoyl-L-threonolactone (**14**) was isolated with a yield of 61% after purification, which is consistent with the yield reported in the literature.<sup>367</sup> It is noteworthy that under the given reaction conditions, selective protection of the 2'-hydroxy position of lactone **13** was observed. Subsequently, silylation was conducted at the 3'-hydroxyl position of **14** with *tert*-butyldiphenylchlorosilane (TBDPS-Cl). Thereby, the glycosyl donor **12** was afforded in a 66% yield.



**Scheme 3: Synthesis of tNaMTP (6).**

For the synthesis of **tNaMTP** (**6**, Scheme 3), the nucleobase precursor **10** was first prepared, which can subsequently be coupled to the glycosyl donor **12**. Therefore,

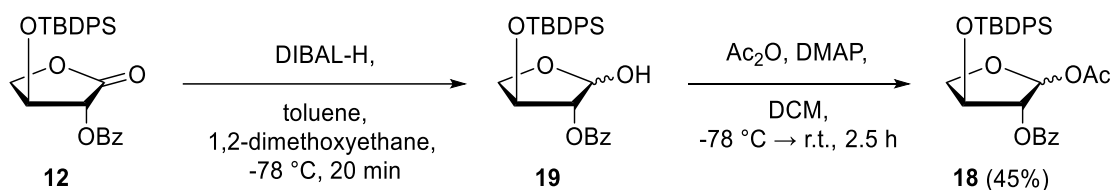
2-methoxynaphthalene (**11**) was converted to 2-bromo-3-methoxynaphthalene (**10**) by bromination in the *ortho*-position of the methoxy group using *n*-butyllithium (*n*-BuLi) and 1,2-dibromoethane. The crude product was recrystallized from *n*-hexanes to afford 2-bromo-3-methoxynaphthalene (**10**) with a moderate yield of 67%. Having obtained both, the glycosyl donor **12** and the nucleobase precursor 2-bromo-3-methoxynaphthalene (**10**), the C-glycosidic bond was established using an adapted protocol from Li *et al.*<sup>161</sup> This was accomplished through *in situ* lithium-halogen exchange of 2-bromo-3-methoxynaphthalene (**10**) with *tert*-butyllithium (*t*-BuLi) and subsequent coupling to the lactone functionality of the glycosyl donor **12**. This was followed by reduction with triethylsilane and boron trifluoride diethyl etherate to generate the protected nucleoside **15**. Subsequently, the TBDPS protecting group of nucleoside **15** was removed with tetrabutylammonium fluoride (TBAF). After purification, nucleoside **16** was obtained with a yield of 20% over two steps, which is in accordance with the literature yield.<sup>369</sup> Nucleoside **16** was then converted to the corresponding triphosphate **17** using a three-step *in situ* reaction cascade based on the work of Yoshikawa and Ludwig.<sup>370–372</sup> This involved the reaction of previously prepared nucleoside **16** with POCl<sub>3</sub>, followed by reaction with tributylammonium pyrophosphate to form a 3'-cyclic triphosphate intermediate. The desired nucleoside-3'-triphosphate **17** was eventually obtained through buffered hydrolysis with 1.0 M aqueous triethylammoniumbicarbonate (TEAB). The crude product was then deprotected at the 2'-position using a saturated (sat.) aqueous (aq.) ammonium hydroxide solution to cleave off the benzoyl (Bz) group. After reversed-phase high-performance liquid chromatography (RP-HPLC) purification, the final product **tNaMTP (6)** was obtained with a yield of 6%.

In conclusion, both the glycosyl donor **12** and the nucleobase precursor **10** were successfully prepared, thus enabling the synthesis of **tNaMTP (6)**. The orthogonally protected glycosyl donor **12** was prepared in two steps with an overall yield of 40%, commencing with the commercially available  $\alpha$ -L-threonolactone (**13**). The nucleobase precursor **10** was generated in one step with a 67% yield, starting from 2-methoxynaphthalene (**11**). Subsequently, with both the glycosyl donor **12** and the nucleobase precursor **10** in hand, **tNaMTP (6)** was successfully synthesized in four consecutive steps with an overall yield of 1%.

### 3.1.1.2 Synthesis of the $\alpha$ -L-threofuranosyl nucleotides bearing canonical nucleobases

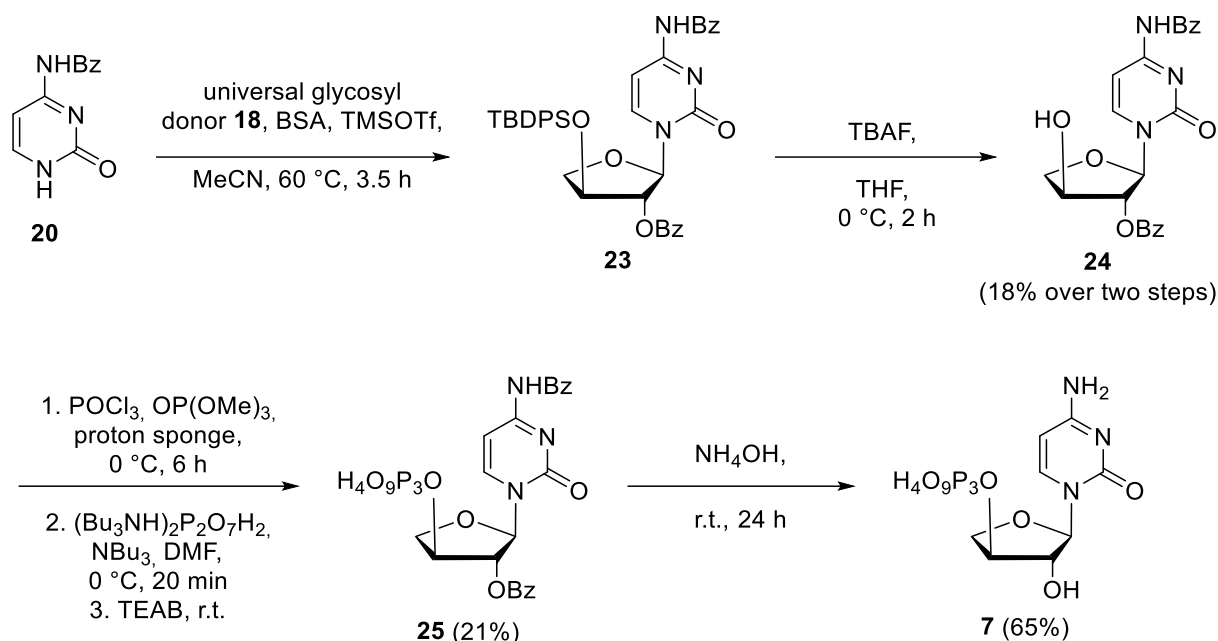
The synthesis of the TNA building blocks bearing the canonical nucleobases has been previously described in the literature using different procedures.<sup>145,146,367</sup> Combining and adapting approaches of these published procedures, the synthesis of the TNA nucleotides was envisioned in a straightforward manner. In brief, this would entail the coupling of the purine and pyrimidine nucleobases to a threofuranosyl donor, followed by subsequent triphosphate synthesis to obtain the TNA nucleotides. Once again, the use of orthogonal protecting groups on the threofuranosyl donor during the synthetic route is crucial for the successful synthesis of the tNTPs. In order to synthesize the TNA nucleotides bearing the canonical nucleobases, the glycosyl donor **12**, which was previously synthesized during the preparation of tNaMTP (**6**), can be utilized once again. However, the glycosyl donor **12** must be converted to its acetylated derivative **18**. This step is necessary, as the acetylated derivative **18** of the glycosyl donor **12** enables the establishment of the *N*-glycosidic bond by Vorbrüggen coupling.<sup>367</sup> This acetylated glycosyl donor **18** can then be used to construct all TNA nucleotides bearing the natural nucleobases, and thus was termed universal glycosyl donor **18**. Furthermore, during the synthesis of the TNA nucleotides bearing canonical nucleobases, a suitable protecting group strategy on the exocyclic functional groups of the nucleobases must be employed to prevent unwanted side reactions. This chapter presents a brief overview of the synthetic pathways employed to generate tCTP (**7**), tATP (**8**), and tGTP (**9**).

Initially, the universal glycosyl donor **18** was prepared in a two-step one-pot reaction starting from previously prepared glycosyl donor **12** (Scheme 4).



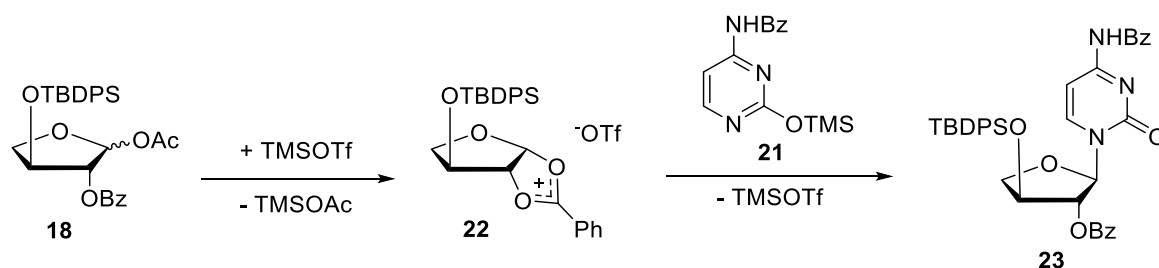
**Scheme 4:** Two-step one-pot synthesis of the universal glycosyl donor **18** starting from glycosyl donor **12**.

According to Sau *et al.*,<sup>367</sup> the lactone functionality of glycosyl donor **12** was reduced to the corresponding lactol **19** using diisobutylaluminium hydride (DIBAL-H). Subsequently, *in situ* acetylation of the anomeric position was performed using acetic anhydride ( $\text{Ac}_2\text{O}$ ) yielding universal glycosyl donor **18** as an anomeric mixture with a 45% yield. Having obtained the universal glycosyl donor **18**, the three desired nucleotides were prepared.



**Scheme 5: Synthesis of tCTP (7).**

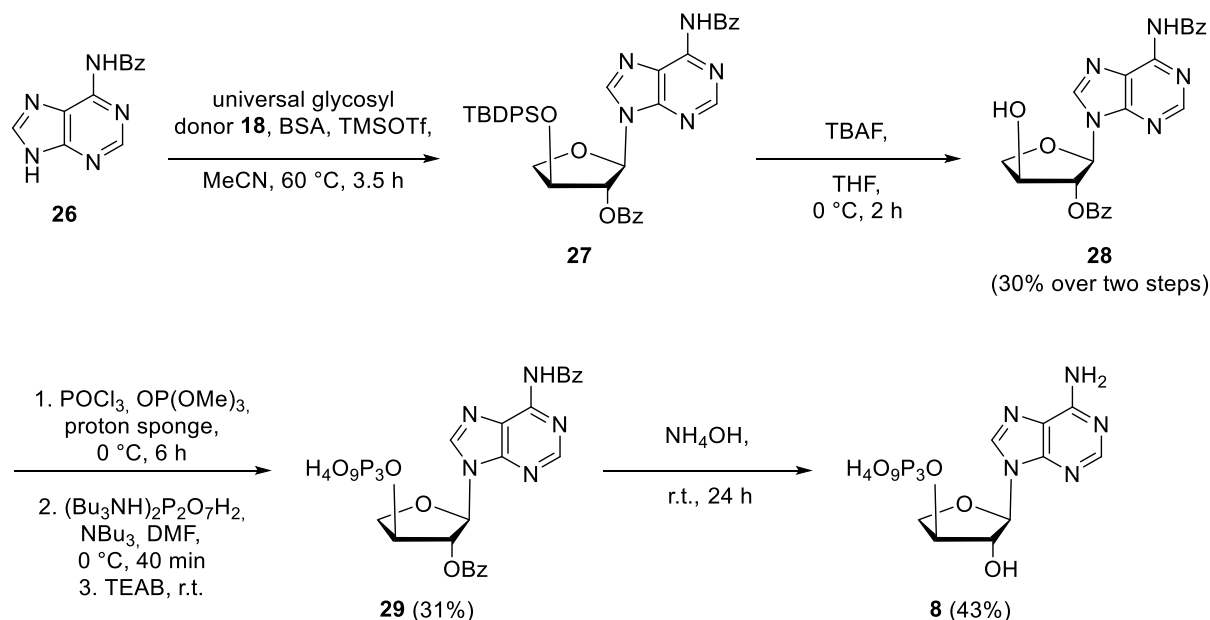
The synthesis of tCTP (7) was adapted from a procedure published by Chaput and coworkers<sup>367</sup> and was accomplished via a four-step synthetic route starting from the previously prepared universal glycosyl donor **18** (Scheme 5). Initially, **18** was reacted with commercially available *N*<sup>4</sup>-benzoyl-protected cytosine **20** under Vorbrüggen glycosylation conditions. This involved heating the threofuranosyl donor **18** and *N*<sup>4</sup>-benzoylcytosine (**20**) in acetonitrile (MeCN) to 60 °C in the presence of *N,O*-bis(trimethylsilyl)acetamide (BSA) to afford the per-silylated nucleobase **21**. Subsequently, trimethylsilyl trifluoromethanesulfonate (TMSOTf) was added as a strong Lewis acid to catalyze the coupling between the acetylated universal glycosyl donor **18** and the silylated nucleobase **21**. Upon addition of TMSOTf, the acetylated universal glycosyl donor **18** undergoes intramolecular cyclization to give 1',2'-dioxolenium ion **22** as an intermediate (Scheme 6). This intermediate blocks the  $\beta$ -face from a nucleophilic attack, which results in high selectivity for the  $\alpha$ -stereoisomer in this reaction.<sup>373</sup>



**Scheme 6: Neighboring group participation causes high selectivity for the  $\alpha$ -stereoisomer during the Vorbrüggen glycosylation.** In the 1',2'-dioxolenium ion intermediate **22**, the  $\beta$ -face is blocked from a nucleophilic attack, resulting in a high selectivity for the  $\alpha$ -stereoisomer.<sup>373</sup>

Subsequently, the crude nucleoside **23** of the previous reaction was treated with TBAF to remove the TBDPS protecting group at the 3'-hydroxy position to give benzoyl-protected cytosine nucleoside **24** with a 18% yield over two steps. Subsequent to this, conversion of **24**

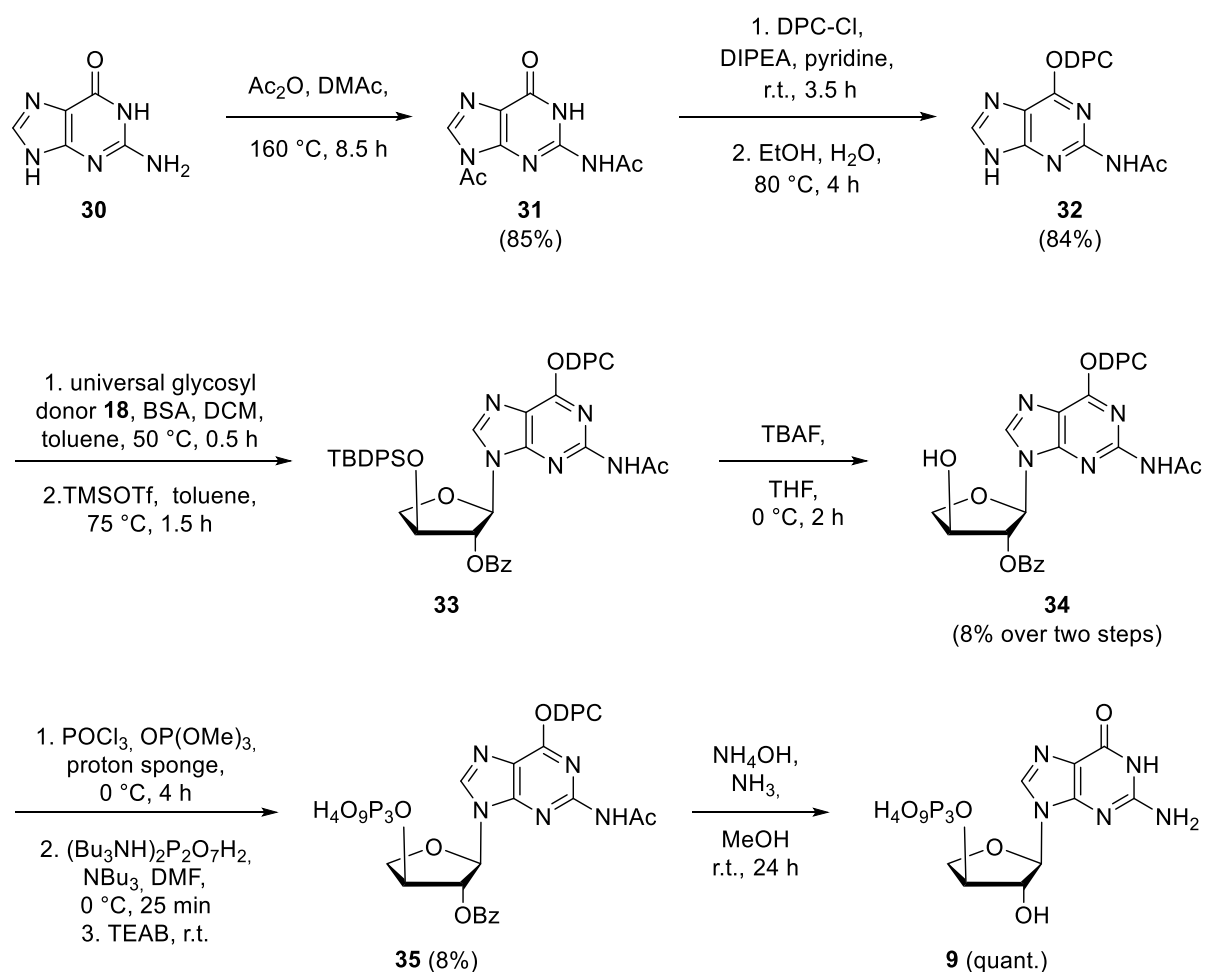
into the corresponding 3'-triphosphate **25** was performed using the same Yoshikawa-Ludwig method as previously presented.<sup>370–372</sup> After RP-HPLC purification, nucleoside 3'-triphosphate **25** was obtained with a 21% yield. Finally, the target triphosphate **7** was obtained by removing the Bz protecting groups at the 2'-hydroxy position and the exocyclic amino group using sat. aq. ammonium hydroxide solution. Thereafter, the target compound tCTP (**7**) was isolated via RP-HPLC, resulting in a 65% yield.



**Scheme 7: Synthesis of tATP (**8**).**

The synthesis of tATP (**8**, Scheme 7) was performed in a manner analogous to the previously described synthesis of tCTP (**7**). Initially, the universal glycosyl donor **18** was coupled to the commercially available *N*<sup>6</sup>-benzoyladenine (**26**) under Vorbrüggen conditions. In this reaction, BSA was again employed as the silylating agent in the first step, while TMSOTf was again subsequently added as the Lewis acid. Aqueous workup gave the crude product of **27**, which was submitted to the next step without further purification. Deprotection of the TBDPS group at the 3'-hydroxyl group was achieved using TBAF, which yielded **28** after purification with a 30% yield over two steps. Following nucleoside formation, triphosphorylation was performed using the Yoshikawa-Ludwig method.<sup>370–372</sup> After RP-HPLC purification, benzoyl-protected 3'-triphosphate **29** was obtained with a yield of 31%. In order to generate the fully deprotected adenosine-3'-triphosphate **8**, the Bz protecting groups were removed using sat. aq. ammonium hydroxide solution. After performing another round of RP-HPLC purification, the desired tATP (**8**) was obtained with a yield of 43%.





**Scheme 8: Synthesis of tGTP (9).**

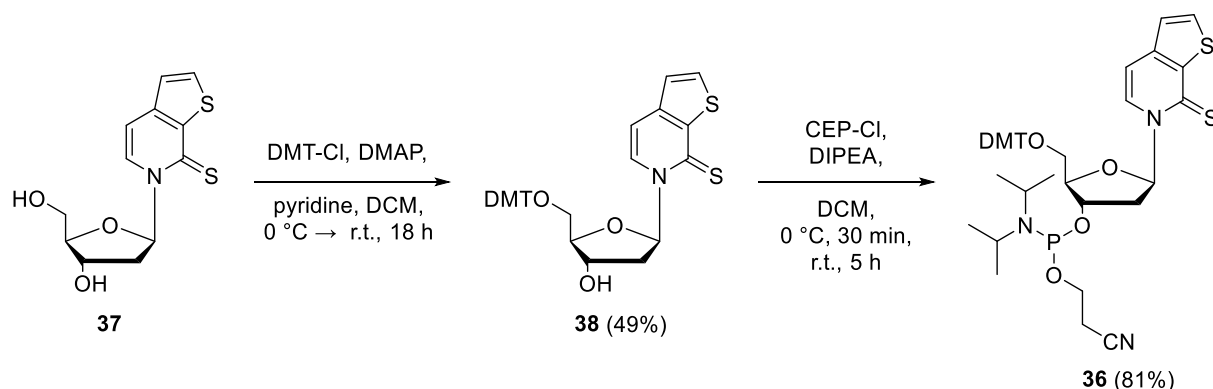
The synthesis of tGTP (**9**) was also adapted from protocols published by the Chaput laboratory (Scheme 8).<sup>367</sup> In order to synthesize tGTP (**9**), protecting groups were initially installed on the guanine nucleobase. In the case of guanine (**30**), it is necessary to protect both the exocyclic amino group and the  $O^6$  oxygen to prevent unwanted side reactions. The protection of the latter is necessary due to the fact that a bulky modification at that position prevents the formation of a glycosyl bond via the  $N^7$  nitrogen by steric hindrance. This facilitates the selective formation of the desired  $N^9$ -regioisomer over the unwanted  $N^7$ -regioisomer during the glycosylation reaction.<sup>374</sup> The installation of protecting groups to guanine (**30**) was carried out according to a procedure from Zou and Robins.<sup>374</sup> In a first step, acetic anhydride was used to attach acetyl groups to the exocyclic amino group and the  $N^9$  nitrogen of guanine (**30**). Therefore, guanine (**30**) was heated in *N,N*-dimethylacetamide (DMAc) and in the presence of acetic anhydride. Upon cooling of the reaction,  $N^{2,9}$ -diacetylguanine (**31**) precipitated out and was afforded with an 85% yield. The bulky diphenylcarbamoyl (DPC) protecting group was then attached to the  $O^6$  oxygen using diphenylcarbamoyl chloride (DPC-Cl) and diisopropylethylamine (DIPEA). Subsequently, the acetyl group at the  $N^9$  nitrogen was selectively removed by heating the residue of the previous step in a mixture of ethanol and water. After cooling of the reaction to ambient temperatures, the desired

*N*<sup>2</sup>-acetyl-*O*<sup>6</sup>-diphenylcarbamoylguanine (**32**) precipitated and was obtained with an 84% yield. The suitably protected guanine **32** was then subjected to the glycosylation reaction with the universal glycosyl donor **18**, applying Vorbrüggen conditions. Analogous to the previous Vorbrüggen reactions, BSA was initially employed as the silylating agent, followed by the addition of TMSOTf as the Lewis acid. However, due to the poor solubility of protected guanine **32** in MeCN, the literature reported the use of a solvent mixture of toluene and 1,2-dichloroethane (DCE) and heating to 75 °C for the silylation step during Vorbrüggen glycosylation.<sup>146,367</sup> However, due to safety concerns regarding the acute toxicity and carcinogenicity of DCE, it was replaced by DCM and the reaction was conducted at 50 °C. Thin layer chromatography (TLC) monitoring of the reaction confirmed complete consumption of the universal glycosyl donor **18**, indicating that the substitution of DCE by DCM may be a viable approach for this reaction. After aqueous workup, nucleoside **33** was obtained as the crude product and was subjected to the next step without further purification. In the subsequent step, the TBDPS group attached to the 3'-hydroxyl group was removed using TBAF. After purification, the pure nucleoside **34** was obtained with a yield of only 8% after the glycosylation and deprotection step, which is significantly lower than the reported literature yield of 69%.<sup>367</sup> The discrepancy between the observed yield and the yield stated in the literature may be attributed to the different solvent system and the lower reaction temperature, which may have influenced the reaction contrary to the initial assumption. Despite the low yield, the obtained amount of nucleoside **34** was sufficient for further triphosphorylation, thus allowing the synthetic route towards the desired target compound tGTP (**9**) to be proceeded. Further investigation of the glycosylation reaction and subsequent desilylation was conducted as part of a side project and is discussed in greater detail in Chapter 3.2.1. Next, nucleoside **34** was converted into the corresponding 3'-triphosphate **35** again using the method based on procedures developed by Yoshikawa and Ludwig.<sup>370-372</sup> After RP-HPLC purification, the 3'-triphosphate **35** was afforded with an 8% yield. Subsequent removal of the protecting groups under basic conditions yielded the target nucleotide tGTP (**9**) in quantitative yield.

In summary,  $\alpha$ -L-threofuranosyl nucleoside triphosphates bearing the canonical nucleobases **C**, **A** and **G** were successfully synthesized. To this end, the glycosyl donor **12** was converted into the universal glycosyl donor **18**, which yielded 45% of the desired compound. Universal glycosyl donor **18** was then employed in subsequent glycosylation reactions to prepare all desired nucleotides. Starting from the universal glycosyl donor **18**, tCTP (**7**) and tATP (**8**) were prepared in four consecutive steps with an overall yield of 3% and 4%, respectively. The synthesis of tGTP (**9**) was achieved in six steps with an overall yield of 4%.

### 3.1.2 Synthesis of the deoxyribofuranosyl TPT3 phosphoramidite

In order to enable primer extension assays to investigate the incorporation of the highly modified t**TPT3**TP (**4**) and t**NaM**TP (**6**) into a nascent TNA strand, it is necessary to synthesize both the TNA nucleotides and the corresponding deoxyribonucleoside phosphoramidites. This is because the required UB-containing DNA templates for TNA primer extensions must be prepared on the solid phase using the UB phosphoramidites. While the **NaM**-bearing deoxyribonucleoside phosphoramidite is commercially available, the **TPT3**-bearing deoxyribonucleoside phosphoramidite is not. Consequently, the **TPT3**-bearing deoxyribonucleoside phosphoramidite must be synthesized in order to facilitate the chemical synthesis of **TPT3**-modified DNA templates and, subsequently, the enzymatic preparation of **NaM**-modified TNA. Based on previously published procedures for the synthesis of other nucleoside phosphoramidites bearing unnatural nucleobases,<sup>206,207</sup> the synthesis of d**TPT3** phosphoramidite **36** was planned in two consecutive steps: Firstly, the d**TPT3** nucleoside **37** was envisioned to be protected at the 5'-OH with a 4,4'-dimethoxytrityl (DMT) group, followed by selective phosphorylation at the 3'-OH to finally generate the suitably protected phosphoramidite **36** for chemical solid phase synthesis of **TPT3**-modified DNA.



**Scheme 9: Synthesis of dTPT3-CEP (36).**

Initially, the DMT-protection of the 5'-hydroxyl group of d**TPT3** nucleoside **37** was performed using DMT-Cl (Scheme 9). After purification, the 5'-DMT-protected nucleoside **38** was obtained in a 49% yield. Subsequently, the phosphorylation was performed using bis(2-cyanoethyl)-*N,N*-diisopropylchlorophosphoramidite (CEP-Cl). The crude product was then purified using a short pad of silica, resulting in the isolation of d**TPT3**-CEP **36** as a diastereomeric mixture with an 81% yield. It is important to note that phosphoramidites are prone to oxidation and hydrolysis. Consequently, special care was taken to work under an inert atmosphere throughout the phosphorylation reaction and subsequent purification process.

The successfully synthesized d**TPT3**-CEP **36** was then handed over to *Ella Biotech* for evaluation of its coupling efficiency. Given the successful outcome of the coupling tests, the

**dTPT3-CEP 36** was employed in chemical solid phase synthesis to prepare **TPT3**-modified DNA templates.

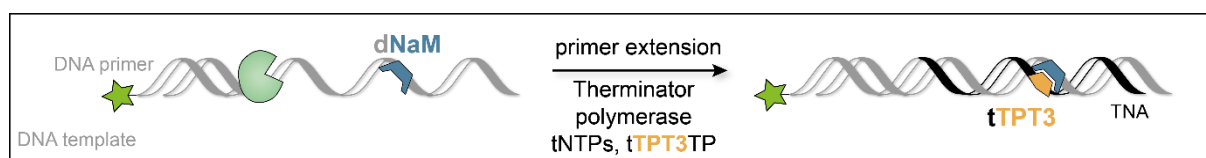
In conclusion, the **dTPT3-CEP 36** was prepared in two steps, with an overall yield of 40%. Moreover, the **dTPT3-CEP 36** was successfully employed in chemical solid phase synthesis, resulting in the production of **TPT3**-modified DNA templates for subsequent TNA primer extension assays. The objective of these assays is to investigate the incorporation of **tNaMTP (6)** into a growing TNA strand.

### 3.1.3 TNA primer extension assays for the preparation of TNA with an expanded genetic alphabet

In the next phase of this project, the synthesized TNA nucleotides bearing the canonical and non-canonical nucleobases were employed in TNA primer extension assays. In this context, **dNaM**- and **dTPT3**-modified DNA templates were employed in primer extension assays to investigate the enzymatic incorporation of **tTPT3TP** and **tNaMTP** into a nascent TNA strand. This chapter presents and discusses the results of primer extension assays for exTNA synthesis using different polymerases and reaction conditions.

#### 3.1.3.1 Towards the enzymatic synthesis of exTNA with Therminator polymerase

Initially, the Therminator polymerase was employed to investigate the incorporation of **tTPT3TP** and **tNaMTP** into TNA. This polymerase has been reported to be an efficient DNA-dependent TNA polymerase<sup>141</sup> and is readily available for purchase by common suppliers. Initial investigations towards the incorporation of **tTPT3TP** into a nascent TNA strand of repeating **tT** residues by primer extension assays (PEx) employing Therminator polymerase were already carried out during my master's thesis.<sup>362</sup> At the time, stalling of the polymerase at the site of the UB occurred and no full-length product was observed even after prolonged reaction times of 24 hours. Consequently, a series of preliminary PEx with varying reaction conditions were performed in order to aim for full-length exTNA products. This included the investigation of different temperatures during exTNA primer extensions, the addition of additives and different **tTPT3TP** concentrations. Using these different reaction conditions, small-scale primer extension assays were conducted (Figure 26), followed by dPAGE analysis. In order to directly track the results of PEx by dPAGE, a 5'-fluorescently labeled primer was used in combination with a fluorescent readout. The DNA primer and template were annealed, and the primer-template complex was then incubated with tNTPs and Therminator polymerase for TNA primer extension.



**Figure 26: Schematic representation of small-scale TNA primer extension assays to prepare tTPT3-modified exTNA with Therminator polymerase.** A fluorescently-labeled DNA primer was extended with a TNA sequence complementary to the NaM-modified DNA template.

Firstly, different temperatures were employed during exTNA primer extension assays. In the literature, temperatures of 55 °C and 75 °C are reported for TNA PEx with Therminator polymerase.<sup>141,144</sup> However, the results obtained by dPAGE showed the presence of truncation products at the site of the UB at both temperatures, with no evidence of full-length exTNA (Figure S1).

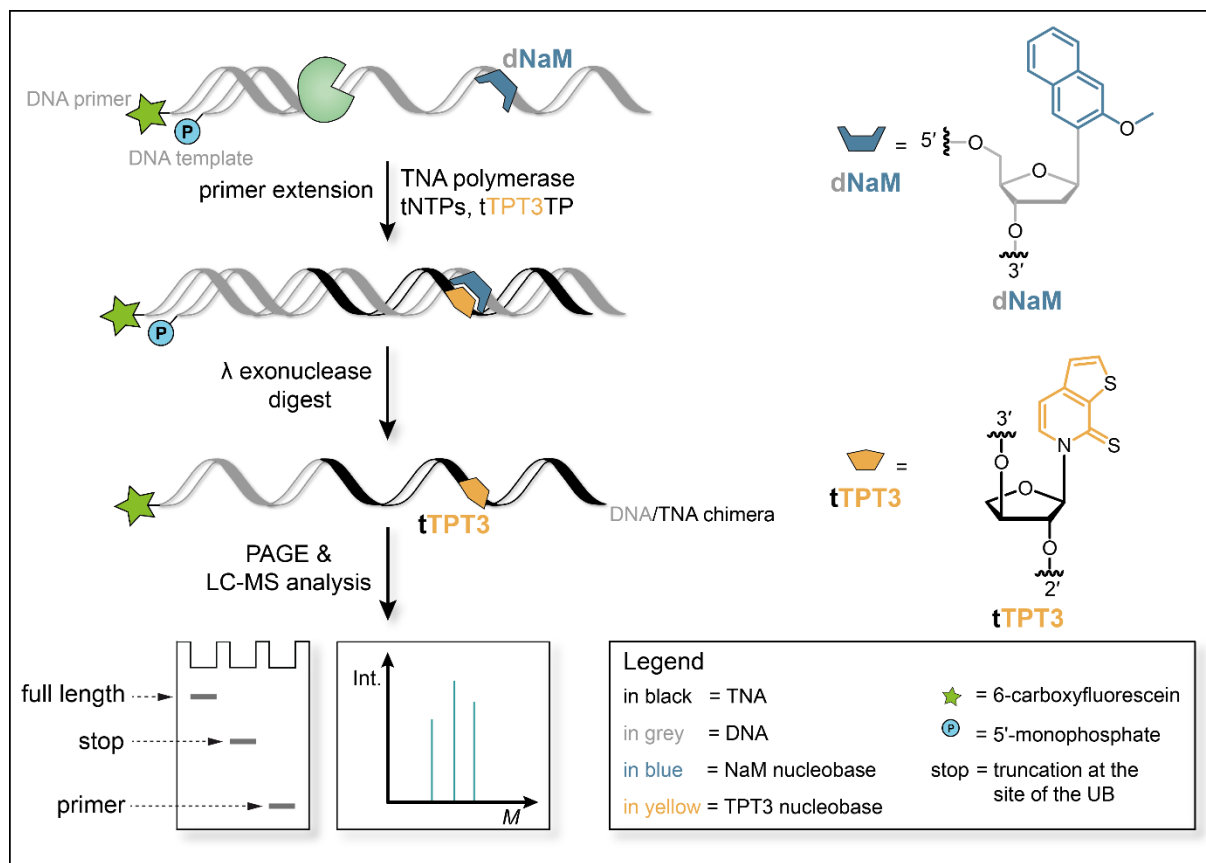
Subsequently, a series of PEx additives were screened to ascertain their impact on exTNA synthesis. These additives included commonly utilized DNA amplification enhancers such as DMSO, spermidine, bovine serum albumin (BSA), and dithiothreitol (DTT). Notably, BSA and DTT have been employed in published protocols for TNA primer extension assays with Therminator polymerase.<sup>143</sup> To elucidate the effect of these PCR enhancers on enzymatic TNA synthesis, different small-scale primer extension assays were set up with different combinations of these additives. The analysis of these PEx by dPAGE again revealed that no full-length exTNA was obtained. However, a slightly higher yield was observed for PEx supplemented with spermidine and DMSO (Figure S2).

Subsequently, the influence of tTPT3TP concentrations on full-length exTNA synthesis was investigated. Therefore, different tTPT3TP concentrations (*i.e.* 0.25 mM, 0.5 mM, 0.75 mM, 1.0 mM) were added to small-scale TNA primer extension assays and the results were evaluated by dPAGE. dPAGE again showed that no full-length product was obtained, regardless of tTPT3TP concentration (Figure S3).

The previously presented approaches, which employed different PEx conditions to obtain full-length TPT3-modified exTNA, were unsuccessful. The dPAGE analysis of the PEx revealed that alterations in the employed conditions did not result in the desired full-length products. However, it is difficult to make statements based on dPAGE whether tTPT3TP was incorporated, resulting in polymerase stalling or whether stalling of the polymerase occurred prior to incorporation of tTPT3TP. To further evaluate this aspect, large-scale PEx were conducted and subsequently analyzed by LC-MS (electrospray ionization, negative mode – ESI<sup>-</sup>).

In order to remove the DNA template prior to the LC-MS measurements, templates with a 5'-phosphate modification were employed from that point forward. This entails the removal of the template from the reaction mixture by  $\lambda$ -exonuclease digestion, thus allowing only the TNA products from PEx to be further analyzed by LC-MS.

Another challenge for the LC-MS analysis of PEx is the occurrence of **dG/tG** mismatches during TNA synthesis, which is a literature-known issue.<sup>138,147</sup> The formation of **dG/tG** mismatches significantly complicates the interpretation of the LC-MS data, as it can lead to a multitude of potential TNA byproducts, rendering a specific molecular weight assignment challenging. Therefore, the subsequent large-scale PEx were conducted using templates comprising the UB nucleosides and solely three of the canonical nucleosides, namely **dA** and **dT**, and either **dG** or **dC**, in order to circumvent the potential formation of **dG/tG** mismatches. More specifically, PEx were conducted using the DNA\_pTemp\_midATG<sup>NaM</sup> and DNA\_pTemp\_midATC<sup>NaM</sup> templates (for the exact sequences, please refer to Chapter 7.5.2). The DNA\_pTemp\_midATG<sup>NaM</sup> sequence is 37 nt in length and contains a **dNaM** modification at position 31. The DNA\_pTemp\_midATC<sup>NaM</sup> sequence is 38 nt in length and contains a **dNaM** modification at position 32.

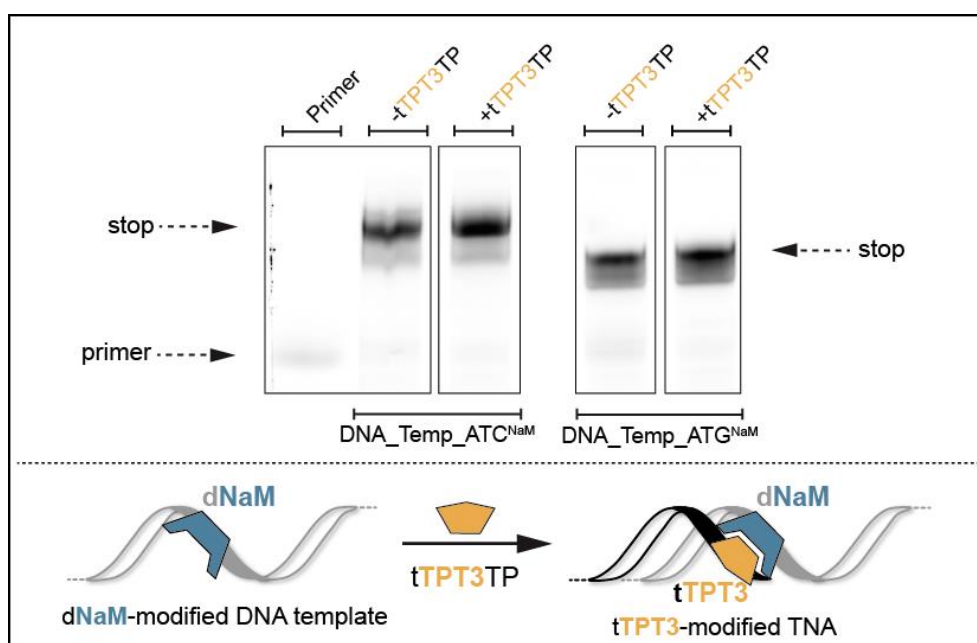


**Figure 27: Schematic representation of the optimized large-scale primer extension assay for the preparation of tTPT3-modified exTNA.** In large-scale primer extension assays, a 5'-fluorescently-labeled DNA primer is elongated with a TNA sequence complementary to the 5'-phosphorylated DNA template. The DNA template is then selectively digested with  $\lambda$ -exonuclease and the resulting DNA/TNA chimera is subjected to dPAGE and LC-MS analysis.

The final set-up for large-scale primer extension assays with subsequent LC-MS analysis involved the annealing of a 20-nucleotide-long 5'-fluorescently labeled DNA primer to the 5'-monophosphorylated DNA templates (Figure 27). Subsequently, the primer-template complexes were incubated for 16 hours at 55 °C in the presence of tTPT3TP, the appropriate

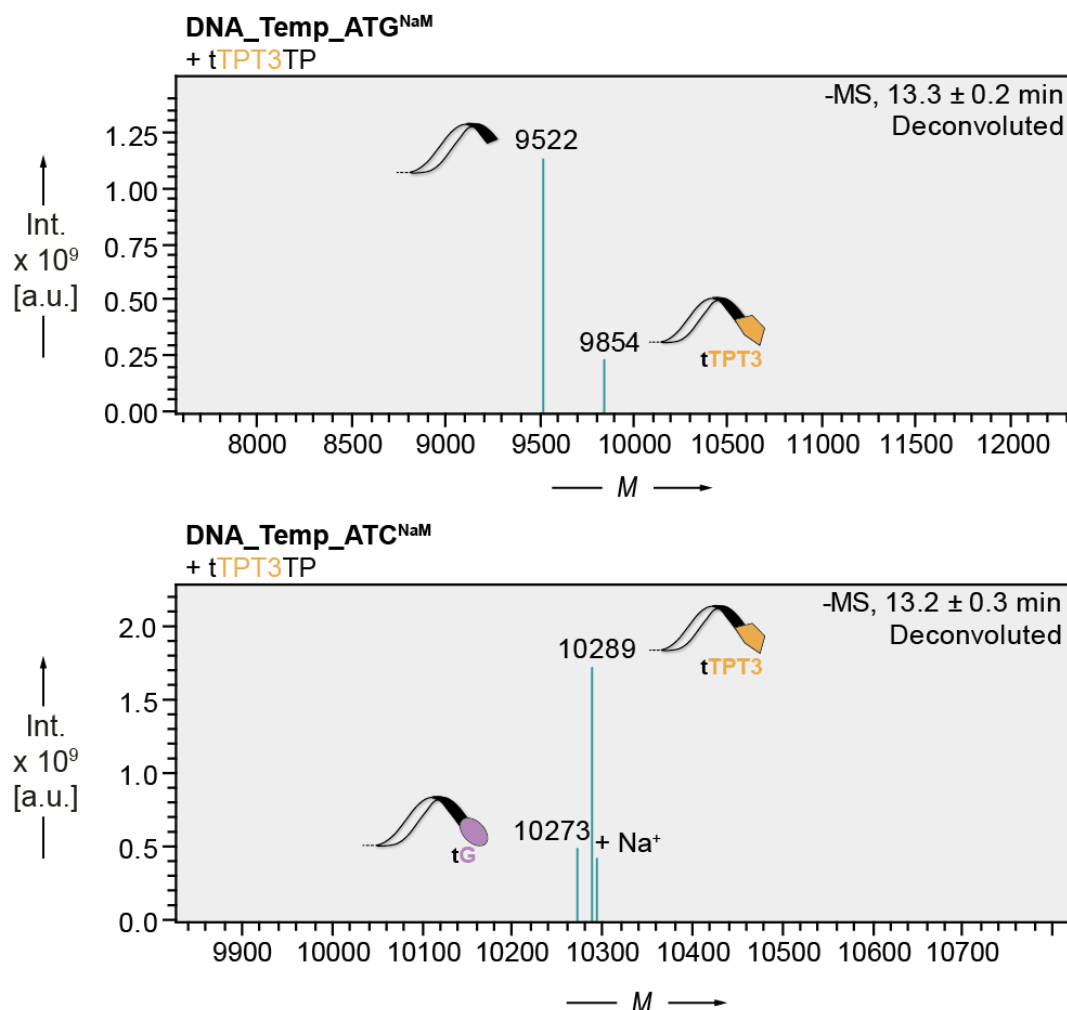
tNTPs dependent on the template utilized, 1.0 mM MnCl<sub>2</sub>, and Terminator polymerase. As a negative control, PEx were performed with the dNaM-modified DNA templates, yet without the addition of the corresponding unnatural tTPT3TP. The negative control serves as a size marker for the stop fragment at the UB position. Prior to LC-MS and dPAGE analysis, the PEx were subjected to digestion with λ-exonuclease to remove the 5'-monophosphorylated DNA templates. The products obtained by PEx with this set-up are DNA/exTNA chimera, in which a TNA sequence is appended to the DNA primer.

As can be observed on dPAGE and in accordance with previous PEx assays, the oligonucleotides obtained from large-scale PEx with tTPT3TP migrated at the same heights as the negative controls (Figure 28). This suggests that the processing of tTPT3TP and the subsequent elongation of the TNA strand did not proceed smoothly.



**Figure 28: dPAGE results of Terminator polymerase-mediated synthesis towards tTPT3-modified exTNA using dNaM-modified DNA templates containing either only dA, dT, and dG or dA, dT, and dC.** Fluorescence scan ( $\lambda_{\text{excitation}} = 477 \text{ nm}$ ,  $\lambda_{\text{emission}} = 535 \text{ nm}$ ) of the 20 % dPAGE analysis of primer extension assays using the DNA\_pTemp\_midATG<sup>NaM</sup> and DNA\_pTemp\_midATC<sup>NaM</sup> DNA templates with the respective TNA nucleotides and tTPT3TP. The negative controls do not contain tTPT3TP. The primer only and negative controls serve as size markers. The complete gel is depicted in Figure S4.

The LC-MS data obtained from these experiments provide corroboration for the observations made on dPAGE. Utilizing the DNA\_pTemp\_midATG<sup>NaM</sup> template, in the presence of tTPT3TP, resulted in molecular weights corresponding to the truncation product before tTPT3TP processing (9522 g mol<sup>-1</sup>) and post-incorporation of tTPT3TP (9854 g mol<sup>-1</sup>; Figure 29, top). As anticipated, the negative control yielded a molecular weight of the stop fragment before tTPT3TP incorporation (9522 g mol<sup>-1</sup>; Figure S5).



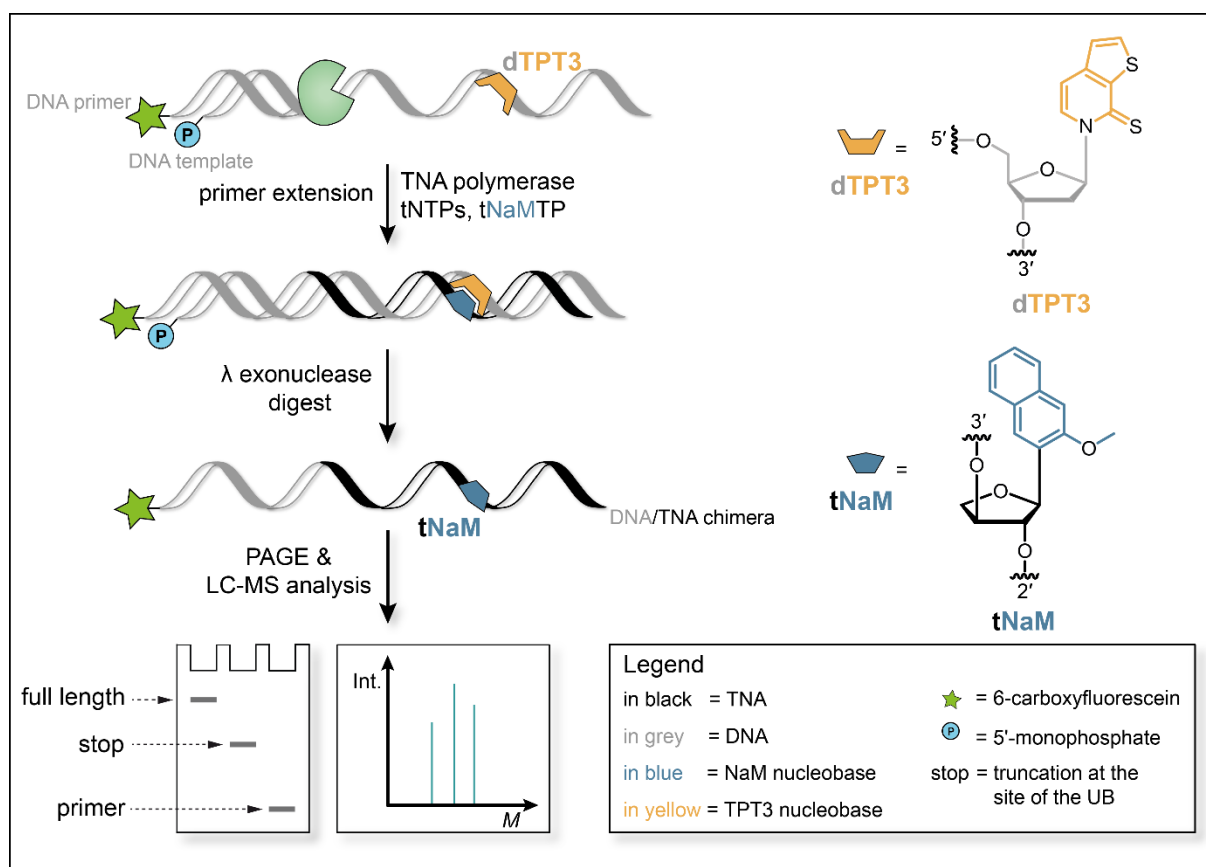
**Figure 29: MS results of Therminator polymerase-mediated synthesis towards tTPT3-modified exTNA using dNaM-modified DNA templates containing either only dA, dT, and dG or dA, dT, and dC.** Top: Deconvoluted ESI<sup>-</sup> mass spectrum of PEx using the DNA\_pTemp\_midATG<sup>NaM</sup> DNA template. The mass spectrum shows the molecular weights corresponding to the stop fragments before and after tTPT3TP incorporation (9522 g mol<sup>-1</sup> and 9854 g mol<sup>-1</sup>, respectively). Bottom: Deconvoluted ESI<sup>-</sup> mass spectrum of PEx using the DNA\_pTemp\_midATC<sup>NaM</sup> DNA template. The mass spectrum shows the molecular weights corresponding to the stop products after incorporation of tTPT3TP (10289 g mol<sup>-1</sup>) and the stop fragment after mispairing of dNaM with tG (10273 g mol<sup>-1</sup>). The complete data set, including ultraviolet (UV) traces and the data of the negative controls, is available in Figures S5 and S6.

It is worth noting that the outcomes were somewhat disparate when the DNA\_pTemp\_midATC<sup>NaM</sup> template was employed. In addition to the molecular weight of the stop fragment following tTPT3TP incorporation (10289 g mol<sup>-1</sup>), misincorporation of tGTP (10273 g mol<sup>-1</sup>) opposite dNaM was also observed (Figure 29, bottom). In contrast, truncation prior to the incorporation of the UB was not observed. The negative control demonstrated the presence of stop fragments following mutagenic readthrough (10273 g mol<sup>-1</sup>, 10572 g mol<sup>-1</sup>; Figure S6), with tGTP being the sole misincorporated substrate observed opposite dNaM. These results indicate that tGTP is the preferred substrate of Therminator polymerase, resulting in its primary misincorporation opposite dNaM. The misincorporation of tGTP



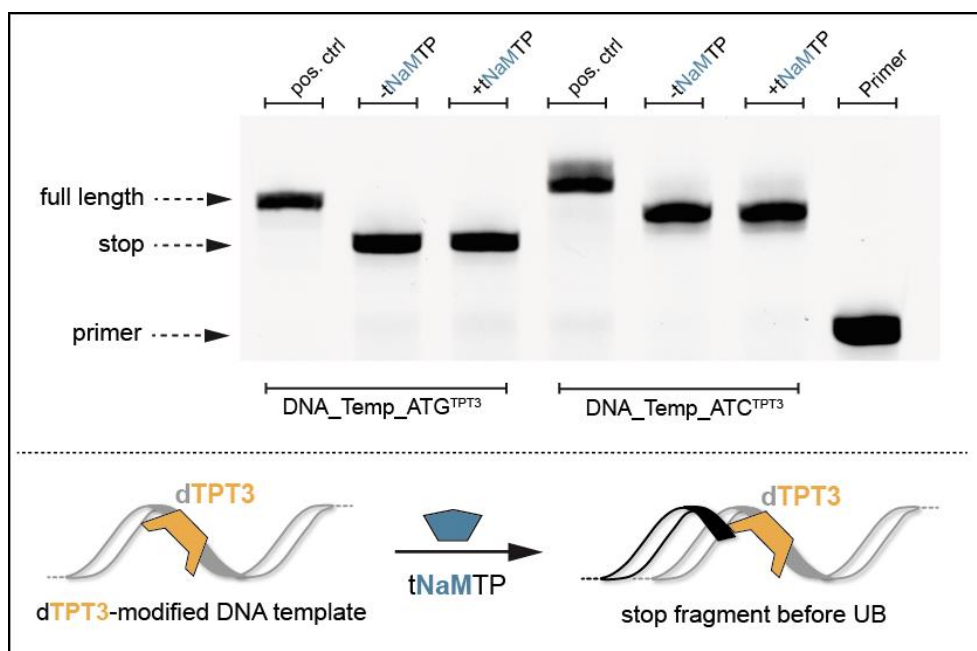
opposite **dNaM** may be attributed to the partial  $\pi$  stacking of the nucleobases in the absence of hydrogen bonds.

The subsequent step was to examine the incorporation of **tNaMTP** by Terminator polymerase. Therefore, the identical templates utilized in the incorporation studies with **tTPT3TP** were employed, albeit bearing a **dTPT3** modification instead of **dNaM** (Figure 30). In more detail, the templates DNA\_pTemp\_midATG<sup>TPT3</sup> and DNA\_pTemp\_midATC<sup>TPT3</sup> were employed (for the exact sequences, please refer to Chapter X). The DNA\_pTemp\_midATG<sup>TPT3</sup> sequence is 37 nt in length and contains one **dTPT3** modification at position 31. The DNA\_pTemp\_midATC<sup>TPT3</sup> sequence is 38 nt in length and contains one **dTPT3** modification at position 32. The previously utilized protocol for large-scale PEx with Terminator polymerase was performed in an analogous manner, using **tNaMTP** instead of **tTPT3TP**. A negative control without the addition of **tNaMTP** was also included. In addition, a positive control primer extension was performed utilizing a DNA template without **dTPT3** and containing exclusively canonical tNTPs. The positive and negative control experiments serve as a size markers for the full-length product and the stop fragment at the site of the UB, respectively. Subsequently, the templates were subjected to digestion by  $\lambda$ -exonuclease, and the resulting DNA/TNA chimera products were evaluated by dPAGE and LC-MS analysis.



**Figure 30: Schematic representation of large-scale primer extension assays for the preparation of **tNaM**-modified exTNA.** In large-scale primer extension assays, a 5'-fluorescently-labeled DNA primer is elongated with a TNA sequence complementary to the 5'-phosphorylated DNA template. The DNA template is then selectively digested with  $\lambda$ -exonuclease and the resulting DNA/TNA chimera is subjected to dPAGE and LC-MS analysis.

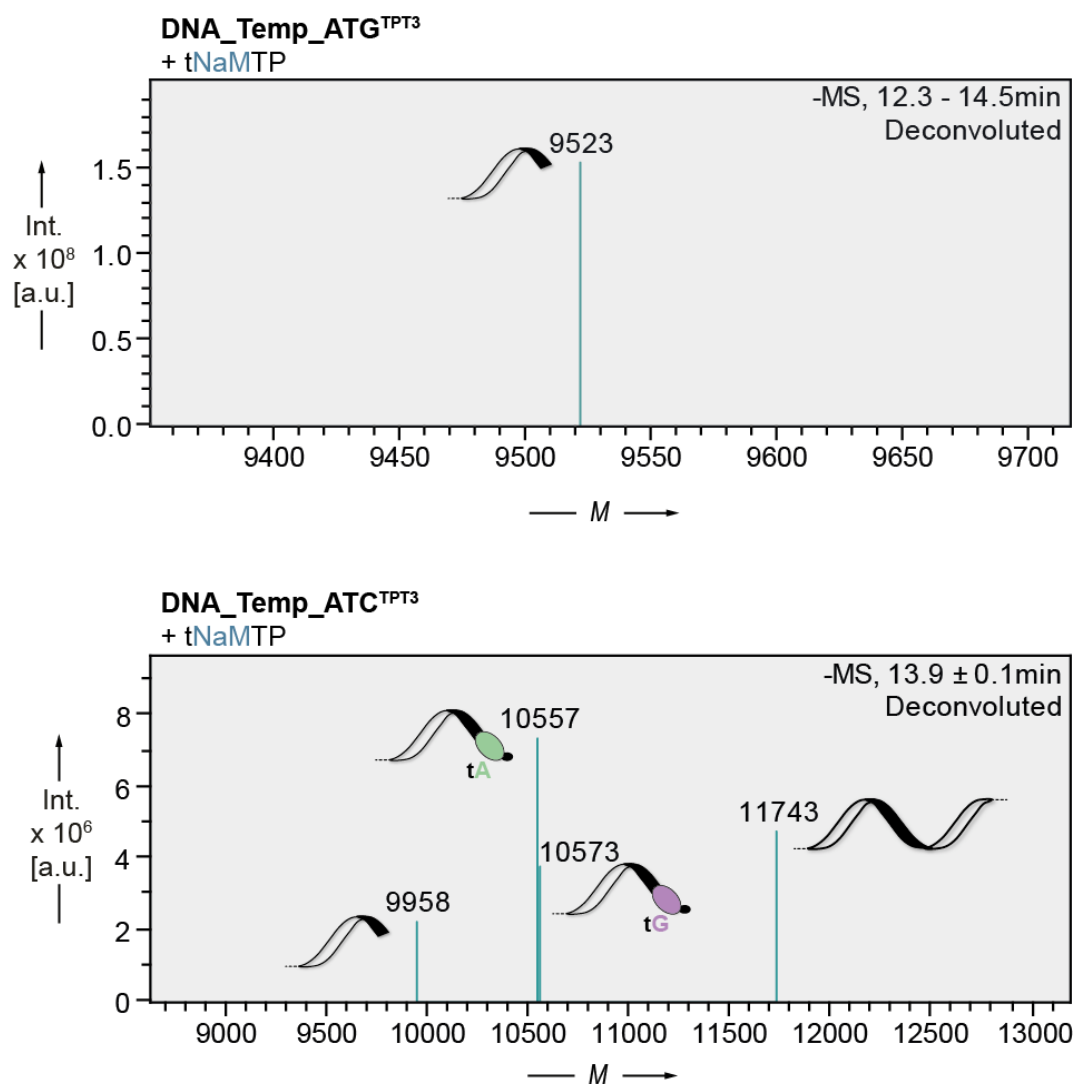
dPAGE analysis of the primer extension assays clearly revealed that no **NaM**-modified full-length product migrating at the same height as the positive control was obtained under the given PEx conditions (Figure 31). The **tNaMTP** containing PEx result in stop products of the same length as the negative controls. When comparing the different PEx sets using either the DNA\_pTemp\_midATG<sup>TPT3</sup> or DNA\_pTemp\_midATC<sup>TPT3</sup> template it is clearly visible that the latter template results in a stop fragment migrating significantly slower compared to the stop fragment obtained by PEx using the DNA\_pTemp\_midATG<sup>TPT3</sup> template.



**Figure 31: dPAGE results of Therminator polymerase-mediated synthesis towards tNaM-modified exTNA using dTPT3-modified DNA templates containing either only dA, dT, and dG or dA, dT, and dC.** Fluorescence scan ( $\lambda_{\text{excitation}} = 477 \text{ nm}$ ,  $\lambda_{\text{emission}} = 535 \text{ nm}$ ) of the 20 % dPAGE analysis of primer extension assays using the DNA\_pTemp\_midATG<sup>TPT3</sup> and DNA\_pTemp\_midATC<sup>TPT3</sup> DNA templates with the respective TNA nucleotides and **tNaMTP**. The negative controls lack **tNaMTP**. The positive controls contain an unmodified template and tNTPs with exclusively canonical nucleobases. The primer only, negative, and positive controls serve as size markers.

To further elucidate the observed difference in the product lengths obtained from the different templates, and to examine whether enzymatic TNA synthesis is aborted before or after **tNaMTP** incorporation, an LC-MS analysis was conducted. In the case of PEx employing the DNA\_pTemp\_midATG<sup>TPT3</sup> template, a molecular weight corresponding to the stop fragment before incorporation of the UB ( $9523 \text{ g mol}^{-1}$ ) was detected, indicating that the Therminator polymerase is unable to process the highly modified **tNaMTP** (Figure 32, top). As anticipated, the same molecular weight was also identified in the negative control of this experiment (Figure S7). LC-MS analysis of PEx using the DNA\_pTemp\_midATC<sup>TPT3</sup> template also revealed a molecular weight corresponding to the stop fragment before **tNaMTP** incorporation ( $9958 \text{ g mol}^{-1}$ ), indicating that the polymerase had stalled at the site of the UB (Figure 32, bottom). Furthermore, molecular weights were observed that correspond to misincorporation of **tG** ( $10573 \text{ g mol}^{-1}$ ) and **tA** ( $10557 \text{ g mol}^{-1}$ ) opposite **dTPT3**, followed by the incorporation of

the next templated **tA** nucleotide (Figure 32). The same results were obtained for the negative control, in which **tNaMTP** was omitted (Figure S8). It appears that the Terminator polymerase exhibits a markedly higher substrate promiscuity when **dTPT3** is present in the template compared to **dNaM**. Moreover, complete skipping of nucleotide incorporation opposite the UB was also observed ( $11743 \text{ g mol}^{-1}$ ) in the negative control (Figure S8).



**Figure 32: MS results of Terminator polymerase-mediated synthesis towards tNaM-modified exTNA using dTPT3-modified DNA templates containing either only dA, dT, and dG or dA, dT, and dC.** Top: Deconvoluted ESI<sup>-</sup> mass spectrum of PEx using the DNA\_pTemp\_midATG<sup>TPT3</sup> DNA template. The mass spectrum shows the molecular weight corresponding to the stop fragment prior to **tNaMTP** incorporation ( $9523 \text{ g mol}^{-1}$ ). Bottom: Deconvoluted ESI<sup>-</sup> mass spectrum of PEx using the DNA\_pTemp\_midATC<sup>TPT3</sup> DNA template. The mass spectrum shows the molecular weights corresponding to the stop fragment before **tNaMTP** incorporation ( $9958 \text{ g mol}^{-1}$ ) as well as products resulting from mispairing of **dTPT3** with **tG** ( $10573 \text{ g mol}^{-1}$ ) and **tA** ( $10557 \text{ g mol}^{-1}$ ) and further extension with the next nucleotide and skipping of the UB ( $11743 \text{ g mol}^{-1}$ ). The complete data set, including UV traces and data of positive and negative controls, is available in Figures S7 and S8.

In conclusion, the presented results demonstrate that Terminator polymerase is unable to synthesize full-length exTNA bearing **tTPT3** or **tNaM**. LC-MS analysis revealed that the incorporation of **tTPT3TP** opposite **dNaM** was to a small extent catalyzed by Terminator

polymerase. Nevertheless, a considerable quantity of truncation product prior to the incorporation of the UB was observed. Moreover, the extension of the nascent TNA strand after **tTPT3TP** incorporation was not catalyzed by Terminator polymerase. In addition, the data obtained revealed that **tGTP** is predominantly misincorporated opposite **dNaM**. In contrast, the incorporation of **tNaMTP** opposite **dTPT3** was not catalyzed by Terminator polymerase. In absence of **tGTP** in PEx, only a stop product was observed prior to the incorporation of **tNaMTP**. In the presence of **tGTP** in PEx, misincorporation of **tG** and **tA** opposite **dTPT3** was evident. This leads to the hypothesis that **tGTP** influences the processivity of Terminator polymerase.

Previous studies on the incorporation of unnatural base pairs into DNA by Terminator polymerase, conducted by the Romesberg laboratory, yielded similar results.<sup>208</sup> For instance, the incorporation of the **dMMO2:d5SICS** UBP, a progenitor of the **TPT3:NaM** UBP, was observed to be catalyzed by Terminator polymerase. Nevertheless, the subsequent extension of DNA after UB incorporation was also less efficient than the processing of the UB itself. The Romesberg group also reported relatively low fidelities for the synthesis of DNA bearing **dMMO2:d5SICS** by Terminator polymerase. The most efficiently formed mispair was **d5SICS/dG**, with **d5SICS/dA** and **d5SICS/dT** being only slightly less efficiently formed.<sup>208</sup>

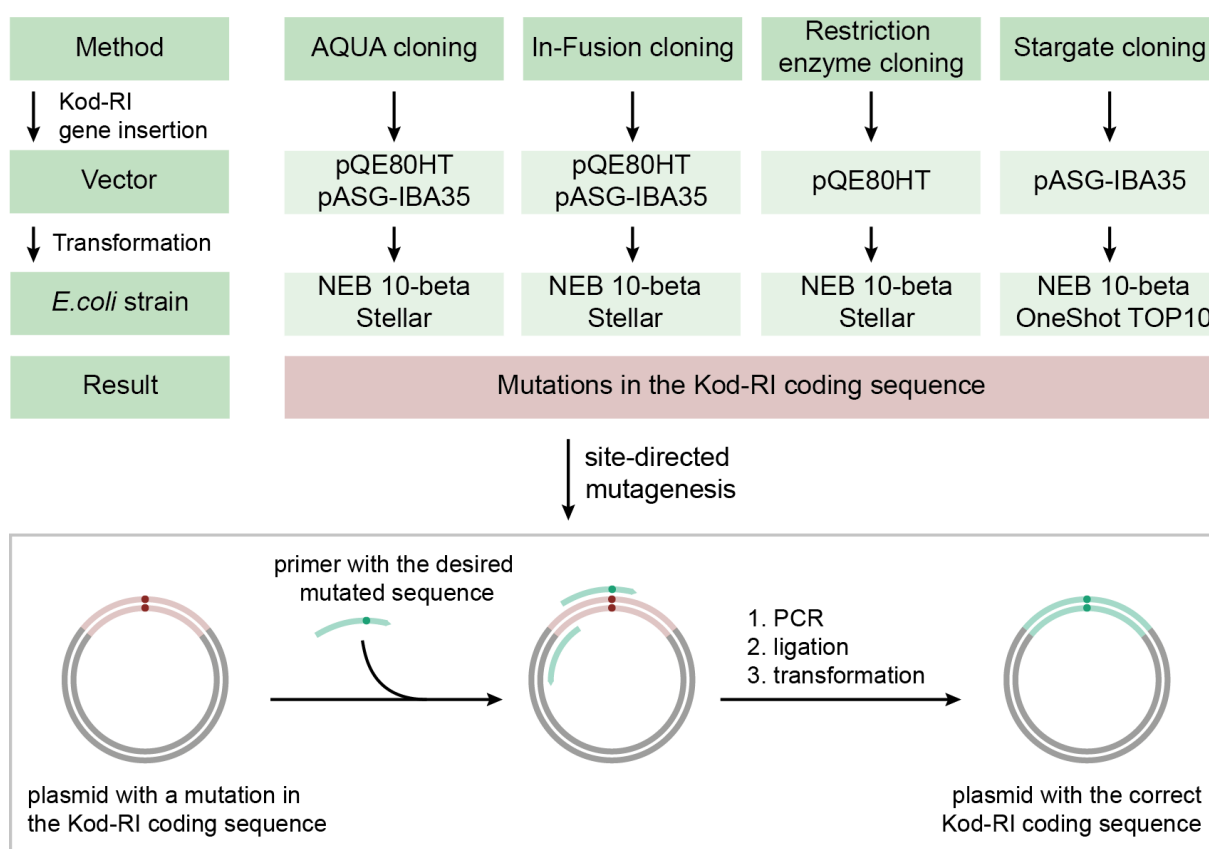
The findings of this chapter highlight the necessity for further research into the enzymatic preparation of exTNA. Despite testing a range of conditions during PEx, the incorporation of both TNA nucleotides bearing the UBs **TPT3** and **NaM** was not sufficiently accommodated by Terminator polymerase. Consequently, subsequent investigations involved the testing of another polymerase with the potential to facilitate exTNA preparation.

### 3.1.3.2 Towards the enzymatic synthesis of exTNA with Kod-RI polymerase

Initial investigations on the incorporation of **tTPT3TP** and **tNaMTP** by the commercially available Terminator TNA polymerase did not yield the desired full-length exTNA. The Terminator polymerase demonstrated a consistent tendency to stall at the site of the UB, predominantly before processing of the highly modified substrates **tNaMTP** and **tTPT3TP**. Given that employing distinct reaction conditions during PEx was insufficient to enzymatically synthesize full-length exTNA with the Terminator polymerase, it was concluded that substituting the polymerase appeared to be a more promising approach. In this context, the engineered Kod-RI TNA polymerase was considered an interesting candidate, as this polymerase is reported to be 5-fold faster and 20-fold more efficient in TNA synthesis compared to the Terminator polymerase.<sup>140,151</sup> Therefore, it was planned to investigate the

potential of the Kod-RI polymerase to process the highly modified t $\text{TPT3TP}$  and t $\text{NaMTP}$  nucleotides.

The Kod-RI polymerase was evolved in the Chaput laboratory and specifically optimized for efficient TNA synthesis.<sup>140</sup> Thus, this tailor-made polymerase is not commercially available. Consequently, the Kod-RI polymerase had to be expressed and purified in-house. The Kod-RI polymerase coding sequence was derived from the published protein sequence,<sup>151</sup> codon-optimized for expression in *E.coli*, and commissioned for synthesis by *Thermo Fisher GeneArt*. In order to enable protein expression, it was necessary to clone the Kod-RI gene into an appropriate expression vector. However, the cloning of the Kod-RI gene into an expression vector proved to be challenging. Consequently, a variety of strategies were employed to construct the desired expression plasmid harboring the Kod-RI coding sequence (Figure 33).



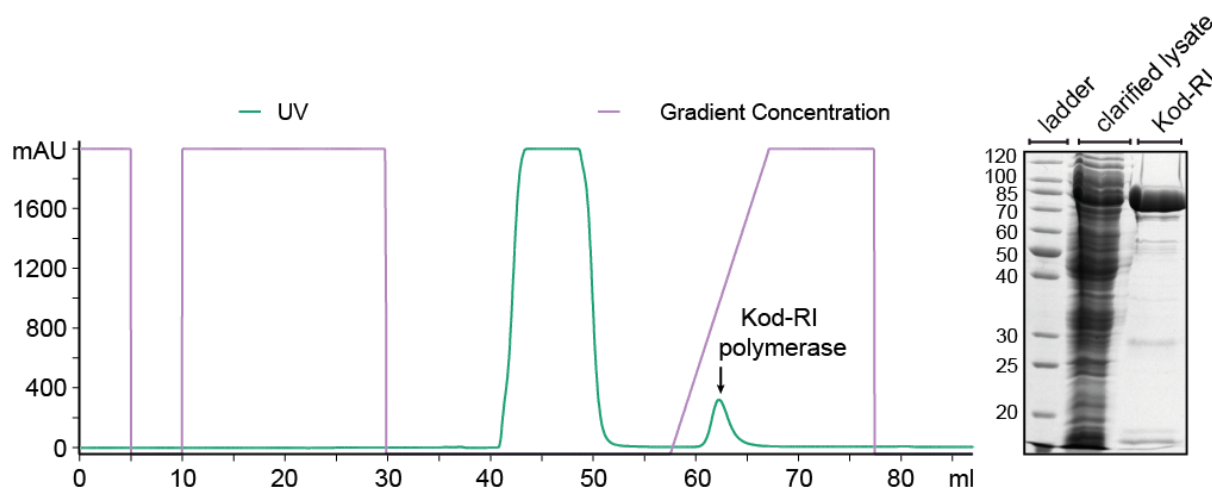
**Figure 33: Overview of the employed cloning methods for the preparation of the expression plasmid harboring the Kod-RI gene.** The four initial cloning methods, including AQUA cloning, In-Fusion cloning, RE cloning, and Stargate cloning, were conducted with different expression vectors and *E.coli* strains. The plasmids isolated after transformation either contained mutations in the Kod-RI coding sequence or only parts of the Kod-RI gene were inserted into the expression vector. The plasmid with the correct Kod-RI coding sequence was finally obtained by site-directed mutagenesis. For further information on the different cloning strategies, please refer to Chapter 5.3.9.

In more detail, four distinct cloning methods were employed: Advanced Quick Assembly (AQUA) cloning<sup>375</sup>, In-Fusion cloning, restriction enzyme (RE) cloning, and Stargate cloning. For further information on the different cloning strategies, please refer to Chapter 5.3.9. In

addition, two expression vectors (pQE80HT, pASG-IBA35) and different *E.coli* strains (NEB10-beta, Stellar, Top10 OneShot) were employed. Nevertheless, each time the Kod-RI coding sequence either contained mutations or only a part of the gene was inserted into the vector.

Finally, a PCR-based site-directed mutagenesis approach using a previously obtained pQE80HT plasmid containing the Kod-RI gene with a single nucleotide deletion was successful in yielding the desired pQE80HT\_Kod-RI construct. For this strategy, a specifically designed primer to introduce the deleted nucleotide during PCR-based plasmid linearization was utilized, followed by 5'-phosphorylation and ligation of the plasmid. Following transformation into competent NEB10-beta *E.colis*, plasmid amplification, and plasmid isolation, the sequence identity was verified by Sanger sequencing. This construct bearing the correct Kod-RI coding sequence was subsequently transformed into NEBExpress® I<sup>q</sup> *E.coli* cells, which are specifically suitable for protein overexpression.

Protein expression and purification were conducted using an adapted protocol published by Nikoomanzar *et al.*<sup>376</sup> In brief, a glycerol stock of NEBExpress® I<sup>q</sup> *E.coli* harboring the pQE80HT\_Kod-RI plasmid was utilized to prepare a starter culture, which was then grown aerobically overnight at 37 °C. Subsequently, the starter culture was used to inoculate the expression culture. Protein expression was induced by the addition of isopropyl- $\beta$ -D-thiogalactoside (IPTG) when the optical density at 600 nm (OD<sub>600</sub>) reached 0.6, followed by incubation at 15 °C overnight. The cells were harvested by centrifugation and lysed using sonication, after which the cell debris was removed by centrifugation. As the wild-type Kod polymerase originates from a hyperthermophilic organism, endogenous *E.coli* proteins were denatured at 80 °C, while the Kod-RI polymerase is not affected by these high temperatures. Subsequently, the denatured endogenous *E.coli* proteins were removed by centrifugation, which significantly simplified the subsequent protein purification process. Immobilized metal affinity chromatography (IMAC) was employed to purify the His-tagged Kod-RI polymerase utilizing a Nickel NTA column (Figure 34).



**Figure 34: Nickel-based IMAC purification of Kod-RI polymerase.** Left: Exemplary UV chromatogram of the affinity purification of Kod-RI polymerase. Right: 10 % SDS-PAGE of the cell lysate before purification and the purified Kod-RI polymerase. The clarified lysate was loaded onto the column and represents the purity of the sample before affinity chromatography. Ladder: PageRuler unstained Protein Ladder, 10-200 kDa. Molecular weight of the target protein: 90 kDa. A 10 % SDS-PAGE of the individual fractions collected during IMAC purification is shown in Figure S9.

Following purification, the polymerase-containing fractions were identified by SDS-PAGE and pooled (Figure S9). Subsequently, the pooled protein fractions were concentrated and transferred to a storage buffer containing 50 % glycerol using centrifugal filters, thus allowing convenient handling and storage at  $-20\text{ }^{\circ}\text{C}$ .

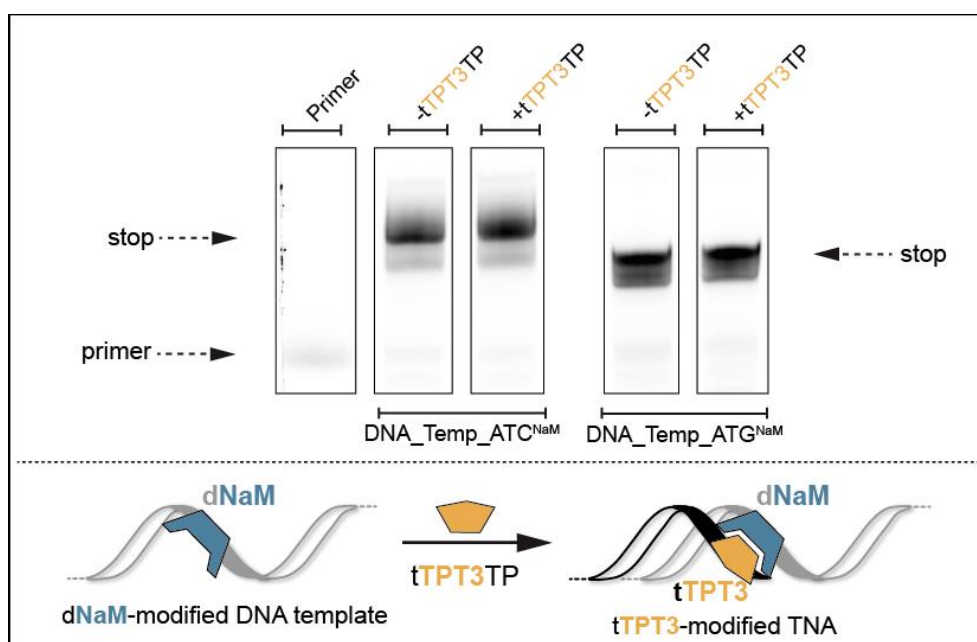
Next, an activity assay was conducted to ascertain the optimal working concentration of the Kod-RI TNA polymerase.<sup>376</sup> Therefore, a two-fold dilution series of the Kod-RI polymerase was prepared, comprising six different concentrations. Subsequently, the aforementioned polymerase concentrations were employed in TNA primer extension assays with tNTPs bearing the canonical nucleobases. Subsequently, the results of the activity assay were evaluated using dPAGE, and the working concentration with the highest quantity of full-length product was ascertained (Figure S10). The protein concentration of the protein stock at the optimal working concentration of the Kod-RI polymerase was determined using the Pierce 660 nm protein assay kit, which is compatible with high amounts of glycerol in the storage buffer. Therefore, a BSA standard curve was prepared (Figure S11) to determine the concentration of the Kod-RI polymerase through interpolation (Table S1). 1544 ng polymerase are required for the efficient extension of 10 pmol of primer. Having purified the Kod-RI TNA polymerase, the processing of tNaMTP and tTPT3TP by this polymerase was investigated using TNA primer extension assays.

First, the incorporation of tTPT3TP by Kod-RI polymerase was investigated by primer extension assays. Given that Kod-RI polymerase, similar to Terminator polymerase, has been reported to have persistent fidelity issues as evidenced by the occurrence of detectable dG/tG mismatch events,<sup>147</sup> the DNA<sub>pTemp\_midATC</sub><sup>NaM</sup> and DNA<sub>pTemp\_midATG</sub><sup>NaM</sup> templates



were used again in PEx (for the exact sequences, please refer to Chapter 7.5.2). Large-scale PEx were conducted according to the previously established protocol, which had been employed during the studies with Therminator polymerase. Therefore, the 5'-monophosphorylated DNA templates were incubated with the 5'-fluorescently labeled DNA primer, **tTPT3TP**, tNTPs, 1.0 mM MnCl<sub>2</sub>, and Kod-RI polymerase. Additionally, negative controls without the addition of **tTPT3TP** were included to serve as size markers for the stop fragment at the position of the UB. Subsequently, the templates were subjected to digestion by λ-exonuclease, and the resulting DNA/TNA chimera products were evaluated by dPAGE and LC-MS analysis.

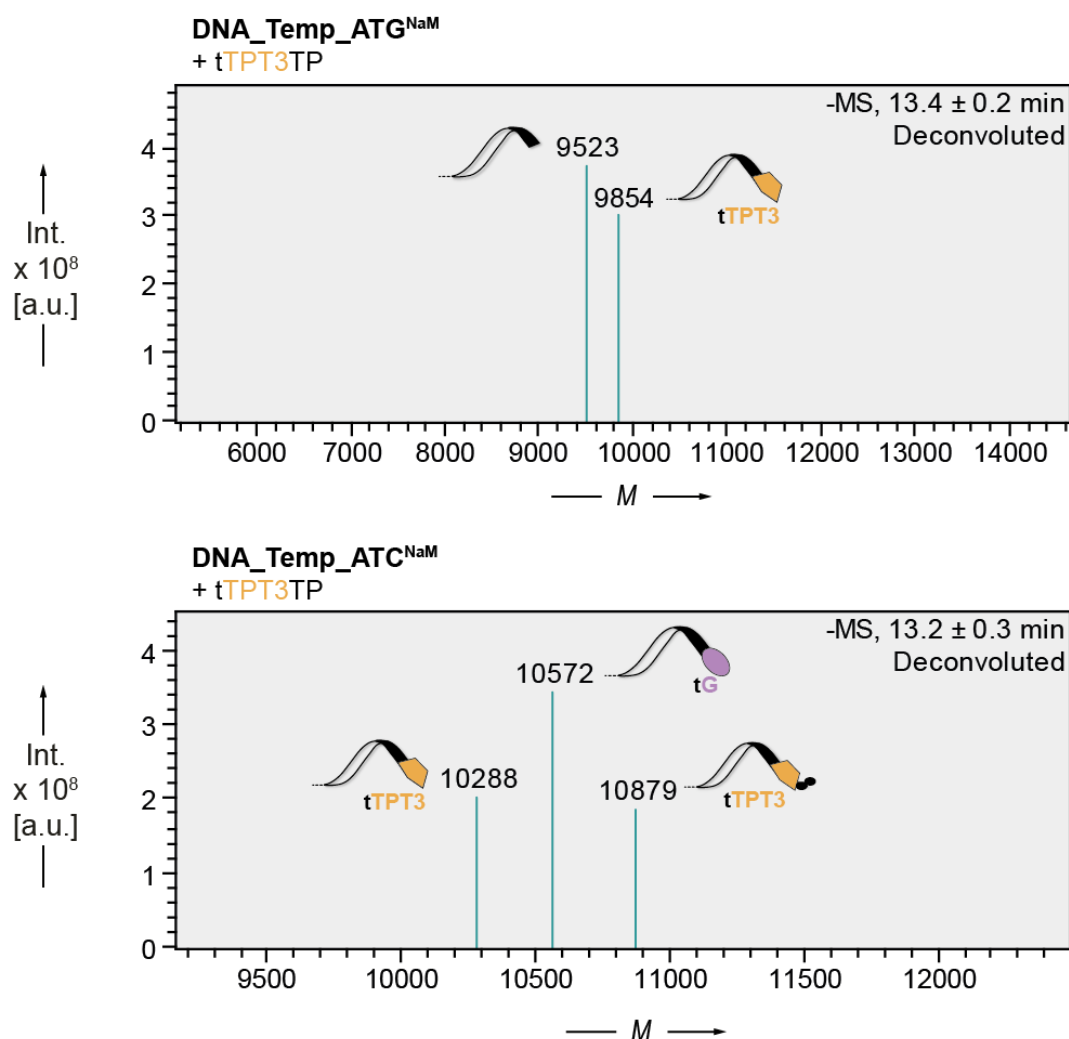
The dPAGE analysis indicates that products obtained by PEx with and without **tTPT3TP** result in product bands migrating at the same height (Figure 35). This suggests that the Kod-RI polymerase, similar to the Therminator polymerase, is unable to synthesize full-length exTNA, and that stalling occurs again at the site of the UB.



**Figure 35: dPAGE results of Kod-RI polymerase-mediated synthesis towards tTPT3-modified exTNA using dNaM-modified DNA templates containing either only dA, dT, and dG or dA, dT, and dC.** Fluorescence scan ( $\lambda_{\text{excitation}} = 477 \text{ nm}$ ,  $\lambda_{\text{emission}} = 535 \text{ nm}$ ) of the 20 % dPAGE analysis of primer extension assays using the DNA\_pTemp\_midATG<sup>NaM</sup> and DNA\_pTemp\_midATC<sup>NaM</sup> DNA templates and the respective TNA nucleotides and **tTPT3TP**. The negative controls do not contain **tTPT3TP**. The primer only and negative controls serve as size markers. The complete gel is depicted in Figure S4.

Further analysis of the obtained products by LC-MS yielded comparable results to those obtained by the Therminator polymerase. In the case of the DNA\_pTemp\_midATG<sup>NaM</sup> template, the molecular weights indicate that the Kod-RI polymerase stalled before and after the incorporation of **tTPT3TP** (9523 g mol<sup>-1</sup> and 9854 g mol<sup>-1</sup>, respectively; Figure 36, top). The negative control comprises solely a molecular weight corresponding to the stop fragment prior to the incorporation of the UB (9522 g mol<sup>-1</sup>; Figure S12).



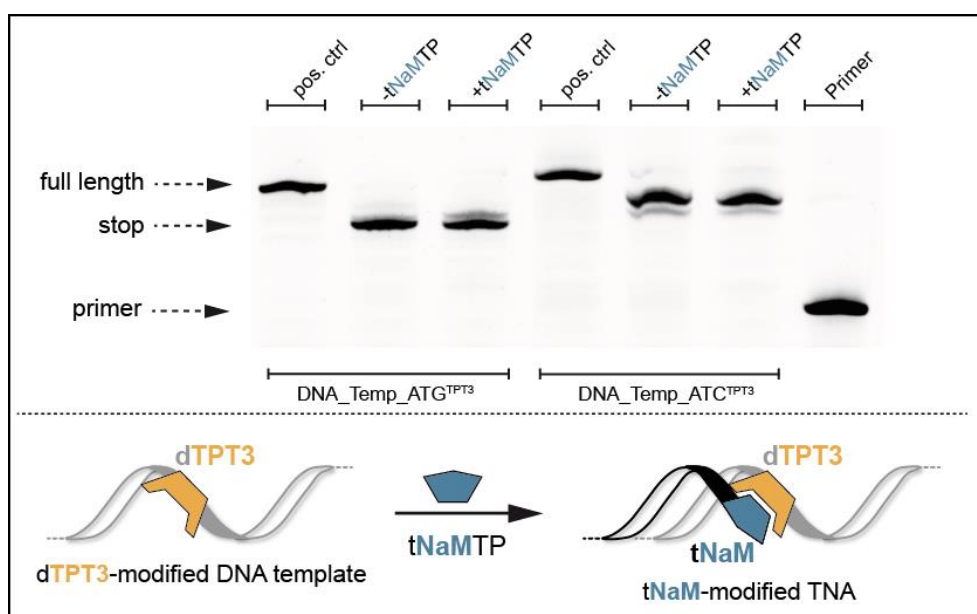


**Figure 36: MS results of Kod-RI polymerase-mediated synthesis towards tTPT3-modified exTNA using dNaM-modified DNA templates containing either only dA, dT, and dG or dA, dT, and dC.** Top: Deconvoluted ESI<sup>-</sup> mass spectrum of PEx using the DNA\_pTemp\_midATG<sup>NaM</sup> DNA template. The mass spectrum shows the molecular weight corresponding to the stop fragment before and after tTPT3TP incorporation (9523 g mol<sup>-1</sup> and 9854 g mol<sup>-1</sup>, respectively). Bottom: Deconvoluted ESI<sup>-</sup> mass spectrum of PEx using the DNA\_pTemp\_midATC<sup>NaM</sup> DNA template. The mass spectrum shows the molecular weights corresponding to the stop product after incorporation of tTPT3TP (10288 g mol<sup>-1</sup>), the stop fragment resulting from the mispairing of dNaM with tG (10572 g mol<sup>-1</sup>), and the stop fragment after the incorporation of tTPT3TP and subsequent extension with two additional nucleotides (10879 g mol<sup>-1</sup>). The complete data set, including UV traces and data of the negative controls, is available in Figures S12 and S13.

Interestingly, when the DNA\_pTemp\_midATC<sup>NaM</sup> template was employed, the Kod-RI polymerase also generated a product that was further elongated with the next two templated nucleotides following tTPT3TP incorporation (10879 g mol<sup>-1</sup>; Figure 36, bottom). In addition, a molecular weight indicative of polymerase stalling following tTPT3TP incorporation (10288 g mol<sup>-1</sup>) was observed. Moreover, a molecular weight corresponding to the stop fragment resulting from misincorporation of tGTP opposite dNaM was also detected (10572 g mol<sup>-1</sup>). The same molecular weight was also observed in the negative control (Figure S13). This indicates that, once more, tGTP is predominantly misincorporated opposite

**dNaM**, and that **tGTP** appears to influence the processivity of the Kod-RI polymerase, as was observed for Terminator DNA polymerase.

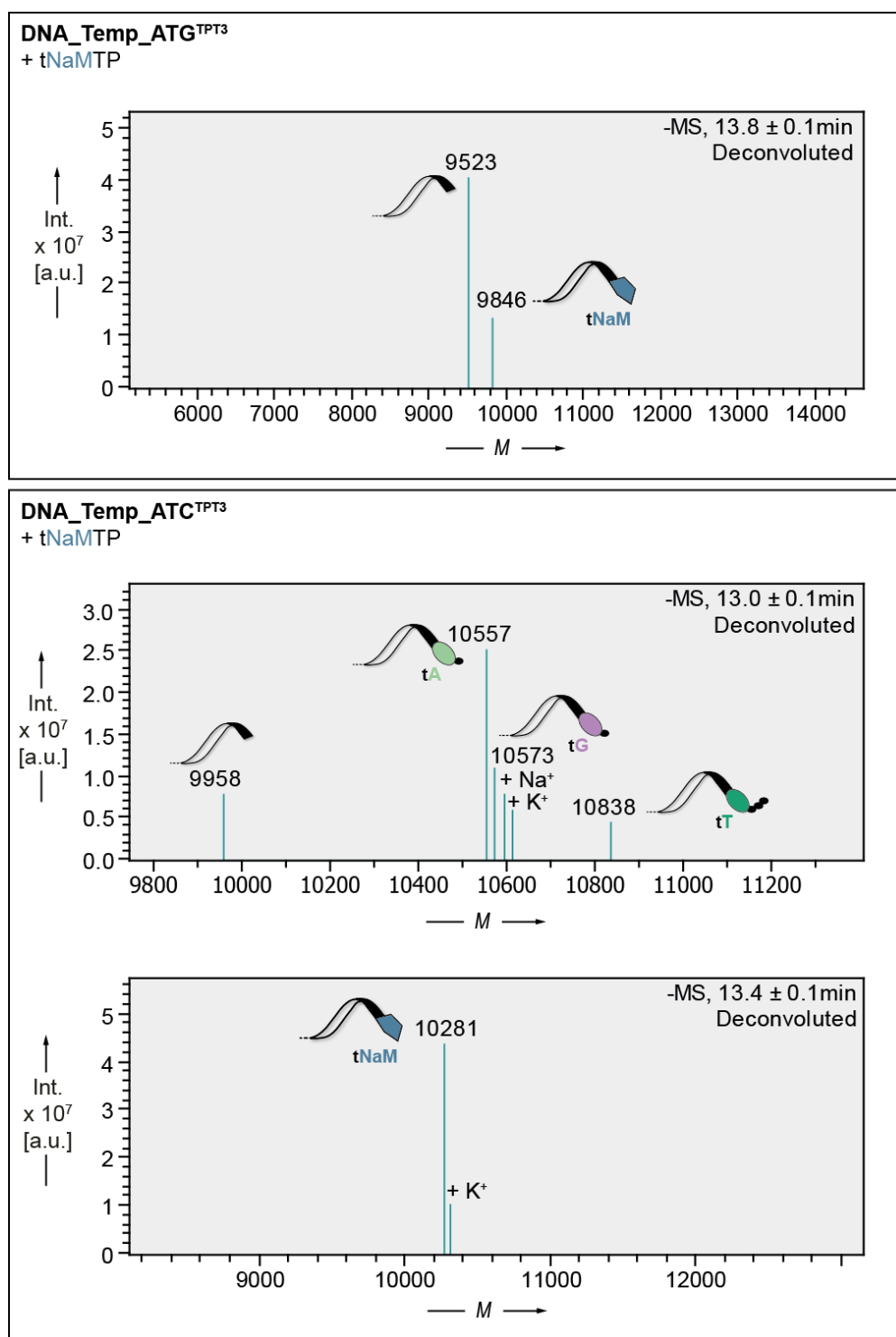
In addition to the investigation of **tTPT3TP** incorporation by Kod-RI polymerase, similar large-scale PEx experiments were also conducted to examine the processing of **tNaMTP** by this polymerase. Therefore, the DNA\_pTemp\_midATG<sup>TPT3</sup> and DNA\_pTemp\_midATC<sup>TPT3</sup> sequences were employed again as templates (for the exact sequences, please refer to Chapter 7.5.2). The previously utilized protocol for large-scale PEx with Kod-RI polymerase was performed in an analogous manner, using **tNaMTP** instead of **tTPT3TP**. Once more, positive and negative control experiments were included as references. Following PEx, the templates were subjected to digestion by  $\lambda$ -exonuclease, and the resulting DNA/TNA chimera products were evaluated by dPAGE and LC-MS analysis.



**Figure 37: dPAGE results of Kod-RI polymerase-mediated synthesis towards tNaM-modified exTNA using dTPT3-modified DNA templates containing either only dA, dT, and dG or dA, dT, and dC.** Fluorescence scan ( $\lambda_{\text{excitation}} = 477 \text{ nm}$ ,  $\lambda_{\text{emission}} = 535 \text{ nm}$ ) of the 20 % dPAGE analysis of primer extension assays using the DNA\_pTemp\_midATG<sup>TPT3</sup> and DNA\_pTemp\_midATC<sup>TPT3</sup> templates and the respective TNA nucleotides and **tNaMTP**. The negative controls do not contain **tNaMTP**. The positive controls contain an unmodified template and tNTPs with exclusively canonical nucleobases. The primer only, negative, and positive controls serve as size markers.

The results obtained by dPAGE demonstrate that **tNaMTP**-containing PEx do not show a product band at the same heights as the positive control, but TNA synthesis rather terminates at the site of the UB for both templates utilized (Figure 37). However, it is noteworthy that PEx containing **tNaMTP** using the DNA\_pTemp\_midATG<sup>TPT3</sup> template showed two distinct product bands, whereas the negative control without **tNaMTP** resulted in only one product band. The second band was observed to migrate slightly higher on the dPAGE compared to the band also present in the negative control. This may indicate partial incorporation of **tNaMTP** followed

by termination of the TNA synthesis. To ascertain whether partial incorporation of **tNaMTP** by Kod-RI polymerase was achieved, LC-MS analysis was conducted (Figure 38).



**Figure 38: MS results of Kod-RI polymerase-mediated synthesis towards tNaM-modified exTNA using dTPT3-modified DNA templates containing either only dA, dT, and dG or dA, dT, and dC.** Top: Deconvoluted ESI<sup>-</sup> mass spectrum of PEx using the DNA\_pTemp\_midATG<sup>TPT3</sup> DNA template. The mass spectrum shows the molecular weight corresponding to the stop fragment before **tNaMTP** incorporation (9523 g mol<sup>-1</sup>) and after **tNaMTP** incorporation (9846 g mol<sup>-1</sup>). Bottom: Deconvoluted ESI<sup>-</sup> mass spectra of PEx using the DNA\_pTemp\_midATC<sup>TPT3</sup> DNA template. The mass spectra show the molecular weights corresponding to the stop fragment before and after **tNaMTP** incorporation (9958 g mol<sup>-1</sup> and 10281 g mol<sup>-1</sup>, respectively) as well as products resulting from mispairing

of **dTPT3** with **tG** (10573 g mol<sup>-1</sup>), **tT** (10838 g mol<sup>-1</sup>), and **tA** (10557 g mol<sup>-1</sup>) followed by further elongation with one to three nucleotides. The complete data set, including UV traces and data of the positive and negative controls, is available in Figures S14 and S15.

Indeed, the mass spectrum revealed two molecular weights, one corresponding to the truncation product before **tNaMTP** incorporation (9523 g mol<sup>-1</sup>) and the other corresponding to the stop product post **tNaMTP** incorporation (9846 g mol<sup>-1</sup>; Figure 38, top). In contrast, the negative control only yielded a molecular weight corresponding to the stop fragment prior to **tNaMTP** incorporation (9523 g mol<sup>-1</sup>, Figure S14). This result indicates that the Kod-RI polymerase can at least partially catalyze the incorporation of **tNaMTP**. By being capable of accepting **tNaMTP** as a substrate, Kod-RI polymerase proves to be superior compared to Terminator polymerase.

The products obtained by PEx using the DNA\_pTemp\_midATC<sup>TPT3</sup> template exhibited the same product band pattern on dPAGE, regardless of the presence of **tNaMTP** (Figure 37). This suggests that PEx containing **tGTP** once again resulted in a greater number of misincorporation events opposite the UB. To substantiate this conclusion, further insight was gained through LC-MS analysis. As anticipated, the LC-MS data corroborated the initial hypothesis that the presence of **tGTP** enhances mismatch events opposite **dTPT3**. In PEx using the DNA\_pTemp\_midATC<sup>TPT3</sup> template, molecular weights corresponding to TNA oligonucleotides with misincorporation of **tG** (10573 g mol<sup>-1</sup>) or **tA** (10557 g mol<sup>-1</sup>) opposite **dTPT3**, followed by the incorporation of the next templated nucleotide, were detected independent of the presence of **tNaMTP** (Figure 38, middle and Figure S15). Moreover, a molecular weight corresponding to an oligonucleotide with a misincorporation of **tT** (10838 g mol<sup>-1</sup>) opposite **dTPT3** and further elongation with the next three nucleotides was identified in both experiments with and without **tNaMTP** (Figure 38, middle and Figure S15). Furthermore, in both PEx, a molecular weight corresponding to the stop fragment prior to **tNaMTP** incorporation (9958 g mol<sup>-1</sup>) was detected (Figure 38, middle and Figure S15). However, the presence of **tNaMTP** in PEx not only resulted in the same products observed as in the negative control, but also resulted in the detection of a molecular weight corresponding to the stop fragment after **tNaMTP** processing (10281 g mol<sup>-1</sup>, Figure 38, bottom). This indicates that the longer TNA oligonucleotides observed on dPAGE from **tNaMTP**-containing PEx were not exclusively the result of mutagenic readthrough.

In summary, the Kod-RI polymerase demonstrated superior performance compared to Terminator polymerase in catalyzing the incorporation of both UB triphosphates. Strikingly, the utilization of Kod-RI polymerase resulted in the observation of further elongation of the nascent TNA strand after **tTPT3TP** incorporation with two additional nucleotides. This phenomenon was not observed with Terminator polymerase. Although Kod-RI polymerase demonstrated the ability to add further nucleotides to **tTPT3** in the TNA strand, a further extension with more than two nucleotides was not facilitated. Moreover, Kod-RI polymerase

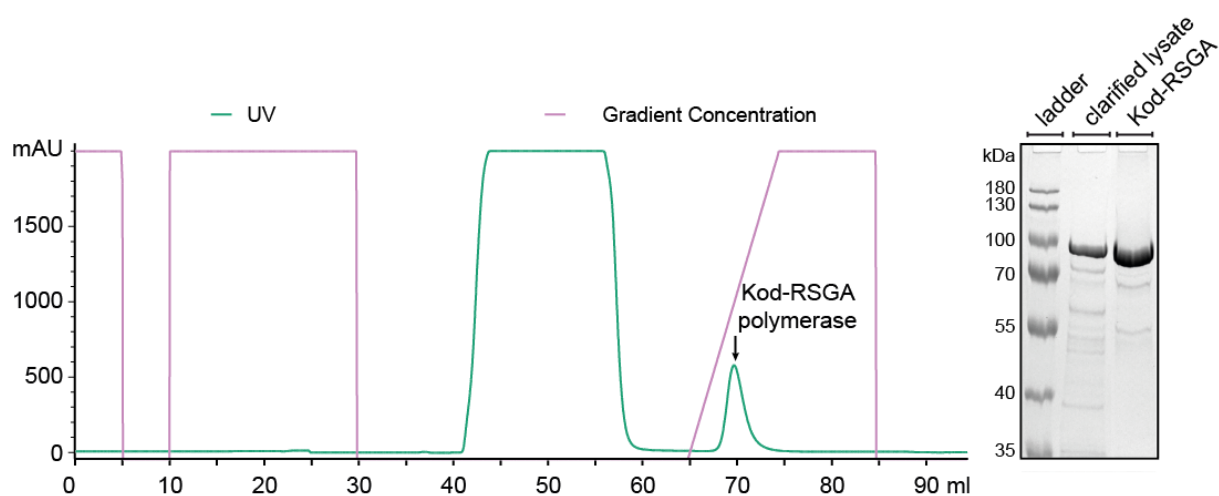
demonstrated the capacity to partially incorporate **tNaMTP**, a phenomenon that was not observed when using Terminator polymerase. Nevertheless, further extension of **tNaM** in the nascent TNA strand was not achieved. Although the Kod-RI polymerase demonstrated superior capabilities for exTNA synthesis compared to the Terminator polymerase, it exhibits two primary drawbacks. Firstly, Kod-RI polymerase frequently misincorporates **tGTP** opposite **dNaM** as well as **tA**, **tG**, and **tT** opposite **dTPT3** in the presence of **tGTP**. Secondly, the Kod-RI polymerase terminates exTNA synthesis in close proximity to the UB. These findings indicate that the Kod-RI polymerase still lacks specificity and efficiency in exTNA synthesis. Consequently, further research was required to address these shortcomings.

### 3.1.3.3 Enzymatic synthesis of exTNA with Kod-RSGA polymerase

The Terminator as well as the Kod-RI polymerase demonstrated only a limited capacity to facilitate the enzymatic incorporation of **tTPT3TP** and **tNaMTP** into a nascent TNA strand in an efficient and faithful manner. The greatest obstacle was the subsequent elongation of the TNA strand following the incorporation of the highly modified triphosphates. Consequently, another Kod polymerase variant, namely the Kod-RSGA polymerase, was considered for further incorporation studies of **tTPT3TP** and **tNaMTP**. The Kod-RSGA polymerase is an engineered manganese-independent TNA polymerase exhibiting a synthesis rate that is 20-fold higher than that of the Kod-RI polymerase.<sup>152</sup> This third-generation TNA polymerase has demonstrated proficiency in processing modified tNTPs with substantial hydrophobic side chains attached to the C<sup>5</sup> carbon of pyrimidine nucleobases.<sup>232–234</sup> It was therefore anticipated that this polymerase may also be able to catalyze the incorporation of TNA nucleotides bearing the hydrophobic UBP **TPT3:NaM**.

As with the Kod-RI polymerase, the laboratory-evolved Kod-RSGA TNA polymerase is not available for commercial purchase. Therefore, it was first necessary to express and purify the Kod-RSGA polymerase. Given the considerable difficulty of cloning the Kod-RI gene, a site-directed mutagenesis approach was chosen for the construction of the plasmid bearing the Kod-RSGA gene. Consequently, the pQE80HT plasmid bearing the Kod-RI gene was taken as a basis and mutated at the desired positions employing a whole plasmid mutagenesis approach to obtain the coding sequence of Kod-RSGA polymerase in the expression vector. In order to change the Kod-RI polymerase coding sequence to the Kod-RSGA coding sequence, four distinct mutations were required.<sup>152</sup> The four desired mutations, N491S, R606G, I664E, and T723A,<sup>152</sup> were sequentially introduced by PCR using primers that were specifically designed to contain mismatches that would insert the desired changes in the coding sequence. Each round of site-directed mutagenesis commenced with an initial PCR

linearization of the complete plasmid and simultaneous targeted mutation using the respective mismatch primer. Subsequently, 5'-phosphorylation and ligation were performed in order to recover the circular plasmid. The constructed plasmid was then transformed into competent NEB 10-beta *E.coli* cells for propagation, and the mutant plasmid was subsequently isolated. Finally, the identity of the coding sequence was verified by Sanger sequencing. After four consecutive rounds, the desired plasmid bearing the Kod-RSGA polymerase coding sequence was successfully obtained. The plasmid was then transformed into competent NEBExpress® I<sup>q</sup> *E.coli* cells, which are optimized for protein expression. The overexpression of the Kod-RSGA polymerase was conducted in a manner analogous to the previously performed protein expression of the Kod-RI polymerase (Chapter 3.1.3.2), followed by nickel-based IMAC purification of the His-tagged protein (Figures 39 and S16).

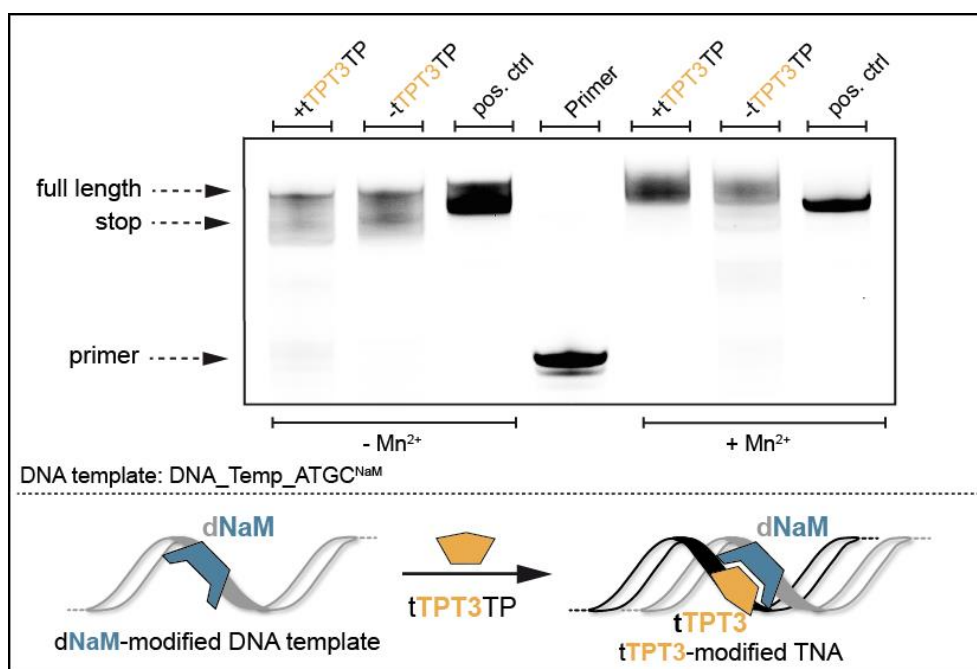


**Figure 39: Nickel-based IMAC purification of Kod-RSGA polymerase.** Left: UV chromatogram of the affinity purification of Kod-RSGA polymerase. Right: 10 % SDS-PAGE of the cell lysate before purification and the purified Kod-RSGA polymerase. The clarified lysate was loaded onto the column and represents the purity of the sample before affinity chromatography. Ladder: PageRuler prestained Protein Ladder, 10-180 kDa. Molecular weight of the target protein: 90 kDa. A 10 % SDS-PAGE of the individual fractions collected during IMAC purification is shown in Figure S16.

Next, the previously described TNA polymerase activity assay was conducted to determine the optimal working concentration of the Kod-RSGA TNA polymerase.<sup>376</sup> The results of the activity assay were evaluated using dPAGE, and the working concentration with the highest quantity of full-length product was ascertained (Figure S17). The concentration of the protein stock at the optimal working concentration of the Kod-RSGA polymerase was once again determined using the Pierce 660 nm protein assay kit (Figure S11 and Table S2). Thereby it was determined that 2253 ng of polymerase efficiently extend 10 pmol of DNA primer. Having purified the Kod-RSGA TNA polymerase, the processing of tNaMTP and tTPT3TP by this polymerase was investigated by TNA primer extension assays.

The initial experiments with Kod-RSGA polymerase were conducted to investigate the processing of tTPT3TP by Kod-RSGA polymerase. As the occurrence of dG/tG mismatches

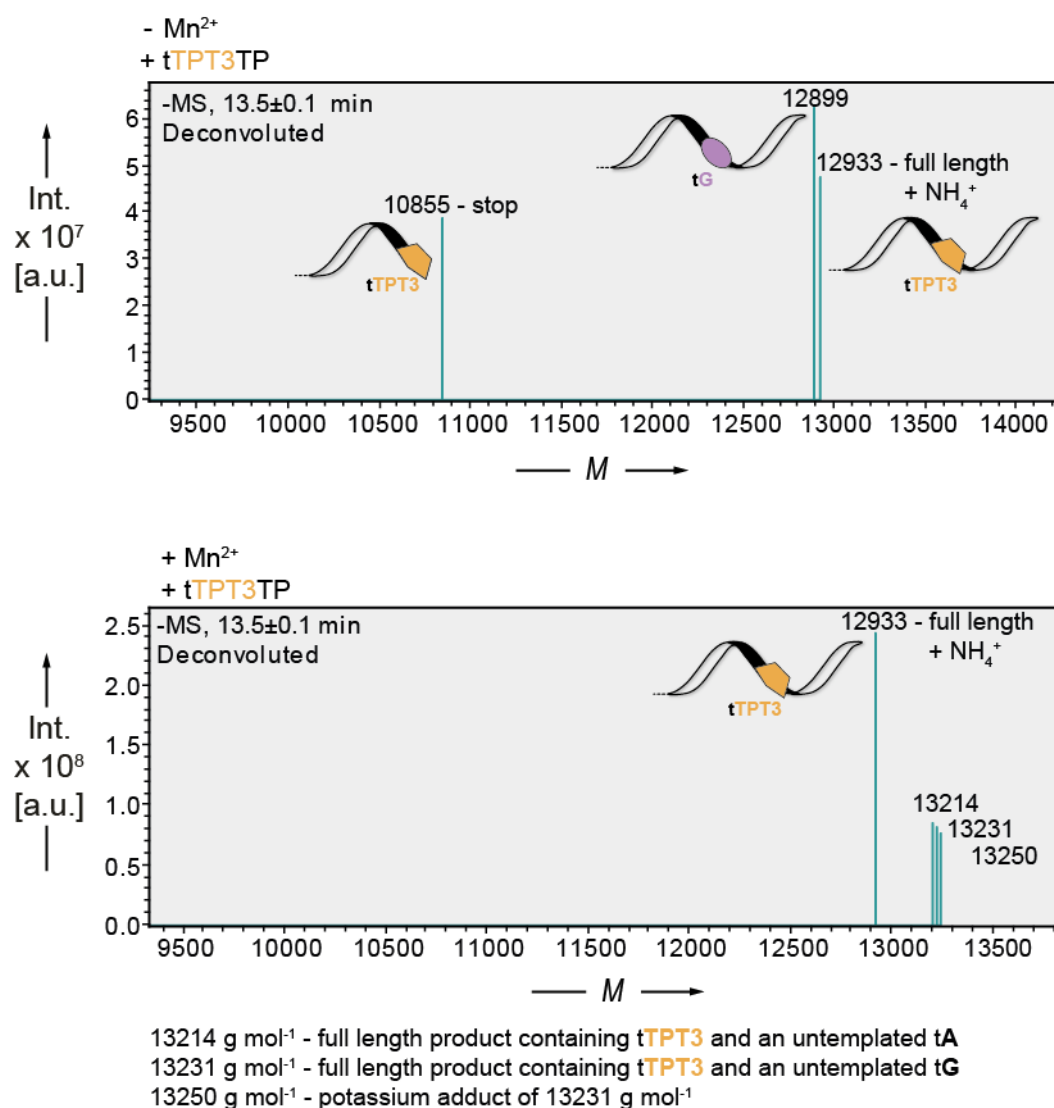
during enzymatic TNA synthesis with Kod-RSGA polymerase was reported to be significantly lower compared to Terminator and Kod-RI polymerase,<sup>152</sup> templates containing all four canonical nucleobases and dNaM were employed. In more detail, the 41-nt-long DNA\_pTemp\_mid2ATGC<sup>NaM</sup> sequence (for the exact sequences, please refer to Chapter 7.5.2) was used as a template bearing one dNaM modification at position 34. Although Kod-RSGA polymerase is described to be Mn<sup>2+</sup>-independent with respect to the processing of  $\alpha$ -L-treofuranosyl triphosphates with canonical nucleobases,<sup>152</sup> experiments with the addition of MnCl<sub>2</sub> were performed. Manganese ions are known to relieve conformational constraints in the active sites of polymerases,<sup>377</sup> thus it was postulated that the addition of MnCl<sub>2</sub> in PEx could promote the processing of tTPT3TP by Kod-RSGA polymerase. Consequently, PEx were performed in parallel without MnCl<sub>2</sub> or with the addition of 1.0 mM MnCl<sub>2</sub>. Therefore, the 5'-fluorescently labeled DNA primer and the 5'-monophosphorylated DNA template were annealed. Subsequently, the primer-template complex was incubated in the presence of tTPT3TP, tNTPs, and Kod-RSGA polymerase at 55 °C for 16 hours. Once again, positive and negative controls were performed to serve as full-length and stop fragment size markers on dPAGE. Prior to LC-MS and dPAGE analysis, the 5'-monophosphorylated DNA templates were subjected to digestion with  $\lambda$ -exonuclease.



**Figure 40: dPAGE results of Kod-RSGA polymerase-mediated synthesis of tTPT3-modified exTNA using a dNaM-modified DNA template.** Fluorescence scan ( $\lambda_{\text{excitation}} = 477 \text{ nm}$ ,  $\lambda_{\text{emission}} = 535 \text{ nm}$ ) of the 20 % dPAGE analysis of primer extension assays with canonical tNTPs and tTPT3TP, with and without MnCl<sub>2</sub> supplementation, using the DNA\_pTemp\_mid2ATGC<sup>NaM</sup> DNA template. The negative controls do not contain tTPT3TP. The positive control contains an unmodified template and tNTPs with exclusively canonical nucleobases. The primer only, negative, and positive controls serve as size markers.

dPAGE consistently yielded comparable results for both sets of PEx experiments, regardless of the manganese ion concentration employed (Figure 40). The primer extensions containing

**tTPT3TP**, as well as the negative controls lacking **tTPT3TP**, exhibited a full-length product band that migrated to a similar height as the positive control on dPAGE (Figure 40). It seems reasonable to propose that in the negative control lacking **tTPT3TP**, misincorporation opposite **dNaM** might occur, resulting in the generation of an incorrect full-length product rather than the expected stop fragment. This outcome prompted the question of whether the full-length product from PEx containing **tTPT3TP** yielded the desired **TPT3**-modified exTNA. To determine whether the observed band represented the desired exTNA full-length product with correct **tTPT3TP** incorporation or was the result of mutagenic readthrough, LC-MS analysis was conducted (Figure 41).



**Figure 41: Comparison of mass spectra obtained from Kod-RSGA polymerase-mediated synthesis of **tTPT3**-modified exTNA using DNA\_pTemp\_midATGC<sup>NaM</sup> with and without MnCl<sub>2</sub> supplementation.** Top: Deconvoluted ESI<sup>-</sup> mass spectrum of PEx without Mn<sup>2+</sup> ions. The mass spectrum shows the molecular weights corresponding to the stop fragment post **tTPT3TP** incorporation (10855 g mol<sup>-1</sup>), a full-length product with misincorporation of **tGTP** opposite **dNaM** (12899 g mol<sup>-1</sup>), and the ammonium adduct of the full-length **tTPT3**-modified exTNA (12933 g mol<sup>-1</sup>). Bottom: Deconvoluted ESI<sup>-</sup> mass spectrum of PEx containing 1.0 mM MnCl<sub>2</sub>. The mass spectrum shows the molecular weights corresponding to the ammonium adduct of the



**TPT3**-modified full-length product (12933 g mol<sup>-1</sup>) and molecular weights of the **TPT3**-modified full-length product with untemplated additional nucleotides (13214 g mol<sup>-1</sup>, 13231 g mol<sup>-1</sup>, 13250 g mol<sup>-1</sup>). The complete data set, including UV traces and spectra of positive and negative controls, is available in Figures S18 and S19.

The LC-MS data from the experiments with and without Mn<sup>2+</sup> confirmed the incorporation of **tTPT3TP** and further elongation of the TNA strand after the **UB**, resulting in the correct full-length product in both cases (12933 g mol<sup>-1</sup> ammonium adduct; Figure 41 and Table 2). However, in the absence of Mn<sup>2+</sup> ions, **dNaM:tG** mismatches (12899 g mol<sup>-1</sup>) and truncation after processing of **tTPT3TP** also occurred (10855 g mol<sup>-1</sup>; Figure 41, top and Table 2). Interestingly, in the presence of Mn<sup>2+</sup>, mutagenic readthrough and stop fragments were not observed (Figure 41, bottom and Table 2). Consequently, it can be postulated that the presence of Mn<sup>2+</sup> ions generally facilitates the proper incorporation of **tTPT3TP** into TNA.

Furthermore, the LC-MS data demonstrated that the omission of **tTPT3TP** in the negative control PEx resulted in a preferential misincorporation of **tGTP** opposite **dNaM**, thereby confirming mutagenic readthrough (12899 g mol<sup>-1</sup>; Table 2 and Figure S18). The misincorporation of **tG** opposite **dNaM** may be attributed to partial  $\pi$ -stacking of the nucleobases in the absence of hydrogen bonds. Strikingly, these results demonstrate the ability of Kod-RSGA polymerase to strongly discriminate against misincorporation of an incorrect nucleotide in the presence of the correct TNA triphosphate and Mn<sup>2+</sup> ions. Notably, the presence of Mn<sup>2+</sup> ions in all samples resulted in the non-templated elongation of the TNA strand with an extra nucleotide (Table 2, Figures S18 and S19), whereas the absence of Mn<sup>2+</sup> ions occasionally led to a TNA oligonucleotide lacking the last nucleotide in the negative controls (Table 2 and Figure S18). A summary of the LC-MS results obtained from PEx using DNA\_pTemp\_mid2ATGC<sup>NaM</sup> is presented in Table 2.

**Table 2: Summary of the LC-MS results obtained from primer extension assays using DNA\_pTemp\_mid2ATGC<sup>NaM</sup>.** The corresponding UV traces and mass spectra are available in Figure 41 and Figures S18 and S19.

PEX conditions	Full-length product	Stop	
		fragment	Mutagenic readthrough
<b>Positive control,</b> <b>- Mn<sup>2+</sup></b>	Yes, 12584 g mol <sup>-1</sup>	No	No
<b>+ tTPT3TP,</b> <b>- Mn<sup>2+</sup></b>	Yes, 12933 g mol <sup>-1</sup> <sup>α</sup>	Yes, 10855 g mol <sup>-1</sup> <sup>β</sup>	Yes, 12899 g mol <sup>-1</sup> (dNaM:tG)
<b>- tTPT3TP,</b> <b>- Mn<sup>2+</sup></b>	No	Yes, 10838 g mol <sup>-1</sup> (dNaM:tG)	Yes, 12582 g mol <sup>-1</sup> (UB skipped <sup>γ</sup> , fl-tA <sup>δ</sup> ) 12600 g mol <sup>-1</sup> (dNaM:tG, fl-tA <sup>δ</sup> ) 12900 g mol <sup>-1</sup> (dNaM:tG)
<b>Positive control,</b> <b>+ Mn<sup>2+</sup></b>	Yes, 12882 g mol <sup>-1</sup> (fl+tA <sup>ε</sup> ) 12898 g mol <sup>-1</sup> (fl+tG <sup>ε</sup> )	No	No
<b>+ tTPT3TP,</b> <b>+ Mn<sup>2+</sup></b>	Yes, 12933 g mol <sup>-1</sup> <sup>α</sup> 13214 g mol <sup>-1</sup> (fl+tA <sup>ε</sup> ) 13231 g mol <sup>-1</sup> (fl+tG <sup>ε</sup> )	No	No
<b>- tTPT3TP,</b> <b>+ Mn<sup>2+</sup></b>	No	No	Yes, 12900 g mol <sup>-1</sup> (dNaM:tG), 13189 g mol <sup>-1</sup> (dNaM:tG, fl+tT <sup>ε</sup> ) 13199 g mol <sup>-1</sup> (dNaM:tG, fl+tA <sup>ε</sup> ) 13215 g mol <sup>-1</sup> (dNaM:tG, fl+tG <sup>ε</sup> )

<sup>α</sup> ammonium adduct

<sup>β</sup> stop fragment post incorporation of tTPT3TP

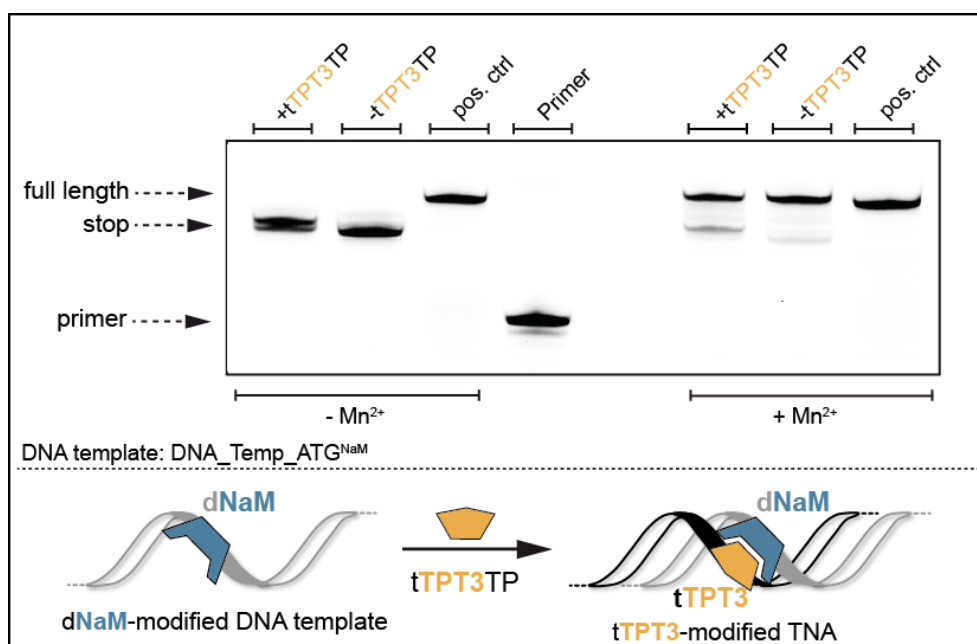
<sup>γ</sup> skipping of the incorporation of a nucleotide opposite to the UB

<sup>δ</sup> the last nucleotide of the full-length product is missing (fl-1)

<sup>ε</sup> addition of one extra untemplated nucleotide to the full-length product (fl+1)

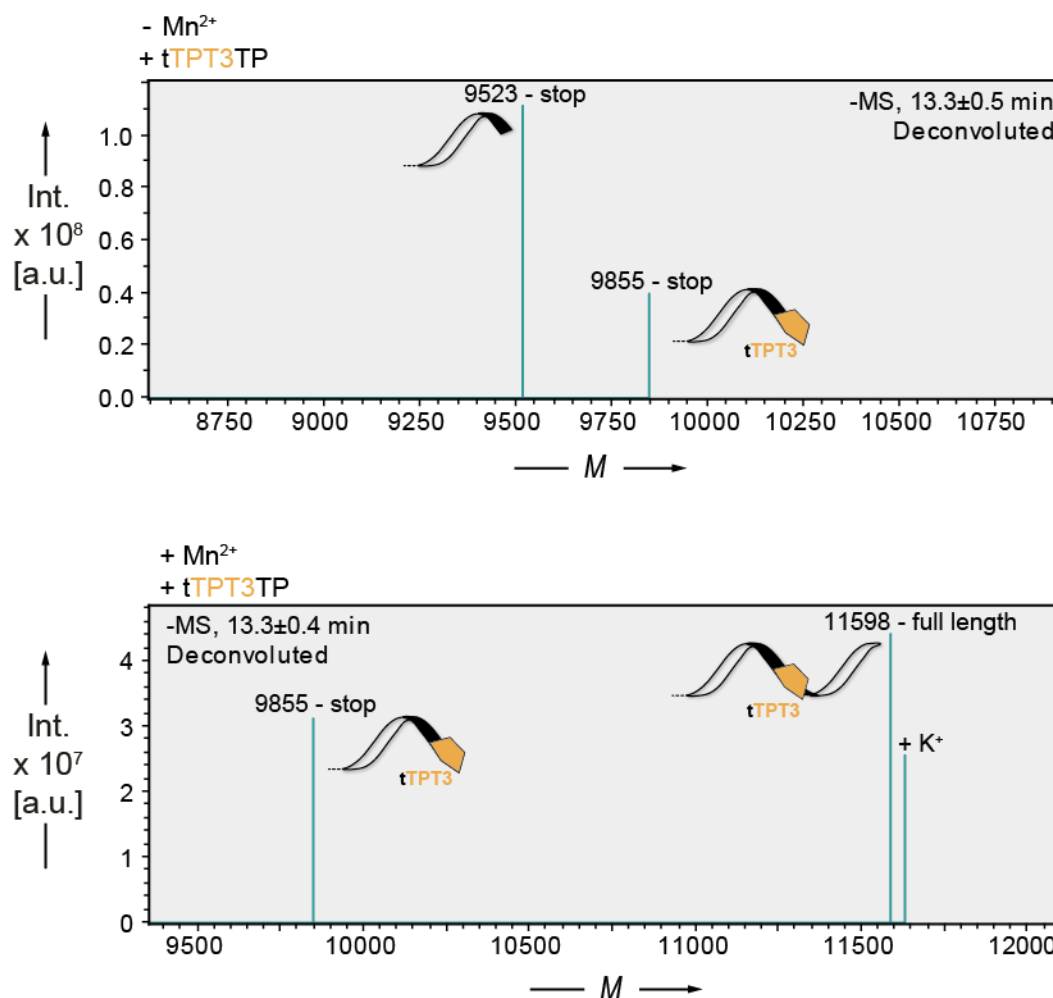
Subsequently, PEX were performed with a DNA template comprising solely the three canonical nucleobases, **A**, **G**, and **T**, as well as the unnatural UB **NaM**. This was done with the objective of elucidating the synthesis of exTNA without the addition of tGTP. This choice was made because misincorporation of tGTP opposite dNaM was observed to be a prevalent phenomenon in the preceding experiment. Moreover, previous studies on enzymatic TNA synthesis have demonstrated that tGTP, in particular, is a challenging substrate for TNA polymerases.<sup>147</sup> Therefore, the dNaM-modified template DNA\_pTemp\_midATG<sup>NaM</sup> (for the exact sequence, please refer to Chapter 7.5.2) was employed in large-scale primer extension experiments conducted in the absence and presence of 1.0 mM MnCl<sub>2</sub>. The employed template

is 37 nt in length and contains the UB **NaM** at position 31. Figure 42 presents the results of the dPAGE analysis of the PEx assays conducted.



**Figure 42: dPAGE results of Kod-RSGA polymerase-mediated synthesis of tTPT3-modified exTNA using a dNaM-modified DNA template containing only dA, dT, and dG.** Fluorescence scan ( $\lambda_{\text{excitation}} = 477 \text{ nm}$ ,  $\lambda_{\text{emission}} = 535 \text{ nm}$ ) of the 20 % dPAGE analysis of primer extension assays with tTPT3TP, canonical tNTPs except tGTP, with and without  $\text{MnCl}_2$  supplementation using the DNA\_pTemp\_midATG<sup>NaM</sup> DNA template. The negative controls do not contain tTPT3TP. The positive controls contain an unmodified template and tNTPs with exclusively canonical nucleobases. The primer only, negative, and positive controls serve as size markers.

In the absence of  $\text{Mn}^{2+}$  ions, the Kod-RSGA polymerase was unable to synthesize full-length exTNA. The results of the dPAGE analysis indicate that the Kod-RSGA polymerase terminated TNA synthesis at two distinct positions, most probably either prior to and post tTPT3TP incorporation (Figure 42, left). Furthermore, the negative control migrated at the expected length of the truncation product at the site of the UB. These results were corroborated by LC-MS data. In PEx containing tTPT3TP, molecular weights corresponding to the stop fragment before the incorporation of tTPT3TP ( $9523 \text{ g mol}^{-1}$ ), and after the incorporation of tTPT3TP ( $9855 \text{ g mol}^{-1}$ ) were detected (Figure 43, top and Table 3). In the negative control, a molecular weight corresponding to the truncation product prior to tTPT3TP incorporation ( $9523 \text{ g mol}^{-1}$ ) was observed (Table 3 and Figure S20). It is noteworthy that in primer extensions without tGTP the incorporation of an incorrect nucleotide opposite dNaM did not occur in the negative control (Table 3 and Figure S20). This is consistent with the previous observation that tGTP is the predominantly misincorporated nucleotide opposite dNaM.



**Figure 43: Comparison of mass spectra obtained from Kod-RSGA polymerase-mediated synthesis of tTPT3-modified exTNA using DNA\_pTemp\_midATG<sup>NaM</sup> with and without MnCl<sub>2</sub> supplementation.** Top: Deconvoluted ESI<sup>-</sup> mass spectrum of PEx without Mn<sup>2+</sup> ions. The mass spectrum shows the molecular weights corresponding to the stop fragment before tTPT3TP incorporation (9523 g mol<sup>-1</sup>) and the stop fragment post tTPT3TP incorporation (9855 g mol<sup>-1</sup>). Bottom: Deconvoluted ESI<sup>-</sup> mass spectrum of PEx containing 1.0 mM MnCl<sub>2</sub>. The mass spectrum shows the molecular weights corresponding to the tTPT3-modified full-length product (11598 g mol<sup>-1</sup>) and the stop fragment post tTPT3TP incorporation (9855 g mol<sup>-1</sup>). The complete data set, including UV traces and mass spectra of the positive and negative controls, is available in Figure S20 and S21.

In contrast, a substantial quantity of full-length product was observed on dPAGE in the presence of Mn<sup>2+</sup> ions (Figure 42, right). It is noteworthy that the negative control containing Mn<sup>2+</sup> also demonstrated the presence of a full-length TNA product on dPAGE, although tTPT3TP was omitted in this PEx. To ascertain whether the full-length product observed in the tTPT3TP-containing reaction resulted from the incorporation of tTPT3TP opposite dNaM or from mutagenic readthrough, LC-MS analysis was employed. The mass spectrum of the tTPT3TP-containing primer extension assay with Mn<sup>2+</sup> confirmed the presence of the molecular weight corresponding to the full-length exTNA oligonucleotide bearing the tTPT3 modification (11598 g mol<sup>-1</sup>) as well as the molecular weight corresponding to the stop fragment after the incorporation of tTPT3TP (9855 g mol<sup>-1</sup>; Figure 42, bottom and Table 3). In

contrast, the mass spectrum of the  $Mn^{2+}$ -containing negative control showed misincorporation of **tTTP** (11558 g mol<sup>-1</sup>) and **tATP** (11566 g mol<sup>-1</sup>) opposite to **dNaM** in the template strand, as well as complete skipping of the incorporation of a nucleotide (11266 g mol<sup>-1</sup>) opposite to the UB (Table 3 and Figure S20). The manganese ions are likely to increase the substrate promiscuity of the Kod-RSGA polymerase, resulting in mutagenic readthrough although **tGTP** is not present in this PEx. Nevertheless, the results of PEx with  $Mn^{2+}$  once again demonstrated that Kod-RSGA is highly adept at discriminating against the incorporation of an incorrect nucleotide, provided that the correct counterpart is present. It is notable that the nontemplated elongation of the TNA strand, which was observed when all tNTPs were present in the PEx, did not occur in absence of **tGTP** (Table 3, Figures S20 and S21). This indicates that **tGTP** is the preferred substrate triphosphate of Kod-RSGA polymerase and affects its processivity. A summary of the LC-MS results obtained from PEx using DNA\_pTemp\_midATG<sup>NaM</sup> is presented in Table 3.

**Table 3: Summary of the LC-MS results obtained from primer extension assays using DNA\_pTemp\_midATG<sup>NaM</sup>.** The corresponding UV traces and mass spectra are available in Figure 43 and Figures S20 and S21.

<b>PEx conditions</b>	<b>Full-length product</b>	<b>Stop fragment</b>	<b>Mutagenic readthrough</b>
<b>Positive control,</b> <b>- Mn<sup>2+</sup></b>	Yes, 11267 g mol <sup>-1</sup>	No	No
<b>+ tTPT3TP,</b> <b>- Mn<sup>2+</sup></b>	No	Yes, 9523 g mol <sup>-1</sup> <sup>α</sup> , 9855 g mol <sup>-1</sup> <sup>β</sup>	No
<b>- tTPT3TP,</b> <b>- Mn<sup>2+</sup></b>	No	Yes, 9523 g mol <sup>-1</sup> <sup>α</sup>	No
<b>Positive control,</b> <b>+ Mn<sup>2+</sup></b>	Yes, 11267 g mol <sup>-1</sup>	No	No
<b>+ tTPT3TP,</b> <b>+ Mn<sup>2+</sup></b>	Yes, 11598 g mol <sup>-1</sup>	Yes, 9855 g mol <sup>-1</sup> <sup>β</sup>	No
<b>- tTPT3TP,</b> <b>+ Mn<sup>2+</sup></b>	No	No	Yes, 11266 g mol <sup>-1</sup> (UB skipped <sup>γ</sup> ), 11558 g mol <sup>-1</sup> (dNaM:tT), 11566 g mol <sup>-1</sup> (dNaM:tA)

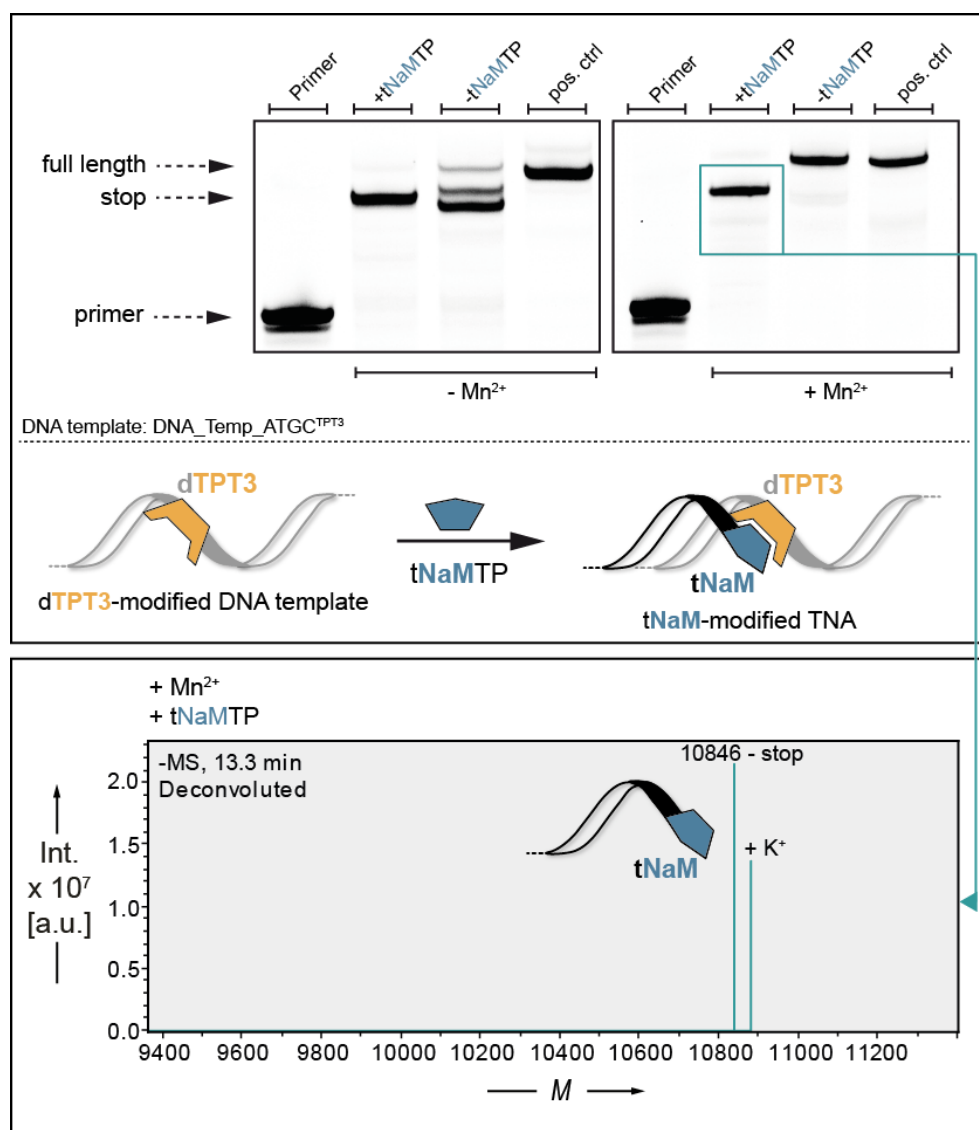
<sup>α</sup> stop fragment before incorporation of **tTPT3TP**

<sup>β</sup> stop fragment post incorporation of **tTPT3TP**

<sup>γ</sup> skipping of the incorporation of a nucleotide opposite to the UB

In addition to the data obtained for the processing of **tTPT3TP** by Kod-RSGA polymerase, PEx were also conducted to investigate the processing of **tNaMTP** by Kod-RSGA polymerase with and without the addition of manganese ions. Consequently, large-scale TNA primer extension assays were conducted using Kod-RSGA polymerase and a **TPT3**-modified DNA template. In

more detail, DNA\_pTemp\_mid2ATGC<sup>TPT3</sup> (for the exact sequence, please refer to Chapter 7.5.2) was used as a template, which is 41 nt in length and contains one **dTPT3** modification at position 34.



**Figure 44: Kod-RSGA polymerase-mediated synthesis of tNaM-modified extNA using a dTPT3-modified DNA template.** Top: Fluorescence scan ( $\lambda_{\text{excitation}} = 477 \text{ nm}$ ,  $\lambda_{\text{emission}} = 535 \text{ nm}$ ) of the 20 % dPAGE analysis of primer extension assays with canonical tNTPs and tNaMTP, with and without  $\text{MnCl}_2$  supplementation, using the DNA\_pTemp\_mid2ATGC<sup>TPT3</sup> DNA template. The negative controls do not contain tNaMTP. The positive control contains an unmodified template and tNTPs with exclusively canonical nucleobases. The primer only, negative and positive controls serve as size markers. Bottom: Deconvoluted ESI- mass spectrum of PEx containing tNaMTP and 1.0 mM  $\text{MnCl}_2$ . The mass spectrum shows the molecular weight corresponding to the stop fragment post tNaMTP incorporation (10846 g mol<sup>-1</sup>). The complete data set, including UV traces and spectra of positive and negative controls, is available in Figures S22 and S23.

dPAGE analysis revealed that the PEx containing a **dTPT3**-modified template and the corresponding **tNaMTP** did not yield the desired full-length product band at the same heights as the positive control (Figure 44). This result indicated that no full-length extNA with **tNaM** modification was obtained, despite the use of different  $\text{Mn}^{2+}$  concentrations. Further analysis

of the DNA/exTNA chimera by LC-MS confirmed the initial findings. A molecular weight corresponding to the stop fragment after **tNaMTP** processing (10846 g mol<sup>-1</sup>) was observed in both cases with and without Mn<sup>2+</sup> supplementation, thereby confirming the initial incorporation of **tNaMTP** into the nascent TNA (Figure 44, bottom, Table 4 and Figure S22). Contrary to expectations, the addition of Mn<sup>2+</sup> ions to the PEx did not result in further elongation of the TNA strand following **tNaMTP** incorporation. Instead, the dPAGE analysis revealed that the negative control, which had been supplemented with MnCl<sub>2</sub>, yielded a full-length product band, in contrast to the expected stop fragment at the site of the UB. This indicates that tNTPs bearing canonical nucleobases have potentially been misincorporated opposite **dTPT3** in the negative control containing Mn<sup>2+</sup>. Further analysis by LC-MS revealed that in the absence of the correct **tNaMTP** nucleotide, misincorporation of **tGTP** (12898 g mol<sup>-1</sup>), **tCTP**, or **tATP** (13157 g mol<sup>-1</sup>) opposite **dTPT3** occurred (Table 4 and Figure S22). This observation once again demonstrates that the Kod-RSGA polymerase is capable of effectively distinguishing between the incorporation of an incorrect nucleotide when the corresponding correct unnatural nucleotide is present. The results of the LC-MS analysis obtained from PEx using DNA\_pTemp\_mid2ATGC<sup>TPT3</sup> are summarized in Table 4.

**Table 4: Summary of the LC-MS results obtained from primer extension assays using DNA\_pTemp\_mid2ATGC<sup>TPT3</sup>.** The corresponding mass spectra are available in Figure 44 and Figures S22 and S23.

<b>PEx conditions</b>	<b>Full-length product</b>	<b>Stop fragment</b>	<b>Mutagenic readthrough</b>
<b>Positive control,</b> <b>- Mn<sup>2+</sup></b>	Yes, 12583 g mol <sup>-1</sup>	No	No
<b>+ tNaMTP,</b> <b>- Mn<sup>2+</sup></b>	No	Yes, 10846 g mol <sup>-1</sup> <sup>α</sup>	No
<b>- tNaMTP,</b> <b>- Mn<sup>2+</sup></b>	No	Yes, 10524 g mol <sup>-1</sup> <sup>β</sup>	No
<b>Positive control,</b> <b>+ Mn<sup>2+</sup></b>	Yes, 12859 g mol <sup>-1</sup> (fl+t <b>C</b> <sup>γ</sup> ) 12882 g mol <sup>-1</sup> (fl+t <b>A</b> <sup>γ</sup> )	No	No
<b>+ tNaMTP,</b> <b>+ Mn<sup>2+</sup></b>	No	Yes, 10846 g mol <sup>-1</sup> <sup>α</sup>	No
<b>- tNaMTP,</b> <b>+ Mn<sup>2+</sup></b>	No	No	Yes, 12898 g mol <sup>-1</sup> ( <b>dTPT3:tG</b> ), 13157 g mol <sup>-1</sup> ( <b>dTPT3:tC/A</b> , fl+t <b>A/tC</b> <sup>γ</sup> )

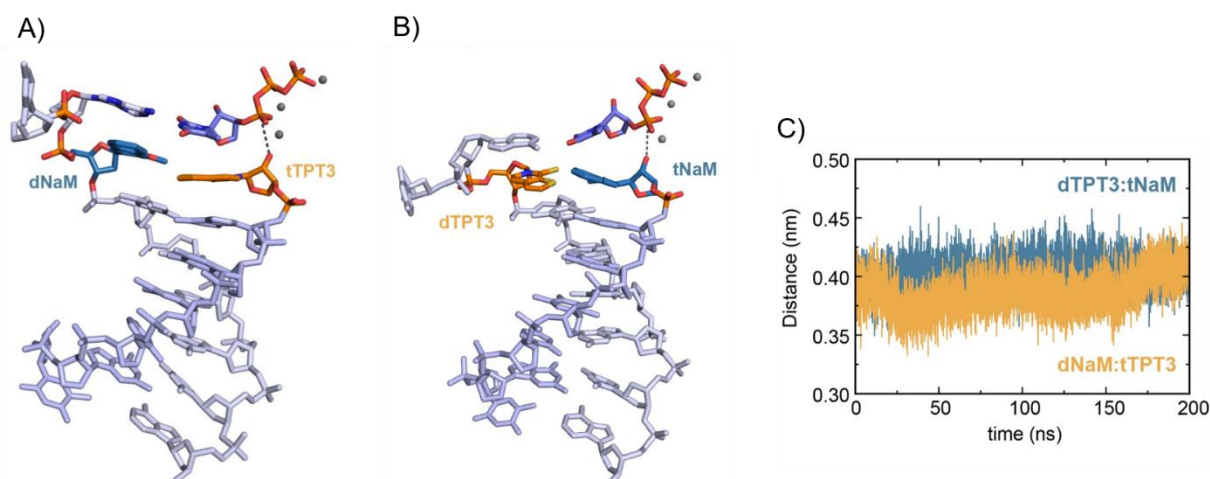
<sup>α</sup> stop fragment post incorporation of **tNaMTP**

<sup>β</sup> stop fragment before incorporation of **tNaMTP**

<sup>γ</sup> addition of one extra untemplated nucleotide to the full-length product (fl+1)

Notably, LC-MS data showed that the addition of manganese ions to the primer extension assay resulted in a nontemplated extension of the TNA strand by one nucleotide in the positive control (Table 4 and Figure S23). Despite the apparent reduction in specificity of Kod-RSGA polymerase by  $Mn^{2+}$ , these ions were not sufficient to promote the formation of full-length **NaM**-modified exTNA.

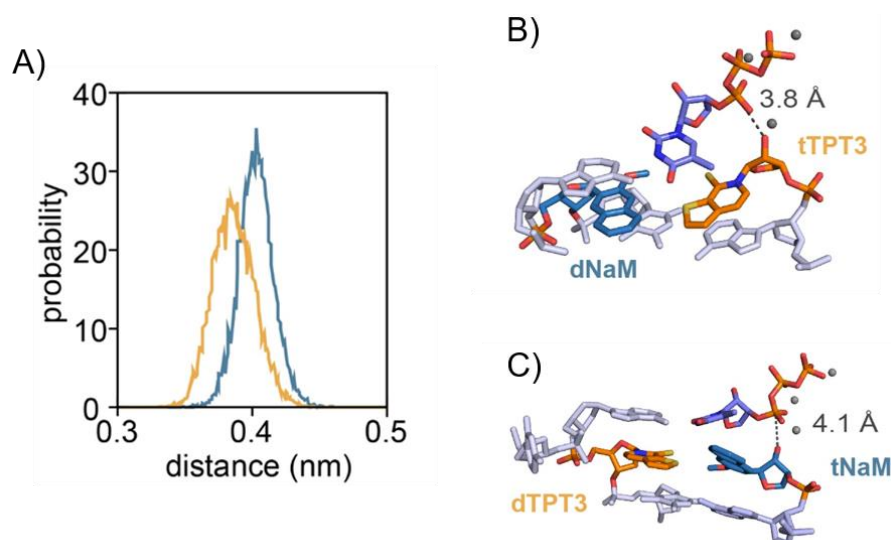
To further illustrate whether the structural requirements in the active site of Kod-RSGA polymerase support the extension of nascent TNA after the incorporation of **tTPT3**TP and block the extension of **tNaM**, all-atom molecular dynamics (MD) simulations were performed. The MD simulations were conducted by Prof. in Dr. Stephanie Kath-Schorr and performed in explicit solvent (TIP3P water model) using the GROMACS<sup>378</sup> software package in combination with the CHARMM36<sup>379</sup> force field. The initial model of Kod-RSGA polymerase in complex with a DNA template and a nascent TNA sequence was constructed based on the crystal structure of Kod-RI polymerase in its closed ternary complex (PDB: 5VU8<sup>151</sup>). For further details on the model construction and TNA parametrization,<sup>380</sup> please refer to the experimental methods in Chapter 5.3.13. The **TPT3:NaM** base pair was positioned at the post-insertion site, designated as -1, either as the **dTPT3:tNaM** or the **dNaM:tTPT3** base pair (Figure 45A and 45B, Figure S24). Stable DNA-TNA duplexes in complex with the polymerase are maintained over 200 ns of simulation, demonstrating the potential for TNA elongation of **tTPT3** and **tNaM**. During the simulation, the distance between the 2'-OH of a nascent TNA strand containing either **tTPT3** or **tNaM** and the 3'  $\alpha$ -phosphate of the next incoming triphosphate was observed over time.



**Figure 45: Representative snapshot (cluster analysis) of a 200 ns MD simulation of the Kod-RSGA polymerase in complex with a DNA template containing either dNaM or dTPT3 and a nascent TNA oligonucleotide containing the respective counterpart tTPT3 or tNaM.** A) The **dNaM:tTPT3** duplex, showing the incoming TNA triphosphate in a suitable geometry for bond formation, promoted by  $Mg^{2+}$  ions (grey spheres). B) The **dTPT3:tNaM** duplex, showing the incoming TNA triphosphate in a suitable geometry for bond formation, promoted by  $Mg^{2+}$  ions (grey spheres). C) Time-dependent evolution of the distances between the 2'-OH of the nascent TNA and the 3'  $\alpha$ -phosphate (triphosphate) in both duplexes containing either the the **dNaM:tTPT3** (orange) or **dTPT3:tNaM** (cyan) UBP over a 200 ns simulation.



When the distance between the 2'-OH of the respective UB in the TNA strand and the 3'  $\alpha$ -phosphate of the incoming triphosphate is plotted against the time, the resulting traces reveal slight differences throughout the simulation (Figure 45C). This discrepancy becomes more pronounced when the distances between the 2'-OH nucleophile of the nascent TNA and the 3'  $\alpha$ -phosphate of the incoming natural TNA triphosphate are compared in a probability density function (Figure 46A). The mean distance between the 2'-OH of **tTPT3** and the 3'  $\alpha$ -phosphate of the incoming natural TNA triphosphate is 3.8 Å (Figure 46B), while this distance is 4.1 Å in the case of **tNaM** (Figure 46C). It is evident that the distance is marginally greater for the duplex with **tNaM** bearing the nucleophilic 2'-OH compared to the duplex with **tTPT3**. The increased distance to the incoming TNA triphosphate observed for the **dTPT3:tNaM** duplex is consistent with the previously obtained experimental observation that Kod-RSGA is stalled after incorporation of **tNaM**. This distance should impede the bond formation and thus further TNA elongation.



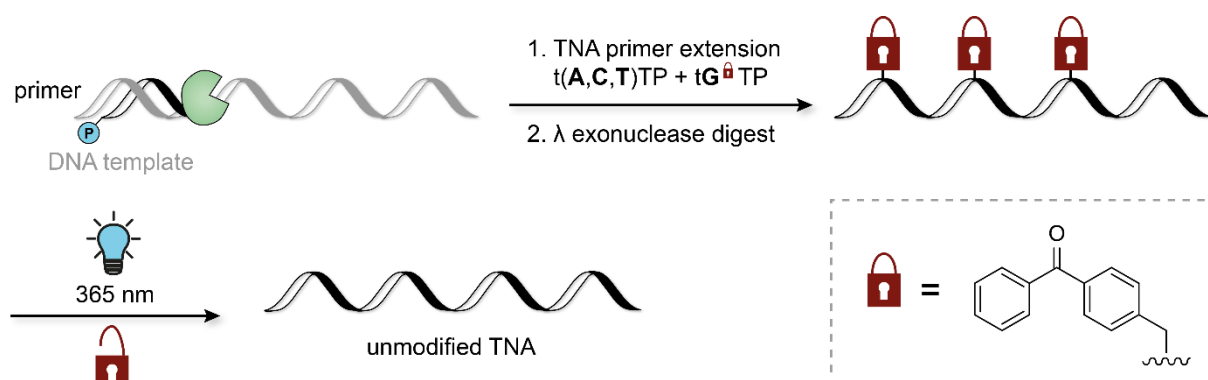
**Figure 46: Average distance of the 2'-OH nucleophile of either **tTPT3** or **tNaM** in a nascent TNA to the 3'  $\alpha$ -phosphate of the incoming natural TNA triphosphate.** A) Graphical representation of the 2'-OH to 3'  $\alpha$ -phosphate distance distribution in both DNA:TNA duplexes containing either **dNaM:tTPT3** (orange) or **dTPT3:tNaM** (cyan) over 200 ns simulation. B) Close up view of the **dNaM:tTPT3** duplex showing the incoming TNA triphosphate in a suitable geometry for bond formation, promoted by  $Mg^{2+}$  ions (grey spheres). The distance between the 2'-OH of the **tTPT3** and the 3'  $\alpha$ -phosphate of the incoming natural TNA triphosphate is 3.8 Å. C) Close up view of the **dTPT3:tNaM** duplex showing the incoming TNA triphosphate in a suitable geometry for bond formation, promoted by  $Mg^{2+}$  ions (grey spheres). The distance between the 2'-OH of the **tNaM** and the 3'  $\alpha$ -phosphate of the incoming natural TNA triphosphate is 4.1 Å.

In conclusion, this study represents the first demonstration of enzymatic XNA (here: TNA) synthesis with an expanded genetic alphabet. While the Terminator polymerase and the Kod-RI polymerase were unable to efficiently synthesize full-length exTNA with an internal UB modification, the engineered TNA polymerase Kod-RSGA demonstrated partial capability of synthesizing an exTNA oligonucleotide from a DNA template containing the unnatural base

pair **TPT3:NaM**. The results presented in this chapter demonstrate that the Kod-RSGA polymerase is capable of efficiently extending a DNA primer with a 21-nucleotide-long **TPT3**-containing TNA sequence from a **dNaM**-containing DNA template. To achieve this, the presence of  $Mn^{+2}$  ions was identified as a crucial factor. Moreover, misincorporation of **tGTP** opposite **dNaM** was observed to small extents. The incorporation of **tNaMTP** opposite a **dTPT3**-containing DNA template was also catalyzed by the polymerase, but subsequent elongation of the nascent TNA strand was not observed. *In silico* analyses of the closed ternary complex of Kod-RSGA polymerase indicate an increased distance between the 2'-OH of the **tNaM** in the TNA strand and the 3'  $\alpha$ -phosphate of the incoming TNA triphosphate compared to when **tTPT3** is in the TNA. This results in a less favorable conformation for bond formation in the case of **tNaM** elongation.

### 3.2 A traceless approach to suppress Hoogsteen base pair formation in enzymatic TNA preparation

To date, the enzymatic synthesis of TNA containing all four canonical nucleobases is limited by its low fidelity. The reduced fidelity is caused by Hoogsteen base pairs formed between dG in the DNA template and an incoming tGTP in the enzyme active site (see Figure 8).<sup>138,147</sup> In a previous study, the Chaput laboratory employed a 7-deaza-7-phenyl tGTP to obstruct the formation of the Hoogsteen base pair (see Figure 8).<sup>148</sup> However, this approach results in the generation of TNA molecules bearing modified guanine residues, which is not always desired. This illustrates the need for a reversible strategy to block the Hoogsteen base pairing site of guanine that would ultimately yield TNA with unmodified guanine nucleobases. Therefore, the attachment of a photocleavable group to the *N*<sup>7</sup> nitrogen of guanine was proposed. Based on previous work on photocleavable modifications on guanosine by the Rentmeister group, the benzophenone (BP) modification was selected as a suitable photocage for this project.<sup>381</sup> The photolabile BP-group is potentially capable of blocking the Hoogsteen base pairing site of tGTP during enzymatic synthesis. Following enzymatic TNA synthesis, the BP group can be removed by light, ultimately resulting in native TNA (Figure 47). Consequently, this approach represents a traceless alternative to current strategies to suppress G/G Hoogsteen base pairing in the enzyme active site during the preparation of TNA oligonucleotides.

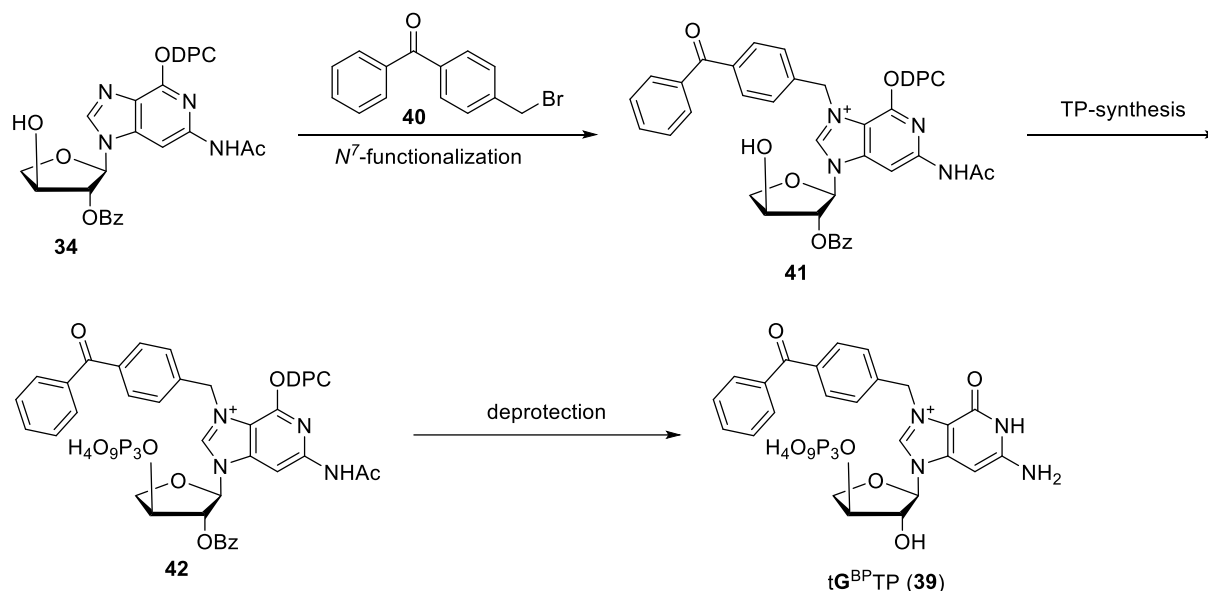


**Figure 47: Proposed strategy for a traceless approach to suppress G/G Hoogsteen base pair formation during enzymatic TNA synthesis.** During TNA primer extension, a benzophenone-modified threoguanosine triphosphate is used instead of tGTP to suppress potential G/G mismatches. After enzymatic preparation of TNA, the benzophenone modification can be removed tracelessly by irradiation with light ( $\lambda = 365$  nm) to release unmodified TNA.

This chapter presents the first synthetic steps towards the development of an *N*<sup>7</sup> benzophenone-photocaged threoguanosine building block. The synthesis towards the benzophenone-modified threoguanosine encountered several synthetic challenges, which are outlined in detail in this chapter.

### 3.2.1 Synthesis towards a threoguanosine building block containing a photolabile $N^7$ -modification

The synthesis of  $N^7$ -BP-modified threoguanosine triphosphate **39** ( $tG^{BP}TP$ ) was envisaged to start with protected  $tG$  **34**, the synthesis of which had already been demonstrated in Chapter 3.1.1.2. The proposed synthetic route is presented in Scheme 10.

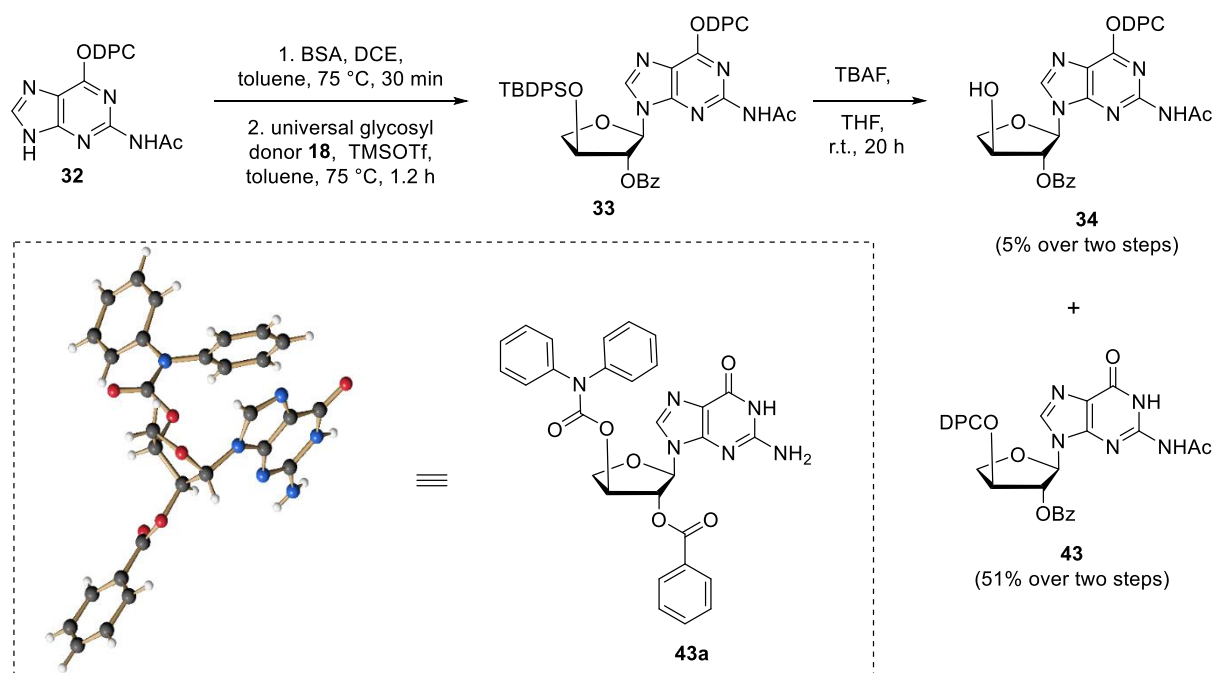


**Scheme 10: Proposed synthetic strategy toward the benzophenone-photocaged threoguanosine triphosphate **39**.** Step 1:  $N^7$ -functionalization of protected threoguanosine **34** with 4-(bromomethyl)benzophenone (BP-Br, **40**). Step 2: Triphosphate synthesis. Step 3: Removal of protecting groups under mild basic conditions to yield the target compound  $tG^{BP}TP$  (**39**).

It was intended to attach the BP group to the  $N^7$  nitrogen of  $tG$  **34** via a nucleophilic substitution with 4-(bromomethyl)benzophenone (BP-Br, **40**) to prepare compound **41** using a similar protocol as reported in the literature for the corresponding ribose building block.<sup>381</sup> Subsequently, 3'-triphosphorylation was foreseen to prepare compound **42**, followed by the removal of the remaining protecting groups under mild basic conditions, with the objective of ultimately yielding the desired  $tG^{BP}TP$  (**39**). It should be noted that parts of the synthetic work presented in this chapter were performed by the students Rose Jordan, Hannah Dochtermann, and Jan van Stiphoudt under my supervision.

Initially, the starting material, protected  $tG$  **34**, was required to be prepared once again, as it had been used up for the preparation of  $tGTP$  (**9**, see Chapter 3.1.1.2). To reiterate, the synthesis of **34** was achieved via a two-step reaction cascade, which included the Vorbrüggen coupling of the suitably protected guanine **32** and the universal glycosyl donor **18**, followed by 3'-desilylation of the crude product **33** using TBAF. As outlined in Chapter 3.1.1.2, in the first attempt to synthesize threoguanosine **34**, a solvent mixture of DCM and toluene was employed in the glycosylation step, diverging from the literature procedure, which utilized a DCE and toluene mixture.<sup>146,367</sup> This was done in order to circumvent the use of the highly toxic and

carcinogenic solvent DCE. However, only 8% of **34** was obtained after desilylation, which is considerably lower than the 69% reported in the literature for this two-step reaction.<sup>367</sup> In order to ascertain whether the solvent change was responsible for the low yield, the reaction was repeated as reported in the literature, this time using DCE instead of DCM at 75 °C (Scheme 11). However, following the subsequent deprotection of the crude product **33** with TBAF, the desired target nucleoside **34** was again obtained with a yield of only 5%. This suggests that the use of DCM can be excluded as a potential explanation for the low yield. In order to ascertain the cause of the observed low yield, the compounds isolated during column chromatography of the crude reaction mixture of **34** were subjected to further investigation. This was done in order to identify any byproducts that might be responsible for the low yield.

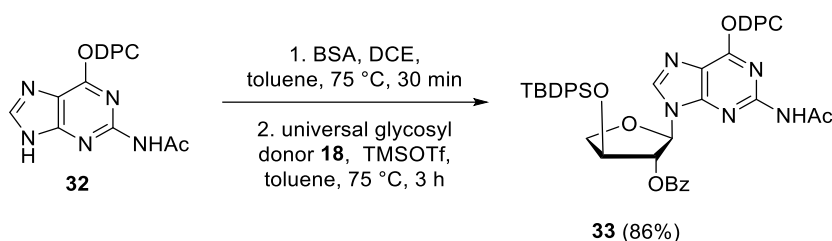


**Scheme 11: Two-step synthesis of threoguanosine **34** and the formed byproduct **43**.** Step 1: Vorbrüggen glycosylation of guanine **32** and universal glycosyl donor **18**. Step 2: TBAF deprotection of the TBDPS protecting group at the 3'-OH of **33**. Dashed box: The crystal structure depicts the formed byproduct **43a**, in which the diphenylcarbamoyl protecting group of the O<sup>6</sup> oxygen of the guanine moiety migrated to the 3'-OH group of the sugar moiety. The labile N<sup>2</sup>-acetyl protecting group of **43** was potentially removed during the prolonged crystallization process resulting in **43a**. The single-crystal X-ray diffraction analysis was performed by Dr. Jörg-Martin Neudörfel.

Consequently, an isolated byproduct with a smaller retention factor ( $R_f$ ) than the desired product **34** was subjected to comprehensive NMR analysis. <sup>1</sup>H NMR analysis of this compound resulted in a spectrum that was similar to that of the literature-known target compound **34**,<sup>367</sup> although with slight differences in the chemical shifts of the signals (Figure S85). Further structural elucidation via two-dimensional NMR spectroscopy yielded inconclusive results. Subsequent single-crystal X-ray diffraction analysis revealed that the diphenylcarbamoyl group had migrated from the O<sup>6</sup> oxygen of the guanine nucleobase to the deprotected 3'-hydroxyl group of the sugar (Scheme 11). This accounted for the discrepancies observed in the <sup>1</sup>H NMR

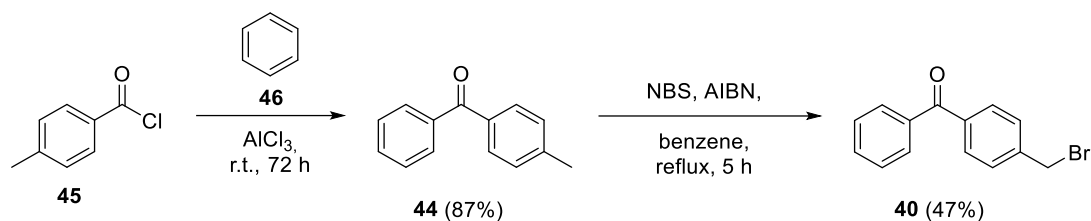
spectrum. The migration of the protecting group was likely facilitated by the weak basicity of TBAF during the deprotection step, resulting in the formation of **43** with a yield of 51% as a byproduct in this reaction. The substantial formation of byproduct **43** in this step may be explained by the fact that the deprotection step with TBAF was performed for 20 hours at ambient temperatures, rather than for 1 hour at 0 °C as reported in the literature.<sup>367</sup> It is notable that the crystal structure did not show the exocyclic *N*<sup>2</sup>-acetyl protecting group of **43**, as observed in the proton NMR spectrum. This suggests the potential removal of the acetyl group during the prolonged crystallization process, resulting in compound **43a** (Scheme 11).

The collective findings indicate that the glycosylation reaction was not the primary cause of the low yield of **34**; rather it was the migration of the diphenylcarbamoyl group during desilylation with TBAF. This was further corroborated by the fact that 3'-*O*-TBDPS-protected nucleoside **33** was obtained after Vorbrüggen glycosylation and column chromatography with an 86% yield when, in later attempts, desilylation was not successively performed *in situ* with the crude reaction mixture (Scheme 12).



**Scheme 12: Vorbrüggen glycosylation of protected guanine **32** with universal glycosyl donor **18**.** The 3'-*O*-TBDPS-protected threoguanosine **33** was isolated with a good yield of 86%.

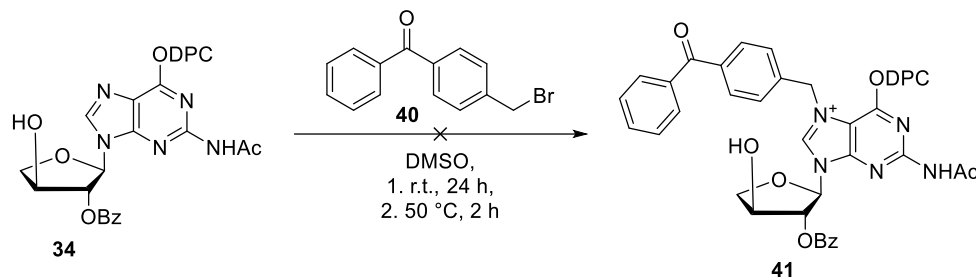
In order to synthesize **tG<sup>BP</sup>TP (39)**, it was necessary to prepare not only the **tG** nucleoside, but also **BP-Br (40)**. Consequently, **BP-Br (40)** was synthesized in two steps (Scheme 13). Firstly, 4-methylbenzophenone (**44**) was prepared under Friedel-Crafts acylation conditions according to a published procedure from Gobbi *et al.* using 4-methylbenzoylchloride (**45**), benzene (**46**), and  $\text{AlCl}_3$ .<sup>382</sup> Upon aqueous workup, the desired product **44** was obtained with a good yield of 87%. Subsequently, Wohl-Ziegler bromination in the benzylic position of **44** with *N*-bromosuccinimide (NBS) was conducted using an adapted procedure published by Tayama *et al.*<sup>383</sup> The desired target compound **BP-Br (40)** was generated with a yield of 47%, which is lower than that reported in the literature (90%<sup>383</sup>). Notably, the pure product **40** was isolated by precipitation rather than by column chromatography, which is a less time consuming method and compensates for the lower yield.



**Scheme 13: Two-step synthesis of 4-(bromomethyl)benzophenone (BP-Br, **40**).** Step 1: Friedel Crafts acylation of 4-methylbenzoylchloride (**45**) with benzene (**46**). Step 2: Wohl-Ziegler bromination of **44** using *N*-bromosuccinimide and azobis(isobutyronitrile) (AIBN).

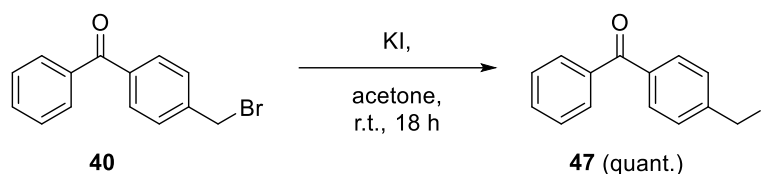
With the differently protected threoguanosines **33**, **34**, and **43** and BP-Br (**40**) in hand, initial synthetic attempts were made towards the BP-photocaged  $\text{tG}^{\text{BP}}\text{TP}$  (**39**) in a manner analogous to the  $N^7$ -protection of ribofuranosylguanine, as reported by Rentmeister *et al.*<sup>381</sup> Notably, all reactions for the attachment of the BP group to threoguanosine were unexceptionally conducted under exclusion of light to prevent cleavage of the photolabile group during the reaction.

Initially,  $N^7$ -functionalization of nucleoside **34** was performed with BP-Br (**40**; Scheme 14). According to literature, nucleoside **34** was stirred with one equivalent of BP-Br (**40**) in DMSO for 24 hours. However, the reaction was not successful and the formation of the desired target compound **41** was not observed. Thus, the reaction mixture was heated to 50 °C for 2 hours, but the desired product **41** was not formed.



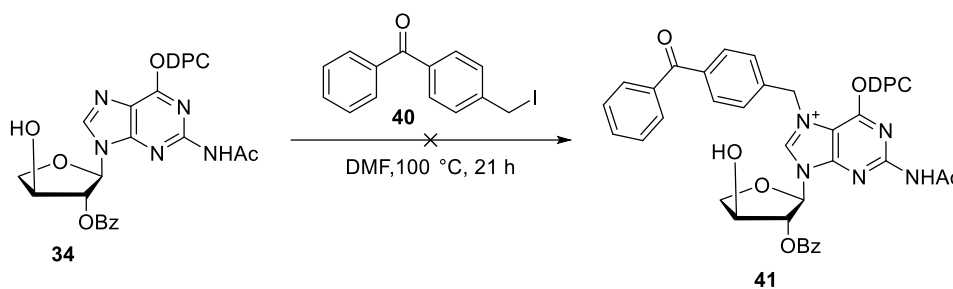
**Scheme 14: First attempt at the  $N^7$ -functionalization of threoguanosine **34** with one equivalent of BP-Br (**40**).** The reaction did not yield the desired BP-modified threoguanosine **41**.

Mechanistic reconsiderations regarding reaction conditions that would accelerate or even enable the desired product formation were made. The reaction proceeds via a nucleophilic substitution: The attachment of the BP-Br (**40**) to guanine necessitates a reaction between the nucleophilic  $N^7$  atom of guanine and the methylene C-atom of the BP-Br (**40**). This results in the liberation of bromide as a leaving group. Since leaving groups play an important role in nucleophilic substitutions, it was intended to replace the bromine with an even better leaving group. Accordingly, the iodo-substituted benzophenone **47** was proposed, which was synthesized via the Finkelstein reaction in a single step, starting from BP-Br (**40**) employing an adapted protocol from Ortiz-Trankina *et al.* (Scheme 15).<sup>384</sup> The desired 4-(iodomethyl)benzophenone (BP-I, **47**) was obtained in quantitative yield.



**Scheme 15: Synthesis of 4-(iodomethyl)benzophenone (BP-I, 47) via the Finkelstein reaction.**

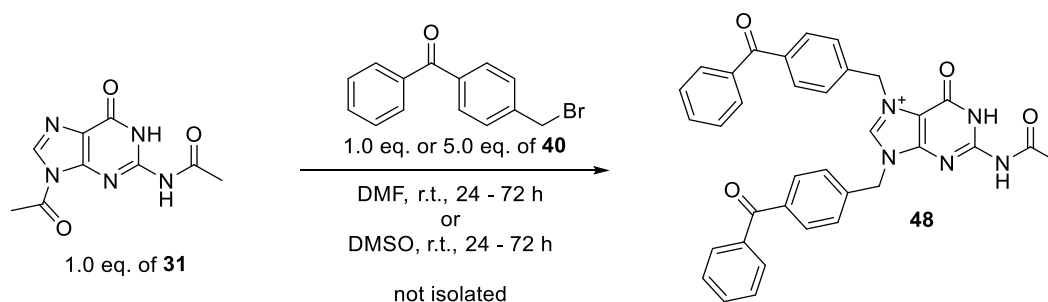
Moreover, DMF was tested as an alternative polar aprotic solvent here, as it was previously employed by Pautus *et al.* in similar reactions to modify the  $N^7$  position of various  $N^9$ -modified guanines.<sup>385</sup> Nevertheless, the utilization of one equivalent of BP-I (**47**) in DMF at 100 °C for the  $N^7$ -functionalization of nucleoside **34** did not result in the desired target compound **41** (Scheme 16).



**Scheme 16: Second attempt at the  $N^7$ -functionalization of threoguanosine **34** with one equivalent of BP-I (**47**). The reaction did not yield the desired BP-modified threoguanosine **41**.**

Subsequently,  $N^{2,9}$ -diacetyl-protected guanine **31** was employed as a simplified test system to examine the impact of increasing the equivalents of BP-Br (**40**) on the reaction (Scheme 17). Additionally, the efficacy of DMSO and DMF as solvents in this reaction was evaluated. Therefore, four small-scale test reactions were conducted, either with **31** and one equivalent of BP-Br (**40**) or with **31** and five equivalents of BP-Br (**40**). The reactions were performed in parallel in DMSO and DMF, respectively. Upon combining all components, a suspension was obtained in all cases, likely due to the poor solubility of guanine **31**. Interestingly, after 24 hours of stirring at ambient temperatures in DMF, the reaction mixture containing five equivalents of BP-Br (**40**) turned into a slightly yellow solution, indicating potential  $N^7$ -functionalization of guanine **31** with BP, which may have increased the solubility of the resulting compound in DMF. The same observation was made after 72 hours of stirring when DMSO was used as a solvent and five equivalents of BP-Br (**40**) were employed. This indicates that the reaction proceeded at a faster rate in DMF. To gain further insight into the reaction with five equivalents of BP-Br (**40**), the crude reaction mixture was subjected to LC-MS analysis. It is noteworthy that a molecular weight corresponding to the double benzylated species **48** was determined in the presence of five equivalents of BP-Br (**40**) in both DMF and DMSO (Figure S25). This indicates that functionalization with benzophenone also occurs at the  $N^9$  nitrogen of guanine, resulting in the formation of the  $N^{7,9}$ -benzophenone-modified byproduct **48**.

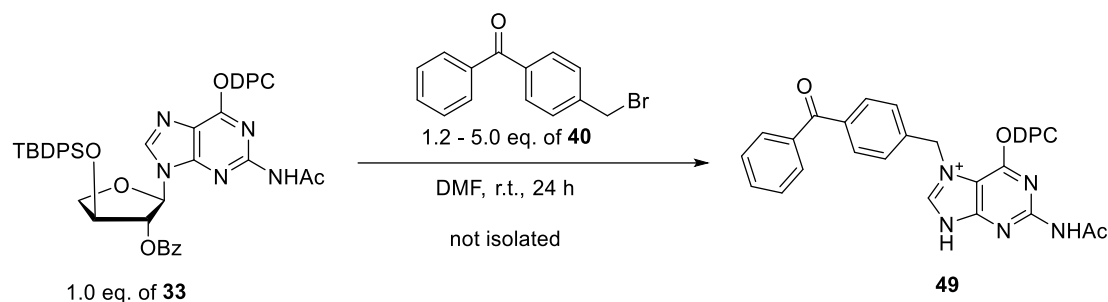




number of equivalents of BP-Br ( <b>40</b> )	solvent	time [h]	reaction product
1.0	DMF	72	no reaction
5.0	DMF	24	<b>48</b>
1.0	DMSO	72	no reaction
5.0	DMSO	72	<b>48</b>

**Scheme 17: Preliminary study on the  $N^7$ -functionalization of  $N^{2,9}$ -diacetyl-protected guanine **31** as a simplified test system with 1.0 and 5.0 equivalents of BP-Br (**40**).  $N^7$ -functionalization of **31** using 5.0 equivalents of BP-Br (**40**) resulted in the formation of  $N^{7,9}$ -benzophenone-modified product **48**. No reaction occurred when 1.0 eq. of BP-Br (**40**) was used. The corresponding LC-MS data is shown in Figure S25.**

Given the ease with which the acetyl group at the  $N^6$  nitrogen of **31** can be removed, it is plausible that the reaction has also occurred at this site. In contrast, similar reaction products were not observed when only one equivalent of BP-Br (**40**) was used. The results obtained from the simplified test system demonstrated that benzophenone functionalization at the  $N^7$  nitrogen of guanine is in general possible when using more than one equivalent of BP-Br (**40**). Furthermore, it was postulated, that the use of threoguanosine would circumvent the formation of the  $N^{7,9}$ -benzophenone-modified byproduct **48**. Consequently, the reaction was repeated with threoguanosine as the starting material. A preliminary study was conducted in which nucleoside **33** was stirred with different amounts of equivalents of BP-Br (**40**; *i.e.*, 1.2, 2.0, 3.0, 4.0, 5.0 eq.) in DMF at ambient temperature for 24 hours (Scheme 18). Subsequently, LC-MS analysis was conducted to examine the reaction progress. As observed in previous experiments, the use of 1.2 equivalents of BP-Br (**40**) is insufficient for successful attachment of the benzophenone group at the  $N^7$  position and only unreacted starting material **33** was identified by mass spectrometry. Interestingly, when two or more equivalents of BP-Br (**40**) were used, a molecular weight that is likely to be that of compound **49** was detected by LC-MS as the main product (Figure S26). This indicates that the BP group was indeed attached to threoguanosine **33**, but that the target compound decomposed, most likely due to depurination.

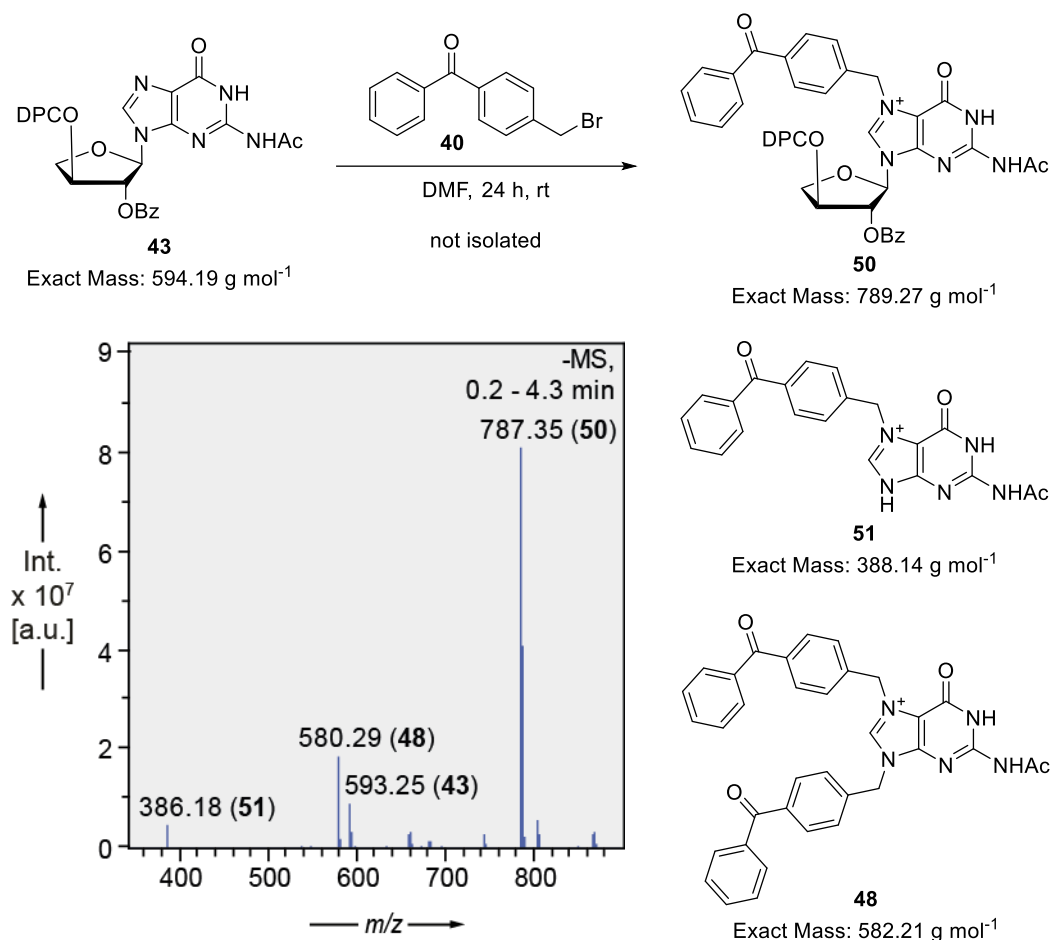


number of equivalents of BP-Br ( <b>40</b> )	reaction product
1.2	no reaction
2.0	<b>49</b> (depurination)
3.0	<b>49</b> (depurination)
4.0	<b>49</b> (depurination)
5.0	<b>49</b> (depurination)

**Scheme 18: Preliminary study on the *N*<sup>7</sup>-functionalization of threoguanosine **33** with different amounts of equivalents of BP-Br (**40**).** *N*<sup>7</sup>-functionalization of **33** using 2.0 or more equivalents of BP-Br (**40**) potentially results in the depurination product **49**. No reaction occurred when 1.2 eq. of BP-Br (**40**) were used. The corresponding LC-MS data is shown in Figure S26.

In addition to the potential decomposition product **49**, small quantities of unreacted starting material were detected by LC-MS, regardless of the quantity of BP-Br (**40**) utilized (Figure S26). The findings of this preliminary study indicated that the reaction is very cumbersome and that further optimization of the reaction is required.

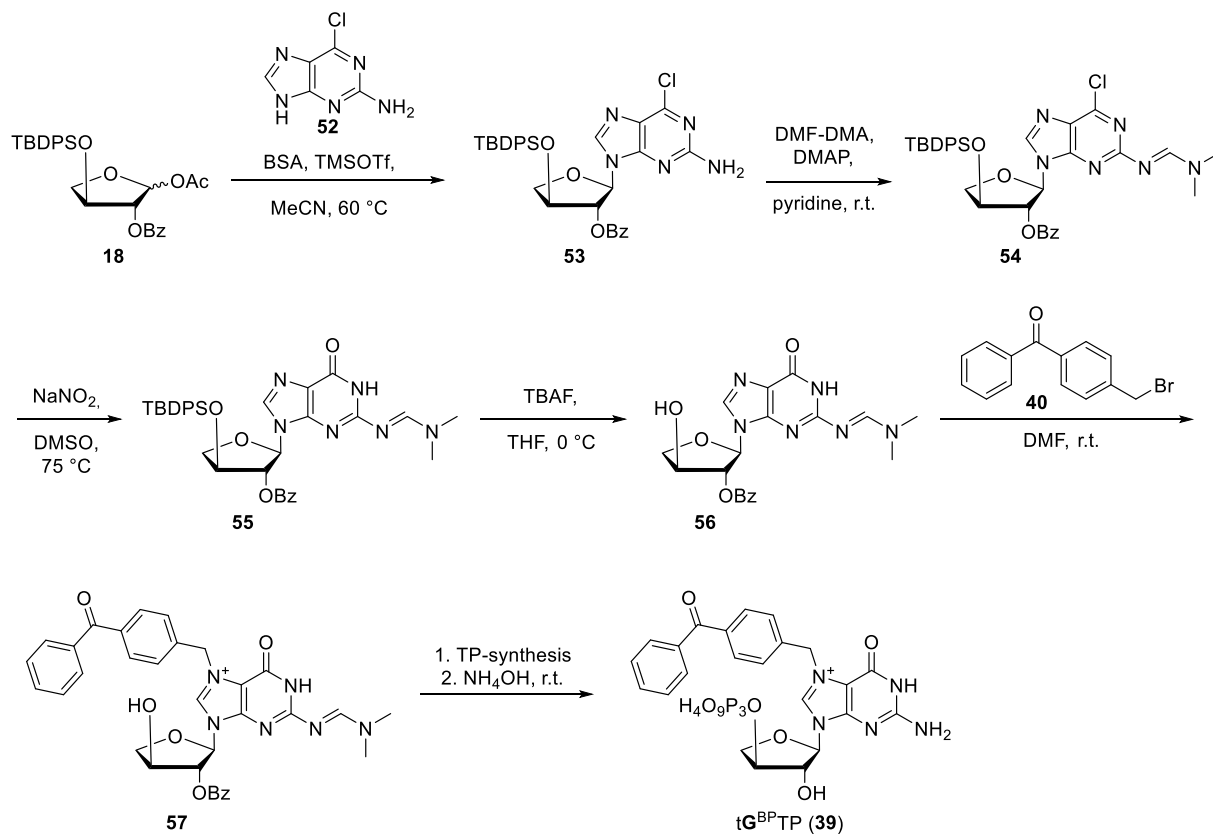
Interestingly, literature procedures for high-yielding functionalizations at the *N*<sup>7</sup> position of guanine are performed without a protecting group at the O<sup>6</sup> oxygen.<sup>381,385,386</sup> Therefore, a test reaction using the available nucleoside byproduct **43**, with the diphenylcarbamoyl group no longer attached to the O<sup>6</sup> oxygen of the guanine nucleobase, was performed (Scheme 19). DMF was employed as the solvent and five equivalents of BP-Br (**40**) were utilized. After 24 hours of stirring at ambient temperatures, the crude reaction was analyzed using LC-MS. The molecular weights obtained by LC-MS analysis indicated the presence of a mixture of compounds corresponding to the desired benzophenone-modified threoguanosine **50**, *N*<sup>7,9</sup>-benzophenone-modified guanine **48**, and depurination product **51**, as well as the starting material **43** (Scheme 19).



**Scheme 19: Products obtained by the benzophenone-functionalization of threoguanosine **43** without the  $O^6$  diphenylcarbamoyl protecting group at the guanine.** The ESI mass spectrum (negative mode) of the crude reaction shows a complex mixture of the starting material **43**, the desired target compound **50**, and the byproducts **48** and **51** obtained by depurination.

Separation of the reaction mixture using analytical RP-HPLC was possible and was thus used to track the reaction progress over time. In an attempt to accelerate the full conversion of the starting material **43** toward the product **50**, stirring of the reaction was continued and the progress was analyzed again after 48 hours using analytical RP-HPLC. However, extending the reaction time only reduced the amount of starting material **43**, but did not result in its complete conversion (Figure S27). Furthermore, the area under the curve (AUC) of the LC chromatograms obtained after 24 and 48 hours of reaction time were compared (Tables S3 and S4). This analysis revealed that the amount of product **50** remained unchanged, while the formation of the byproducts **48** and **51** increased over time. Thus, a longer reaction time did not result in an overall increase in product formation. Nevertheless, the presence of the desired product **50** was demonstrated, indicating that the guanine protecting groups may affect the stability of the  $N^7$ -benzophenone-modified threoguanosine and thus also the success of this reaction. In this small-scale test reaction, the 3'-OH group of the sugar was masked with the migrated diphenylcarbamoyl protecting group, thus precluding further 3'-triphosphorylation to yield tG<sup>BP</sup>TP (**39**).

The results obtained demonstrate that future attempts should focus on the  $N^7$ -functionalization of threoguanosine toward  $tG^{BP}TP$  (**39**) in the absence of protecting groups on the  $O^6$  oxygen of guanine and the 3'-OH of the threose sugar moiety. A proposed strategy for the synthesis of  $tG^{BP}TP$  (**39**) that meets the aforementioned requirements is presented in Scheme 20.



**Scheme 20: Proposed synthetic pathway to synthesize  $tG^{BP}TP$  (**39**) with the  $N^7$ -functionalization performed on threoguanosine **56** without the diphenylcarbamoyl protecting group at the  $O^6$  oxygen of guanine.**

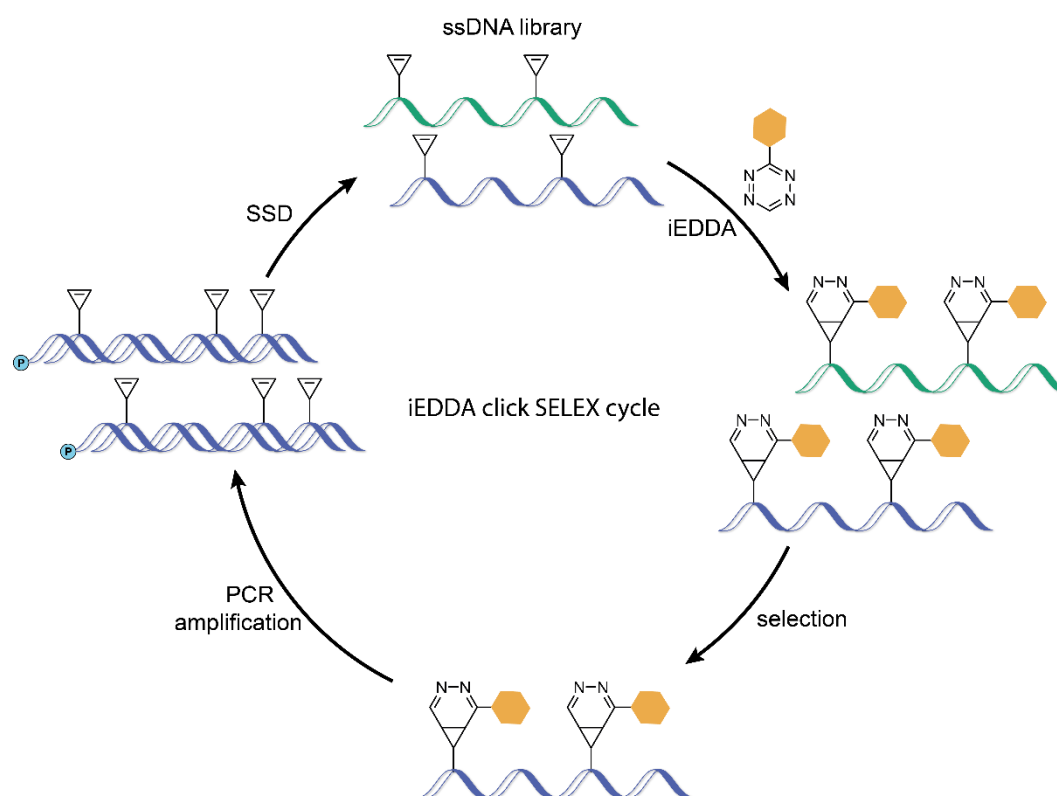
The proposed strategy is inspired by a recently published route for an optimized  $tGTP$  synthesis developed by Sarma *et al.*, which is performed without the bulky DPC protecting group at the  $O^6$  oxygen of guanine.<sup>387</sup> The proposed new route towards  $tG^{BP}TP$  (**39**) would also start with the universal glycosyl donor **18**, which would be coupled to 6-chloropurine-2-amine (**52**, 6Cl-G) under Vorbrüggen conditions to give nucleoside **53**. As reported by Sarma *et al.*, the desired  $N^9$  regioisomer with the correct stereoconfiguration is selectively obtained via this route.<sup>387</sup> Furthermore, 6Cl-G is commercially available, circumventing the laborious, multi-step synthesis of the protected guanine **32**, which was required for the previous synthetic route. Following the glycosylation reaction, it is envisaged to protect the exocyclic amino group with  $N,N'$ -dimethylformamide dimethyl acetal (DMF-DMA), yielding nucleoside **54**. The protecting group is required in order to avoid cross-reactivity during the later performed triphosphorylation reaction. Subsequently, the nucleobase is intended to be converted into guanine to generate compound **55**. The reported reaction conditions for this step include stirring of nucleoside **54** with  $NaNO_2$  in DMSO at 75 °C.<sup>387</sup> Subsequently, the TBDPS protecting group at the 3'-OH is aspired to be removed using TBAF to yield compound **56**. Thereafter, the attachment of the

BP-Br (**40**) to the  $N^7$  nitrogen of the threoguanosine **56** is planned to prepare nucleoside **57**, followed by selective 3'-triphosphorylation using standard protocols.<sup>370-372</sup> The final removal of the 2'-OH and exocyclic NH<sub>2</sub> protecting groups is reported to be accomplished with sat. aq. ammonium hydroxide solution,<sup>387</sup> which would ultimately yield tG<sup>BP</sup>TP (**39**). If one considers the synthesis of the protected guanine that was necessary for the previous route, the proposed strategy would not increase the overall number of reaction steps required to synthesize the final building block **39**, rendering it a viable synthetic route for upcoming approaches.

In summary, a variety of reaction conditions were examined to achieve the synthesis of benzophenone-protected threoguanosine. It was demonstrated that the attachment of the benzophenone moiety at the  $N^7$  position of threoguanosine is generally feasible, although the yield is relatively low due to incomplete conversion of the starting material and partial depurination of the product during the synthesis. The synthetic steps performed provided valuable insights and revealed several key aspects for future synthetic approaches towards tG<sup>BP</sup>TP (**39**). The benzophenone functionalization of the  $N^7$  nitrogen of threoguanosine requires the use of a greater number of equivalents of BP-Br (**40**) in comparison to the literature procedure for the corresponding ribose building block.<sup>381</sup> Moreover, the reaction rate was found to be faster when DMF was used as the solvent compared to DMSO. Additionally, the presence of protecting groups on the guanine nucleobase of threoguanosine was found to impede the success of the reaction. Therefore, it would be beneficial for future work to consider the  $N^7$ -functionalization step on an unprotected threoguanosine following the proposed synthetic route.

### 3.3 Towards cyclopropene-modified nucleotides for an iEDDA click-SELEX approach

This project is a collaboration with the Mayer research group (University of Bonn), which is strongly involved in aptamer selection through the click-SELEX approach. In click-SELEX, EdUTP (**3**) is utilized to prepare alkyne-modified DNA libraries. During the SELEX process, the resulting alkyne-modified DNAs are then further functionalized with azide-containing chemical entities via CuAAC to increase the chemical diversity of the resulting clickmers.<sup>253</sup> However, only recently, researchers from the Mayer laboratory have identified a significant issue: Even after stringent purification of the oligonucleotides enriched during click-SELEX, the copper ions used to catalyze the functionalization step via the CuAAC reaction are not completely removed. They found evidence that clickmers selected by CuAAC-based click-SELEX fold into three-dimensional structures depending on these residual copper ions, which is very likely to influence the three-dimensional folding of the clickmer. Consequently, if a clickmer is selected *in vitro* and is ultimately synthesized by chemical solid-phase synthesis for *in vivo* applications, the folding of the clickmer in the absence of copper ions may deviate from its original three-dimensional structure during the selection process. A different three-dimensional structure of the clickmer will most likely result in the loss of its ability to bind the target, thereby depleting its function and possibly causing off-target effects. Therefore, the concept of establishing a click-SELEX approach based on a copper-free bioorthogonal click reaction, such as the iEDDA reaction, was developed. In order to enable click-SELEX based on the iEDDA reaction, it is necessary to perform DNA library preparation with a nucleotide that carries a suitable reactive handle, for example cyclopropene. This allows downstream functionalization of DNA with tetrazine-containing modifications using the iEDDA reaction.<sup>261</sup> A schematic representation of an iEDDA-based click-SELEX approach is depicted in Figure 48.



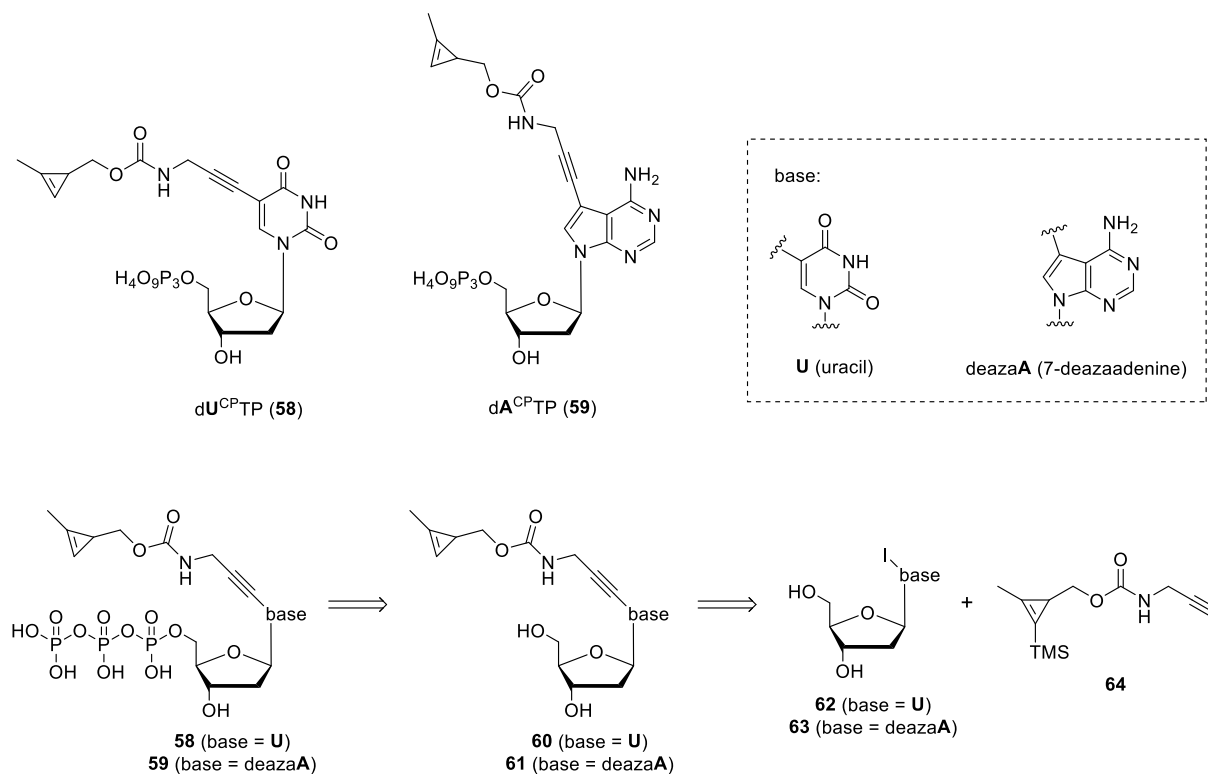
**Figure 48: Schematic representation of the envisioned iEDDA-based click-SELEX process.** In iEDDA-based click-SELEX, a cyclopropene-modified ssDNA library is functionalized with a tetrazine-conjugated modification. The modified sequences are then incubated with the target of interest. The unbound sequences are removed and the bound sequences are recovered (selection). The recovered sequences are amplified by PCR using a CP-modified nucleotide and one 5'-phosphorylated primer (PCR amplification). In the next step, single strand displacement is performed (SSD). Thereby, the dsDNA library is converted into an ssDNA library by  $\lambda$ -exonuclease digestion of the 5'-phosphorylated DNA strand. The resulting ssDNA library can then be submitted into the next iEDDA-based click-SELEX cycle.

This chapter presents the first steps towards the ultimate goal of an iEDDA-based click-SELEX approach for the selection of novel clickmers without interfering copper ions. This includes the syntheses of cyclopropene-modified deoxyadenosine and deoxyuridine triphosphates (Chapter 3.3.1). Furthermore, preliminary experiments on the polymerase processing of cyclopropene-modified deoxyadenosine triphosphate are presented, as well as the downstream functionalization of the obtained cyclopropene-modified DNA (Chapter 3.3.2). Moreover, evaluation of the thermal stability of the cyclopropene modification by NMR and LC MS is outlined (Chapter 3.3.3).

### 3.3.1 Synthesis of cyclopropene-modified deoxyribonucleotides

Cyclopropene-modified nucleotides permit the implementation of a click-SELEX approach based on the iEDDA click reaction. However, in contrast to EdUTP (**3**), cyclopropene-modified nucleotides are not commercially available and have to be synthesized in-house. Given that polymerases readily accommodate modifications at the  $C^5$  position of pyrimidine nucleobases and at the  $C^7$  position of 7-deazapurine nucleobases,<sup>163</sup> these positions appear suitable for the attachment of a cyclopropene moiety.

In the course of this study, the synthesis of two distinct methylcyclopropene (CP)-modified nucleotides was envisaged:  $C^5$  methylcyclopropene-modified 2'-deoxyuridine triphosphate ( $dU^{CP}TP$ , **58**) and  $C^7$  methylcyclopropene-modified 7-deaza-2'-deoxy-adenosine triphosphate ( $dA^{CP}TP$ , **59**). The methylcyclopropene modification was selected as it has demonstrated enhanced stability compared to a simple cyclopropene group under aqueous conditions.<sup>365</sup> Based on previously published procedures for the assembly of these cyclopropene-modified nucleoside triphosphates, the synthesis of both  $dU^{CP}TP$  (**58**) and  $dA^{CP}TP$  (**59**) was planned analogously.<sup>214,263,363</sup> As in these previously published procedures, the methylcyclopropene moiety, was intended to be attached to the corresponding nucleotide via a terminal alkyne-based linker, which connects the CP moiety via a carbamate bridge to the nucleotide. The retrosynthetic strategy for  $dU^{CP}TP$  (**58**) and  $dA^{CP}TP$  (**59**) is outlined in Scheme 21.



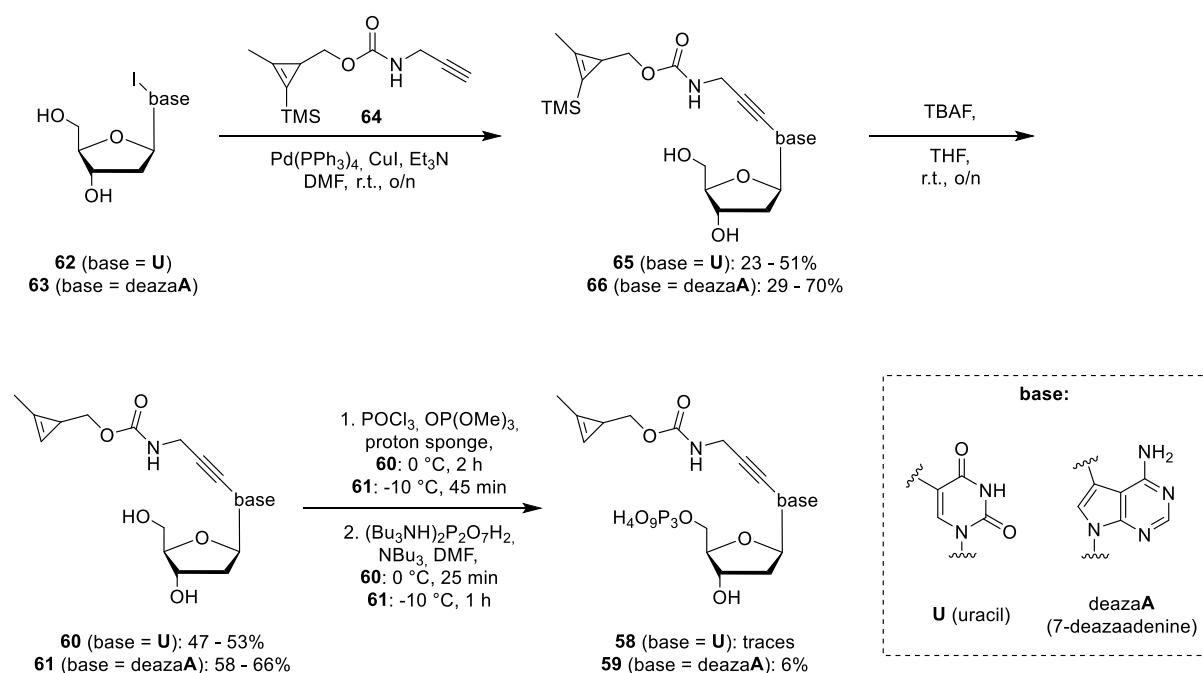
**Scheme 21: Retrosynthetic considerations providing access to the target compounds  $dU^{CP}TP$  (**58**) and  $dA^{CP}TP$  (**59**).** The retrosynthesis of both compounds proceeds identically, and thus the nucleobases are shown in a generalized manner as “base”. The nucleobases are depicted in the dashed box.



In essence, retrosynthetic considerations involve the disassembly of dU<sup>CP</sup>TP (**58**) and dA<sup>CP</sup>TP (**59**) into the corresponding methylcyclopropene-modified nucleosides **60** and **61**. The triphosphate can later be reintroduced by standard triphosphorylation reactions for nucleosides.<sup>370</sup> Further fragmentation of the CP-modified nucleosides **60** and **61** yield the corresponding C<sup>5</sup>-iodinated deoxyuridine **62** or C<sup>7</sup>-iodinated 7-deazaadenosine **63** and the methylcyclopropene carbamate alkyne linker **64**. The resulting nucleoside synthons **62** and **63** are commercially available can be joined with **64** by a Sonogashira reaction to give methylcyclopropene-modified deoxyuridine **60** and methylcyclopropene-modified deoxyadenosine **61**.

The methylcyclopropene carbamate alkyne linker system **64**, which allows straightforward attachment to iodinated nucleosides, was developed by Dr. Frank Eggert, a former member of the Kath-Schorr group, and was kindly reproduced by Sarwar Aziz for use in this project.<sup>214</sup>

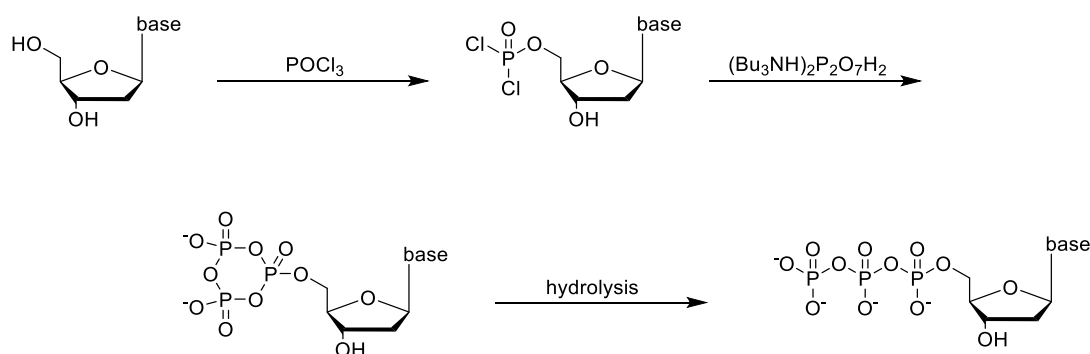
To construct dU<sup>CP</sup>TP (**58**) and dA<sup>CP</sup>TP (**59**), a three-step synthesis was performed starting from commercially available 5-iodo-2'-deoxyuridine (**62**) and 7-deaza-7-iodo-2'-deoxyadenosine (**63**), respectively (Scheme 22).



**Scheme 22: Synthetic route towards cyclopropene-modified nucleoside triphosphates dU<sup>CP</sup>TP (**58**) and dA<sup>CP</sup>TP (**59**).** The synthesis was performed analogously, and thus the nucleobases are shown in a generalized manner as “base”. The nucleobases are depicted in the dashed box.

In the first step, the trimethylsilyl (TMS)-protected methylcyclopropene carbamate alkyne linker **64** was coupled to either iodinated nucleosides **62** or **63** via a Sonogashira cross-coupling reaction in the presence of Et<sub>3</sub>N, CuI, and tetrakis(triphenylphosphine)palladium(0) as a catalyst. The TMS protection of the cyclopropene moiety is of critical importance in this step, as cyclopropenes and mono-substituted cyclopropenes are highly attractive substrates

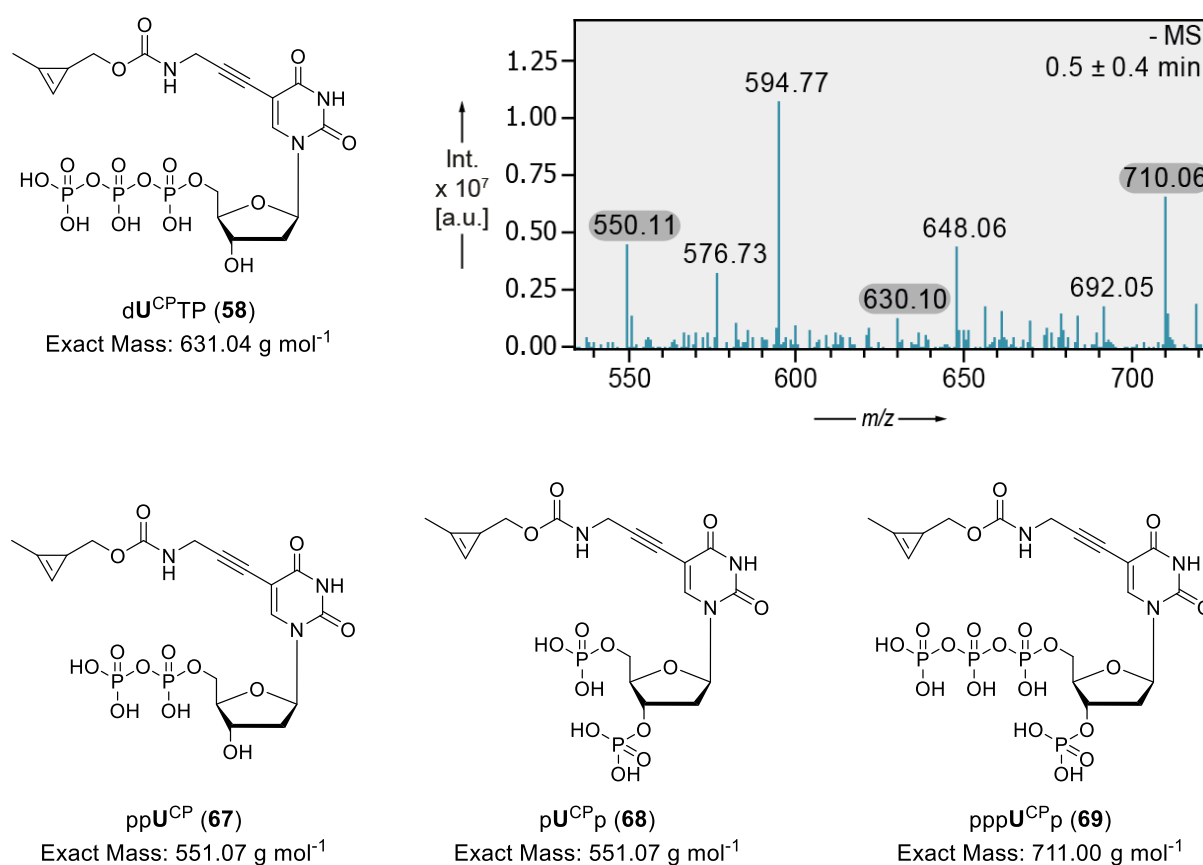
for transition metals such as palladium.<sup>214,388</sup> The Sonogashira reaction was performed three times with each iodinated nucleoside during the course of this study, with moderate yields between 23 - 51% of TMS-**U**<sup>CP</sup> (**65**) and 29 - 70% of TMS-**A**<sup>CP</sup> (**66**). The obtained yields are in accordance with those reported in the literature.<sup>214,363</sup> The wide range of yields can be explained by the initial use of tetrakis(triphenylphosphine)palladium(0) that was already in use and stored for an extended period of time. This has most likely resulted in the oxidation of the triphenylphosphine to triphenylphosphine oxide, inactivating the catalyst. This decomposition of the catalyst was also evidenced by the color of the catalyst, which was more greenish than the usual yellow color. Replacing the old batch of Pd(PPh<sub>3</sub>)<sub>4</sub> catalyst with a fresh batch resulted in increased reaction yields in the subsequent reactions. Subsequently, the TMS protecting group was removed from nucleosides **65** and **66** using TBAF. This deprotection reaction was performed twice with each nucleoside during the course of this study, yielding the CP-modified nucleosides d**U**<sup>CP</sup> (**60**) and d**A**<sup>CP</sup> (**61**) with moderate yields of 47 - 58% and 53 - 66%, respectively. The obtained compounds **60** and **61** were then converted to the corresponding triphosphates **58** and **59** using a three-step *in situ* reaction cascade based on the work of Yoshikawa and Ludwig (Scheme 23).<sup>370-372</sup> In general, this method involves the reaction of (deoxy)nucleosides with POCl<sub>3</sub> to give the corresponding nucleoside-5'-phosphorodichloridate, followed by reaction with tributylammonium pyrophosphate to the 5'-cyclic triphosphate. Finally, the desired nucleoside-5'-triphosphate is obtained through buffered hydrolysis.<sup>389</sup>



**Scheme 23: Schematic representation of the Yoshikawa and Ludwig method for the generation of nucleoside 5'-triphosphates.**<sup>389</sup>

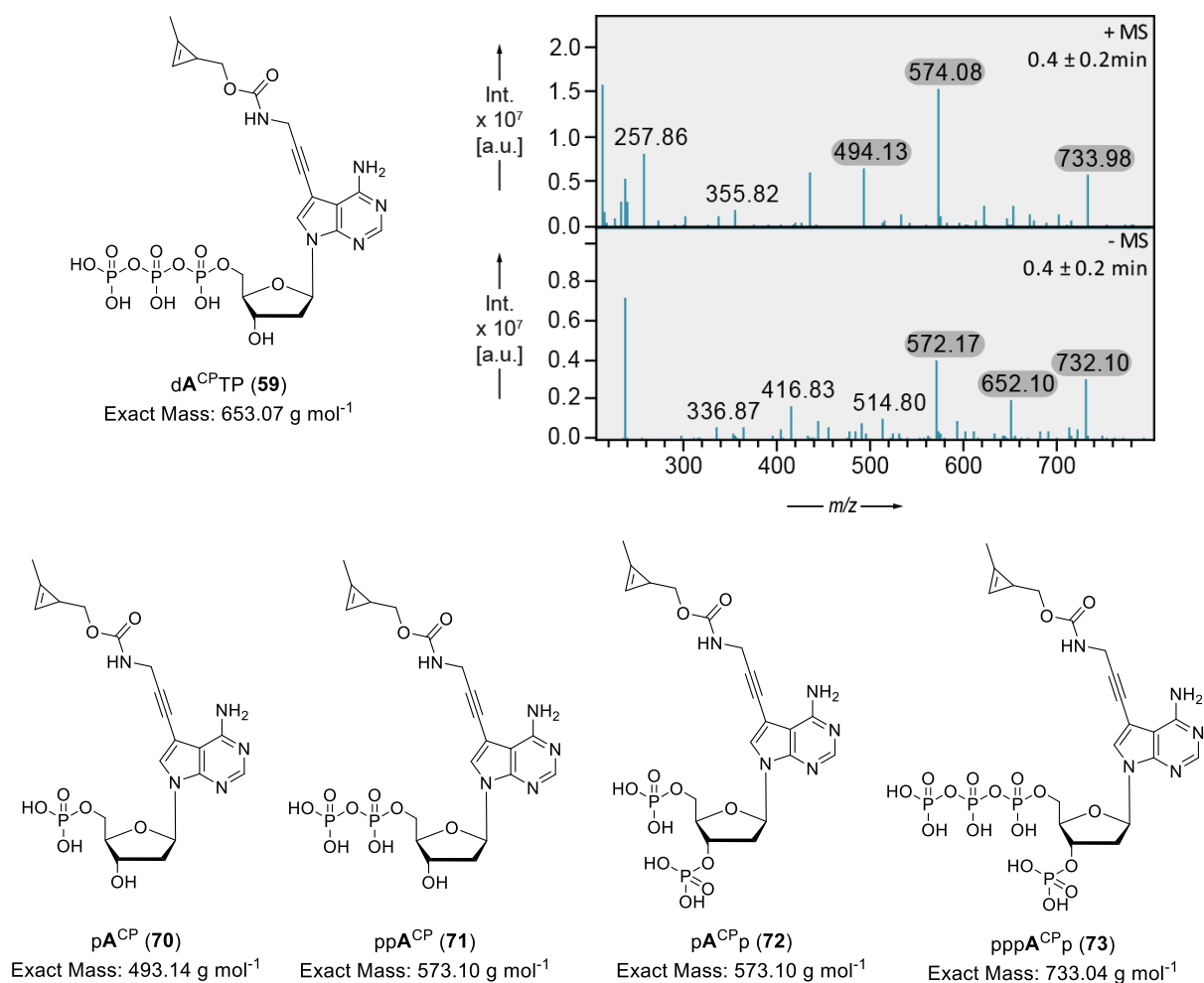
The advantage of this method is that, according to the literature, no protecting group is required on the secondary 3'-hydroxyl group of deoxynucleosides, as this position is less favored compared to the more accessible primary 5'-hydroxyl group.<sup>389</sup> However, LC-MS analysis of the crude reaction mixtures from the triphosphorylation reactions of **U**<sup>CP</sup> (**60**) and **A**<sup>CP</sup> (**61**) revealed only small amounts of the corresponding desired triphosphates **58** and **59**, and the formation of several highly phosphorylated species, including byproducts with one, two, and four phosphate groups. In more detail, LC-MS analysis of the triphosphorylation reaction of the **U**<sup>CP</sup> nucleoside **60** revealed that only trace amounts of the desired **U**<sup>CP</sup> 5'-triphosphate **58** were

formed, and molecular weights corresponding to di- and tetraphosphorylated species were observed. Given the determined molecular weights, the proposed structures for the formed byproducts are likely to be  $\text{dU}^{\text{CP}}\text{-5'}$ -diphosphate ( $\text{ppU}^{\text{CP}}$ , **67**),  $\text{dU}^{\text{CP}}\text{-3',5'}$ -bisphosphate ( $\text{pU}^{\text{CPp}}$ , **68**), and  $\text{dU}^{\text{CP}}\text{-3'}$ -monophosphate-5'-triphosphate ( $\text{pppU}^{\text{CPp}}$ , **69**). The proposed structures for the byproducts formed are depicted in Figure 49. It is important to note that the structures postulated here are only conjectures based on the premise that the primary 5'-OH group usually has a higher reactivity compared to the 3'-OH group.<sup>389</sup> To unambiguously confirm the hypothesized structures of these byproducts, NMR would be required to assign the positions of the phosphate groups with certainty by their characteristic chemical shifts.<sup>390,391</sup>



**Figure 49: Analysis of the crude reaction mixture of  $\text{dU}^{\text{CP}}\text{-TP}$  (**58**) obtained after triphosphate synthesis.** Top left: Target compound  $\text{dU}^{\text{CP}}\text{-TP}$  (**58**). Top right: ESI mass spectrum (negative mode) obtained by LC-MS analysis of the crude reaction mixture. The UV chromatogram can be found in Figure S28. Bottom: Proposed byproducts that potentially formed during the triphosphate synthesis of  $\text{dU}^{\text{CP}}\text{-TP}$  (**58**).

Although the abundance of the desired 5'-triphosphate **59** is higher in the case of the triphosphorylation reaction with  $\text{dA}^{\text{CP}}$  nucleoside **61**, several byproducts were also formed. The molecular weights determined by LC-MS indicate the presence of mono-, di-, and tetraphosphorylated species in addition to the desired  $\text{dA}^{\text{CP}}$  triphosphate **59**. It can be reasonably assumed that these molecular weights can be attributed to the formation of  $\text{dA}^{\text{CP}}\text{-5'}$ -monophosphate ( $\text{pA}^{\text{CP}}$ , **70**),  $\text{dA}^{\text{CP}}\text{-5'}$ -diphosphate ( $\text{ppA}^{\text{CP}}$ , **71**),  $\text{dA}^{\text{CP}}\text{-3',5'}$ -bisphosphate ( $\text{pA}^{\text{CPp}}$ , **72**) and  $\text{dA}^{\text{CP}}\text{-3'}$ -monophosphate-5'-triphosphate ( $\text{pppA}^{\text{CPp}}$ , **72**, Figure 50).



**Figure 50: Analysis of the crude reaction mixture of dA<sup>CP</sup>TP (59) after triphosphate synthesis.** Top left: Target compound dA<sup>CP</sup>TP (59). Top right: ESI mass spectra (alternating mode) obtained by LC-MS analysis of the crude reaction mixture. The UV chromatogram can be found in Figure S28. Bottom: Proposed byproducts that potentially formed during the triphosphate synthesis of dA<sup>CP</sup>TP (59).

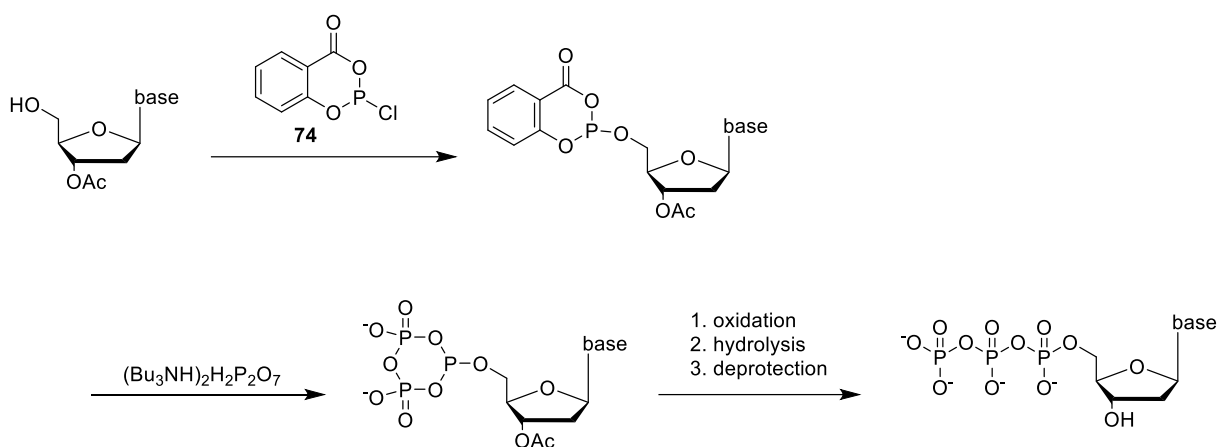
Similar byproducts were also observed by Borsenberger *et al.* and Gillerman *et al.* employing the same one-pot triphosphorylation method for deoxyribonucleosides and ribonucleosides, respectively.<sup>392,393</sup> The literature suggests different mechanisms for the formation of these byproducts during triphosphate synthesis using the Yoshikawa-Ludwig method.<sup>392,394</sup> The formation of monophosphate byproducts as in case of pA<sup>CP</sup> (70) can be explained by direct hydrolysis of the corresponding 5'-dichlorophosphate intermediate. The formation of nucleoside diphosphates involves the unwanted hydrolysis of the pyrophosphate, which can then react with the 5'-dichlorophosphate intermediate, possibly giving the diphosphate byproducts 67 and 71. The reaction of POCl<sub>3</sub>, which is used in excess in this reaction, with the less favored 3'-hydroxyl group of the 5'-dichlorophosphate intermediate, could result in the bisphosphate byproducts 68 and 72 as well as the nucleoside-3'-monophosphate-5'-triphosphates 69 and 73.<sup>392,394</sup> Notably, Knoblauch *et al.* reported that highly phosphorylated byproducts are particularly formed when bulky substituents are attached to the nucleobase.<sup>390</sup>

Attempts to isolate the desired  $dU^{CP}TP$  (**58**) from the crude reaction mixture using RP-HPLC were unsuccessful. Even when a very long and flat gradient was employed, the product eluted in the same peak as the di- and tetra-phosphorylated species **67**, **68**, and **69**. In the case of  $A^{CP}TP$  (**59**), separation of the desired product **59** from the byproducts by RP-HPLC was laborious and baseline separation was not achieved. Nevertheless, traces of reasonably pure product **59** were isolated according to LC-MS data. However, the  $^1H$  NMR spectrum revealed primarily the presence of triethylamine and peaks corresponding to the product **59** were difficult to discern.

The separation of reaction mixtures containing several different phosphorylated species has also been reported in the literature to be a challenging task. It is often reported that standalone RP-HPLC is inadequate because the highly polar phosphate species elute rapidly in one peak.<sup>392,395</sup> Consequently, ion exchange is frequently employed to separate the various phosphorylated species.<sup>395</sup> However, despite this, RP-HPLC is still necessary subsequent to ion exchange in order to remove inorganic phosphate salts that are formed during the reaction. The two-step purification of the crude product using this coupled ion exchange and RP-HPLC technique is lengthy and laborious, which is likely to result in a decrease in yield due to the relatively low chemical stability of nucleotides.<sup>396</sup> Therefore, the primary goal should be to directly prevent the formation of byproducts, and thus optimization of the triphosphorylation reaction was pursued. Of course, special care was taken to exclude moisture from the reaction. This was achieved by even more stringent drying of all starting materials and equipment used. In addition, the reaction was optimized with respect to temperature and time based on the findings of Borsenberger *et al.* and Gillerman *et al.*<sup>392,393</sup> In the initial attempt at triphosphate synthesis of  $dU^{CP}TP$  (**58**) and  $dA^{CP}TP$  (**59**), the addition of  $POCl_3$  in the first step of the procedure resulted in the reaction with both the 5'- and 3'-hydroxyl groups. Consequently, in the second attempt, the reaction temperature was lowered from 0 °C to -10 °C in order to direct the reaction to the more accessible 5'-position. Moreover, the reaction time of this first step was reduced to 45 minutes, a notable decrease from the previously applied 120 minutes. This adjustment was implemented to prevent the reaction of  $POCl_3$  with the 3'-hydroxyl group over time, a phenomenon that is possible due to the excess of  $POCl_3$  employed. Moreover, in the initial attempt at triphosphate synthesis, the addition of tributylammonium pyrophosphate in the second step of the reaction did not result in complete conversion of the nucleoside-5'-phosphorodichloridate intermediate to the desired nucleoside-5'-triphosphate. Hence, the reaction time was extended from 30 minutes to 60 minutes. This adaptation should facilitate the reaction between the nucleoside-5'-phosphorodichloridate and the pyrophosphate, thus promoting the formation of the desired 5'-triphosphate, while simultaneously reducing the amount of the 5'-monophosphate byproduct **70**. Additionally, the temperature in the second step was also reduced from 0 °C to -10 °C, as this has been reported

to result in a further decrease in byproduct formation.<sup>392</sup> Following hydrolysis of the 5'-cyclic triphosphate intermediate using TEAB buffer, LC-MS analysis of the crude reaction mixtures was conducted. LC-MS analysis of the crude product resulting from the triphosphorylation of dU<sup>CP</sup> nucleoside **60** did not reveal the presence of a molecular weight corresponding to the desired product **58** or any of the previously detected side products. In contrast, the LC-MS analysis of the crude product from the triphosphorylation reaction of dA<sup>CP</sup> nucleoside **61** confirmed that the detected molecular weight is consistent with the target mass of dA<sup>CP</sup>TP (**59**), with the sole byproduct being the dA<sup>CP</sup>-5'-monophosphate **70**. Notably, the separation of the mono- and triphosphates **70** and **59** was achieved by RP-HPLC using a long and flat gradient, resulting in the successful isolation of pure dA<sup>CP</sup>TP (**59**) with a yield of 6%.

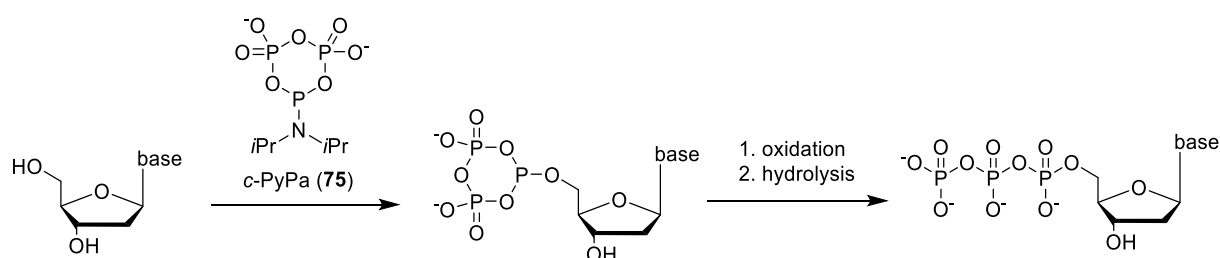
The disparate outcomes observed in the triphosphate synthesis of dU<sup>CP</sup>TP (**58**) and dA<sup>CP</sup>TP (**59**) demonstrate that the success of the triphosphate synthesis is contingent upon the substrate, and that there is no universal protocol for the synthesis of nucleoside triphosphates.<sup>396</sup> In future work, it would be beneficial to further optimize the reaction conditions for the synthesis of dU<sup>CP</sup>TP (**58**) and to test alternative methods for triphosphate synthesis. Potential alternative procedures for the nucleoside triphosphate synthesis include the Ludwig-Eckstein method, which is arguably the most popular, and the recently developed Jessen method.<sup>397–399</sup>



**Scheme 24:** Schematic representation of the popular Ludwig-Eckstein nucleoside triphosphate synthesis approach.<sup>400</sup>

In the Ludwig-Eckstein method, salicyl phosphorochloridate (**74**) is employed to react with the 5'-OH group of nucleosides, thereby yielding an activated phosphite intermediate (Scheme 24). As phosphorylation with salicylic phosphorochloridate (**74**) is not regioselective for the 5'-OH, the secondary 3'-OH and, in the case of ribonucleosides, also the 2'-OH group must be equipped with protecting groups. In the second step, tributylammonium pyrophosphate is then added, which replaces the salicylic acid, resulting in the formation of a 5'-deoxycyclotriphosphate ester. The iodine-mediated oxidation followed by hydrolysis ultimately yields the desired triphosphate.<sup>400</sup> The use of protecting groups prevents the

formation of byproducts, which also significantly simplifies the purification and isolation of the 5'-triphosphate. However, in order to selectively install protecting groups to the secondary alcohols, first the primary 5'-OH group needs to be tritylated. Following the subsequent protection of the secondary hydroxyl groups of the sugar using an orthogonal protection group strategy, the 5'-position must be selectively unmasked in order to allow triphosphate synthesis. Once the triphosphate has been successfully synthesized, the protecting groups at the 2'-OH and 3'-OH must be removed using mild conditions in order to maintain the integrity of the compound.<sup>389</sup> These multiple protection and deprotection reactions add further steps to the overall synthetic route, which most likely results in a decrease in the final overall yield.



**Scheme 25:** Schematic representation of the nucleoside triphosphate synthesis approach using the cyclic pyrophosphoryl P-amidite approach developed by the Jessen research group.<sup>398</sup>

Another valuable method for the synthesis of nucleoside triphosphates is a recently developed procedure by Jessen and co-workers (Scheme 25),<sup>398</sup> which has only recently been implemented in our laboratory. In this method, cyclic pyrophosphoryl P-amidite (c-PyPa, **75**) is employed as the phosphitylating reagent, resulting in the 5'-deoxycyclotriphosphate ester. This intermediate is subsequently oxidized and hydrolysed to give the 5'-triphosphate. It is anticipated that the sterically demanding c-PyPa (**75**) will react selectively with the best accessible 5'-hydroxyl group, even in the case of unprotected nucleosides. Moreover, the simultaneous introduction of all three phosphorus atoms of the triphosphate moiety should also result in a reduction of the formation of the monophosphate byproduct.<sup>397</sup> Therefore, the latter method appears to be more suitable in the context of this work and should be preferentially tested to synthesize dU<sup>CP</sup>TP in future work, as it has the potential to reduce byproduct formation without adding further steps to the synthetic route.

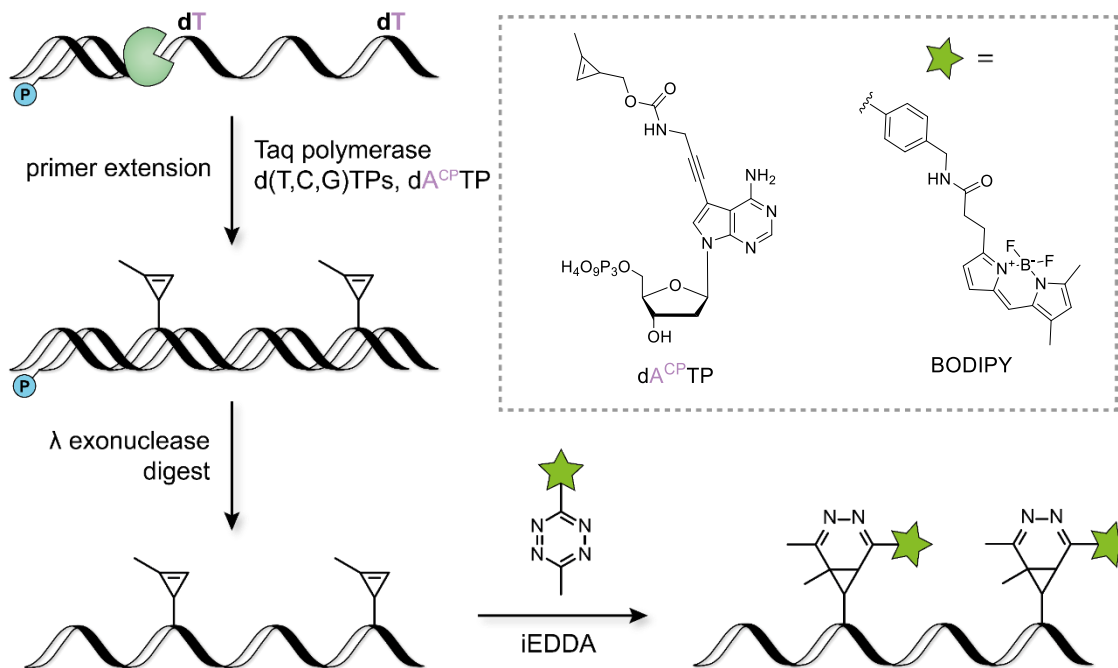
In summary, both nucleosides, dU<sup>CP</sup> (**60**) and dA<sup>CP</sup> (**61**) were successfully prepared in two steps with satisfactory overall yields of 41% and 34%, respectively. Furthermore, the synthesis of dA<sup>CP</sup>TP (**59**) was achieved with a 6% yield for the triphosphorylation reaction after meticulous optimization of the reaction conditions, resulting in sufficient quantities for further *in vitro* testing. Conversely, approaches to optimize the triphosphorylation reaction towards dU<sup>CP</sup>TP (**58**) were unsuccessful. Consequently, upcoming research should focus on further adjusting and optimizing the reaction conditions for the preparation of dU<sup>CP</sup>TP (**58**). Overall, the results demonstrate that the success of the triphosphorylation is highly dependent on the

substrate, and that no universal protocol can be applied to each nucleoside. Instead, tailored reaction conditions must be employed for each nucleoside substrate.

Following the successful synthesis of  $\text{dA}^{\text{CP}}\text{TP}$  (**59**), a series of preliminary experiments was conducted. The objective was to gain initial insight into the polymerase processing and subsequent downstream functionalization of  $\text{dA}^{\text{CP}}\text{TP}$  (**59**), as well as to evaluate the thermal stability of the cyclopropene modification.

### 3.3.2 Enzymatic preparation of cyclopropene-modified DNA and downstream iEDDA click functionalization

Two key factors are essential for the success of an iEDDA-based click-SELEX approach: Firstly, the employed nucleotide bearing the reactive CP-handle must be accepted as a substrate by cognate polymerases; and secondly, the reactive handle must successfully undergo the downstream iEDDA click functionalization after incorporation into DNA. Therefore, primer extension assays were conducted to prepare CP-modified DNA ( $\text{DNA}^{\text{CP}}$ ). Subsequently, the  $\text{DNA}^{\text{CP}}$  obtained was subjected to an iEDDA click reaction with a tetrazine-fluorophore conjugate. The attachment of the fluorescent label allows for the selective visualization of the CP-modified DNA, thereby indicating whether the incorporation of  $\text{dA}^{\text{CP}}\text{TP}$  by a polymerase and the downstream functionalization of the CP moiety function properly. The experimental setup for those preliminary experiments is presented in Figure 51.

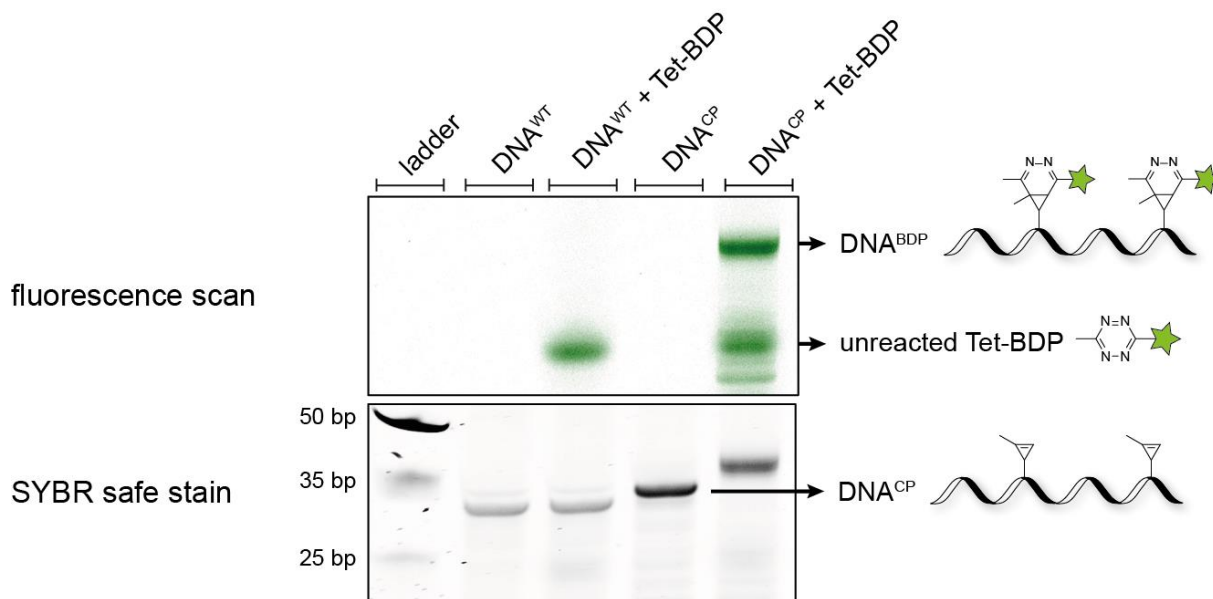


**Figure 51: Schematic representation of preliminary experiments to test the polymerase processing of  $\text{dA}^{\text{CP}}\text{TP}$  and its downstream functionalization via iEDDA.** Primer extension assays were performed with quantitative substitution of  $\text{dATP}$  with  $\text{dA}^{\text{CP}}\text{TP}$  using Taq polymerase. After  $\lambda$  exonuclease digestion of the 5'-monophosphate-modified template strand,  $\text{DNA}^{\text{CP}}$  was modified with 6-methyl-tetrazine-BODIPY via iEDDA.



Primer extension assays with Taq polymerase were performed using a 40 nt long, 5'-phosphate-modified DNA template (DNA\_pTemp\_mid2ATGC; for the exact sequence information, please refer to Chapter 7.5.2) and a 20 nt long DNA primer. The canonical dATP was fully substituted by dA<sup>CP</sup>TP to achieve quantitative incorporation of the CP-modified nucleotide opposite each dT. The 20 nt long templating region bears five randomly distributed, non-consecutive dTs. This would result in five dA<sup>CP</sup> modifications in the 40 nt full-length product strand. A control experiment was also conducted using only canonical, unmodified dATPs to prepare unmodified DNA (DNA<sup>WT</sup>) as a size marker. Following completion of the primer extension reactions, the 5'-phosphate-modified template was subjected to digestion with  $\lambda$ -exonuclease and purification via centrifugal filters. This approach ensures that the template does not distort the visualization of the unmodified control DNA during SYBR safe post-staining. Thereafter, the purified products were subjected to fluorescence click labeling in order to evaluate the successful incorporation of dA<sup>CP</sup>TP. Both DNA<sup>CP</sup> and DNA<sup>WT</sup> were incubated with a 50-fold excess of 6-methyl-tetrazine-BODIPY (Tet-BDP) and were analyzed by 20 % dPAGE and fluorescence scanning. Following visualization of the fluorescent bands, SYBR safe post-staining was performed to visualize unmodified DNA.

The DNA sequences prepared with dA<sup>CP</sup>TP showed a fluorescent signal, indicating successful processing of the modified nucleotide by Taq polymerase and effective downstream functionalization via iEDDA reaction (Figure 52). As expected, the unmodified DNA<sup>WT</sup> control did not show a fluorescent signal and was only visible after SYBR post-staining. Notably, in addition to the bands corresponding to the Tet-BDP click-labeled DNA<sup>CP</sup> product (DNA<sup>BDP</sup>), faster migrating and strongly fluorescent bands were visible (Figure S29). A click labeling control experiment performed with dA<sup>CP</sup>TP and Tet-BDP revealed that the faster migrating bands correspond to the click-labeled nucleotide, suggesting that dA<sup>CP</sup>TP is not sufficiently removed by purification with centrifugal filters after primer extensions. Moreover, DNA<sup>CP</sup>, and even more so, DNA<sup>BDP</sup>, migrate slightly slower than unmodified DNA<sup>WT</sup>. This is in accordance with expectations, as the size and quantity of modifications can have a significant effect on the migration behavior of such a short DNA sequence during dPAGE.



**Figure 52: 20 % dPAGE analysis of primer extension assays with dA<sup>CP</sup>TP using Taq polymerase and further downstream functionalization of DNA<sup>CP</sup> with Tet-BDP via iEDDA.** The upper image depicts the fluorescence scan ( $\lambda_{\text{excitation}} = 477 \text{ nm}$ ,  $\lambda_{\text{emission}} = 535 \text{ nm}$ ) and the lower image displays the UV visualization after SYBR safe post-staining. Ladder: ultra low range (ULR) ladder. The complete gel, including a Tet-BDP only control, is shown in Figure S29.

It is notable, that the length of the prepared DNA<sup>WT</sup> and DNA<sup>CP</sup>, as determined with the utilized dsDNA ladder, is shorter than the expected size of 40 nt. However, LC-MS analysis of the DNA<sup>WT</sup> and DNA<sup>CP</sup> samples yielded molecular weights corresponding to the respective full-length products (Figure S30). The differing migration behavior of the sample in comparison to the ladder may be attributed to the utilization of disparate buffers for sample and ladder preparation. The samples were prepared using a formamide/urea-based colorless PAGE loading buffer, while the ready-to-use ladder is supplied in a Tris-HCl-based buffer. The shielding effects of different ions and ion concentrations in the buffers can significantly influence the migration of DNA in the electric field.<sup>401</sup>

Interestingly, when OneTaq polymerase was employed in primer extension assays under identical reaction conditions, the outcomes were less reliable. In both experiments with and without dA<sup>CP</sup>TP, dPAGE analysis revealed the formation of numerous shorter truncation fragments (Figure S31). This led to the assumption that OneTaq polymerase appears to be less suitable for primer extension assays compared to Taq polymerase and further optimization may be necessary to achieve comparable results.

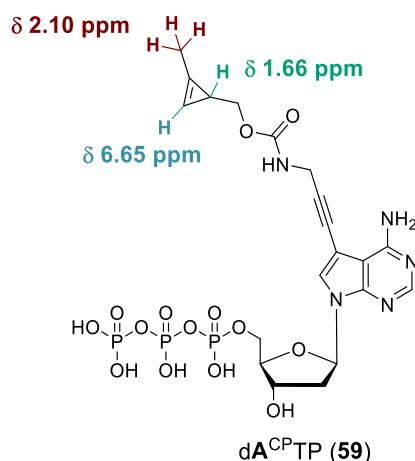
In conclusion, dA<sup>CP</sup>TP was accepted by Taq polymerase as a substrate and was successfully incorporated into DNA. The downstream functionalization of DNA<sup>CP</sup> was also successful, as indicated by a strong fluorescent signal after iEDDA click reaction with Tet-BDP. Therefore, the presented experiments represent a preliminary investigation into the potential of an iEDDA-based click-SELEX approach with dA<sup>CP</sup>TP. Future work should include further experiments to assess the suitability of dA<sup>CP</sup>TP for the use in PCR amplifications. Additionally,

it is essential to determine whether DNA that has been further functionalized with distinct chemical entities via the CP moiety can serve as a template for PCR amplification. Both aspects are crucial for the PCR amplification of modified and enriched oligonucleotides during click-SELEX. However, it is first necessary to investigate the stability of the CP-linker moiety to ensure that the modification can withstand a wide range of temperatures. This is important because the modification will be exposed to elevated and changing temperatures during a SELEX cycle, such as during PCR amplification and elution of the enriched oligonucleotides from the target.

### 3.3.3 Stability studies of the cyclopropene-carbamate modification

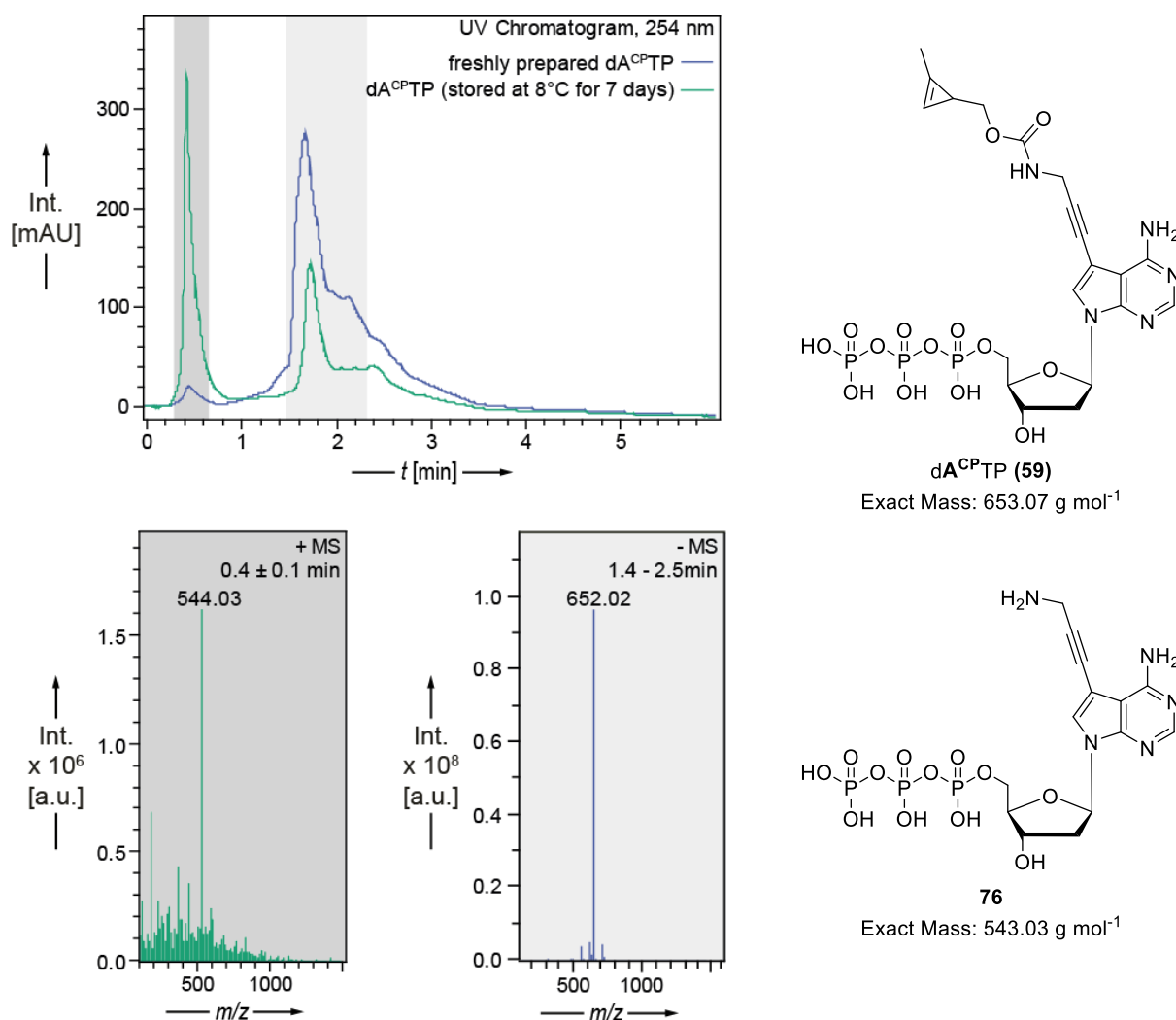
It is well known, that not only the reactivity but also the stability of the cyclopropene moiety is significantly influenced by its substituents.<sup>365,402</sup> Therefore, it is important to investigate the stability of the synthesized **dA<sup>CP</sup>TP (59)**. For reliable iEDDA-based click-SELEX, it is of utmost importance that the CP-linker system is stable under the conditions that are applied during the SELEX procedure. In particular, this includes the thermal stability of the CP-linker system, which will be exposed to elevated and changing temperatures during PCR amplification and during the elution of the bound sequences from the target. Additionally, the CP-linker system should be stable over multiple freeze-thaw cycles and at room temperature to allow convenient handling during the performance of consecutive rounds of SELEX. Only when the CP-linker system is stable under these conditions, it can be guaranteed that the oligonucleotides will be reliably and quantitatively functionalized in each selection cycle. Any changes in the modification pattern can affect the folding and function of the aptamer sequences, which may result in the loss of functional sequences or the enrichment of false positive sequences. Therefore, quantitative functionalization is important to achieve a reproducible functionalization pattern in each SELEX cycle and to avoid ineffective aptamer selection.

In order to investigate the stability of the CP-linker, <sup>1</sup>H NMR and LC-MS analysis of **dA<sup>CP</sup>TP (59)** was performed at different time points. Therefore, <sup>1</sup>H NMR was initially performed with freshly prepared **dA<sup>CP</sup>TP (59)**.



**Figure 53: Structure of dA<sup>CP</sup>TP (59) with the assigned <sup>1</sup>H NMR chemical shifts of the protons from the cyclopropene moiety.** The corresponding <sup>1</sup>H NMR spectrum is shown in Figure S32.

The <sup>1</sup>H NMR spectrum of freshly prepared dA<sup>CP</sup>TP (59) corresponded to the expected chromatogram, showing distinct signals for the characteristic protons of the cyclopropene moiety. The protons of the methyl substituent at the cyclopropene ring were assigned to a singlet at δ 2.10 ppm (red), the proton attached to the cyclopropene double bond was assigned to a signal at δ 6.65 ppm (green), and the proton at the carbon that carries the linker was assigned to a signal at δ 1.66 ppm (blue; Figures 53 and S32). It was not possible to reliably assign the signals corresponding to the two methylene groups of the linker moiety due to overlap with the signals of the sugar. After recording the spectrum of the freshly prepared dA<sup>CP</sup>TP (59), the NMR sample was stored at -20 °C for 24 hours, then thawed and remeasured. The newly recorded spectrum indicated that the signals previously assigned to the characteristic protons of the cyclopropene moiety vanished and several novel, not assignable peaks occurred (Figure S32). This suggests instability and decomposition of the CP-modification. In order to further elucidate the decomposition products, LC-MS analysis was performed (Figure 54). Initially, LC-MS analysis was conducted with the freshly prepared dA<sup>CP</sup>TP (59) compound. The UV trace at 254 nm of the LC chromatogram of freshly prepared dA<sup>CP</sup>TP (59) showed a single broad main peak with a maximum intensity of 280 mAU and a retention time between 1.5 and 2.5 minutes. The mass spectrum in the negative mode from this time frame indicated a molecular weight of 652.02 g mol<sup>-1</sup>, corresponding to the target compound 59.

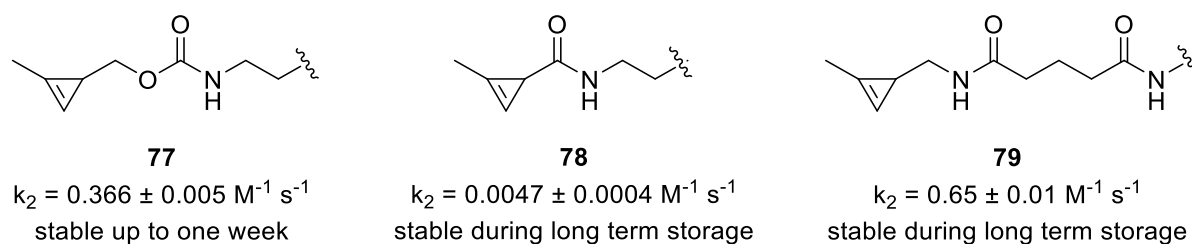


**Figure 54: Stability evaluation of dA<sup>CPTP</sup> (59) by LC-MS analysis.** Top left: UV chromatograms ( $\lambda = 254$  nm) of the freshly prepared dA<sup>CPTP</sup> (59, blue) and the same dA<sup>CPTP</sup> (59) sample (green) after storage at 8 °C for 7 days. Bottom left: ESI mass spectra of dA<sup>CPTP</sup> stored at 8 °C for 7 days (green, positive mode) and freshly prepared dA<sup>CPTP</sup> (blue, negative mode). On the right: Target compound dA<sup>CPTP</sup> (59) and its corresponding molecular weight, as well as the proposed decomposition product 76 and its corresponding molecular weight.

Secondly, LC-MS measurements were conducted with the same dA<sup>CPTP</sup> (59) sample after one week storage at 8 °C (Figure 54). The obtained spectrum showed a significant alteration in the UV trace compared to the previously recorded spectrum of freshly prepared dA<sup>CPTP</sup>(59). The previous dominant peak of the UV trace diminished significantly, reaching an intensity of 140 mAU and a new, distinct peak emerged with a retention time spanning between 0.3 min and 0.7 min. The corresponding mass spectrum of this time frame revealed a molecular weight of 544.01 g mol<sup>-1</sup> in the positive mode. This molecular weight could be assigned to nucleotide 76 with a propargylamine moiety attached to the nucleobase, indicating instability and decomposition of the CP-carbamate moiety. Proposedly, decomposition of the CP and cleavage of the carbamate moiety followed by subsequent decarboxylation of the carbamic acid intermediate would result in the formation of this propargylamine functionality.

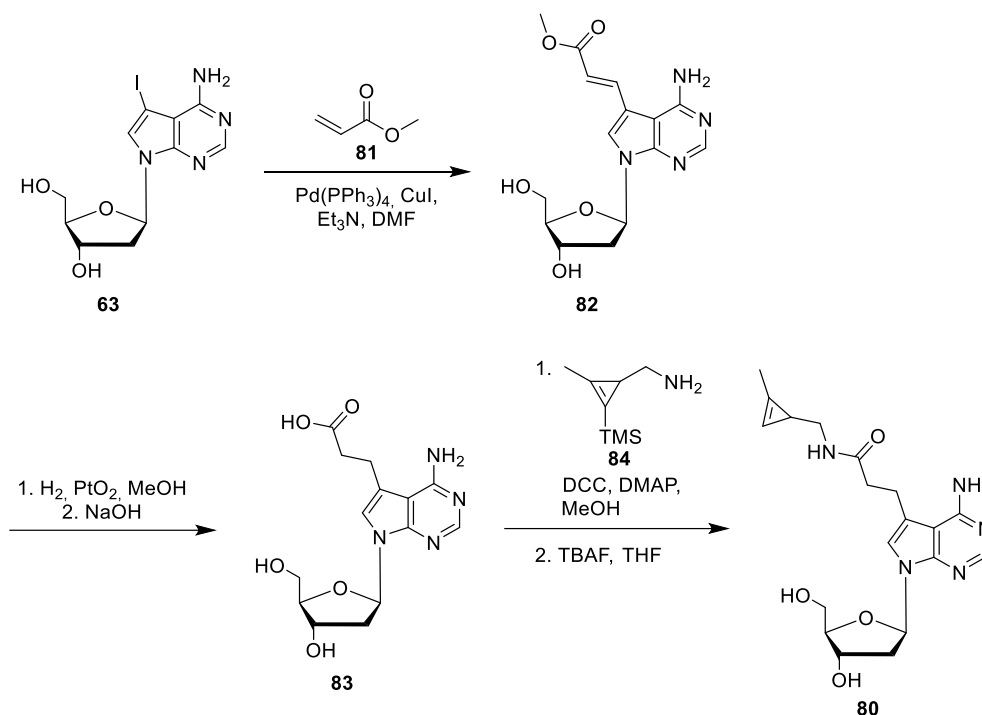
Nevertheless, in order to be able to make detailed and certain statements about the cleavage mechanism, it would be necessary to perform in-depth NMR studies.

The decomposition of a similar methylcyclopropene carbamate linker **77** has also been reported in the literature by Devaraj and colleagues (Figure 55).<sup>364,403</sup> Their results indicated that the carbamate moiety is stable for up to one week, but decomposes after multiple freeze-thaw cycles and longer storage times. Due to its ease of preparation, high reactivity, and decent stability, the carbamate-based linker system is indeed very interesting and suitable for many applications. However, in experiments that require quantitative functionalization of the oligonucleotide despite long incubation times at high temperatures, such as in an iEDDA-based click-SELEX approach, the carbamate linker is not a suitable option. Therefore, it is necessary to consider alternative linker systems. In this case, it is essential to find a balance between reactivity and stability, as these two factors do not always go hand in hand. A number of cyclopropene substituents have been investigated in terms of stability and reactivity by now.<sup>365,402</sup> Based on the available literature, an amide-based linker system appears to be a very promising option (Figure 55). The amide-based linkers **78** and **79** have been shown to exhibit enhanced stability compared to the carbamate version **77**.<sup>364,403</sup>



**Figure 55: Comparison of carbamate- and amide-substituted cyclopropenes: kinetics ( $k_2$ ) of the iEDDA reaction with differently substituted CPs and their stability.**<sup>364,403</sup>

However, the design of the amide linker must be done carefully. Electron-withdrawing groups, such as the carbonyl function of the amide, significantly reduce the reactivity of the cyclopropene moiety in the iEDDA reaction (Figure 55).<sup>364</sup> While the CP-amide linker **78** is more stable than the CP-carbamate linker **79**, it possesses a significantly reduced reactivity. However, when the amide functionality was positioned vice versa, as in the case of the CP-amide linker **79**, an improved stability and even a two-fold higher reaction rate was demonstrated compared to the carbamate linker **77** (Figure 55).<sup>403</sup> Consequently, future work should focus on the synthesis of a nucleotide bearing such an amide-linked CP moiety, with the aim of generating a suitable modification for the planned iEDDA click-SELEX approach. Scheme 26 illustrates a novel CP-amide-modified deazaA nucleoside **80** and its proposed synthesis.



**Scheme 26: Proposed synthetic route towards the cyclopropene-modified deazaA nucleoside 80 containing an amide linker.** The CP-amide-modified nucleoside **80** is expected to provide increased stability and higher reactivity compared to dA<sup>CP</sup>TP (**59**), which contains a carbamate linker.

Briefly, according to the published protocols by Kajiyama *et al.*, methyl acrylate **81** could be coupled to 7-deaza-7-iodo-2'-deoxyadenosine (**63**) under Sonogashira conditions to prepare nucleoside **82**. Subsequent hydrogenation over PtO<sub>2</sub> and deprotection with NaOH would give carboxy-modified nucleoside **83**.<sup>404</sup> Next, a similar approach to that described by Ganz *et al.*<sup>405</sup> could be employed to couple the carboxy group of **83** to cyclopropene amine **84** using *N,N'*-dicyclohexylcarbodiimide (DCC) as an activating agent. Finally, the TMS deprotection with TBAF would ultimately yield CP-amide-modified deazaadenosine **80** for subsequent triphosphate synthesis. The synthesis of cyclopropene amine **84**, which is utilized in this synthetic route, is literature known and could be achieved in four consecutive steps starting from trimethylsilylpropyne.<sup>214,364</sup>

The synthetic procedure towards CP-amide modified compound **80** is more laborious and less straightforward than that employed for the preparation of dA<sup>CP</sup>TP (**59**). The carbamate alkyne linker **64** used in this study is more universal and can be readily attached to any iodinated nucleoside, whereas the CP-amide modification is constructed stepwise on the nucleoside and not prior to attachment. Nevertheless, the stability, which is most likely provided by the novel linker, makes this the preferred route to eventually enable iEDDA-based click-SELEX in the future.

In summary, NMR and LC-MS studies revealed that the CP-carbamate moiety of dA<sup>CP</sup>TP (**59**) is susceptible to decomposition upon freeze-thaw cycles and prolonged storage. These findings indicate that the CP-carbamate linker is not suitable for experiments involving

prolonged incubation at elevated temperatures. However, the CP-carbamate-modified nucleotide  $dA^{CPTP}$  (**59**) would be exposed to harsh reaction conditions, including elevated temperatures, during PCR amplification and elution of the oligonucleotides from the target during each click-SELEX cycle. Therefore, the stability of the linker system is crucial for reproducible iEDDA click-SELEX experiments and successful clickmer selection. Consequently, a novel CP-amide-modified nucleotide (**80**) was proposed as a potential solution to the aforementioned issues, most likely offering improved stability and reactivity. This nucleotide is expected to be a valuable addition to future work aimed at developing an iEDDA-based click-SELEX approach.



### 3.4 Investigation of the interaction between mZBP1 and Z-RNA

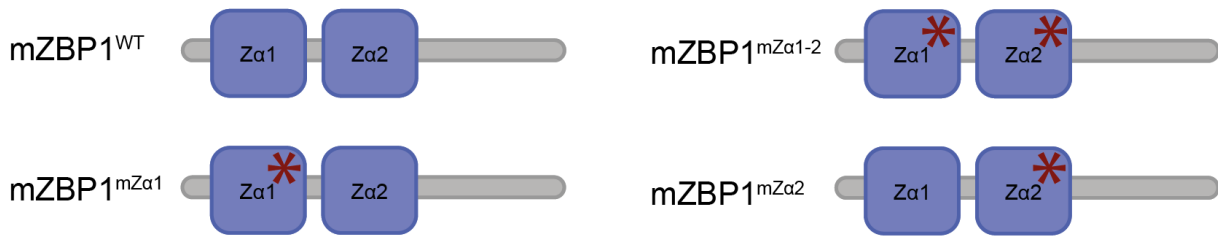
This project is a collaboration with the Pasparakis research group (CECAD Köln), which is engaged in research on Z-nucleic acids and Z-nucleic acid binding proteins *in vivo*.

ZBP1-mediated signaling has been demonstrated to play a pivotal role in the immune response of a host upon the sensing of viral Z-RNA.<sup>351,406,407</sup> Various research groups have investigated the ZBP1-triggered necroptosis pathway in the context of viral infections using *in vivo* knock-out studies.<sup>350,353–355,357</sup> The published results demonstrate that the Z $\alpha$ 2 domain of ZBP1 is essential for virus-induced cell death, while the Z $\alpha$ 1 domain alone is insufficient to trigger cell death.<sup>350,353–355,357</sup> Based on these findings, the function of the Z $\alpha$ 1 domain appears to be redundant and the functional significance of the two Z $\alpha$  domains in tandem orientation remains elusive. Therefore, the functional significance of the two tandem Z $\alpha$  domains for ligand binding was investigated in a simplified *in vitro* system. In this context, it was planned to investigate whether ZBP1 can also bind A-RNA and induce the A to Z transition. In addition, the binding of the two Z $\alpha$  domains was planned to be studied in an isolated manner to investigate their interdependencies and contribution to ligand binding.

This chapter presents the first steps towards comprehensive molecular binding studies of mZBP1 and its RNA ligand *in vitro*. Therefore, the expression and purification of wild-type mZBP1 and three mZBP1 mutants with loss-of-function mutations in either the Z $\alpha$ 1, Z $\alpha$ 2, or both domains were performed (Chapter 3.4.1). Moreover, preliminary results from an electrophoretic mobility shift assay to determine binding of mZBP1<sup>WT</sup> to m<sup>8</sup>G-modified Z-RNA and A-RNA are presented (Chapter 3.4.2). In addition, microscale thermophoresis (MST) was tested as an alternative method to EMSA and employed for preliminary molecular interaction studies between mZBP1<sup>WT</sup> and Z-RNA.

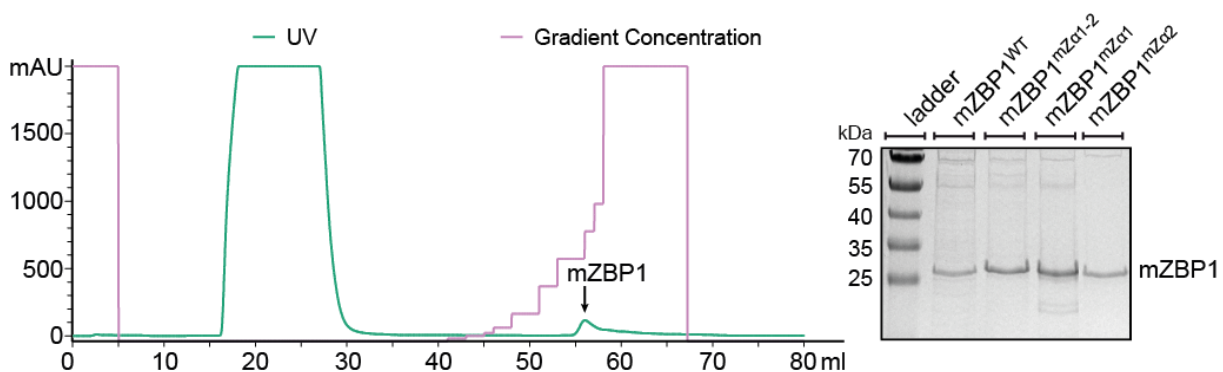
#### 3.4.1 Expression and purification of mZBP1 variants

The four desired ZBP1 variants include the murine wild-type ZBP1 (mZBP1<sup>WT</sup>) and three different loss-of-function mutants, as illustrated in Figure 56. The initial design of the three mZBP1 mutants included the introduction of either a loss-of-function mutation in the Z $\alpha$ 1 domain (mZBP1<sup>mZ $\alpha$ 1</sup>), the Z $\alpha$ 2 domain (mZBP1<sup>mZ $\alpha$ 2</sup>), or both Z $\alpha$  domains (mZBP1<sup>mZ $\alpha$ 1-2</sup>; Figure 56). The mutations N45D and Y49A result in a non-functional Z $\alpha$ 1 domain, while the mutations N121D and Y125A result in a non-functional Z $\alpha$ 2 domain. These mutations substitute key amino acids that are essential for binding Z-nucleic acids.<sup>408</sup> Notably, the designed mZBP1 mutants are truncated versions of the actual proteins and lack the RHIM domains.



**Figure 56: mZBP1<sup>WT</sup> and the three mutants mZBP1<sup>mZα1</sup>, mZBP1<sup>mZα2</sup>, and mZBP1<sup>mZα1-2</sup>.** The red asterisks indicate a loss-of-function mutation in the respective Zα domain.

The pET28a(+) expression plasmids, which contain the desired coding sequences of mZBP1<sup>WT</sup> and the three distinct mZBP1 mutants with N- and C-terminal His-tags, were procured from *Twist Bioscience*. The expression of these constructs and the purification of the proteins were conducted using an adapted protocol based on the published procedures from Nichols *et al.* and Placido *et al.*<sup>409,410</sup> Initially, the four plasmids were transformed into NEB T7 Express *lysY/lq* competent *E.coli* cells, which are suitable for protein overexpression. Then, starter cultures for each protein variant were prepared using glycerol stocks of NEB T7 Express *lysY/lq* *E.coli* harboring the plasmid with the respective mZBP1 variant. The starter cultures were cultivated aerobically overnight at a temperature of 37 °C. Subsequently, the starter cultures were used to inoculate the expression cultures. Protein expression was induced by the addition of IPTG at a point of time when the OD<sub>600</sub> of the expression culture reached 0.6. This was followed by incubation of the induced expression cultures at 15 °C overnight. Subsequently, the cells were harvested by centrifugation and lysed using sonication. Following this, the cell debris was removed by centrifugation. The His-tagged mZBP1 variants were isolated by IMAC purification, which involved the use of a cobalt-based resin. Following purification, the protein-containing fractions were identified by SDS-PAGE and pooled (Figures S33, S34, S35 and S36). Figure 57 shows one exemplary chromatogram from IMAC purification of mZBP1<sup>mZα1</sup> and an SDS-PAGE of the pooled fractions of all expressed and purified mZBP1 variants.

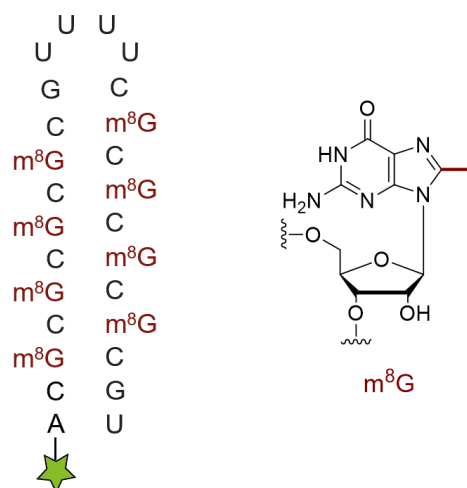


**Figure 57: Cobalt-based IMAC purification of mZBP1 variants.** Left: UV chromatogram of the affinity purification of mZBP1<sup>mZα1</sup> as a representative example for the purification of all mZBP1 variants. Right: 15 % SDS-PAGE of the pooled fractions of the expressed and purified mZBP1 variants. Ladder: PageRuler Prestained Protein Ladder, 10-180 kDa. Molecular weight of target proteins: 24 kDa. The complete SDS-PAGE is shown in Figure S37.

Subsequently, the pooled protein fractions were concentrated and transferred to a storage buffer containing 50 % glycerol and DTT using centrifugal filters. However, during the protein concentration and buffer exchange steps, the proteins aggregated and precipitated. This phenomenon may have been caused by the formation of a concentration gradient in the vicinity of the membrane, which resulted in a localized extremely high concentration of the protein at the filter membranes. Consequently, a different buffer exchange method was selected, this time employing dialysis. During dialysis, the concentration and buffer exchange proceed at a slow rate over time. Dialysis was therefore deemed an appropriate method for buffer exchange, with the proteins successfully transferred into a buffer suitable for protein storage at -20 °C. Subsequently, the protein concentration of the distinct samples was quantified. For subsequent quantitative EMSA evaluation, precise determination of the protein concentrations is of major importance. Consequently, the Pierce 660 nm protein quantification kit was employed as it is compatible with high glycerol content and DTT. A BSA standard curve was prepared (Figure S11) and the protein concentration of the samples containing the different mZBP1 variants was determined by interpolation (Table S5). Once the mZBP1 protein samples had been obtained, interaction studies with Z-RNA and A-RNA were conducted.

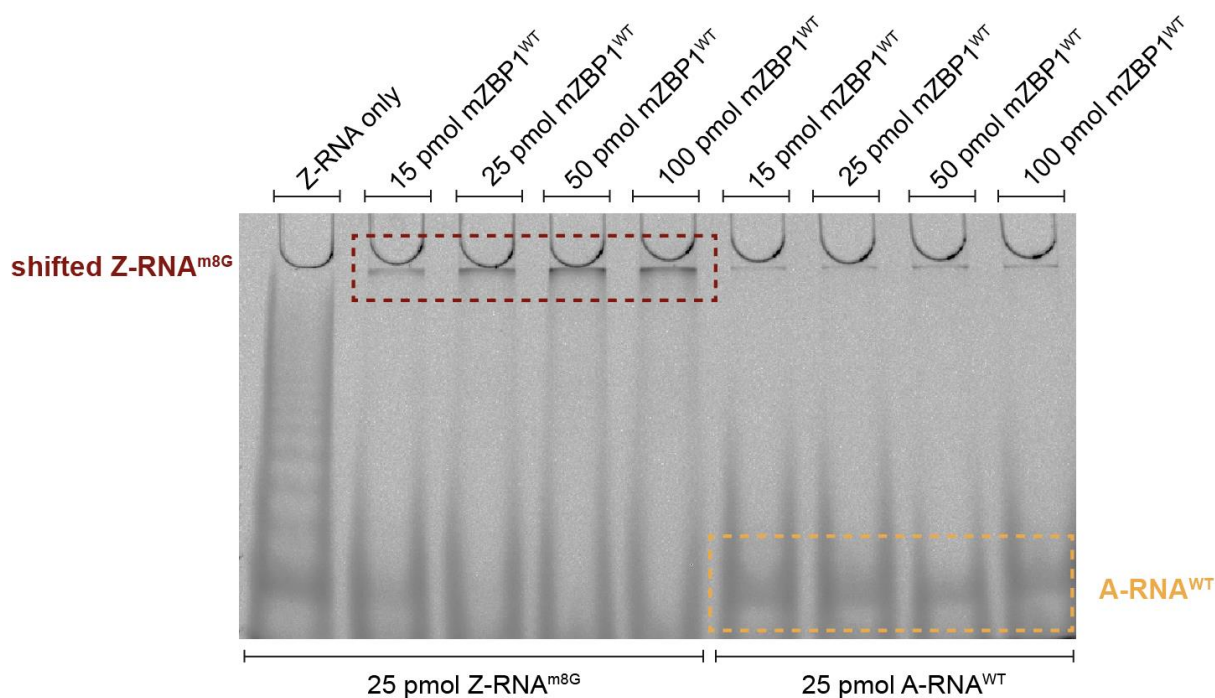
### 3.4.2 *In vitro* interaction studies of Z-RNA and mZBP1

Following the successful expression and purification of the four distinct mZBP1 variants, interaction studies were aspired using an electrophoretic mobility shift assay. EMSA is a technique that enables the detection of protein-nucleic acid interactions under close-to-native conditions.<sup>411</sup> The method is based on the principle that the electrophoretic mobility of a protein-nucleic acid complex is generally lower than that of free nucleic acids. Consequently, the binding of the protein to the nucleic acid results in a slower migration of the protein-nucleic acid complex on nPAGE, causing an upward shift of the oligonucleotide on the gel.<sup>411</sup> To facilitate the straightforward detection of this shift, the oligonucleotide ligand must be fluorescently labeled. Consequently, a fluorescently labeled Z-RNA was prepared. Therefore, a synthetically prepared 5'-amino-modified 26 nt long **GC**-rich RNA hairpin containing m<sup>8</sup>**G** modifications was labeled with AlexaFluor 488 (AF488) NHS-ester (Figure 58; please also refer to Chapter 7.5.2 for the exact sequence). The m<sup>8</sup>**G**-residues serve to stabilize the Z-conformation of RNA at low salt concentrations, thus enabling interaction studies under close-to-native conditions *in vitro*. The identical fluorescently-labeled RNA sequence lacking m<sup>8</sup>**G** modifications served as the A-RNA<sup>WT</sup> control. This setup allows for a comparative analysis of the interactions between the mZBP1 variants and Z-RNA<sup>m<sup>8</sup>G</sup> versus A-RNA<sup>WT</sup>.



**Figure 58: Fluorescently-labeled GC-rich RNA hairpin containing  $\text{m}^8\text{G}$  modifications to stabilize the Z-conformation.** The  $\text{m}^8\text{G}$  modification is depicted on the right.

Initially, the minimum detectable amount of AF488-labeled Z-RNA $^{\text{m}^8\text{G}}$  on nPAGE was determined through titration experiments (Figure S38). A quantity of 25 pmol of AF488-labeled Z-RNA $^{\text{m}^8\text{G}}$  was identified as an appropriate amount for further investigation by EMSA. In the initial EMSA experiment, the interaction of fully functional mZBP1 $^{\text{WT}}$  with AF488-labeled Z-RNA $^{\text{m}^8\text{G}}$  and A-RNA $^{\text{WT}}$  was investigated to elucidate whether mZBP1 $^{\text{WT}}$  can bind to A-RNA and induce a change from the A- to the Z-conformation. Therefore, fixed quantities of AF488-labeled Z-RNA $^{\text{m}^8\text{G}}$  and A-RNA $^{\text{WT}}$  were incubated with titrating amounts of mZBP1 $^{\text{WT}}$ , followed by nPAGE analysis (Figure 59).



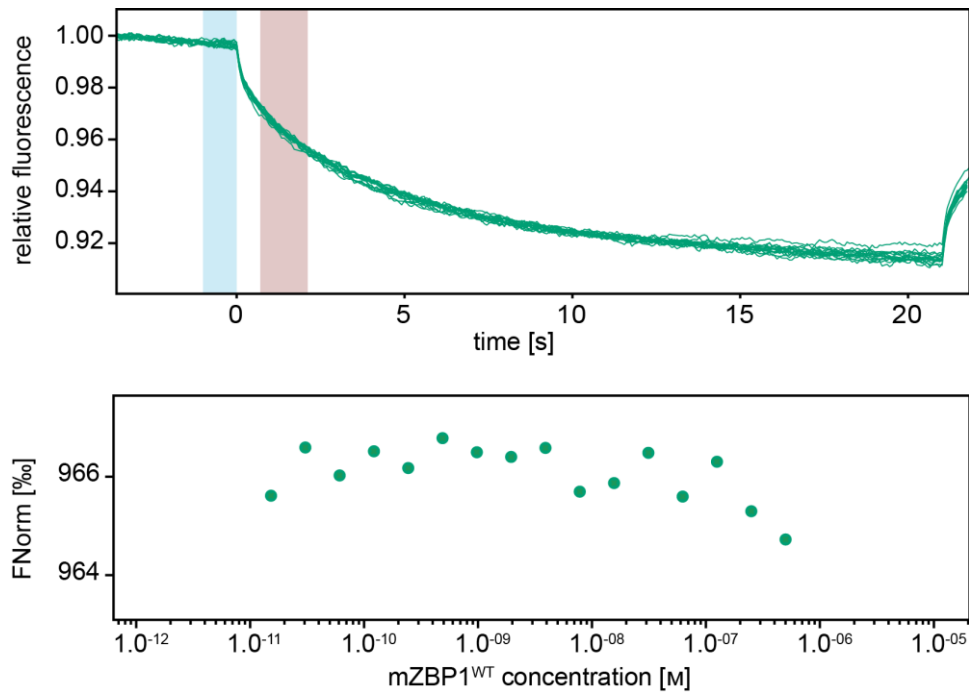
**Figure 59: 8 % nPAGE of comparative interaction studies of mZBP1 $^{\text{WT}}$  with Z-RNA $^{\text{m}^8\text{G}}$  and A-RNA $^{\text{WT}}$ .** A fixed amount of 25 pmol of fluorescently-labeled Z-RNA $^{\text{m}^8\text{G}}$  or A-RNA $^{\text{WT}}$  was incubated with titrating amounts of mZBP1 $^{\text{WT}}$  (*i.e.* 15 pmol, 25 pmol, 50 pmol and 100 pmol). Visualization was performed by fluorescence scanning ( $\lambda_{\text{excitation}} = 477 \text{ nm}$ ,  $\lambda_{\text{emission}} = 535 \text{ nm}$ ). The complete gel is shown in Figure S39.

The results obtained by nPAGE demonstrate that mZBP1 binds to Z-RNA<sup>m8G</sup>, but not A-RNA<sup>WT</sup>. This is evidenced by the observation of a distinct fluorescent band in the wells containing samples with Z-RNA<sup>m8G</sup> and mZBP1<sup>WT</sup>. Moreover, the signal of free Z-RNA<sup>m8G</sup> decreased concomitantly, particularly with increasing amounts of mZBP1<sup>WT</sup>. A comparable upward shift of the fluorescent signal was not observed for samples containing A-RNA<sup>WT</sup> without m<sup>8</sup>G-modifications. Consequently, it can be concluded that two tandem Z $\alpha$  domains are not able to bind to A-RNA<sup>WT</sup> with the same affinity as Z-RNA<sup>m8G</sup> to convert A-RNA into Z-RNA.

It is important to note, however, that the mZBP1<sup>WT</sup> exhibited only minimal migration into the nPAGE. To further elucidate the reason for this observation, the isoelectric point (pI) of the protein was considered. A range of theoretical pIs was calculated using different computational tools available online (Chapter 5.1), with values between 7.04 and 7.83 being obtained. This indicates that the protein is likely to have a net charge of zero under the utilized nPAGE conditions at a pH of 8, which precludes its migration in the electric field. Notably, increasing the pH used for nPAGE to 8.9 did also not result in the migration of the protein into the gel.

To overcome the challenge of the protein not migrating into the native PAGE, microscale thermophoresis was chosen as an alternative method to determine the molecular interaction between mZBP1 and Z-RNA<sup>m8G</sup> under close-to-native conditions. MST is an immobilization-free technique used to detect molecular interactions by quantifying the directed thermophoretic movement of fluorescently labeled molecules in response to a temperature gradient.<sup>412</sup> In this method, one biomolecule is fluorescently labeled and incubated with varying concentrations of an unlabeled second biomolecule. Subsequently, the samples are loaded into capillaries, and a temperature gradient is applied. Then, the temperature-related change in relative fluorescence is measured to determine the interaction. Upon interaction between the two molecules, properties such as charge, size, and hydration shell of the molecules change affecting the mobility of the molecules and, in turn, the temperature-dependent fluorescence intensity of the labeled probe. This temperature-related intensity change is employed to quantify the interaction between the two molecules.<sup>412</sup>

In order to conduct MST measurements, a fixed concentration of 20 nM fluorescently-labeled Z-RNA<sup>m8G</sup> was initially incubated with varying concentrations of mZBP1<sup>WT</sup> (*i.e.* ~0.15 nM to ~500 nM). The concentration range was selected based on a previously published apparent dissociation constant ( $K_d$ ) of ZBP1 for Z-DNA, which was calculated to be  $35 \pm 9$  nM.<sup>340</sup> This value was considered as an initial reference point for the herein performed MST experiments. Subsequently, the temperature-related intensity change of the fluorescent oligonucleotide was measured. Contrary to expectations, the results obtained from MST measurements were inconclusive and no thermophoresis signal could be detected (Figure 60), at least with the applied experimental settings.



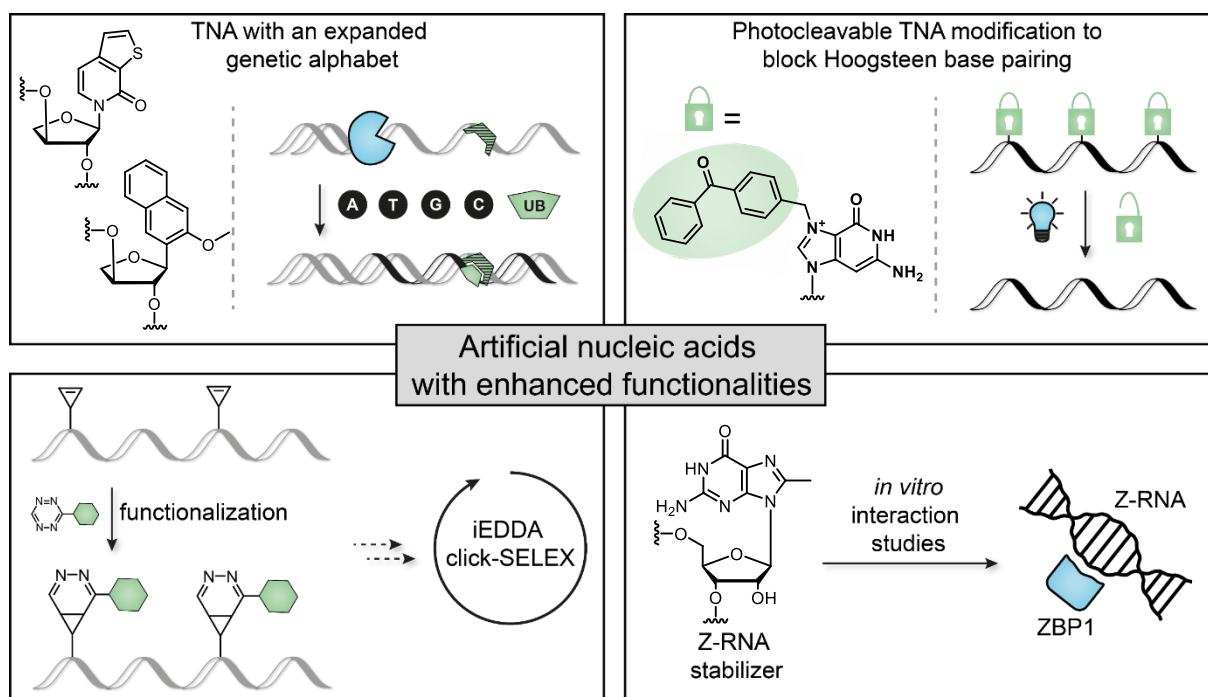
**Figure 60: Microscale thermophoresis binding curves for AF-488 labeled Z-RNA<sup>m8G</sup> with mZBP1<sup>WT</sup>.** Top: Raw fluorescent time traces. Bottom: The change in thermophoresis is expressed as the change in the normalized fluorescence (FNorm), which is defined as  $F_{hot}/F_{cold}$ . The blue (cold) and red (hot) shaded areas in the raw fluorescent data indicate the regions used to average the F-values to calculate the relative fluorescence intensities plotted in the lower graph. The pretest data is shown in Figure S40.

The inconclusive results may be attributed to the strong fluorescent background of the Z-RNA<sup>m8G</sup> sample, which is likely due to the presence of residual free AF488 NHS-ester that was not completely removed during purification following the click reaction. This excess of free fluorophore was also observed on nPAGE during EMSA (Figure S39) and may disturb MST measurements.

In summary, EMSA demonstrated that mZBP1<sup>WT</sup> specifically binds to m<sup>8</sup>G-modified Z-RNA, but not to A-RNA<sup>WT</sup>. Thus, the function of the two tandem Z $\alpha$  domains is most likely not to bind double stranded A-RNA and convert it into Z-RNA. Therefore, upcoming research should focus on studying the interaction between Z-RNA and the mZBP1 variants that have mutations in their Z $\alpha$  domains. Molecular interaction studies with these mutants can provide valuable insights into the interdependencies of the Z $\alpha$  domains and their significance for ligand binding. The results presented in this chapter should be taken into account when planning further optimization of experiments.

## 4 Conclusion and outlook

Within the framework of this thesis a variety of chemical modifications were employed to alter the functionalities of nucleic acids. Four research projects were conducted, each focusing on a distinct aspect of the functionalization of nucleic acids (Figure 61). The objective of these projects was either the development of these functionalized nucleic acids for diverse applications or their utilization in advanced studies.



**Figure 61: Overview of the four research projects on artificial nucleic acids with enhanced functionalities that were conducted within the framework of this thesis.** Top left: Enzymatic preparation of TNA with an expanded genetic alphabet using the **TPT3:NaM** UBP. This project involved the synthesis of highly modified unnatural TNA nucleotides. Top right: A photocleavable and traceless approach to block dG/tG Hoogsteen base pairing during enzymatic TNA synthesis. For this purpose, a benzophenone-modified threoguanosine building block was synthesized. Bottom left: Polymerase-mediated synthesis of methylcyclopropene-modified DNA and downstream functionalization to increase the chemical diversity of DNA for an iEDDA-based click-SELEX approach. Therefore, a C<sup>7</sup>-methylcyclopropene-modified 7-deaza-2'-deoxyadenosine triphosphate was synthesized. Bottom right: *In vitro* interaction studies between Z-RNA and Z-DNA binding protein 1. Here, a C<sup>8</sup>-methylated guanine building block was used to stabilize Z-RNA under close-to-native conditions.

The main project of this thesis aimed at the development and enzymatic preparation of a novel artificial nucleic acid with a dual modification pattern (Figure 61, top left). The aspired combination of the nuclease-resistant threose as an alternative sugar moiety with the **TPT3:NaM** UBP as a third base pair to generate exTNA represents a significant extension of the scope of currently available nucleic acid analogs, with an unprecedented degree of artificial modifications. Such an exTNA has great potential for future *in vitro* selection of nuclease-resistant high-affinity aptamers for therapeutic and diagnostic applications.

For the first time, the enzymatic synthesis of a TNA sequence with an internal **tTPT3** modification was demonstrated using the Kod-RSGA TNA polymerase. This is the first example of a xeno nucleic acid with an expanded genetic alphabet, representing a pivotal advancement in the field of artificial nucleic acid research. Moreover, the incorporation of **tNaMTP** into TNA by the Kod-RSGA polymerase was demonstrated, although limited by polymerase stalling. In order to achieve these results, primer extension assays were performed and a series of reaction conditions were screened to identify those that facilitate enzymatic exTNA synthesis. Two additional polymerases, namely Kod-RI and Terminator polymerases, were also tested to assess their capability to process **tTPT3TP** and **tNaMTP**. However, in comparison to the Kod-RSGA polymerase, these polymerases demonstrated a relatively low efficiency in catalyzing the incorporation of **tTPT3TP** and **tNaMTP**, as well as the further elongation of the nascent TNA strand. Additional significant achievements within this project, which made enzymatic exTNA production possible in the first place, included the successful synthesis of the novel compounds **tNaMTP** and **dTPT3-CEP**. The **dTPT3** phosphoramidite was required for the solid-phase synthesis of the employed **TPT3**-modified DNA templates to investigate the **tNaMTP** incorporation by TNA polymerases. Moreover, the synthesis of TNA nucleotides bearing cytosine, adenine, and guanine was successfully accomplished.

Despite these advancements, certain limitations persist. The specificity of the Kod-RSGA polymerase, which has been identified as the most effective in incorporating the highly modified UB triphosphates, is partially limited. The Kod-RSGA polymerase exhibits a certain degree of ineffectiveness in discriminating between the correct incorporation of **tTPT3TP** or misincorporation of **tGTP** opposite **dNaM** in the templating strand. Moreover, the preparation of **tNaM**-modified exTNA remains challenging, with the currently tested polymerases exhibiting difficulty in elongating the nascent TNA strand post **tNaMTP** incorporation. *In silico* analyses of the closed ternary complex of Kod-RSGA polymerase revealed a less favorable conformation for bond formation with **tNaM** compared to **tTPT3**. This is due to an increased distance between the 2'-OH of the **tNaM** in the TNA strand and the 3'  $\alpha$ -phosphate of the incoming TNA triphosphate.

The introduction of a third unnatural, hydrophobic UBP into TNA enables the selective introduction of novel binding modalities beyond the regular hydrogen bonding of natural nucleobases. This broadens the addressable target spectrum of artificial aptamers in comparison to those solely composed of natural nucleobases. However, for the selection of exTNA aptamers, the experimental setup needs to be adapted by using TNA instead of DNA primers. This would result in exTNA aptamers that are entirely composed of a TNA backbone. Furthermore, polymerase evolution would be crucial to overcome current shortcomings of enzymatic exTNA synthesis. To this end, crystallization or NMR studies could provide further insight into the current challenges of exTNA synthesis and allow for the formulation of rational



conclusions for further polymerase engineering. Conversely, a synthetic approach could be employed to address the limitations of incorporating **tNaMTP**. This entails replacing the **NaM** nucleobase with the UB **PTMO**, an alternative counterpart to **TPT3**.<sup>212</sup> **PTMO** has a smaller ring size compared to **NaM**, and thus it would be beneficial to investigate whether its incorporation and the further elongation of the nascent TNA strand is more efficiently catalyzed by polymerases. A preliminary *in silico* analysis with the **PTMO** building block could provide valuable insights for subsequent experimental approaches. In the longer term, future research should also concentrate on identifying a reverse transcriptase that is capable of transcribing exTNA back into DNA, thus enabling exTNA aptamer selection. Potential candidates for testing include Bst DNA polymerase<sup>139,148</sup> and SuperScript II polymerase<sup>138</sup>, which have been reported as the most efficient TNA reverse transcriptases. An alternative approach would be the use of linked genotype-phenotype display methods for aptamer selection, as demonstrated by Szostak, Chaput, and colleagues.<sup>144,303,304</sup> This strategy may obviate the necessity for an exTNA reverse transcriptase.

In conclusion, the presented pioneering work on exTNA paves the way for the development of innovative exTNA-based therapeutics and diagnostic tools. In particular, the selection and evolution of nuclease-resistant exTNA aptamers with an enhanced chemical diversity has great potential due to a broader targeting range.

In the context of the second project, the objective was to attach a photocleavable group to the *N*<sup>7</sup> position of **tGTP** to obstruct Hoogsteen base pairing during enzymatic TNA synthesis. The formation of **dG/tG** Hoogsteen base pairs during polymerase-mediated synthesis of TNA currently represents the greatest source of error.<sup>138,147</sup> To enhance the fidelity of enzymatic TNA synthesis, the utilization of 7-deaza-7-modified **tGTP** substrates to block the Hoogsteen base pairing site of **G** has been reported.<sup>148</sup> However, these modifications yield modified TNA, which is not always desired. During this project, the concept of a cleavable modification was conceived for the same purpose. In this context, a strategy based on a photolabile modification to the *N*<sup>7</sup> position of the guanine nucleobase was elaborated (Figure 61, top right). A literature search identified benzophenone as a suitable photolabile modification for the *N*<sup>7</sup> position of guanine.<sup>381</sup> During the synthesis towards benzophenone-modified threoguanosine triphosphate, various synthetic challenges were encountered and circumvented. Finally, a benzophenone-modified threoguanosine building block was isolated, but the isolated amount was not sufficient for its further use. To address challenges of the current synthetic strategy, an alternative synthetic route was proposed, which should be considered during future work. Looking further ahead, upcoming work should also include investigations regarding the processing of **tG<sup>BPTP</sup>** by TNA polymerases and the photoinduced removal of the

benzophenone photocage. Subsequently, fidelity measurements as reported by Chaput and coworkers<sup>138,148,152</sup> would be necessary to evaluate the capability of the benzophenone group to suppress **G/G** Hoogsteen base pairing.

In addition, alternative photolabile groups may be considered in the context of this work.<sup>413</sup> The attachment of a different photolabile group may proceed more smoothly, resulting in higher yields and less byproduct formation. Additionally, less sterically demanding photolabile groups, such as smaller aryl ketones, may also be superior substrates for TNA polymerases.

Although this work is still in its early stages, it serves to highlight the pivotal aspects of the synthesis of a photocaged threoguanosine species. This photocleavable and traceless approach to obstruct Hoogsteen base pairing presents an innovative alternative to currently reported permanent modifications and warrants further exploration. The insights obtained during the course of this study will eventually guide further synthetic efforts towards a benzophenone-modified **tG<sup>BP</sup>TP** building block.

Despite the presented strategies to enhance the functionality of TNA, the third project aimed to augment the chemical diversity of DNA through the use of iEDDA click chemistry. It was aspired to develop an iEDDA-based click-SELEX method for the selection of clickmers. The iEDDA reaction is characterized by unmatched kinetics and excellent bioorthogonality and has been excessively employed for the attachment of reporter groups to study various biomolecules *in vitro* and *in vivo*.<sup>213,215,264,267</sup> However, its utilization for the selection of clickmers has not yet been considered. A copper-free functionalization method during click-SELEX is superior to the currently used CuAAC reaction and could overcome current shortcomings of clickmer selection. This includes folding of clickmers in dependence of copper ions during the *in vitro* selection process, resulting in non-native folding of the aptamer. In this project, the strategy for an iEDDA-based click-SELEX approach was elaborated and first steps towards its implementation were performed. (Figure 61, bottom left). Therefore, a cyclopropene-modified adenosine triphosphate was successfully synthesized. Furthermore, it was demonstrated that **dA<sup>CP</sup>TP** was accepted as a substrate by Taq DNA polymerase to prepare cyclopropene-modified DNA. The utilization of a tetrazine-fluorophore conjugate as a trackable test modification demonstrated the successful functionalization of cyclopropene-modified DNA post-enzymatic preparation.

Nevertheless, the current limitations to utilize the prepared **dA<sup>CP</sup>TP** in iEDDA-based click-SELEX include the low stability of the employed cyclopropene linker system. Decomposition of the cyclopropene-carbamate linker system was observed, rendering it unsuitable for click-SELEX. Consequently, the design of the reactive handle was reconsidered, culminating in a novel cyclopropene-amide linker system that promises an improved stability and, at the same time, an even faster reactivity of the attached cyclopropene. Future work

should focus on the synthesis of the proposed compound, followed by the investigation of its incorporation efficiency, functionalization, and stability.

The presented results establish a foundation for a copper-free click-SELEX approach and provide valuable insights for future strategies that will potentially enable iEDDA-based click-SELEX.

Lastly, the fourth project sought to elucidate the interaction between ZBP1 and RNA in its A- and Z-conformation *in vitro* (Figure 61, bottom right). Results of current *in vivo* knock-out studies suggest that the Z $\alpha$ 2 domain is the primary driver of the ZBP1-mediated necroptosis pathway *in vivo*,<sup>350,353–355,357</sup> while the Z $\alpha$ 1 domain appears to be redundant. Consequently, a simplified *in vitro* setting was established during the course of this work to investigate the functional significance of the two tandem Z $\alpha$  domains for ligand binding. Accordingly, wild-type mZBP1 and three mZBP1 mutants were successfully expressed and purified by IMAC. The mutants either contained a loss-of-function mutation in the Z $\alpha$ 1 domain, the Z $\alpha$ 2 domain, or both Z $\alpha$  domains. To facilitate the interaction studies of the mZBP1 mutants with the unstable structure of Z-RNA *in vitro*, a **GC**-rich hairpin RNA modified with m<sup>8</sup>**G** residues was employed. This modification markedly stabilizes Z-RNA, thus enabling the performance of interaction studies with Z-RNA under close-to-native conditions. For visualization purposes during the interaction studies, the m<sup>8</sup>**G**-modified Z-RNA was fluorescently labeled via NHS-ester chemistry.

Initially, an electrophoretic mobility shift assay was employed to investigate the potential conversion of A-RNA to Z-RNA upon binding of the wild-type mZBP1. The results demonstrated that the wild-type mZBP1 binds to Z-RNA, but not A-RNA, indicating that mZBP1 cannot induce a conformational change in RNA. Notably, the employed EMSA method proved to be rather unsuitable for the intended interaction studies, as the protein only insufficiently migrated into the nPAGE. Therefore, microscale thermophoresis was tested as an alternative method. However, the high background fluorescence of the Z-RNA sample caused by residual fluorophore from the click-labeling reaction interfered with this method, giving inconclusive results. Consequently, future work should focus on the fluorescent labeling of Z-RNA on the solid support during its chemical synthesis. This would ideally result in a pure Z-RNA ligand, enabling molecular interaction studies using MST. The forthcoming work should investigate the binding affinities of the mZBP1 mutants to Z-RNA to elucidate the interdependencies of the Z $\alpha$  domains and their contribution to ligand binding.

In summary, the preliminary results of the molecular interaction studies of wild-type mZBP1 with Z-RNA and A-RNA provide important insights for further optimization of the experimental setup and the employed methods to facilitate future efforts.

In conclusion, this work provides insight into a diverse range of potential applications of chemical modifications to nucleic acids, which can be used to enhance their functionalities (Figure 62). The development of a highly modified exTNA involved the use of artificial nucleic acid building blocks, which resulted in a nucleic acid analog with novel properties, including increased nuclease resistance and new binding modalities. Such an artificial nucleic acid has the potential to be used for therapeutic and diagnostic applications. Moreover, an effort was made to install a transient chemical modification on threoguanosine using a photolabile group in order to influence the fidelity of the enzymatic preparation of TNA. In addition, it was aspired to employ iEDDA click chemistry for the augmentation of the chemical diversity of DNA. The suitability of a cyclopropene-carbamate linker was investigated for its utilization in an iEDDA-based click-SELEX approach. This paves the way for the future possibility of DNA clickmer selection without the need for a copper-catalyzed functionalization step. Finally, it was demonstrated that chemical modifications can be employed to alter the conformation of nucleic acids and stabilize the relatively transient and rare Z-conformation, thus enabling *in vitro* binding studies with Z-NA binding proteins. In summary, the collective achievements of the various projects provide a solid foundation for future efforts that may enable the desired versatile applications of the functionalized artificial nucleic acids.

## 5 Materials and methods

Chapter 5 provides detailed descriptions of all necessary materials, equipment, and procedures used in this interdisciplinary study, including chemical and biochemical methods.

### 5.1 Software and data analysis

Software applications for data processing and evaluation as well as databases used in this study are depicted in Tables 5 and 6.

**Table 5: Software applications used in this study.**

<b>Software</b>	<b>Versions</b>	<b>Source</b>
ApE – A Plasmid Editor	3.0.6	<i>M. Wayne Davis</i>
ChemDoodle 2D	12.6.0	<i>iChemLabs, LLC.</i>
ChemDraw Prime	22.2	<i>PerkinElmer</i>
Compass DataAnalysis	4.2, 5.2	<i>Bruker</i>
Creative Suite	3-5	<i>Adobe</i>
Expasy Compute pI/MW tool	-	<a href="https://web.expasy.org/compute_pi/">https://web.expasy.org/compute_pi/</a>
Expasy ProtParam tool	-	<a href="https://web.expasy.org/protparam/">https://web.expasy.org/protparam/</a>
LaLign	-	<a href="https://www.ebi.ac.uk/jdispatcher/psa/align">https://www.ebi.ac.uk/jdispatcher/psa/align</a>
MestReNova	14.3.2, 6.0.2	<i>Mestrelab Research S.L.</i>
MO.Affinity Analysis	2.3	<i>Nano Temper Technologies</i>
MO.Control	1.6	<i>Nano Temper Technologies</i>
Mongo Oligo Mass Calculator	2.06	<a href="http://mass.rega.kuleuven.be/mongo.htm">http://mass.rega.kuleuven.be/mongo.htm</a>
Office	2013, 2019	<i>Microsoft</i>
OligoCalc: Oligonucleotide Properties Calculator	3.2.7	<a href="http://biotools.nubic.northwestern.edu/OligoCalc.html">http://biotools.nubic.northwestern.edu/OligoCalc.html</a>
Prot pi Protein Tool	-	<a href="https://www.protpi.ch/Calculator/ProteinTool">https://www.protpi.ch/Calculator/ProteinTool</a>
QIAGEN Ligation Calculator	-	<a href="https://www.qiagen.com/us/applications/enzymes/tools-and-calculators/ligation-calculator">https://www.qiagen.com/us/applications/enzymes/tools-and-calculators/ligation-calculator</a>
Snapgene	7.2.0	<i>GSL Biotech LLC</i>
SnapGene Viewer	5.3.2	<i>GSL Biotech LLC</i>
Unicorn	1.3	<i>GE Healthcare, Cytiva</i>

**Table 6: Databases used in this study.**

Database	Source
UniProt	<a href="https://www.uniprot.org/">https://www.uniprot.org/</a>
AlphaFold Protein Structure Database	<a href="https://alphafold.ebi.ac.uk/">https://alphafold.ebi.ac.uk/</a>
RCSB PDB	<a href="https://www.rcsb.org/">https://www.rcsb.org/</a>
CAS SciFinder <sup>n</sup>	<a href="https://scifinder-n.cas.org">https://scifinder-n.cas.org</a>

## 5.2 Chemical methods

### 5.2.1 General chemical methods and equipment

#### Chemicals and suppliers

Unless otherwise stated, all chemicals and solvents employed in synthetic procedures were used directly without further purification as purchased from established suppliers (*Acros Organics, Alfa Aesar, Fisher Scientific, Honeywell Fluka, VWR, Merck, Sigma-Aldrich, Carbosynth, Carl Roth, TCI, ABCR, Berry & Associates*).

#### General remarks

For drying of organic layers over  $\text{MgSO}_4$  or  $\text{Na}_2\text{SO}_4$  anhydrous salts were used.

Thin layer chromatography for tracking of reaction progresses and identification of target compound fractions was performed using TLC silica gel 60 F<sub>254</sub> aluminium sheets from *Merck*. Visualization was carried out using UV light ( $\lambda = 254 \text{ nm}$ ). Compounds without significant UV absorbance were visualized with  $\text{KMnO}_4$  staining reagent (42 mM  $\text{KMnO}_4$ , 482 mM  $\text{K}_2\text{CO}_3$ , 21 mM  $\text{NaOH}$ ) followed by heat drying.

Preparative column chromatography was carried out manually using silica gel (*Geduran Si 60 0.063-0.200 mm, Merck* or *Silica gel, 0.035-0.070 mm, Thermo Scientific*) as stationary phase. Reaction yields were calculated from the limiting starting material (1.0 eq.).

Triethylammonium bicarbonate buffer was prepared by bubbling  $\text{CO}_2$  through a 1.0 M mixture of HPLC grade  $\text{NEt}_3$  with ultrapure water ( $\text{ddH}_2\text{O}$ ) until a pH of approximately 7.5 – 8.0 was reached.

Fourier-transform infrared (FT-IR) spectroscopy was performed on a *Beckman Coulter DU 800*.

#### Schlenk technique

All procedures referred to as “under an atmosphere of argon”, “in a dry, argon flushed flask”, or similar were carried out using standard Schlenk technique (*i.e.* air- and moisture-free). Glassware was heat-dried under high vacuum and flushed with dry argon gas (N46 - 99,996 %,

*Air Liquide*; BIP 6.0 - 99.9999%, *AirProduct*) using a Schlenk line before setting up the reaction. Syringes and needles were purged with argon gas three times before use.

## 5.2.2 NMR spectroscopy

Nuclear magnetic resonance spectra were recorded by the institutional core facilities at room temperature on a *Bruker Avance I 300*, *Bruker Avance I 400*, *Bruker Avance I 500*, *Bruker Avance II+ 600*, *Bruker Avance III 500*, *Bruker Avance III HD 500 MHz Prodigy*, or *Bruker Avance III HD 700 MHz Cryo*.

Chemical shifts ( $\delta$ ) are given in parts per million (ppm) and spectra were calibrated to the respective residual solvent signal according to literature values<sup>414</sup>:  $\delta(\text{CDCl}_3) = 7.26$  ppm ( $^1\text{H}$ ), 77.16 ppm ( $^{13}\text{C}$ );  $\delta(\text{DMSO-}d_6) = 2.50$  ppm ( $^1\text{H}$ ), 39.52 ppm ( $^{13}\text{C}$ );  $\delta(\text{D}_2\text{O}) = 4.79$  ppm ( $^1\text{H}$ );  $\delta(\text{CD}_3\text{OD}) = 3.31$  ppm ( $^1\text{H}$ ), 49.00 ppm ( $^{13}\text{C}$ ),  $\delta(\text{DMF-}d_7) = 8.03$  ppm ( $^1\text{H}$ ), 34.89 ppm ( $^{13}\text{C}$ ). Reported coupling constants ( $J$ , [Hz]) were calculated from apparent signal positions in first order approximation.

In case of overlapping or ambiguous signals, "specific assignment given wherever possible" is stated at the beginning of NMR evaluation results. In these cases, the most likely atom number is given first followed by other(s) that possibly could be the origin of the signal.

## 5.2.3 HPLC analysis and purification

High performance liquid chromatography (HPLC) was applied to monitor reaction progress, analyze purity of substances, and isolate the product fractions from complex mixtures. Analytic runs and micro purifications were performed on an *Agilent 1260 Infinity II* HPLC system (*Agilent Technologies*) equipped with a multi wavelength detector (MWD) and a fluorescence detector using an *EC 150/4.6 Nucleodur 100-5 C18ec* column (*Macherey Nage*) and 0.1 % (w/v) ammonium acetate as solvent A and acetonitrile (MeCN) as solvent B.

Preparative HPLC purifications were either performed on an *Agilent 1200 Series* HPLC system (*Agilent Technologies*) with a *Gemini NX-C18 AXIA* column (5  $\mu\text{m}$ , 110  $\text{\AA}$ , 75  $\times$  30 mm, *phenomenex*) or on a *puriFlash 5.250P* (*Interchim*) with a *PFB5C18XS-150/212* column (*Interchim*) using 0.1 M triethylammonium bicarbonate buffer (pH = 7.5 – 8.0) as solvent A and acetonitrile as solvent B.

### 5.2.4 HPLC-MS spectrometry

High performance liquid chromatography in combination with mass spectrometry (HPLC-MS) was performed to monitor reaction progress and to ensure purity and integrity of target compounds.

HPLC-MS measurements were either performed on an *HTC esquire* (*Bruker Daltonic*) in combination with an *Agilent 1100 Series* HPLC system (*Agilent Technologies*) or on an *amaZon SL* mass spectrometer (*Bruker Daltonics*) in combination with an *Elute SP* HPLC system (*Bruker Daltonics*).

As a stationary phase a *Zorbax StableBond* (80 Å, C18, 2.1 × 50 mm, 5 µm) column (*Agilent Technologies*) was employed. In general, all compounds were submitted to HPLC-MS using either 0.1 % (v/v) formic acid or 0.1 % (w/v) ammonium acetate as solvent A and acetonitrile as solvent B. Depending on the solubility of the compounds, appropriate gradients in a range from 5 % → 100 % B and 60 % → 100 % B in 20 min at a flow rate of 0.4 mL min<sup>-1</sup> were applied. Triphosphates were generally submitted to HPLC-MS using 0.1 % (w/v) ammonium acetate as solvent A and acetonitrile as solvent B applying a gradient from 5 % → 100 % B in 20 min at a flow rate of 0.4 mL min<sup>-1</sup>. LC-MS data were interpreted using *Compass DataAnalysis 4.2* or *5.2*.

### 5.2.5 GC-MS spectrometry

Gas chromatography in combination with mass spectrometry (GC-MS) was performed to monitor reaction progress and in order to ensure integrity of target compounds.

GC-MS measurements were performed on a *Hewlett Packard HP 6890* gas chromatograph with a *HP 5974* mass detector and H<sub>2</sub> as the carrier gas. In general, the method starts at 50 °C with an initial hold for 5 min, followed by a 20°C per min gradient to 280 °C with a final hold of 10 min.

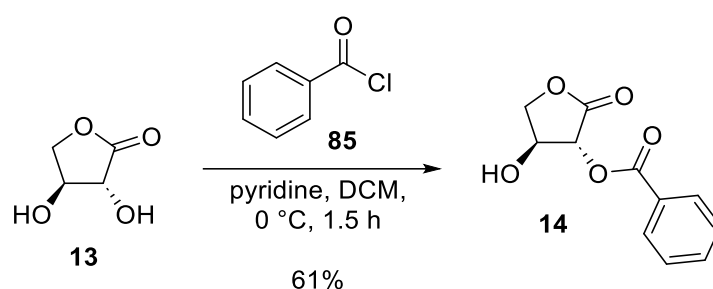
### 5.2.6 HR MS spectrometry

High resolution mass spectrometry (HR MS) was performed to accurately verify compounds, in particular novel compounds. High resolution mass spectra were recorded by institutional core facilities on either an *Exactive GC Orbitrap mass spectrometer* (*Thermo Fisher Scientific*) using electron ionization (EI) or on an *LTQ Orbitrap XL mass spectrometer* (*Thermo Fisher Scientific*) using either electrospray ionization (ESI) or atmospheric pressure chemical ionization (APCI).



## 5.2.7 Chemical syntheses

### 5.2.7.1 Synthesis of 2-O-Benzoyl-L-threonolactone (**14**)



#### Experimental Procedure:

Lactone **14** was prepared according to a protocol of Sau *et al.*<sup>367</sup>

To a stirred solution of  $\alpha$ -L-threonolactone (**13**, 1000 mg, 8.26 mmol, 1.0 eq.) in anhydrous DCM (18 mL) and anhydrous pyridine (2 mL), benzoyl chloride (**85**, 0.8 mL, 9.3 mmol, 1.1 eq.) was added dropwise under an atmosphere of argon at 0 °C. Stirring of the mixture was continued for 1.5 h at 0 °C. The crude reaction mixture was then sequentially washed with 1.0 M HCl (3  $\times$  10 mL) and with sat. aq. NaCl (10 mL). The combined aqueous layers were back extracted with DCM (2  $\times$  10 mL). The combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was removed *in vacuo*. The obtained crude product was purified by column chromatography (EtOAc/*c*-Hex, 1/1, *v/v*, *R<sub>f</sub>* = 0.52). 2-O-benzoyl-L-threonolactone (**14**) was isolated as a white solid (1100 mg, 5.0 mmol, 61%).

#### Analytatics:

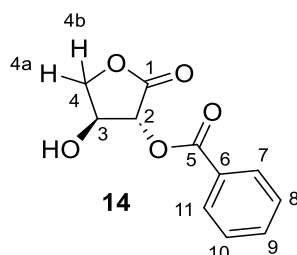
**Chemical Formula:** C<sub>11</sub>H<sub>10</sub>O<sub>5</sub>.

**Molecular Weight [g mol<sup>-1</sup>]:** 222.20.

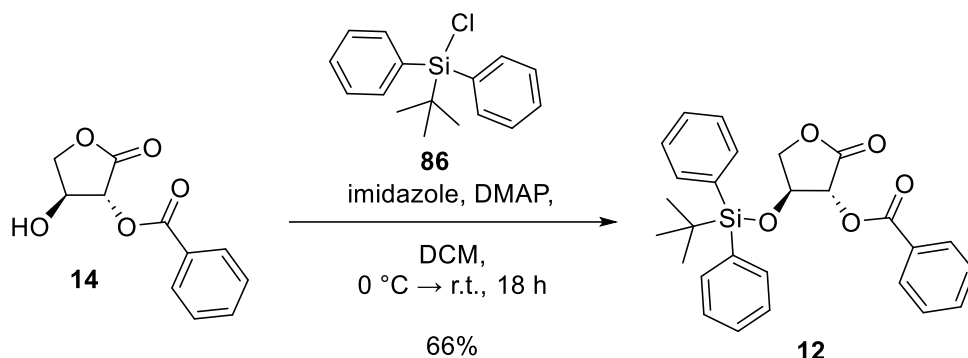
**<sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD, r.t.):**  $\delta$  [ppm] = 8.12 – 8.05 (m, 2H; H-7, H-11), 7.70 – 7.63 (m, 1H; H-9), 7.57 – 7.52 (m, 2H; H-8, H-10), 5.66 (d, <sup>3</sup>*J*<sub>H-2/H-3</sub> = 7.5 Hz, 1H; H-2), 4.78 (q, <sup>3</sup>*J*<sub>H-3/H-2</sub> = 7.5 Hz, <sup>3</sup>*J*<sub>H-3/H-4a</sub> = 7.5 Hz, <sup>3</sup>*J*<sub>H-3/H-4b</sub> = 7.5 Hz, 1H; H-3), 4.59 (dd, <sup>2</sup>*J*<sub>H-4b/H-4a</sub> = 9.1 Hz, <sup>3</sup>*J*<sub>H-4b/H-3</sub> = 7.5 Hz, 1H; H-4b/ H-4a), 4.13 (dd, <sup>2</sup>*J*<sub>H-4a/H-4b</sub> = 9.1 Hz, <sup>3</sup>*J*<sub>H-4a/H-3</sub> = 7.5 Hz, 1H; H-4a/ H-4b).

**<sup>13</sup>C-NMR (126 MHz, CD<sub>3</sub>OD, r.t.):**  $\delta$  [ppm] = 171.56 (C-1), 165.33 (C-5), 133.53 (C-arom.), 129.48 (C-arom.), 128.81 (C-arom.), 128.36 (C-arom.), 75.05 (C-2), 70.28 (C-3), 70.11(C-4).

**HR-MS (APCI<sup>+</sup>):** calculated for [M + H]<sup>+</sup> (C<sub>11</sub>H<sub>11</sub>O<sub>5</sub><sup>+</sup>): 223.0601; found: *m/z* = 223.0600.



### 5.2.7.2 Synthesis of 2-O-benzoyl-3-O-tert-butyl-diphenylsilyl-L-threofuranose (**12**)



#### Experimental Procedure:

Glycosyl donor **12** was prepared according to a protocol of Sau *et al.*<sup>367</sup>

In a dry, argon flushed flask, 2-O-benzoyl-L-threonolactone (**14**, 1100 mg, 4.95 mmol, 1.0 eq.) was suspended in DCM (10 mL). In a second flask, DMAP (43 mg, 0.35 mmol, 0.07 eq.) and imidazole (673 mg, 9.9 mmol, 2.0 eq.) were combined under argon atmosphere. Then, the solution of **14** was transferred into the second flask and the mixture was stirred until completely dissolved. The solution was cooled down to 0 °C and *tert*-butyldiphenylchlorosilane (**87**, 1.29 mL, 5.0 mmol, 1.0 eq.) was added dropwise. The reaction mixture was allowed to warm to ambient temperature and stirred overnight. The solvent was evaporated under reduced pressure and the residue was dissolved in *n*-hexane (10 mL). The organic layer was sequentially washed with 1.0 M HCl (10 mL), H<sub>2</sub>O (10 mL), and sat. aq. NaCl (10 mL). The combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was evaporated *in vacuo* to give glycosyl donor **12** (1510 mg, 3.27 mmol, 66%; EtOAc/*c*-Hex, 1/3, v/v, *R*<sub>f</sub> = 0.86) as a colorless oil.

#### Analytically:

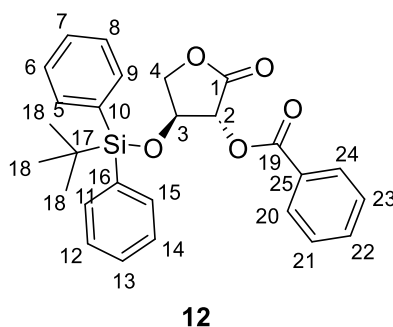
**Chemical Formula:** C<sub>27</sub>H<sub>26</sub>O<sub>5</sub>Si.

**Molecular Weight [g mol<sup>-1</sup>]:** 460.60.

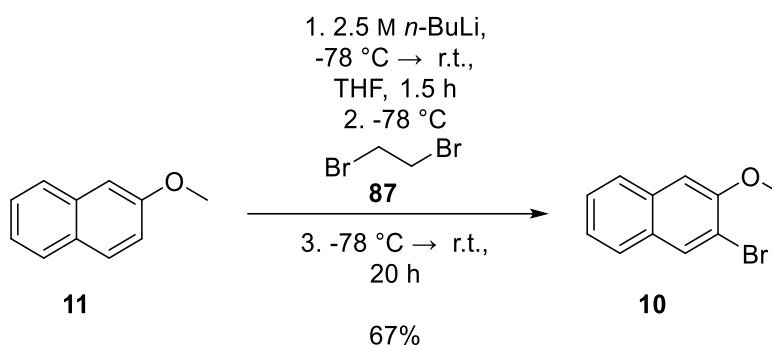
**<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>, r.t.):** δ [ppm] = 8.00 – 7.92 (m, 2H; H-arom.), 7.63 – 7.59 (m, 2H; H-arom.), 7.65 – 7.56 (m, 2H; H-arom.), 7.49 – 7.42 (m, 3H; H-arom.), 7.41 – 7.35 (m, 3H; H-arom.), 7.23 – 7.16 (m, 2H; H-arom.), 5.78 (d, <sup>3</sup>J<sub>H-2/H-3</sub> = 7.6 Hz, 1H; H-2), 4.79 (q, <sup>3</sup>J<sub>H-3/H-2</sub> = 7.5 Hz, <sup>3</sup>J<sub>H-3/H-4a</sub> = 7.5 Hz, <sup>3</sup>J<sub>H-3/H-4b</sub> = 7.5 Hz, 1H; H-3), 4.17 (dd, <sup>2</sup>J<sub>H-4b/H-4a</sub> = 9.1 Hz, <sup>3</sup>J<sub>H-4b/H-3</sub> = 7.5 Hz, 1H; H-4b/H-4a), 4.09 (dd, <sup>2</sup>J<sub>H-4a/H-4b</sub> = 9.1 Hz, <sup>3</sup>J<sub>H-4a/H-3</sub> = 7.5 Hz, 1H; H-4a/H-4b), 1.05 (s, 9H; H-18).

**<sup>13</sup>C-NMR (126 MHz, CDCl<sub>3</sub>, r.t.):** δ [ppm] = 170.17 (C-1), 165.16 (C-19), 135.66 (C-arom.), 132.68 (C-arom.), 131.97 (C-arom.), 130.26 (C-arom.), 128.54 (C-arom.), 128.22 (C-arom.), 128.02 (C-arom.), 75.01 (C-2), 72.16 (C-3), 70.16 (C-4), 26.81 (C-18), 19.11 (C-17).

**HR-MS (ESI<sup>+</sup>):** calculated for [M + Na]<sup>+</sup> (C<sub>27</sub>H<sub>28</sub>NaO<sub>5</sub>Si<sup>+</sup>): 483.1598; found: *m/z* = 483.1593.



### 5.2.7.3 Synthesis of 2-bromo-3-methoxynaphthalene (**10**)



#### Experimental Procedure:

Synthesis of **10** was carried out according to a published procedure by Niimi *et al.*<sup>366</sup>

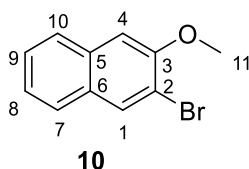
Under an atmosphere of argon, a solution of *n*-buthyllithium in *n*-hexanes (2.5 M, 26 mL, 65 mmol, 1.03 eq.) was added to a solution of 2-methoxynaphthalene (**11**, 10 g, 63 mmol, 1.0 eq) in THF (70 mL) at -78 °C. The mixture was warmed to r.t. and stirred for 1.5 h. After cooling to -78°C, 1,2-dibromoethane (**87**, 5.8 mL, 67 mmol, 1.06 eq) was added to the solution at this temperature. The resulting mixture was warmed to r.t. and again stirred for 20 h. Complete conversion of the starting material was validated by TLC. The reaction mixture was poured into sat. aq. NH<sub>4</sub>Cl solution (30 mL) and was extracted with Et<sub>2</sub>O (3 × 30 mL). The combined organic layers were washed with sat. aq. NaCl solution (3 × 20 mL), dried over MgSO<sub>4</sub>, and concentrated *in vacuo*. The reddish residue was purified by recrystallization from *n*-hexane to give 3-bromo-2-methoxynaphthalene (**10**, 10.08 g, 42.5 mmol, 67%) as a pale yellow solid.

#### Analytcs:

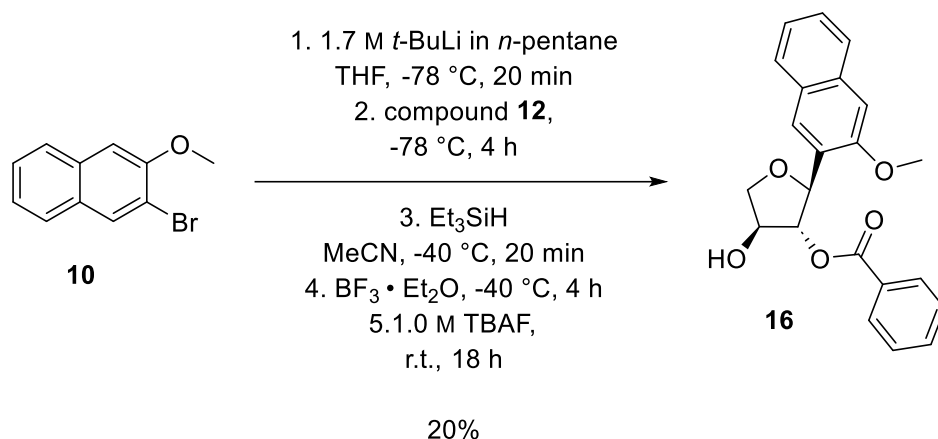
**Chemical Formula:** C<sub>11</sub>H<sub>9</sub>BrO.

**Molecular Weight [g mol<sup>-1</sup>]:** 237.10.

**<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, r.t.):** δ [ppm] = 8.06 (s, 1H; H-1), 7.77 – 7.65 (m, 2H; H-10, H-4), 7.46 (ddd, <sup>3</sup>J<sub>H-8/H-9</sub> = 8.2 Hz, <sup>3</sup>J<sub>H-8/H-7</sub> = 6.9 Hz, <sup>4</sup>J<sub>H-8/H-10</sub> = 1.4 Hz, 1H; H-8/H-9), 7.37 (ddd, <sup>3</sup>J<sub>H-9/H-8</sub> = 8.1 Hz, <sup>3</sup>J<sub>H-9/H-10</sub> 6.9 Hz, <sup>3</sup>J<sub>H-9/H-7</sub> 1.3 Hz, 1H; H-9/H-8), 7.16 (s, 1H; H-4), 4.00 (s, 3H; H-11).



#### 5.2.7.4 Synthesis of 1-(2'-O-benzoyl- $\alpha$ -L-threofuranosyl)-2-methoxynaphthalene (**16**)



#### Experimental Procedure:

The following reaction was performed using an adapted protocol from Li *et al.*<sup>161</sup>

Under an atmosphere of argon, *tert*-butyllithium in *n*-pentane (1.7 M, 2.4 mL, 0.9 mmol, 1.8 eq.) was added dropwise to a solution of 2-bromo-3-methoxynaphthalene (**10**, 1.0 g, 4.2 mmol, 2.0 eq.) in anhydrous THF (18 mL) at -78 °C. The resulting mixture was stirred for 20 min at -78 °C and lactone **12** (967 mg, 2.1 mmol, 1.0 eq.) in anhydrous THF (7 mL) was added slowly. After addition, the reaction mixture was stirred at -78 °C for another 4 h. The reaction was monitored by TLC (*c*-Hex/EtOAc, 4/1, *v/v*) and quenched by the addition of saturated NH<sub>4</sub>Cl solution (20 mL). After warming to r.t., the reaction was extracted with EtOAc (3 × 30 mL) and the organic layer was washed with sat. aq. NaCl (40 mL), dried over anhydrous MgSO<sub>4</sub> and the solvent was evaporated under reduced pressure. The crude product (1.75 g) was obtained as a yellow oil and subjected to the next step without further purification. Under an atmosphere of argon, the residue (1.75 g, 2.83 mmol, 1.0 eq) was dissolved in MeCN (30 mL) and the solution was cooled to -40 °C. Et<sub>3</sub>SiH was added (0.8 mL, 4.81 mmol, 1.7 mL) and stirring was continued for 20 min followed by the addition of BF<sub>3</sub> · Et<sub>2</sub>O (0.4 mL, 3.11 mmol, 1.1 eq.). The resulting mixture was stirred at -40 °C for 4 h, then warmed to room temperature and quenched by the addition of sat. aq. Na<sub>2</sub>CO<sub>3</sub> solution (18 mL). The reaction mixture was extracted with EtOAc (3 × 30 mL) and the organic layer was washed with sat. aq. NaCl solution (40 mL), dried over anhydrous MgSO<sub>4</sub>, filtered and the solvent was removed under reduced pressure. The crude product was purified by silica gel column chromatography (EtOAc/*n*-hexane, 1/2, *v/v*) to afford silyl-protected nucleoside **15** with small

impurities. The residue was dissolved in THF (5 mL) under argon atmosphere and TBAF in THF (1.0 M, 1.7 mL, 1.7 mmol) was added dropwise and the mixture was stirred for 18 h at r.t. The reaction was evaporated to dryness and the residue was redissolved in EtOAc (10 mL). The organic layer was sequentially washed with HCl (1.0 M, 5 mL), H<sub>2</sub>O (5 mL) and sat. aq. NaCl solution (5 mL), dried over anhydrous MgSO<sub>4</sub> and the solvent was evaporated *in vacuo*. The residue was purified by silica column chromatography (EtOAc/*c*-Hex, 1/1, *v/v*, *R<sub>f</sub>* = 0.65) and the desired compound **16** was obtained as a white foam (147 mg, 0.4 mmol, 20%).

Analytics (specific assignment given whenever possible):

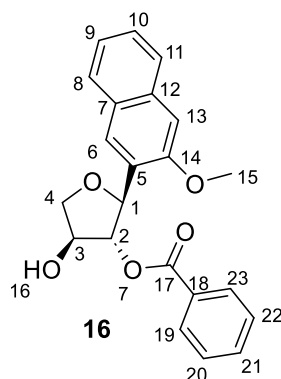
**Chemical Formula:** C<sub>22</sub>H<sub>20</sub>O<sub>5</sub>.

**Molecular Weight [g mol<sup>-1</sup>]:** 364.40.

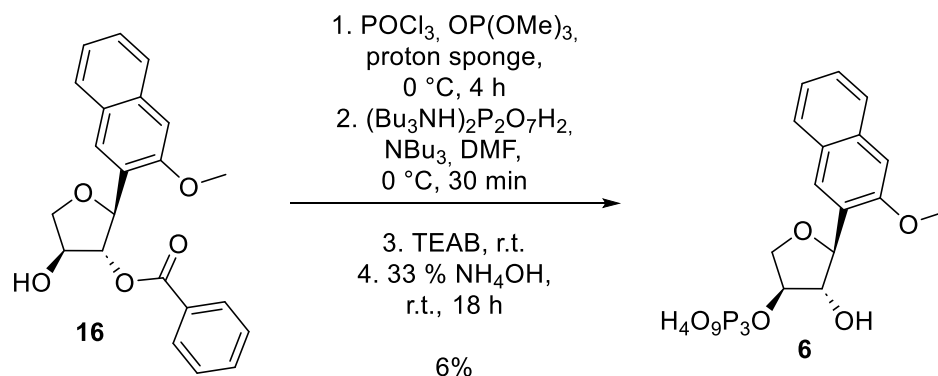
**<sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>, r.t.):** δ [ppm] = 8.09 (s, 1H; H-6), 7.84 (d, <sup>3</sup>*J*<sub>H-11/H-10</sub> = 8.0 Hz, 1H; H-11), 7.70 (d, <sup>3</sup>*J*<sub>H-8/H-9</sub> = 8.2 Hz, 1H; H-8), 7.44 – 7.40 (m, 3H; H-9, H-arom.), 7.38 – 7.33 (m, 2H; H-10, H-arom.), 7.15 – 7.09 (m, 2H; H-arom.), 7.04 (s, 1H; H-13), 5.68 (dd, <sup>3</sup>*J*<sub>H-1/H-2</sub> = 4.5 Hz, 1.2 Hz, 1H; H-1), 5.64 (dd, <sup>3</sup>*J*<sub>H-2/H-1</sub> = 4.5 Hz, 1.5 Hz, 1H; H-2), 4.55 (t, *J* = 4.7 Hz, 1H; H-3), 4.48 (dd, <sup>2</sup>*J*<sub>H-4b/H-4a</sub> = 9.7 Hz, 5.4 Hz, 1H; H-4b/H-4a), 3.99 (dd, <sup>2</sup>*J*<sub>H-4a/H-4b</sub> = 9.7 Hz, 3.7 Hz, 1H; H-4a/H-4b), 3.85 (s, 3H; H-15).

**<sup>13</sup>C-NMR (151 MHz, CDCl<sub>3</sub>, r.t.):** δ [ppm] = 166.54 (C-17), 154.69 (C-14), 134.00 (C-7), 133.20 (C-arom.), 129.86 (C-arom.), 129.50 (C-arom.), 129.47 (C-5), 128.77 (C-arom.), 128.63 (C-12), 128.31 (C-arom.), 127.87 (C-11), 126.93 (C-9), 126.48 (C-8), 126.47 (C-6), 126.16 (C-7), 123.80 (C-10), 104.75 (C-13), 81.34 (C-2), 77.90 (C-1), 77.23 (C-3), 72.88 (C-4), 55.49 (C-15).

**HR-MS (EI<sup>+</sup>, 70 eV):** calculated for [M<sup>+</sup>] (C<sub>22</sub>H<sub>20</sub>O<sub>5</sub>): 364.13053; found: *m/z* = 364.13033.



### 5.2.7.5 Synthesis of 1-( $\alpha$ -L-threofuranos-3'-triphosphate-1'-yl)-2-methoxynaphthalene (**6**)



#### Experimental Procedure:

Triphosphate synthesis was performed according to a protocol published by Srivatsan and Tor.<sup>370</sup>

Nucleoside **16** (75 mg, 0.21 mmol, 1.0 eq.) and 1,8-bis(*N,N*-dimethylamino)naphthalene (proton sponge, 36 mg, 0.17 mmol, 0.8 eq.) were dried under high vacuum overnight and were under argon atmosphere dissolved in trimethyl phosphate (2.31 mL, freshly distilled, stored over 3 Å molecular sieves and under argon) and cooled with an ice bath.  $\text{POCl}_3$  (48  $\mu\text{L}$ , 0.53 mmol, 2.5 eq., freshly distilled and stored under argon) was added dropwise at  $0\text{ }^\circ\text{C}$  and the reaction was stirred under ice cold conditions for 4 h. Subsequently, tributylamine (0.53 mL, 2.2 mmol, 10.5 eq., freshly distilled and stored under argon) and a solution of bis-tri-*n*-butylammonium pyrophosphate in dry DMF (0.5 M, 2.3 mL, 1.8 mmol, 5.4 eq.) were added simultaneously in one portion under ice-cooling. The mixture was stirred at  $0\text{ }^\circ\text{C}$  for 30 min and quenched by the addition of TEAB buffer (1.0 M, 15 mL, pH = 8.0). Volatiles were evaporated under high vacuum at room temperature. Coevaporation of the crude product with  $\text{H}_2\text{O}$  (20 mL) was performed. The crude product **17** was lyophilized overnight and was subjected to the next step without further purification.

In a sealed tube, the residue of triphosphate **17** was combined with saturated  $\text{NH}_4\text{OH}$  (33 %, 10 mL). The reaction was stirred at r.t. for 18 h, followed by evaporation to dryness. The crude product was purified via HPLC (15 %  $\rightarrow$  32 % MeCN in 0.1 M aq. TEAB buffer in 19 min, flow rate = 20.0 mL  $\text{min}^{-1}$ ). After HPLC purification, triphosphate **6** was again coevaporated with  $\text{H}_2\text{O}$  (2  $\times$  10 mL). Purified triphosphate **6** (9.6 mg, 0.014 mmol, 6%) was obtained as the 2  $\times$  triethylammonium (TEA) salt as a yellow foam.

#### Analytcs:

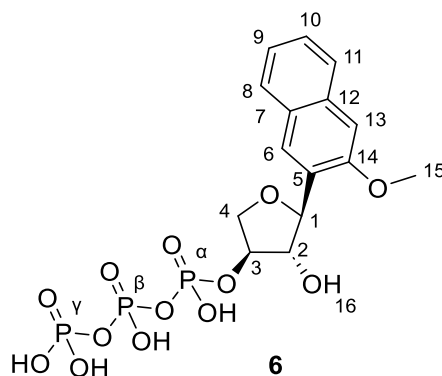
**Chemical Formula:**  $\text{C}_{15}\text{H}_{19}\text{O}_{13}\text{P}_3$ .

**Molecular Weight [g  $\text{mol}^{-1}$ ]:** 500.23.

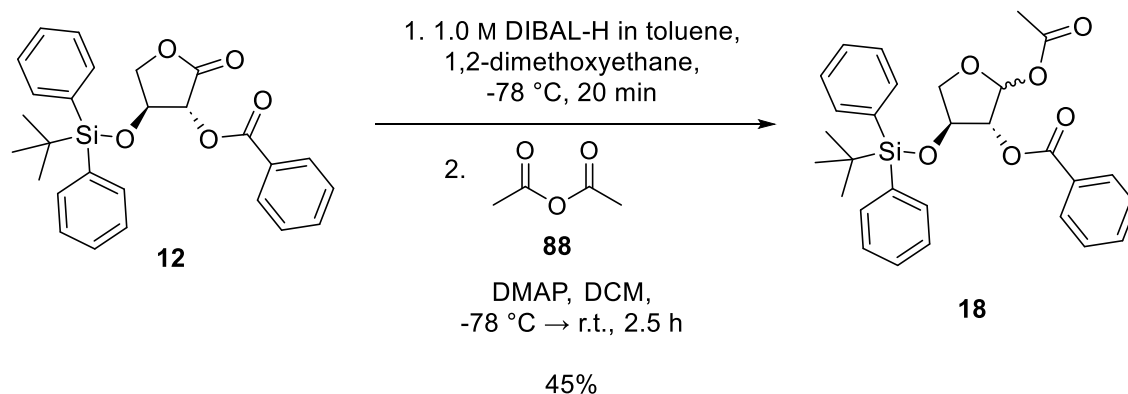
**$^1\text{H}$  NMR (500 MHz,  $\text{D}_2\text{O}$ , r.t.):**  $\delta$  [ppm] = 7.97 (s, 1H; H-6), 7.93 – 7.85 (m, 2H; H-11, H-8), 7.53 (t,  $^3J_{\text{H-10/H-9}}$  = 7.6 Hz, 1H; H-10), 7.44 (t,  $^3J_{\text{H-9/H-10}}$  = 7.8 Hz, 1H; H-9), 7.39 (s, 1H; H-13), 5.53 (d,  $^3J_{\text{H-1/H-2}}$  = 3.3 Hz, 1H; H-1), 4.92 (dd,  $J$  = 8.7 Hz,  $^3J_{\text{H-3/H-4b}}$  = 4.2 Hz, 1H; H-3), 4.83 (d,  $^3J_{\text{H-2/H-1}}$  = 3.2 Hz, 1H; H-2), 4.45 (dd,  $^2J_{\text{H-4b/H-4a}}$  = 10.2 Hz,  $^3J_{\text{H-4b/H-3}}$  = 4.3 Hz, 1H; H-4b/H-4a), 4.19 (d,  $^2J_{\text{H-4a/H-4b}}$  = 10.5 Hz, 1H; H-4a/H-4b), 4.00 (s, 3H; H-15).

**$^{31}\text{P}$ -NMR (202 MHz,  $\text{D}_2\text{O}$ , r.t.):**  $\delta$  [ppm] = -10.92 (d,  $J_{\text{P-}\gamma/\text{P-}\beta}$  = 20.1 Hz; P- $\gamma$ ), -12.28 (d,  $J_{\text{P-}\alpha/\text{P-}\beta}$  = 20.2 Hz; P- $\alpha$ ), -23.38 (t,  $J_{\text{P-}\beta/\text{P-}\alpha}$  = 20.1 Hz; P- $\beta$ ).

**HR-MS (ESI $^+$ ):** calculated for  $[\text{M} - \text{H}]^-$  ( $\text{C}_{15}\text{H}_{18}\text{O}_{13}\text{P}_3$ ): 498.99657; found:  $m/z$  = 498.99874.



### 5.2.7.6 Synthesis of 1-O-acetyl-2-O-benzoyl-3-O-tert-butylidiphenylsilyl-L-threofuranose (**18**)



#### Experimental Procedure:

Universal glycosyl donor **18** was prepared according to a protocol of Sau *et al.*<sup>367</sup>

In a dry, argon flushed flask, glycosyl donor **12** (1510 mg, 3.27 mmol, 1.0 eq.) was dissolved in 1,2-dimethoxyethane (11 mL) and cooled down to -78 °C. To the stirred reaction mixture, a solution of DIBAL-H in toluene (1.0 M, 4.24 mL, 4.24 mmol, 1.3 eq.) was added dropwise and the mixture was stirred for additional 20 min at -78 °C. TLC validated the consumption of starting material (EtOAc/*c*-Hex, 1/3, *v/v*). A premade solution containing  $\text{Ac}_2\text{O}$  (**89**, 1.66 mL, 16.3 mmol, 5 eq.) and DMAP (597 mg, 4.89 mmol, 1.5 eq.) in DCM (2 mL) was then added dropwise at -78 °C. After 20 min, the reaction mixture was allowed to stir for 2.5 h while warming to room temperature. The mixture was then diluted with *n*-hexanes (10 mL) and

poured into cold aq. HCl solution (1.0 M, 20 mL) while stirring. The organic layer was separated and washed sequentially with H<sub>2</sub>O (10 mL), sat. aq. NaHCO<sub>3</sub> (2 × 10 mL) and sat. aq. NaCl (20 mL). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was removed *in vacuo* to give 1131 mg of crude product as a yellow oil. The crude product was then suspended in 10 % EtOAc in *n*-hexanes (10 mL), passed through a short pad of silica and washed out with 10 % EtOAc in *n*-hexanes (600 mL). An anomeric mixture of universal glycosyl donor **18** (732 mg, 1.45 mmol, 45%; EtOAc/*n*-Hex, 1/9, *v/v*, *R<sub>f</sub>* = 0.45, 0.41) was obtained as a colorless oil.

### Analytcs:

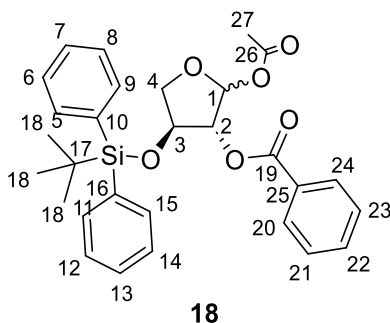
**Chemical Formula:** C<sub>27</sub>H<sub>26</sub>O<sub>5</sub>Si.

**Molecular Weight [g mol<sup>-1</sup>]:** 460.60.

**<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>, r.t.):** δ [ppm] = 8.00 – 7.92 (m, 2H; H-arom.), 7.63 – 7.59 (m, 2H; H-arom.), 7.65 – 7.56 (m, 2H; H-arom.), 7.49 – 7.42 (m, 3H; H-arom.), 7.41 – 7.35 (m, 3H; H-arom.), 7.23 – 7.16 (m, 2H; H-arom.), 5.78 (d, <sup>3</sup>J<sub>H-2/H-3</sub> = 7.6 Hz, 1H; H-2), 4.79 (q, <sup>3</sup>J<sub>H-3/H-2</sub> = 7.5 Hz, <sup>3</sup>J<sub>H-3/H-4a</sub> = 7.5 Hz, <sup>3</sup>J<sub>H-3/H-4b</sub> = 7.5 Hz, 1H; H-3), 4.17 (dd, <sup>2</sup>J<sub>H-4b/H-4a</sub> = 9.1 Hz, <sup>3</sup>J<sub>H-4b/H-3</sub> = 7.5 Hz, 1H; H-4b/H-4a), 4.09 (dd, <sup>2</sup>J<sub>H-4a/H-4b</sub> = 9.1 Hz, <sup>3</sup>J<sub>H-4a/H-3</sub> = 7.5 Hz, 1H; H-4a/H-4b), 1.88 (s, 3H; H-27), 1.05 (s, 9H; H-18).

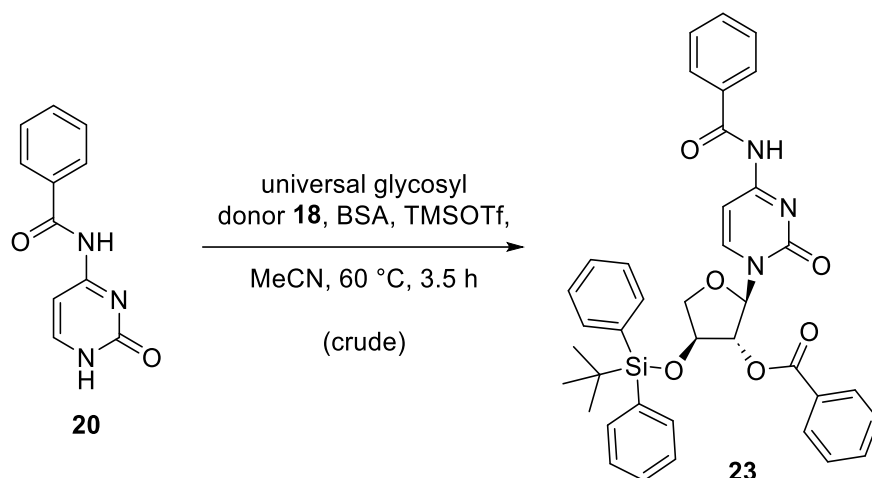
**<sup>13</sup>C-NMR (126 MHz, CDCl<sub>3</sub>, r.t.):** δ [ppm] = 170.17 (C-1), 165.16 (C-19), 135.66 (C-arom.), 132.68 (C-arom.), 131.97 (C-arom.), 130.26 (C-arom.), 128.54 (C-arom.), 128.22 (C-arom.), 128.02 (C-arom.), 75.01 (C-2), 72.16 (C-3), 70.16 (C-4), 26.81 (C-18), 19.11 (C-17).

**HR-MS (ESI<sup>+</sup>):** calculated for [M + Na]<sup>+</sup> (C<sub>27</sub>H<sub>28</sub>NaO<sub>5</sub>Si<sup>+</sup>): 483.1598; found: *m/z* = 483.1593.





### 5.2.7.7 *N*<sup>4</sup>-benzoyl-1-(2'-*O*-benzoyl-3'-*O*-tert-butylidiphenylsilyl- $\alpha$ -L-threofuranosyl)cytosine (**23**)



#### Experimental Procedure:

Nucleoside **23** was prepared according to a protocol of Zhang *et al.*<sup>146</sup>

In a dry, argon flushed flask, universal glycosyl donor **18** (1100 mg, 2.2 mmol, 1.0 eq.) and *N*<sup>4</sup>-benzoylcytosine (**20**, 517 mg, 2.4 mmol, 1.1 eq.) were dissolved in anhydrous MeCN (10 mL). *N,O*-bis(trimethylsilyl)acetamide (1.04 mL, 4.8 mmol, 2.2 eq.) was added and the solution was stirred for 40 min at 60 °C. Subsequently, TMSOTf (0.6 mL, 3.5 mmol, 1.6 eq.) was added dropwise and stirring was continued for another 3 h at 60 °C. The mixture was cooled to r.t., diluted with EtOAc (20 mL), and poured into cold sat. aq. NaHCO<sub>3</sub> solution (15 mL) while stirring. Sat. aq. NaCl solution (20 mL) was added and the organic layer was separated and washed with sat. aq. NaCl (3 × 5 mL), dried over MgSO<sub>4</sub>, and concentrated under reduced pressure. The crude nucleoside **23** (1345 mg, 2.0 mmol, 91%; EtOAc/*c*-Hex, 1/1, *v/v*, *R*<sub>f</sub> = 0.6) was obtained as a white foam and identity of the product was verified via HR-MS. Compound **23** was submitted to the next step without further purification.

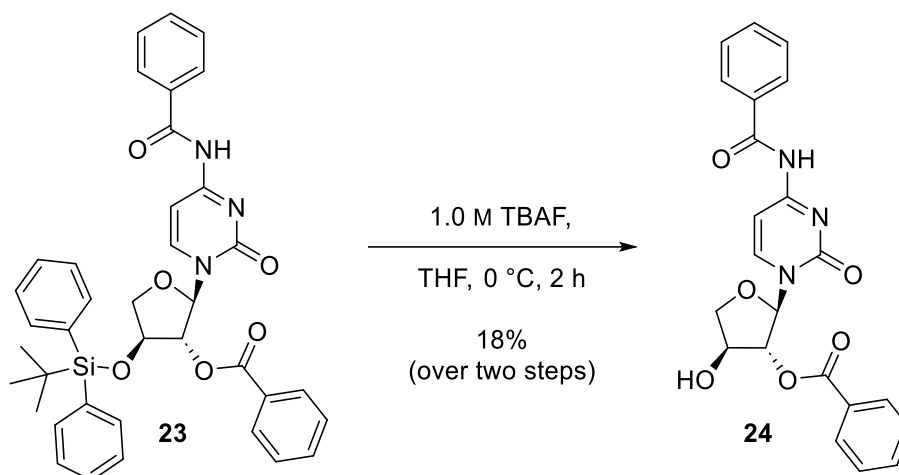
#### Analytics:

**Chemical Formula:** C<sub>38</sub>H<sub>37</sub>N<sub>3</sub>O<sub>6</sub>Si.

**Molecular Weight [g mol<sup>-1</sup>]:** 659.81.

**HR-MS (ESI<sup>+</sup>):** calculated for [M + H]<sup>+</sup> (C<sub>38</sub>H<sub>38</sub>N<sub>3</sub>O<sub>6</sub>Si<sup>+</sup>): 660.2524; found: *m/z* = 660.2525.

### 5.2.7.8 Synthesis of *N*<sup>4</sup>-benzoyl-1-(2'-*O*-benzoyl- $\alpha$ -L-threofuranosyl)cytosine (**24**)



#### Experimental Procedure:

TBDPS deprotection was performed according to a protocol of Sau *et al.*<sup>367</sup>

Under an atmosphere of argon, tetrabutylammonium fluoride in THF (1.0 M, 2.45 mL, 2.45 mmol) was added dropwise to a stirred solution of crude nucleoside **23** (1345 mg, 2.04 mmol, 1.0 eq.) in THF (7.5 mL) at 0 °C. The reaction mixture was stirred for 2 h while allowing to warm to r.t. Subsequently, the solvent was removed *in vacuo* and the residue was dissolved in EtOAc (10 mL). The organic layer was washed with H<sub>2</sub>O (2 × 10 mL) and sat. aq. NaCl (10 mL). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure. The crude product was purified using column chromatography (EtOAc/*c*-Hex, 2/1, *v/v*) to afford pure nucleoside **24** (168 mg, 0.4 mmol, 18% over two steps; EtOAc/*c*-Hex, 2/1, *v/v*, *R*<sub>f</sub> = 0.53) as a white solid.

#### Analytics:

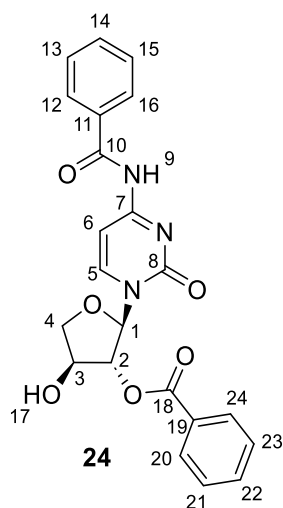
**Chemical Formula:** C<sub>22</sub>H<sub>19</sub>N<sub>3</sub>O<sub>6</sub>.

**Molecular Weight [g mol<sup>-1</sup>]:** 421.41.

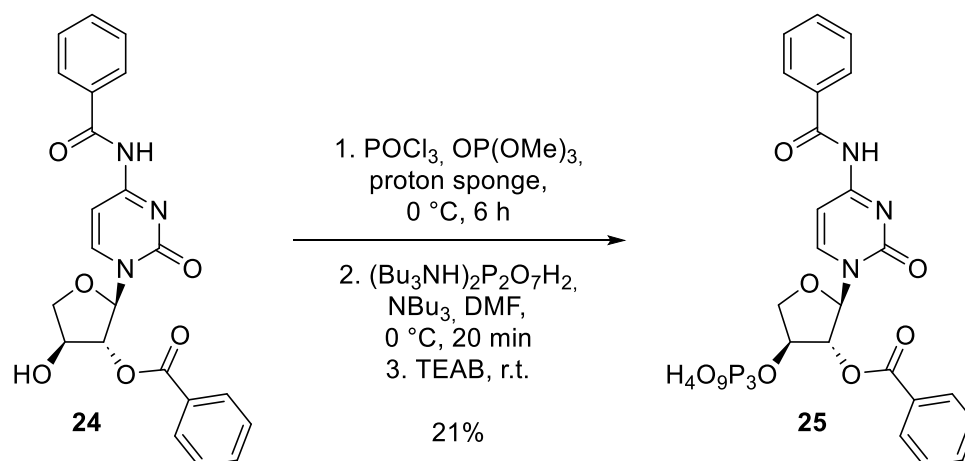
**<sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>, r.t.):**  $\delta$  [ppm] = 11.28 (s, 1H; H-9), 8.21 (d, 7.5 Hz, 1H; H-arom.), 8.07 – 7.99 (m, 4H; H-arom.), 7.76 – 7.67 (m, 1H; H-5), 7.67 – 7.59 (m, 1H; H-arom.), 7.60 – 7.55 (m, 3H; H-arom.), 7.56 – 7.48 (m, 2H; H-6, H-arom.), 5.99 (d, <sup>3</sup>J<sub>H-1/H-2</sub> = 1.0 Hz, 1H; H-1), 5.82 (d, *J* = 3.0 Hz, 1H; H-17), 5.39 (s, 1H; H-2), 4.40 – 4.25 (m, 3H; H-4, H-3).

**<sup>13</sup>C NMR (126 MHz, DMSO-*d*<sub>6</sub>, r.t.):**  $\delta$  [ppm] = 172.86 (C-10), 165.83 (C-18), 164.52 (C-7), 164.44 (C-8), 145.64 (C-5), 133.85 (C-arom.), 132.70 (C-arom.), 129.56 (C-arom.), 129.48 (C-arom.), 129.00 (C-6), 128.87 (C-arom.), 128.86 (C-arom.), 128.43 (C-arom.), 128.42 (C arom.), 90.88 (C-1), 81.52 (C-2), 76.54 (C-4), 73.34, 72.41 (C-3).

**HR-MS (ESI):** calculated for [M + H]<sup>+</sup> (C<sub>22</sub>H<sub>20</sub>N<sub>3</sub>O<sub>6</sub><sup>+</sup>): 422.1347; found: *m/z* = 422.1342.



### 5.2.7.9 Synthesis of *N*<sup>4</sup>-benzoyl-1-(2'-*O*-benzoyl- $\alpha$ -L-threofuranos-3'-triphosphate-1'-yl)cytosine (**25**)



#### Experimental Procedure:

Triphosphate synthesis was performed according to a protocol published by Srivatsan and Tor.<sup>370</sup>

Nucleoside **24** (85 mg, 0.2 mmol, 1.0 eq.) and proton sponge (35 mg, 0.16 mmol, 0.8 eq.) were dried under high vacuum overnight, dissolved in trimethylphosphate (1.37 mL, freshly distilled and stored over 3 Å molecular sieves and under argon) under argon atmosphere and cooled with an ice bath. POCl<sub>3</sub> (46 μL, 0.5 mmol, 2.5 eq., freshly distilled and stored under argon) was added dropwise at 0 °C and the reaction was stirred under ice cold conditions for 6 h. Subsequently, tributylamine (0.5 mL, 2.1 mmol, 10.5 eq., freshly distilled and stored under argon) and a solution of bis-tri-*n*-butylammonium pyrophosphate in dry DMF (0.5 M, 2.2 mL, 1.1 mmol, 5.4 eq.) were added simultaneously in one portion under ice-cooling. The mixture was stirred at 0 °C for 20 min and quenched by the addition of TEAB buffer (1.0 M, 20 mL, pH = 8.0). Volatiles were evaporated under high vacuum at room temperature. Coevaporation of the crude product with H<sub>2</sub>O (2 × 20 mL) was performed. The crude product was lyophilized

overnight before purification via HPLC (0 % → 40 % MeCN in 0.1 M aq. TEAB buffer in 6 min, flow rate = 40.0 mL min<sup>-1</sup>) was performed. After HPLC purification, the residue was again coevaporated with H<sub>2</sub>O (2 × 15 mL) and lyophilized overnight. Purified triphosphate **25** (49 mg, 0.042 mmol, 21%) was obtained as the 5 × triethylammonium salt as a faint yellow foam.

#### Analytcs:

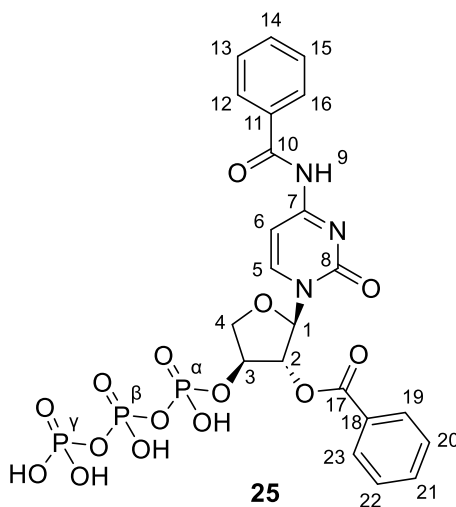
**Chemical Formula:** C<sub>22</sub>H<sub>22</sub>N<sub>3</sub>O<sub>15</sub>P<sub>3</sub>.

**Molecular Weight [g mol<sup>-1</sup>]:** 661.35.

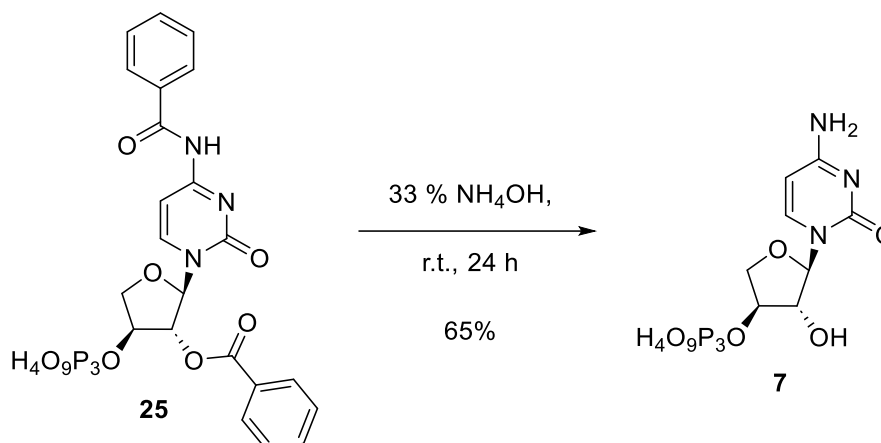
**<sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O, r.t.):** δ [ppm] = 8.36 (d, <sup>3</sup>J<sub>H-5/H-6</sub> = 7.5 Hz, 1H; H-5), 8.02 – 8.07 (m, 2H; H-arom.), 7.89 – 7.84 (m, 2H; H-arom), 7.68 – 7.60 (m, 2H; H-arom), 7.54 – 7.46 (m, 4H; H-arom.), 7.43 (d, <sup>3</sup>J<sub>H-6/H-5</sub> = 7.5 Hz, 1H; H-6), 6.10 (d, <sup>3</sup>J<sub>H-1/H-2</sub> = 1.2 Hz, 1H; H-1), 5.62 (q, <sup>3</sup>J<sub>H-2/H-1</sub> = 1.2 Hz, 1H; H-2), 5.05 (dd, J = 8.4 Hz, <sup>3</sup>J<sub>H-3/H-4a</sub> = 3.5 Hz, 1H; H-3), 4.80 (d, <sup>2</sup>J<sub>H-4b/H-4a</sub> = 10.8 Hz, 1H; H-4b/H-4a), 4.43 (dd, <sup>2</sup>J<sub>H-4a/H-4b</sub> = 10.9 Hz, <sup>3</sup>J<sub>H-4a/H-3</sub> = 3.5 Hz, 1H; H-4a/H-4b).

**<sup>31</sup>P NMR (162 MHz, D<sub>2</sub>O, r.t.):** δ [ppm] = -6.57 (d, <sup>2</sup>J<sub>P-γ/P-β</sub> = 19.9 Hz; P-γ), -13.15 (d, <sup>2</sup>J<sub>P-α/P-β</sub> = 19.9 Hz; P-α), -22.63 (t, <sup>2</sup>J<sub>P-β/P-α</sub> = 19.9 Hz, <sup>2</sup>J<sub>P-β/P-γ</sub> = 19.9 Hz; P-β).

**HR-MS (ESI):** calculated for [M - H]<sup>-</sup> (C<sub>22</sub>H<sub>21</sub>N<sub>3</sub>O<sub>15</sub>P<sub>3</sub><sup>-</sup>): 660.0191; found: *m/z* = 660.0195.



### 5.2.7.10 Synthesis of 1-( $\alpha$ -L-threofuranos-3'-triphosphate-1'-yl)cytosine (**7**)



#### Experimental Procedure:

Debenzoylation was performed according to a protocol of Zou *et al.*<sup>145</sup>

In a sealed tube, triphosphate **25** (49 mg, 0.16 mmol, 1.0 eq.) was dissolved in saturated  $\text{NH}_4\text{OH}$  (33 %, 10 mL). The reaction was stirred at r.t. for 24 h. Volatiles were evaporated under high vacuum at room temperature. The crude product was redissolved in  $\text{H}_2\text{O}$  (5 mL) lyophilized overnight before purification via HPLC (0 %  $\rightarrow$  40 % MeCN in 0.1 M aq. TEAB buffer in 6 min, flow rate = 40.0 mL  $\text{min}^{-1}$ ) was performed. After HPLC purification, the residue was coevaporated with  $\text{H}_2\text{O}$  (2  $\times$  5 mL) and lyophilized overnight. The triphosphate **7** (23.6 mg, 0.028 mmol, 65%) was obtained as the 4  $\times$  triethylammonium salt as a colorless foam.

#### Analytics (specific assignment given whenever possible):

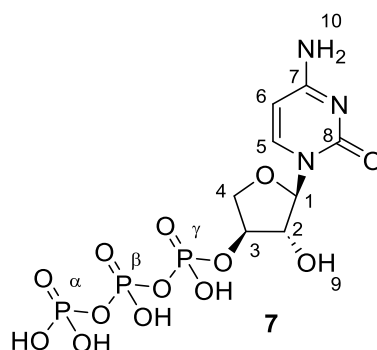
**Chemical Formula:**  $\text{C}_8\text{H}_{14}\text{N}_3\text{O}_{13}\text{P}_3$ .

**Molecular Weight [g  $\text{mol}^{-1}$ ]:** 453.13.

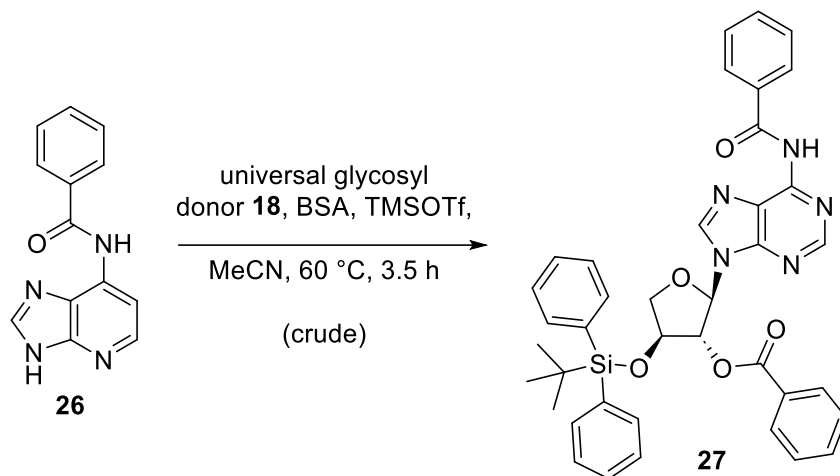
**$^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}$ , r.t.):**  $\delta$  [ppm] = 7.83 (d,  $^3J_{\text{H-5/H-6}} = 7.5$  Hz, 1H; H-5), 6.07 (d,  $^3J_{\text{H-6/G-5}} = 7.5$  Hz, 1H; H-6), 5.86 (s, 1H; H-1), 4.56 (d,  $^2J_{\text{H-4b/H-4a}} = 10.6$  Hz, 1H; H-4b/H-4a), 4.48 (s, 1H; H-2/H-3), 4.35 (dd,  $^2J_{\text{H-4a/H-4b}} = 10.8$  Hz, 3.5 Hz, 1H; H-4a/H-4b).

**$^{31}\text{P}$ -NMR (202 MHz,  $\text{D}_2\text{O}$ , r.t.):**  $\delta$  [ppm] = -10.46 (d,  $^2J_{\text{P-}\gamma/\text{P-}\beta} = 20.0$  Hz; P- $\gamma$ ), -12.98 (d,  $^2J_{\text{P-}\alpha/\text{P-}\beta} = 20.0$  Hz; P- $\alpha$ ), -23.34 (t,  $^2J_{\text{P-}\beta/\text{P-}\alpha}, ^2J_{\text{P-}\beta/\text{P-}\gamma} = 20.0$  Hz; P- $\beta$ ).

**HR-MS (ESI):** calculated for  $[\text{M} - \text{H}]^-$  ( $\text{C}_8\text{H}_{13}\text{N}_3\text{O}_{13}\text{P}_3^-$ ): 451.9667; found:  $m/z = 451.9661$ .



### 5.2.7.11 Synthesis of *N*<sup>6</sup>-benzoyl-1-(2'-*O*-benzoyl-3'-*O*-tert-butyl-diphenylsilyl- $\alpha$ -L-threofuranosyl)adenine (**27**)



#### Experimental Procedure:

Nucleoside **27** was prepared according to a protocol from Zhang *et al.*<sup>146</sup>

In a dry, argon flushed flask, universal glycosyl donor **18** (1100 mg, 2.2 mmol, 1.0 eq.) and *N*<sup>6</sup>-benzoyladenine (**26**, 574 mg, 2.4 mmol, 1.1 eq.) were dissolved in anhydrous MeCN (15 mL). *N,O*-bis(trimethylsilyl)acetamide (1.18 mL, 4.8 mmol, 2.2 eq.) was added and the solution was stirred for 40 min at 60 °C. Subsequently, TMSOTf (0.6 mL, 3.5 mmol, 1.6 eq.) was added dropwise and stirring was continued for another 3 h at 60 °C. The mixture was cooled to r.t., diluted with EtOAc (20 mL), and poured into cold sat. aq. NaHCO<sub>3</sub> solution (15 mL) while stirring. The organic layer was separated and washed with H<sub>2</sub>O (2 × 10 mL), followed by sat. aq. NaCl (10 mL), dried over MgSO<sub>4</sub>, and concentrated under reduced pressure. The crude nucleoside **27** (1345 mg, 2.0 mmol, EtOAc/*c*-Hex, 1/1, *v/v*, *R*<sub>f</sub> = 0.5) was obtained as a white foam and identity of the product was verified via HR-MS. Compound **27** was submitted to the next step without further purification.

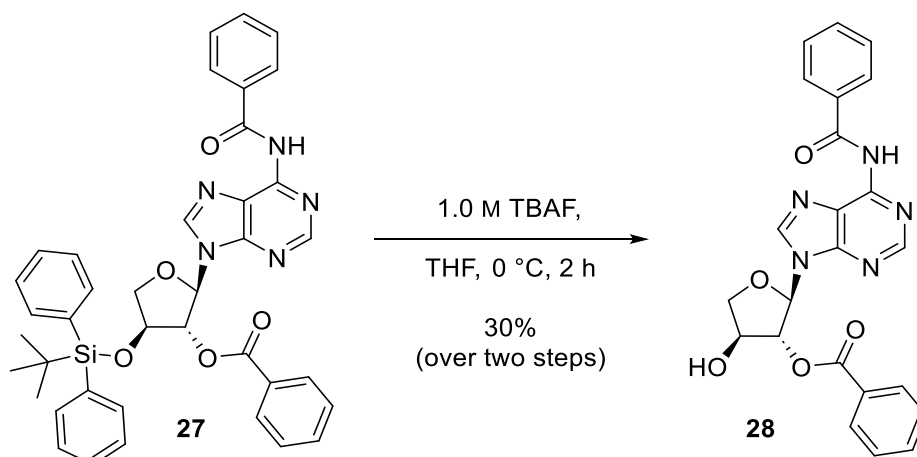
#### Analytics:

**Chemical Formula:** C<sub>39</sub>H<sub>37</sub>N<sub>5</sub>O<sub>5</sub>Si.

**Molecular Weight [g mol<sup>-1</sup>]:** 683.84.

**HR-MS (ESI<sup>+</sup>):** calculated for [M + H]<sup>+</sup> (C<sub>39</sub>H<sub>38</sub>N<sub>5</sub>O<sub>5</sub>Si<sup>+</sup>): 684.2637; found: *m/z* = 684.2635.

### 5.2.7.12 Synthesis of *N*<sup>6</sup>-benzoyl-1-(2'-*O*-benzoyl- $\alpha$ -L-threofuranosyl)adenine (**28**)



#### Experimental Procedure:

TBDPS deprotection was performed according to a protocol of Sau *et al.*<sup>367</sup>

Under an atmosphere of argon, TBAF in THF (1.0 M, 2.75 mL, 2.75 mmol) was added dropwise to a stirred solution of crude nucleoside **27** (1565 mg, 2.29 mmol, 1.0 eq.) in THF (7 mL) at 0 °C. The reaction mixture was stirred for 2 h while allowing to warm to r.t. Subsequently, the solvent was removed *in vacuo* and the residue was dissolved in EtOAc (10 mL). The organic layer was washed with H<sub>2</sub>O (2 × 10 mL) and sat. aq. NaCl (10 mL). The aq. layer was back extracted with EtOAc (2 × 5 mL). The combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure to give the crude nucleoside **28** as faint yellow oil. The crude product was purified using column chromatography (EtOAc/*c*-Hex, 2/1, *v/v* → EtOAc) to afford pure nucleoside **28** (299 mg, 0.67 mmol, 30% over two steps; EtOAc, *R<sub>f</sub>* = 0.57) as a white solid.

#### Analytcs (specific assignment given whenever possible):

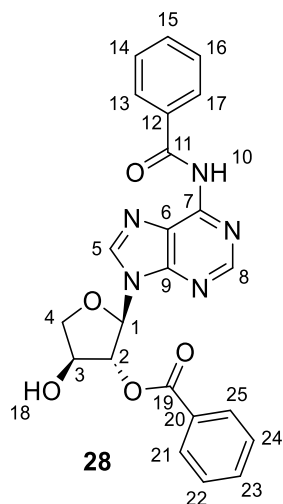
**Chemical Formula:** C<sub>23</sub>H<sub>19</sub>N<sub>5</sub>O<sub>5</sub>.

**Molecular Weight [g mol<sup>-1</sup>]:** 445.44.

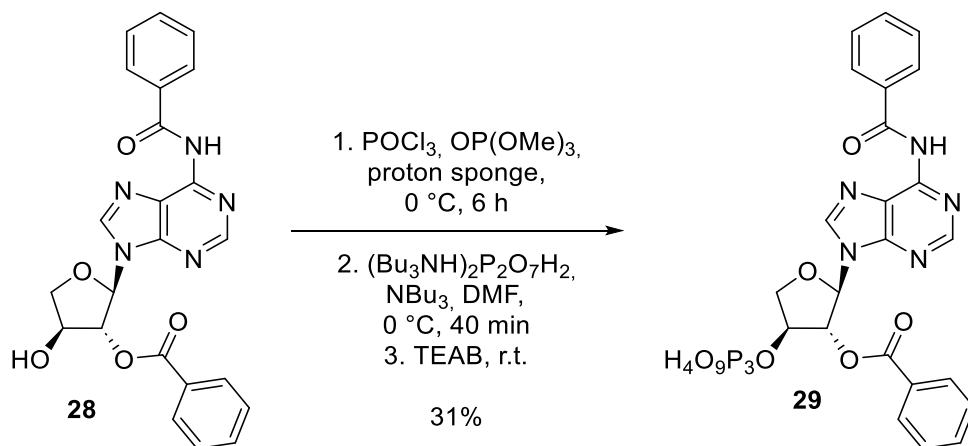
**<sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>, r.t.):**  $\delta$  [ppm] = 11.25 (s, 1H; H-10), 8.77 (s, 1H; H-5/H-8), 8.64 (s, 1H; H-8/H-5), 8.01 – 8.02 (m, 4H; H-arom.), 7.75 – 7.67 (m, 1H; H-arom.), 7.68 – 7.62 (m, 1H; H-arom.), 7.61 – 7.53 (m, 4H; H-arom.), 6.42 (d, <sup>3</sup>*J*<sub>H-1/H-2</sub> = 2.2 Hz, 1H; H-1), 6.11 (d, *J* = 3.9 Hz, 1H; H-18), 5.77 (t, <sup>3</sup>*J*<sub>H-2/H-1</sub> = 2.2 Hz, 1H; H-2), 4.63 – 4.57 (m, 1H; H-3), 4.32 – 4.28 (m, 2H; H-4).

**<sup>13</sup>C NMR (126 MHz, DMSO-*d*<sub>6</sub>, r.t.):**  $\delta$  [ppm] = 165.62 (C-11), 164.82 (C-19), 151.77 (C-9), 151.71 (C-8), 150.39 (C-7), 143.03 (C-5), 133.91 (C-arom.), 133.34 (C-6), 132.44 (C-arom.), 129.56 (C-arom.), 128.84 (C-arom.), 128.72 (C-arom.), 128.48 (C-arom.), 128.46 (C-arom.), 125.52 (C-arom.), 87.50 (C-1), 82.15 (C-2), 74.89 (C-4), 72.88 (C-3).

**HR-MS (ESI):** calculated for [M + H]<sup>+</sup> (C<sub>23</sub>H<sub>20</sub>N<sub>5</sub>O<sub>5</sub><sup>+</sup>): 446.1459; found: *m/z* = 446.1456.



### 5.2.7.13 Synthesis of *N*<sup>6</sup>-benzoyl-1-(2'-*O*-benzoyl- $\alpha$ -L-threofuranos-3'-triphosphate-1'-yl)adenine (**29**)



#### Experimental Procedure:

Triphosphate synthesis was performed according to a protocol published by Srivatsan and Tor.<sup>370</sup>

Nucleoside **28** (150 mg, 0.34 mmol, 1.0 eq.) and proton sponge (58 mg, 0.27 mmol, 0.8 eq.) were dried under high vacuum overnight, dissolved under an atmosphere of argon in trimethyl phosphate (2.31 mL, freshly distilled and stored over 3 Å molecular sieves and under argon) and cooled with an ice bath. POCl<sub>3</sub> (77  $\mu$ L, 0.84 mmol, 2.5 eq., freshly distilled and stored under argon) was added dropwise at 0 °C and the reaction was stirred under ice cold conditions for 6 h. Subsequently, tributylamine (0.84 mL, 3.5 mmol, 10.5 eq., freshly distilled and stored under argon) and a solution of bis-tri-*n*-butylammonium pyrophosphate in dry DMF (0.5 M, 3.6 mL, 1.8 mmol, 5.4 eq.) were added simultaneously in one portion under ice-cooling. The mixture was stirred at 0 °C for 40 min and quenched by the addition of TEAB buffer (1.0 M, 20 mL, pH = 8.0). Volatiles were evaporated under high vacuum at room temperature. Coevaporation of the crude product with H<sub>2</sub>O (30 mL) was performed. The crude product was lyophilized overnight before purification via HPLC (0 %  $\rightarrow$  40 % MeCN in 0.1 M aq. TEAB buffer



in 6 min, flow rate = 40.0 mL min<sup>-1</sup>) was performed. After HPLC purification, residue was again coevaporated with H<sub>2</sub>O (2 × 20 mL) and lyophilized overnight. Purified triphosphate **29** (145 mg, 0.1 mmol, 31%) was obtained as the 7 × triethylammonium salt as a faint yellow foam.

Analytcs (specific assignment given whenever possible):

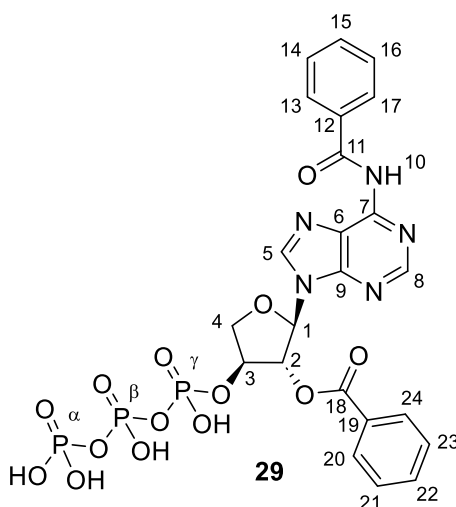
**Chemical Formula:** C<sub>23</sub>H<sub>22</sub>N<sub>5</sub>O<sub>14</sub>P<sub>3</sub>.

**Molecular Weight [g mol<sup>-1</sup>]:** 685.37.

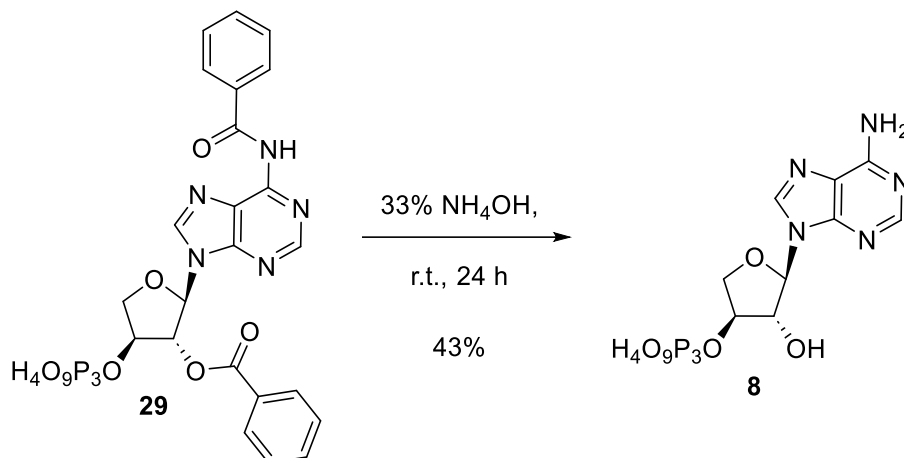
**<sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O, r.t.):** δ [ppm] = 8.74 (s, 1H; H-5/H-8), 8.67 (s, 1H; H-8/H-5), 8.16 – 8.09 (m, 2H; H-arom.), 8.07 – 8.00 (m, 2H; H-arom.), 7.77 – 7.70 (m, 1H; H-arom.), 7.69 – 7.63 (m, 1H; H-arom.), 7.61 – 7.55 (m, 4H; H-arom.), 6.50 (d, <sup>3</sup>J<sub>H-1/H-2</sub> = 1.7 Hz, 1H; H-1), 5.88 (d, <sup>3</sup>J<sub>H-2/H-1</sub> = 1.7 Hz, 1H; H-2), 5.41 – 5.33 (m, 1H; H-3), 4.78 – 4.76 (m, 1H; H-4b/H-4a), 4.55 (dt, <sup>2</sup>J<sub>H-4a/H-4b</sub> = 10.8 Hz, 4.1 Hz, 1H; H-4a/H-4b).

**<sup>31</sup>P NMR (162 MHz, D<sub>2</sub>O, r.t.):** δ [ppm] = -6.29 (d, <sup>2</sup>J<sub>P-γ/P-β</sub> = 21.1 Hz; P-γ), -12.71 (d, <sup>2</sup>J<sub>P-α/P-β</sub> = 19.1 Hz; P-α), -22.54 (t; <sup>2</sup>J<sub>P-β/P-α</sub>, <sup>2</sup>J<sub>P-β/P-γ</sub> = 20.1 Hz; P-β).

**HR-MS (ESI):** calculated for [M - H] (C<sub>23</sub>H<sub>21</sub>N<sub>5</sub>O<sub>14</sub>P<sub>3</sub>): 684.0303; found: *m/z* = 684.0308.



### 5.2.7.14 Synthesis of 1-( $\alpha$ -L-threofuranos-3'-triphosphate-1'-yl)adenine (**8**)



#### Experimental Procedure:

Debenzoylation was performed according to a protocol of Zou *et al.*<sup>145</sup>

In a sealed tube, **29** (145 mg, 0.21 mmol, 1.0 eq.) was dissolved in saturated  $\text{NH}_4\text{OH}$  (33 %, 10 mL). The reaction was stirred at r.t. for 24 h. Volatiles were evaporated under high vacuum at room temperature. The crude product was redissolved in  $\text{H}_2\text{O}$  (5 mL) lyophilized overnight before purification via HPLC (0 %  $\rightarrow$  40 % MeCN in 0.1 M aq. TEAB buffer in 6 min, flow rate = 40.0 mL  $\text{min}^{-1}$ ) was performed. After HPLC purification, the residue was coevaporated with  $\text{H}_2\text{O}$  (2  $\times$  5 mL) and lyophilized overnight. Due to minor impurities, a second HPLC purification (0 %  $\rightarrow$  15 % MeCN in 0.1 M aq. TEAB buffer in 15 min, flow rate = 40.0 mL  $\text{min}^{-1}$ ) was performed. Product containing fractions were combined and the solvent was removed *in vacuo*. The residue was again coevaporated with  $\text{H}_2\text{O}$  (2  $\times$  5 mL) and lyophilized overnight. Pure triphosphate **8** (37.5 mg, 0.043 mmol, 43%) was obtained as the 4  $\times$  triethylammonium salt as a colorless foam.

#### Analytics (specific assignments given whenever possible):

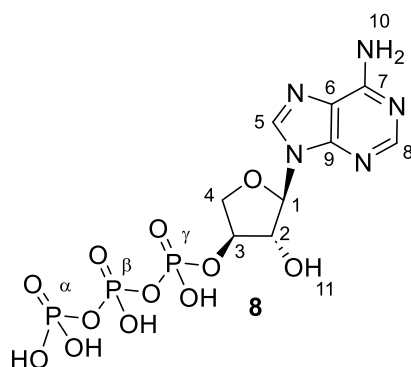
**Chemical Formula:**  $\text{C}_9\text{H}_{18}\text{N}_5\text{O}_{12}\text{P}_3$ .

**Molecular Weight [g  $\text{mol}^{-1}$ ]:** 477.16.

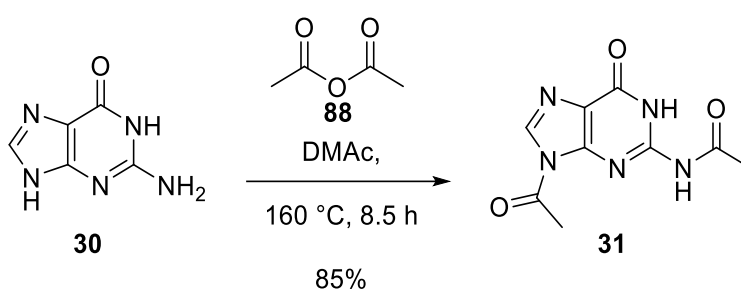
**$^1\text{H}$  NMR (500 MHz,  $\text{D}_2\text{O}$ , r.t.):**  $\delta$  [ppm] = 8.40 (s, 1H; H-5/H-8), 8.21 (s, 1H; H-8/H-5), 6.10 (d,  $J$  = 1.5 Hz, 1H; H-1), 5.01 – 4.93 (m, 1H; H-2/H-3/H-4a/H-4b), 4.93 – 4.90 (m, 1H; H-3/H-2/H-4a/H-4b), 4.48 – 4.40 (m, 2H; H-4a/H-4b/H-3/H-2).

**$^{31}\text{P}$ -NMR (202 MHz,  $\text{D}_2\text{O}$ , r.t.):**  $\delta$  [ppm] = -9.00 (d,  $^2J_{\text{P-}\gamma/\text{P-}\beta}$  = 20.5 Hz), -12.71 (d,  $^2J_{\text{P-}\alpha/\text{P-}\beta}$  = 20.5 Hz), -23.09 (t,  $^2J_{\text{P-}\beta/\text{P-}\alpha}$ ,  $^2J_{\text{P-}\beta/\text{P-}\gamma}$  = 20.5 Hz).

**HR-MS (ESI):** calculated for  $[\text{M} - \text{H}]^-$  ( $\text{C}_9\text{H}_{17}\text{N}_5\text{O}_{12}\text{P}_3$ ): 475.9779; found:  $m/z$  = 475.9768.



### 5.2.7.15 Synthesis of *N*<sup>2,9</sup>-diacetylguanine (**31**)



#### Experimental Procedure:

This reaction was carried out according to a procedure published by Zou and Robins.<sup>374</sup>

In a flame-dried, argon-flushed flask, guanine (**30**; 3 g, 0.02 mol, 1.0 eq.) was suspended in dry *N,N*-dimethylacetamide (DMAc, 25 mL, 0.27 mol, 13 eq.) and Ac<sub>2</sub>O (**89**; 5 mL, 0.05 mol, 2.5 eq.) and was stirred at 160 °C for 8.5 h. The brown solution was cooled down to r.t. overnight without stirring. A pale-yellow solid precipitated out and was filtered off and washed with EtOH (3 × 30 mL). The solid was recrystallized in ethanol (EtOH) and *N*<sup>2,9</sup>-diacetylguanine (**31**; 3.99 g, 0.017 mol, 85 %; DCM/MeOH, 9/1, *v/v*, *R*<sub>f</sub> = 0.51) was obtained as a white solid.

#### Analytcs:

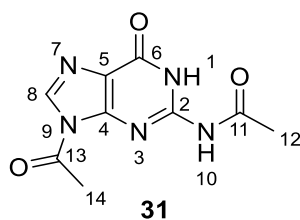
**Chemical Formula:** C<sub>9</sub>H<sub>9</sub>N<sub>5</sub>O<sub>3</sub>.

**Molecular Weight [g mol<sup>-1</sup>]:** 235.20.

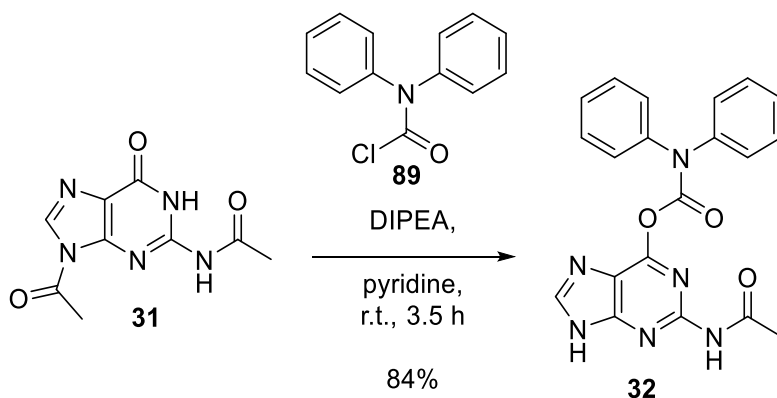
**<sup>1</sup>H NMR (499 MHz, DMSO-*d*<sub>6</sub>, r.t.):** δ [ppm] = 12.14 (br, 1H; H-10), 11.84 (br, 1H; H-1), 8.45 (s, 1H; H-8), 2.82 (s, 3H; H-14), 2.21 (s, 3H; H-12).

**<sup>13</sup>C NMR (75 MHz, DMSO-*d*<sub>6</sub>, r.t.):** δ [ppm] = 216.71 (C-6), 174.19 (C-11), 168.45 (C-13), 148.31 (C-5), 137.55 (C-8), 121.96 (C-4), 25.09 (C-14), 24.26 (C-12).

**FT-IR:** 3148 (w), 3113 (w, ν(N1-H)), 3057 (w), 3005 (w), 2951 (w), 1744 (m), 1701 (m), 1674 (s, ν(C11=O)), 1599 (s, ν(C13=O)), 1555 (s, ν(C6=O)), 1483 (m), 1396 (s), 1371 (s), 1339 (s), 1317 (m), 1229 (s), 1198 (s), 1144(m), 1123 (m), 1065 (w), 1040 (m), 1001 (m), 976 (w), 951 (m), 874 (w), 800 (s), 781 (s), 737 (s), 723 (m), 689 (s), 633 (s), 596 (m), 569 (w).



### 5.2.7.16 Synthesis of *N*<sup>2</sup>-acetyl-*O*<sup>6</sup>-diphenylcarbamoylguanine (**32**)



#### Experimental Procedure:

This reaction was carried out according to a procedure published by Zou and Robins.<sup>374</sup> In a dry, argon flushed flask, diphenylcarbamoyl chloride (**90**; 4.33 g, 18.7 mmol, 1.1 eq.) was added in one portion to a stirred suspension of *N*<sup>9</sup>-diacetylguanine (**31**; 4.00 g, 17 mmol, 1.0 eq.) in dry pyridine (80 mL) and DIPEA (5.92 mL, 34 mmol, 2.0 eq.). The reaction was stirred at ambient temperature for 3.5 h. H<sub>2</sub>O (20 mL) was added and stirring was continued for 15 min. Evaporation of the solvent was performed *in vacuo*, followed by coevaporation with toluene (3 × 25 mL). The purple, brownish residue was suspended in EtOH/H<sub>2</sub>O (1/1, *v/v*, 200 mL) and heated to 80 °C for 4 h. After cooling the reaction mixture, precipitated solid was filtered off, washed extensively with EtOH (3 × 100 mL) and dried using high vacuum. Protected guanine **32** (5.6 g, 14.3 mmol, 84%) was obtained as a purple solid.

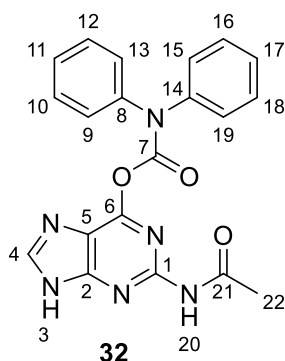
#### Analytcs:

**Chemical Formula:** C<sub>20</sub>H<sub>16</sub>N<sub>6</sub>O<sub>3</sub>.

**Molecular Weight [g mol<sup>-1</sup>]:** 388.39.

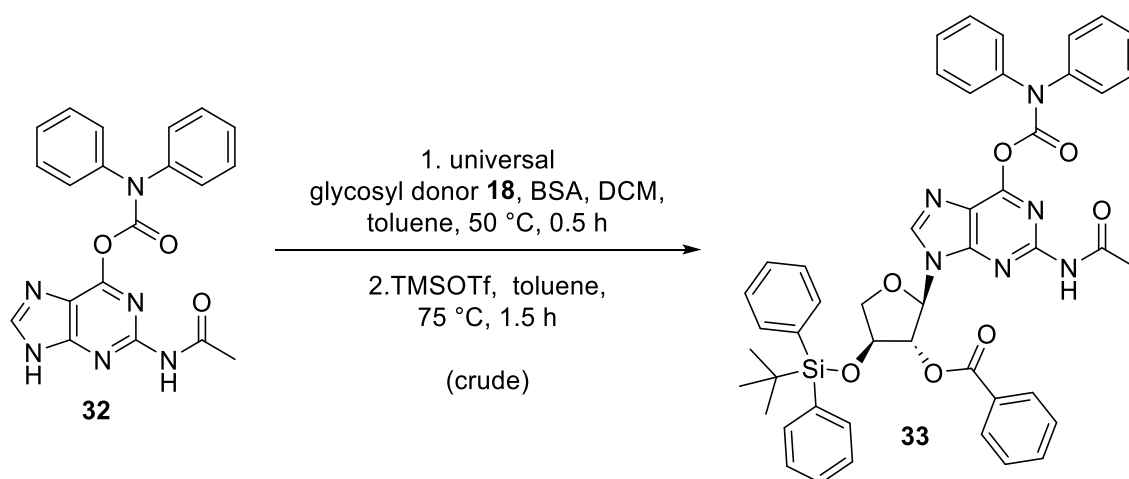
**<sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>, r.t.):** δ [ppm] = 13.56 (s, 1H; H-3), 10.61 (s, 1H; H-4), 8.44 (s, 1H; H-20), 7.55 – 7.40 (m, 8H; H-arom.), 7.32 (m, 2H; H-arom.), 2.16 (s, 3H; H-22).

**HR-MS (ESI<sup>+</sup>):** calculated for [M + H]<sup>+</sup> (C<sub>20</sub>H<sub>17</sub>N<sub>6</sub>O<sub>3</sub><sup>+</sup>): 389.1395; found: *m/z* = 389.1361.



### 5.2.7.17 Synthesis of of *N*<sup>2</sup>-acetyl-*O*<sup>6</sup>-diphenylcarbamoyl-9-(2'-*O*-benzoyl-3'-*O*-tert-butylidiphenylsilyl- $\alpha$ -L-threofuranosyl)-guanine (**33**)

#### Method 1:

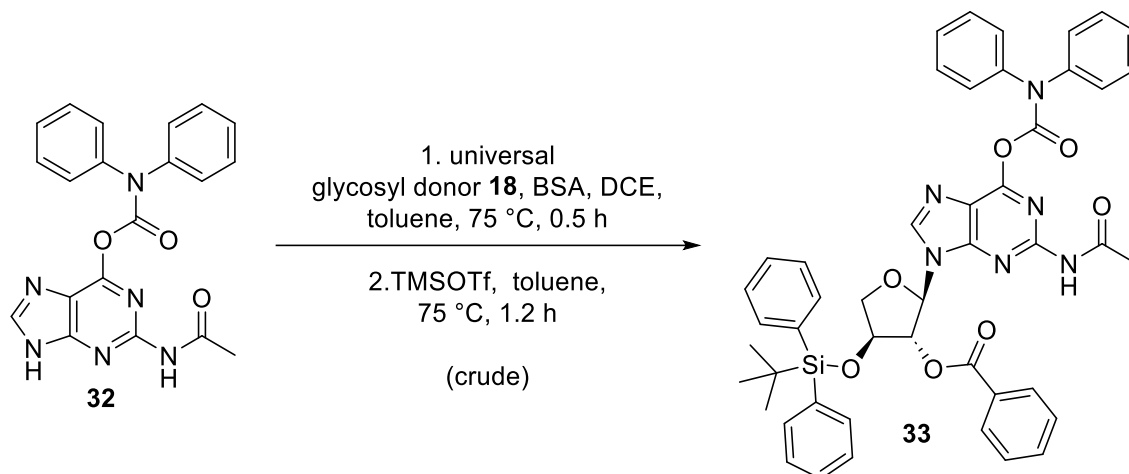


#### Experimental Procedure:

This procedure was adapted from protocols reported by Sau *et al.* and Zhang *et al.*<sup>146,367</sup> Under an atmosphere of argon, guanine **32** (668 mg, 1.71 mmol, 1.2 eq.) was dissolved in a mixture of anhydrous dichloromethane (5.3 mL) and toluene (10.6 mL). *N,O*-bis(trimethylsilyl)acetamide (1.05 mL, 4.28 mmol, 3.0 eq.) was added and the suspension was stirred for 30 min at 50 °C turning into a purple solution. The solvent was removed under reduced pressure, and the residue was dissolved in anhydrous toluene (5 mL). Universal glycosyl donor **18** (732 mg, 1.45 mmol, 1.0 eq.) in anhydrous toluene (5 mL) was added dropwise and the solution was heated to 75°C. TMSOTf (0.53 mL, 3.08 mmol, 1.8 eq.) was then added dropwise and stirring at 75°C was continued for another 1.5 h, after which TLC (EtOAc/*c*-Hex, 1/1, *v/v*) showed complete consumption of universal glycosyl donor **18**. The reaction mixture was cooled to ambient temperature, diluted with EtOAc (25 mL) and poured into sat. aq. NaHCO<sub>3</sub> solution (12 mL) while stirring, resulting in a purple suspension. The suspension was filtered, and the organic layer was separated and washed with H<sub>2</sub>O (15 mL) and sat. aq. NaCl (15 mL), dried over Na<sub>2</sub>SO<sub>4</sub>, and concentrated under reduced pressure. The

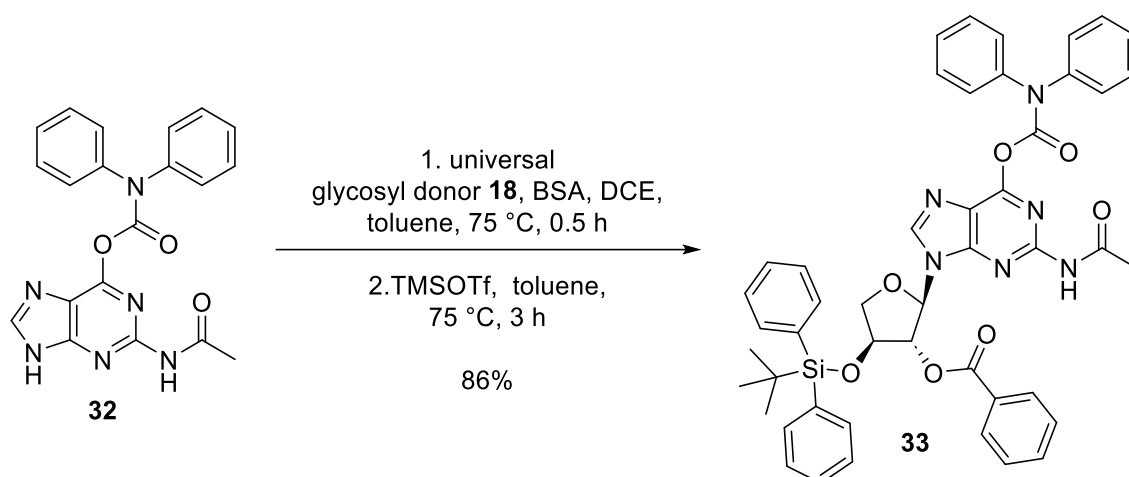
crude product **33** (991 mg, 1.19 mmol) was obtained as a purple, brownish foam and was used directly in the next step without further purification.

### Method 2:



### Experimental Procedure:

This procedure was conducted according to protocols of Sau *et al.* and Zhang *et al.*<sup>146,367</sup> Under an atmosphere of argon, guanine **32** (465 mg, 1.19 mmol, 1.2 eq.) was added to a mixture of anhydrous 1,2-dichloroethane (3.5 mL) and toluene (7 mL). *N,O*-bis(trimethylsilyl)acetamide (0.73 mL, 2.98 mmol, 3.0 eq.) was added and the suspension was stirred for 30 min at 75°C turning into a purple solution. The solvent was removed under reduced pressure, and the residue was dissolved in anhydrous toluene (5 mL). Universal glycosyl donor **18** (510 mg, 1.01 mmol, 1.0 eq.) in anhydrous toluene (5 mL) was added dropwise and the solution was heated to 75°C. TMSOTf (0.37 mL, 2.14 mmol, 2.1 eq.) was then added dropwise and stirring at 75 °C was continued for another 1.2 h, after which TLC (EtOAc/*c*-Hex, 1/1, *v/v*) showed complete consumption of the universal glycosyl donor **18**. The reaction mixture was cooled to ambient temperature, diluted with EtOAc (25 mL) and poured into sat. aq. NaHCO<sub>3</sub> solution (10 mL) while stirring, resulting in a purple suspension. The suspension was filtered, and the organic layer was separated and washed with H<sub>2</sub>O (10 mL) and sat. aq. NaCl (2 × 10 mL), dried over Na<sub>2</sub>SO<sub>4</sub>, and concentrated under reduced pressure. The crude product **33** (916 mg, 1.10 mmol) was obtained as a purple, brownish foam and was used directly in the next step without further purification.

**Method 3:****Experimental Procedure:**

This procedure was conducted according to protocols of Sau *et al.* and Zhang *et al.*<sup>146,367</sup>

Under an atmosphere of argon, guanine **32** (593 mg, 1.53 mmol, 1.2 eq.) was dissolved in a mixture of anhydrous 1,2-dichloroethane (10 mL) and toluene (5 mL). *N,O*-bis(trimethylsilyl)acetamide (1.05 mL, 4.28 mmol, 3.0 eq.) was added dropwise and the suspension was stirred for 30 min at 75 °C turning into a purple solution. The solvent was removed under reduced pressure, and the residue was dissolved in anhydrous toluene (8 mL). universal glycosyl donor **18** (656 mg, 1.30 mmol, 1.0 eq.) in anhydrous toluene (8 mL) was added dropwise and the solution was heated to 75°C. TMSOTf (0.47 mL, 2.75 mmol, 2.1 eq.) was then added dropwise and stirring at 75 °C was continued for 3 h, after which TLC (EtOAc/*c*-Hex, 1/1, *v/v*) showed complete consumption of universal glycosyl donor **18**. The reaction mixture was cooled to ambient temperature, diluted with EtOAc (40 mL) and poured into sat. aq. NaHCO<sub>3</sub> solution (50 mL) while stirring, resulting in a purple suspension. The suspension was filtered, and the organic layer was separated and washed with H<sub>2</sub>O (60 mL) and sat. aq. NaCl (60 mL), dried over MgSO<sub>4</sub>, and concentrated under reduced pressure. The crude product was purified using column chromatography (EtOAc/*c*-Hex, 1/2, *v/v*) to afford pure **33** (1.08 mg, 1.32 mmol, 86%) as a colorless foam.

**Analytcs:**

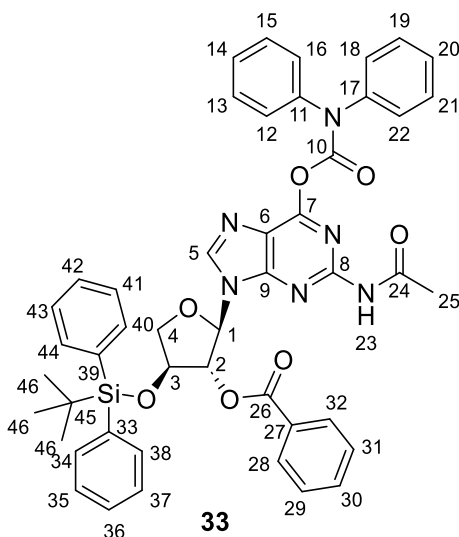
**Chemical Formula:** C<sub>47</sub>H<sub>44</sub>N<sub>6</sub>O<sub>7</sub>Si.

**Molecular Weight [g mol<sup>-1</sup>]:** 832.99.

**<sup>1</sup>H NMR (500 MHz, DMF-*d*<sub>7</sub>, r.t.):** δ [ppm] = 10.70 (s, 1H; H-23), 8.69 (s, 1H; H-5), 7.93 (d, *J* = 7.3 Hz, 2H; H-arom.), 7.71 (dd, *J* = 9.1, 7.8 Hz, 3H; H-arom), 7.65 – 7.41 (m, 16H; H-arom.), 7.35 (t, *J* = 7.3 Hz, 4H; H-arom.), 6.42 (d, <sup>3</sup>*J*<sub>H-1/H-2</sub> = 2.9 Hz, 1H; H-1), 6.28 (t, <sup>3</sup>*J*<sub>H-2/H-1</sub> = 2.9 Hz, 1H; H-2), 4.86 (d, *J* = 4.0 Hz, 1H; H-3), 4.64 (dd, <sup>2</sup>*J*<sub>H-4b/H-4a</sub> = 9.2 Hz, 4.8 Hz, 1H; H-4b/H-4a), 4.24 (dd, <sup>2</sup>*J*<sub>H-4a/H-4b</sub> = 9.2 Hz, 5.5 Hz, 1H; H-4a/H-4b), 2.35 (s, 3H; H-25), 1.00 (s, 9H; H-46).

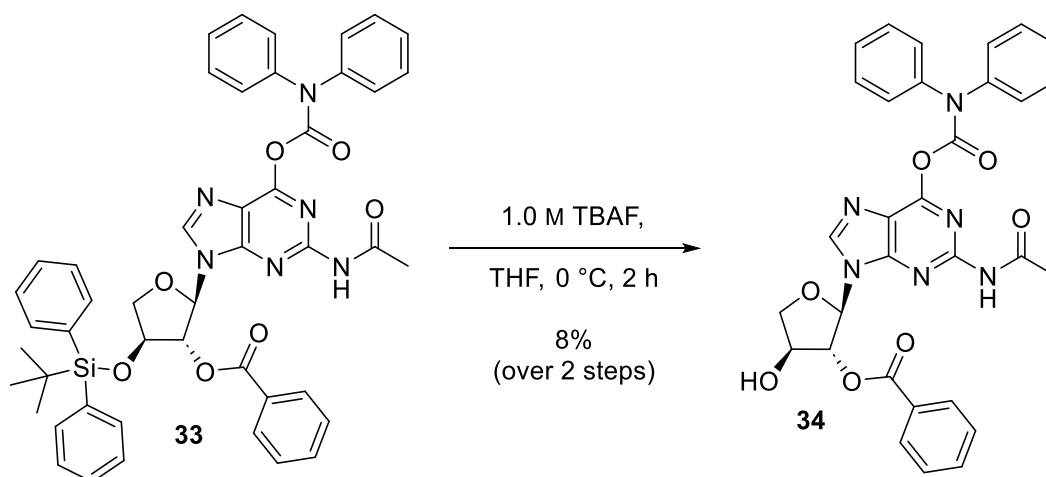
**$^{13}\text{C}$  NMR (126 MHz, DMF- $d_7$ , r.t.):**  $\delta$  [ppm] = 170.27 (C-24), 166.14 (C-26), 157.19 (C-10), 155.62 (C-9), 154.05 (C-8), 151.55 (C-7), 143.32 (C-arom.), 136.83 (C-5), 136.70 (C-arom.), 134.94 (C-arom), 133.79 (C-arom), 133.58 (C-arom), 131.39 (C-arom.), 131.23 (C-arom.), 130.74 (C-arom.), 130.49 (C-arom.), 130.04 (C-arom), 129.85 (C-arom.), 129.16 (C-arom.), 129.00 (C-arom.), 121.81 (C-6), 89.33 (C-1), 82.77 (C-2), 76.80 (C-3), 75.42 (C-4), 27.74 (C-45), 27.46 (C-46), 25.63 (C-25).

**HR-MS (ESI<sup>+</sup>):** calculated for  $[\text{M} + \text{H}]^+$  ( $\text{C}_{47}\text{H}_{45}\text{N}_6\text{O}_7\text{Si}$ ): 833.3114; found:  $m/z$  = 833.3125.



### 5.2.7.18 Synthesis of *N*<sup>2</sup>-acetyl-*O*<sup>6</sup>-diphenylcarbamoyl-9-(2'-*O*-benzoyl- $\alpha$ -L-threofuranosyl)guanine (**34**)

#### Method 1:



#### Experimental Procedure:

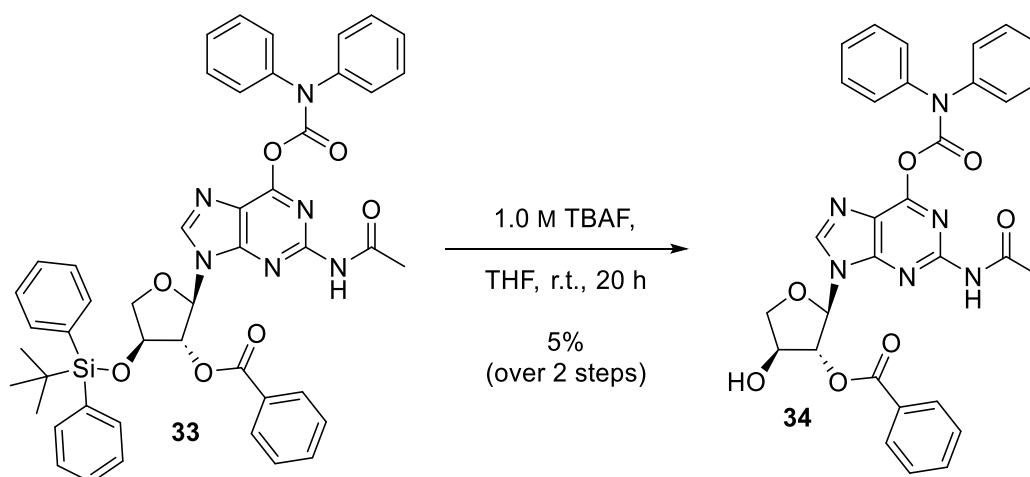
TBDPS deprotection was performed according to Sau *et al.*<sup>367</sup>

Under an atmosphere of argon, tetrabutylammonium fluoride in THF (1.0 M, 1.4 mL, 1.4 mmol) was added dropwise to a stirred solution of nucleoside **33** (Method 1; 991 mg) dissolved in THF (5 mL) at 0 °C. The reaction mixture was stirred for 2 h at 0 °C and the solvent was



subsequently removed *in vacuo* and the residue was dissolved in EtOAc (15 mL). The organic layer was sequentially washed with HCl (1.0 M, 10 mL), H<sub>2</sub>O (10 mL) and sat. aq. NaCl (10 mL), dried over Na<sub>2</sub>SO<sub>4</sub>, and concentrated under reduced pressure to give the crude nucleoside **34** as faint brown oil. The crude product was purified using column chromatography (EtOAc/*c*-Hex, 2/1, *v/v*) to afford pure nucleoside **34** (64 mg, 0.1 mmol, 8% over two steps; EtOAc/*c*-Hex, 2/1, *v/v*, *R*<sub>f</sub> = 0.58) as a white solid.

### Method 2:



### Experimental Procedure:

TBDPS deprotection was performed adapting a protocol by Yang *et al.*<sup>415</sup>

Under an atmosphere of argon, tetrabutylammonium fluoride in THF (1.0 M, 4.4 mL, 4.4 mmol) was added dropwise to a stirred solution of crude nucleoside **33** (Method 2; 916 mg) dissolved in THF (5 mL). The reaction mixture was stirred for 20 h at ambient temperatures and the solvent was subsequently removed *in vacuo* and the residue was dissolved in EtOAc (15 mL). The organic layer was sequentially washed with HCl (1.0 M, 10 mL), H<sub>2</sub>O (10 mL) and sat. aq. NaCl (10 mL), dried over Na<sub>2</sub>SO<sub>4</sub>, and concentrated under reduced pressure to give the crude nucleoside **34** as faint brown oil. The crude product was purified using column chromatography (EtOAc/*c*-Hex, 2/1, *v/v* → EtOAc → MeOH) to afford pure nucleoside **34** (35 mg, 0.06 mmol, 5% over two steps; EtOAc/*c*-Hex, 2/1, *v/v*, *R*<sub>f</sub> = 0.58) as a white solid.

### Analytcs (specific assignment given whenever possible):

**Chemical Formula:** C<sub>31</sub>H<sub>26</sub>N<sub>6</sub>O<sub>7</sub>.

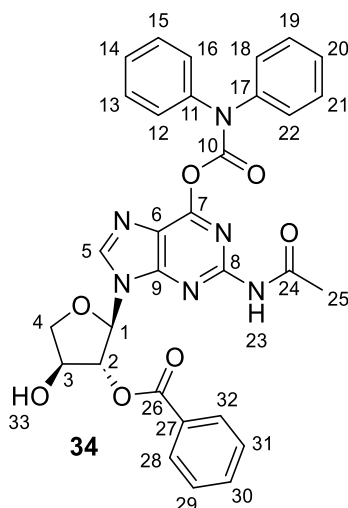
**Molecular Weight [g mol<sup>-1</sup>]:** 594.58.

**<sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD, r.t.):** δ [ppm] = 8.55 (s, 1H; H-5), 8.04 – 7.97 (m, 2H; H-23, H-arom.), 7.63 – 7.57 (m, 1H; H-arom.), 7.53 – 7.42 (m, 6H; H-arom.), 7.37 (m, 5H; H-arom.), 7.26 (m, 2H; H-arom.), 6.38 (d, <sup>3</sup>J<sub>H-1/H-2</sub> = 1.7 Hz, 1H; H-1), 5.69 (t, <sup>3</sup>J<sub>H-2/H-1</sub> = 1.7 Hz, <sup>3</sup>J<sub>H-2/H-3</sub> = 1.7 Hz, 1H; H-2), 4.59 (ddd, <sup>3</sup>J<sub>H-3/H-4b</sub> = 4.6 Hz, <sup>3</sup>J<sub>H-3/H-4a</sub> = 2.7, <sup>3</sup>J<sub>H-3/H-2</sub> = 1.7 Hz, 1H; H-3), 4.37

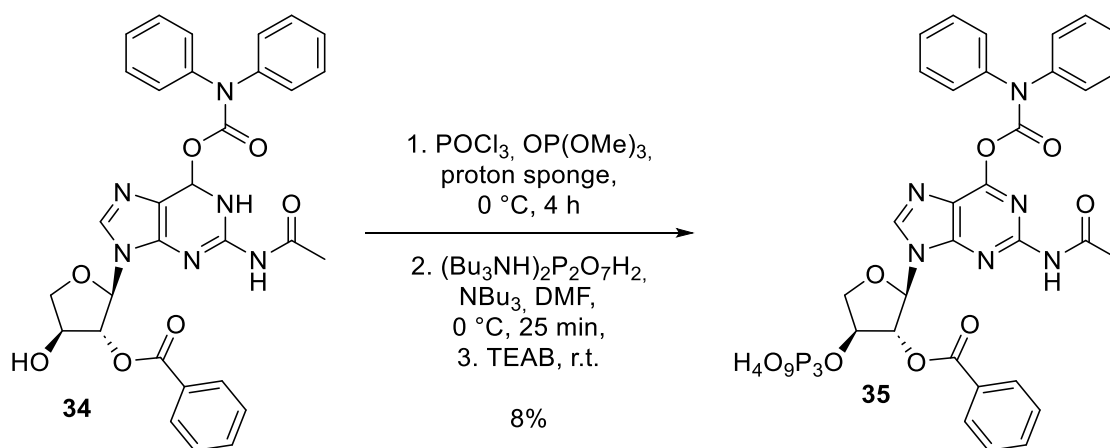
(dd,  $^2J_{H-4a/H-4b} = 9.8$  Hz,  $^3J_{H-4a/H-3} = 2.7$  Hz, 1H; H-4a/H-4b), 4.31 (dd,  $^2J_{H-4b/H-4a} = 9.8$  Hz,  $^3J_{H-4b/H-3} = 4.5$  Hz, 1H; H-4b/H-4a), 2.23 (s, 3H; H-25).

$^{13}\text{C}$  NMR (126 MHz,  $\text{CD}_3\text{OD}$ , r.t.): 172.92 (C-24), 166.80 (C-26), 156.93 (C-10), 155.56 (C-9), 153.73 (C-8), 152.10 (C-7), 145.76 (C-5), 143.17 (C-arom.), 134.81 (C-arom.), 130.78 (C-arom.), 130.28 (C-arom.), 130.23 (C-arom.), 129.68 (C-arom.), 128.54 (C-arom.), 128.30 (C-arom.), 121.79 (C-6), 90.25 (C-1), 84.24 (C-2), 76.69 (C-4), 75.27 (C-3), (C-25).

HR-MS (ESI<sup>+</sup>): calculated for  $[\text{M} + \text{H}]^+$  ( $\text{C}_{31}\text{H}_{27}\text{N}_6\text{O}_7^+$ ): 595.1936; found:  $m/z = 595.1935$ .



### 5.2.7.19 Synthesis of *N*<sup>2</sup>-acetyl-*O*<sup>6</sup>-diphenylcarbamoyl-9-(2'-*O*-benzoyl- $\alpha$ -L-threofuranos-3'-triphosphate-1'-yl)guanine (35)



#### Experimental Procedure:

Triphosphate synthesis was performed according to a protocol published by Srivatsan and Tor.<sup>370</sup>

Nucleoside **34** (57 mg, 0.095 mmol, 1.0 eq.) and proton sponge (16 mg, 0.076 mmol, 0.8 eq.) were dried under high vacuum overnight, dissolved under an atmosphere of argon in trimethylphosphate (0.65 mL, freshly distilled and stored over 3 Å molecular sieves and under argon) and cooled with an ice bath.  $\text{POCl}_3$  (22  $\mu\text{L}$ , 0.24 mmol, 2.5 eq., freshly distilled and

stored under argon) was added dropwise at 0 °C and the reaction was stirred under ice cold conditions for 4 h. Subsequently, tributylamine (0.24 mL, 1.0 mmol, 10.5 eq., freshly distilled and stored under argon) and a solution of bis-tri-*n*-butylammonium pyrophosphate in dry DMF (0.5 M, 1.04 mL, 0.52 mmol, 5.4 eq.) were added simultaneously in one portion under ice-cooling. The mixture was stirred at 0 °C for 25 min and quenched by the addition of TEAB buffer (1.0 M, 13 mL, pH = 8.0). Volatiles were evaporated under high vacuum at room temperature. Coevaporation of the crude product with H<sub>2</sub>O (3 × 10 mL) was performed. The crude product was lyophilized overnight before purification via HPLC (15 % → 40 % MeCN in 0.1 M aq. TEAB buffer in 15 min, flow rate = 40.0 mL min<sup>-1</sup>) was performed. After HPLC purification, the residue was again coevaporated with H<sub>2</sub>O (3 × 10 mL) and lyophilized overnight. Purified triphosphate **35** (9 mg, 0.0079 mmol, 8%) was obtained as the 3 × triethylammonium salt as faint redish foam.

Analytcs (specific assignment given whenever possible):

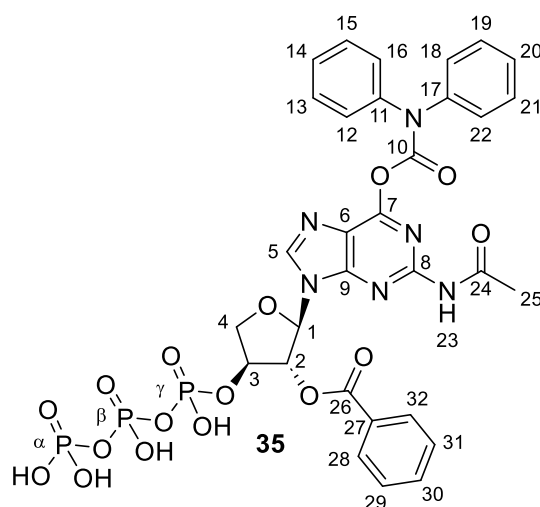
**Chemical Formula:** C<sub>31</sub>H<sub>29</sub>N<sub>6</sub>O<sub>16</sub>P<sub>3</sub>.

**Molecular Weight [g mol<sup>-1</sup>]:** 834.52.

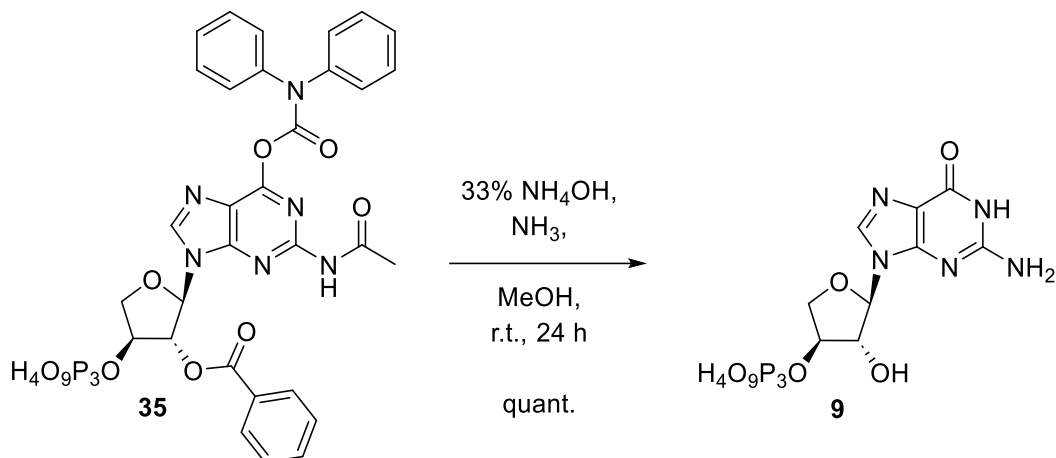
**<sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O, r.t.):** δ [ppm] = 8.58 (s, 1H; H-5), 8.03 – 7.90 (m, 2H; H-arom.), 7.64 – 7.56 (m, 1H; H-arom.), 7.54 – 7.25 (m, 12H; H-arom.), 6.39 (d, <sup>3</sup>J<sub>H-1/H-2</sub> = 2.0 Hz, 1H; H-1), 5.80 (d, <sup>3</sup>J<sub>H-2/H-1</sub> = 2.0 Hz, 1H; H-2), 5.26 – 5.19 (m, 1H; H-3), 4.64 (d, <sup>2</sup>J<sub>H-4a/H-4b</sub> = 10.7 Hz, 1H; H-4a/H-4b), 4.44 (dd, <sup>2</sup>J<sub>H-4b/H-4a</sub> = 10.7 Hz, <sup>3</sup>J<sub>H-4b/H-3</sub> = 4.0 Hz, 1H; H-4b/H-4a), 2.16 (s, 3H; H-25).

**<sup>31</sup>P NMR (162 MHz, D<sub>2</sub>O, r.t.):** δ [ppm] = -10.90 (d, <sup>2</sup>J<sub>P-γ/P-β</sub> = 19.7 Hz; P-γ), -12.86 (d, <sup>2</sup>J<sub>P-α/P-β</sub> = 19.7 Hz; P-α), -23.28 (t, <sup>2</sup>J<sub>P-β/P-α</sub>, <sup>2</sup>J<sub>P-β/P-γ</sub> = 19.7 Hz; P-β).

**HR-MS (ESI):** calculated for [M - H]<sup>-</sup> (C<sub>31</sub>H<sub>28</sub>N<sub>6</sub>O<sub>16</sub>P<sub>3</sub>): 833.0780; found: *m/z* = 833.0786.



### 5.2.7.20 Synthesis of 1-( $\alpha$ -L-threofuranos-3'-triphosphate-1'-yl)guanine (**9**)



#### Experimental Procedure:

Deprotection was performed according to a protocol of Zou *et al.*<sup>145</sup>

In a sealed tube, **35** (9 mg, 0.01 mmol, 1.0 eq.) was dissolved in saturated  $\text{NH}_4\text{OH}$  (33 %, 2 mL) and ammonia in MeOH (7.0 N, 0.2 mL). The reaction was stirred at r.t. for 24 h. Volatiles were evaporated under high vacuum at 25 °C. The crude product was redissolved in  $\text{H}_2\text{O}$  (5 mL) lyophilized overnight before purification via HPLC (0 %  $\rightarrow$  40 % MeCN in 0.1 M aq. TEAB buffer in 15 min, flow rate = 40.0 mL  $\text{min}^{-1}$ ) was performed. After HPLC purification, the residue was coevaporated with  $\text{H}_2\text{O}$  (3  $\times$  5 mL) and lyophilized overnight. Pure **9** (8 mg, 0.01 mmol, quant.) was obtained as the 3  $\times$  triethylammonium salt as a white solid.

#### Analytics (specific assignments given whenever possible):

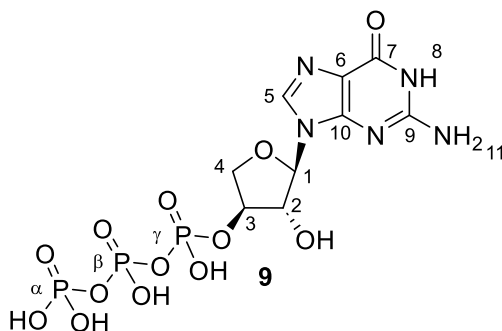
**Chemical Formula:**  $\text{C}_9\text{H}_{14}\text{N}_5\text{O}_{13}\text{P}_3$ .

**Molecular Weight [g  $\text{mol}^{-1}$ ]:** 493.15.

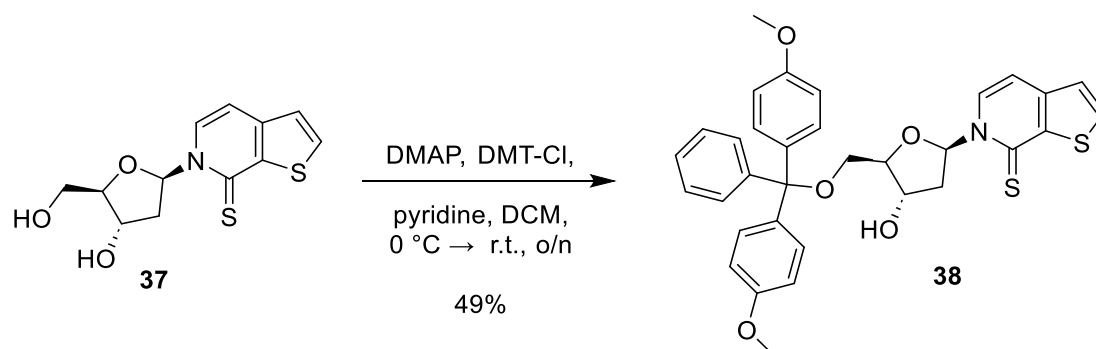
**$^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}$ , r.t.):**  $\delta$  [ppm] = 8.06 (s, 1H; H-5), 5.94 (d,  $J$  = 2.1 Hz, 1H; H-1), 4.99 – 4.87 (m, 2H; H-2, H-3), 4.49 – 4.35 (m, 2H; H-4a, H-4b).

**$^{31}\text{P}$ -NMR (162 MHz,  $\text{D}_2\text{O}$ , r.t.):**  $\delta$  [ppm] = -9.90 (s; P- $\gamma$ ), -12.61 (d,  $^2J_{\text{P-}\alpha/\text{P-}\beta}$  = 20.3 Hz; P- $\alpha$ ), -23.17 (t,  $^2J_{\text{P-}\beta/\text{P-}\alpha}$ ,  $^2J_{\text{P-}\beta/\text{P-}\gamma}$  = 20.2 Hz; P- $\beta$ ).

**HR-MS (ESI):** calculated for  $[\text{M} - \text{H}]^-$  ( $\text{C}_9\text{H}_{13}\text{N}_5\text{O}_{13}\text{P}_3$ ): 491.9728; found:  $m/z$  = 491.9728.



### 5.2.7.21 Synthesis of 6-{5-[bis-(4-methoxy-phenyl)-phenyl-methoxymethyl]-4-hydroxy-tetrahydro-furan-2-yl}-6H-thieno[2,3-c]pyridine-7-thione (**38**)



#### Experimental Procedure:

The procedure for the preparation of DMT-protected nucleoside **38** was adapted from Serebryany and Beigelman.<sup>416</sup>

Under an atmosphere of argon, dTPT3 nucleoside **37** (466 mg, 1.64 mmol, 1.0 eq.) and DMAP (181 mg, 1.48 mmol, 0.9 eq.) were dissolved in anhydrous pyridine (12 mL). In an argon flushed tube, 4,4'-dimethoxytrityl chloride (DMT-Cl, 1226 mg, 3.62 mmol, 2.2 eq.) was dissolved in anhydrous dichloromethane (2 mL) and added dropwise to the previously prepared solution at 0°C. Stirring of the reaction mixture was continued for 30 min. Subsequently, the reaction was warmed to ambient temperatures and stirred overnight. The solvent was removed *in vacuo* and the residue was purified by column chromatography (DCM/MeOH, 95/5). The product **38** (471 mg, 0.8 mmol, 49%, DCM/MeOH, 95/5, *v/v*,  $R_f = 0.37$ ) was obtained as a yellow foam.

#### Analytics (specific assignment given whenever possible):

**Chemical Formula:** C<sub>33</sub>H<sub>31</sub>NO<sub>5</sub>S<sub>2</sub>.

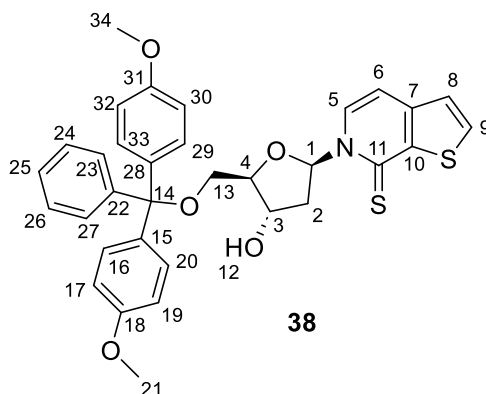
**Molecular Weight [g mol<sup>-1</sup>]:** 585.73.

**<sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>, r.t.):** δ [ppm] = 8.40 (d, <sup>3</sup>J<sub>H-5/H-6</sub> = 7.2 Hz, 1H; H-5), 7.77 (d, <sup>3</sup>J<sub>H-9/H-8</sub> = 5.2 Hz, 1H; H-9), 7.47 – 7.43 (m, 2H; H-arom.), 7.36 – 7.33 (m, 3H; H-arom.), 7.31 – 7.27 (m, 3H; H-1, H-arom.), 7.25 – 7.24 (m, 1H; H-arom.), 7.22 (d, <sup>3</sup>J<sub>H-8/H-9</sub> = 5.3 Hz, 1H; H-8), 7.18 – 7.16 (m, 1H; H-arom.), 6.86 – 6.83 (m, 4H; H-arom.), 6.81 (d, <sup>3</sup>J<sub>H-6/H-5</sub> = 7.2 Hz, 1H; H-6), 4.60 (q, <sup>3</sup>J<sub>H-3/H-2</sub> = 6.4 Hz, 1H; H-3), 4.16 – 4.10 (m, 1H; H-4), 3.79 (s, 6H; H-21, H-34), 3.65 (dd, <sup>2</sup>J<sub>H-13b/H-13a</sub> = 10.8 Hz, 3.1 Hz, 1H; H-13b/H-13a), 3.50 (dd, <sup>2</sup>J<sub>H-13a/H-13b</sub> = 10.8 Hz, <sup>3</sup>J<sub>H-13a/H-4</sub> = 3.3 Hz, 1H; H-13a/H-13b), 2.83 (m, 1H; H-2b/H-2a), 2.35 (ddd, <sup>2</sup>J<sub>H-2a/H-2b</sub> = 14.0 Hz, <sup>3</sup>J<sub>H-2a/H-3</sub> = 6.5 Hz, <sup>3</sup>J<sub>H-2a/H-1</sub> = 4.8 Hz, 1H; H-2a/H-2b).

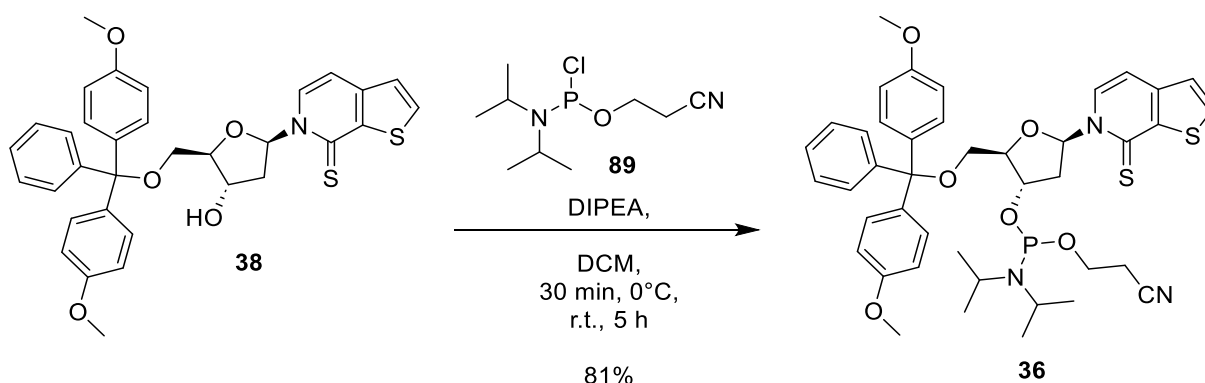
**<sup>13</sup>C-NMR (151 MHz, CDCl<sub>3</sub>, r.t.):** δ [ppm] = 172.34 (C-11), 158.78 (C-18, C-31.), 145.54 (C-10), 144.58 (C-arom.), 139.32 (C-7), 137.53 (C-9), 135.65 (C-28/C-15), 135.60 (C-28/C-15), 130.27 (C-5), 129.26 (C-arom.), 128.32 (C-arom.), 128.13 (C-arom.), 127.97 (C-arom.), 127.89

(C-arom.), 127.20 (C-arom.), 124.35 (C-8), 113.29 (C-arom.), 109.00 (C-6), 90.65 (C-1), 87.11 (C-14), 86.31 (C-4), 70.54 (C-3), 62.35 (C-13), 55.40 (C-34/C-21), 55.38 (C-21/C-34), 41.97 (C-2).

**HR-MS (ESI<sup>+</sup>):** calculated for [M + Na]<sup>+</sup> (C<sub>33</sub>H<sub>31</sub>NNaO<sub>5</sub>S<sub>2</sub><sup>+</sup>): 608.15359; found: *m/z* = 608.15389.



### 5.2.7.22 Synthesis of diisopropyl-phosphoramidous acid 2-[(bis-(4-methoxyphenyl)(phenyl)methoxy)methyl]-5-(7-thioxo-7H-thieno[2,3-c]pyridin-6-yl)-tetrahydro-furan-3-yl ester 2-cyano-ethyl ester (36)



#### Experimental Procedure:

The following procedure is based on previously reported syntheses of other nucleoside phosphoramidites with unnatural nucleobases by Leconte *et al.* and Matsude *et al.*<sup>206,207</sup>

In a dry, argon flushed schlenk flask, nucleoside **38** (465 mg, 0.8 mmol, 1.0 eq.) was dissolved in anhydrous DCM (4 mL) and *N,N*-diisopropylethylamine (1.39 mL, 8.0 mmol, 10 eq.) was added. The reaction mixture was degassed by two consecutive freeze-pump-thaw cycles. Bis(2-cyanoethyl)-*N,N*-diisopropylchlorophosphoramidite (**89**; 0.447 mL, 2.0 mmol, 2.5 eq.) was added and the reaction was stirred for 30 min at 0 °C. Subsequently, the reaction was warmed to ambient temperature and stirring was continued for 5 h. The solvent was removed under reduced pressure and the residue was filtered through a short pad of silica

(EtOAc/c Hex, 10/1, v/v). The product **36** (507 mg, 0.65 mmol, 81%) was obtained as an off-white foam.

Analytics (specific assignment given whenever possible):

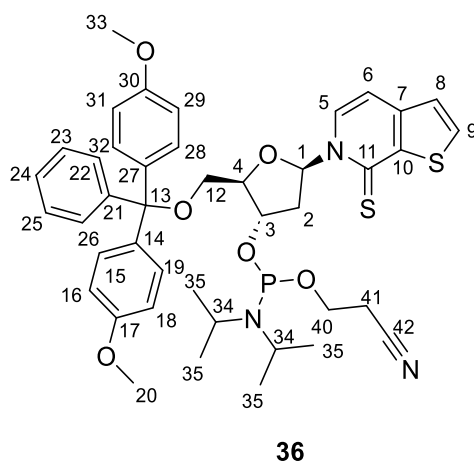
**Chemical Formula:** C<sub>42</sub>H<sub>48</sub>N<sub>3</sub>O<sub>6</sub>P<sub>2</sub>S.

**Molecular Weight [g mol<sup>-1</sup>]:** 758.95.

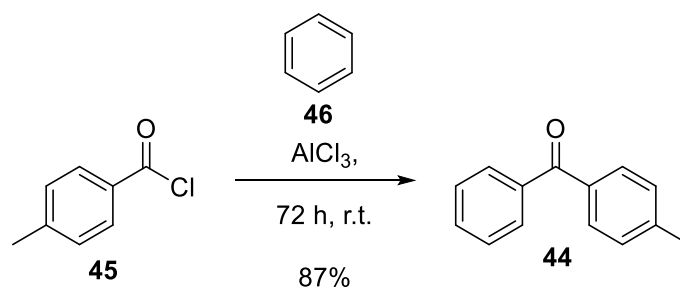
**<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>, r.t.):** δ [ppm] = 8.49 (d, <sup>3</sup>J<sub>H-5/H-6</sub> = 7.2 Hz, 0.5 H; H-5), 8.41 (d, <sup>3</sup>J<sub>H-5/H-6</sub> = 7.2 Hz, 0.5 H; H-5'), 7.78 (m, 1H; H-9), 7.51 – 7.40 (m, 2H; H-arom.), 7.40 – 7.27 (m, 7H; H-1; H-arom.), 7.22 (d, <sup>3</sup>J<sub>H-8/H-9</sub> = 5.3 Hz, 1H; H-8), 6.88 – 6.82 (m, 4H; H-arom.), 6.77 – 6.7 (m, 1H, H-6), 4.84 – 4.74 (m, 0.5H; H-3), 4.75 – 4.64 (m, 0.5H; H-3'), 4.27 – 4.20 (m, 1H; H-4), 3.86 – 3.74 (m, 7H; H-20, H-33, H-40), 3.71 – 3.42 (m, 4H; H-34, H-40, H12a, H-12b), 3.03 – 2.85 (m, 1H; H-2b/H-2a), 2.62 (t, <sup>3</sup>J<sub>H-41/H-40</sub> = 6.6 Hz, 1H; H-41b/ H-41a), 2.46 – 2.29 (m, 2H; H-2a/H-2b, H-41a/ H-41b), 1.17 (m, 9H; H-35), 1.05 (d, <sup>3</sup>J<sub>H-35/H-34</sub> = 6.7 Hz, 3H; H-35).

**<sup>31</sup>P-NMR (162 MHz, CDCl<sub>3</sub>, r.t.):** δ [ppm] = 149.53 , 148.70.

**Annotation:** The phosphoramidite **36** contains a chiral phosphorous (III) center and therefore exists as a mixture of two diastereomers resulting in splitted proton signals. The ratio of the two diastereomers determined by <sup>31</sup>P NMR spectrum amounts approximately 1:1. Single protons for which the splitting of a proton signal is observed are marked as 0.5 H-X and 0.5 H-X'.



### 5.2.7.23 Synthesis of 4-methylbenzophenone (**44**)



Experimental Procedure:

Synthesis of benzophenone **44** was performed according to a protocol of Gobbi *et al.*<sup>382</sup>

$\text{AlCl}_3$  (3.0 g, 22.5 mmol, 1.5 eq.) was suspended in benzene (**46**, 8 mL). The suspension was cooled to 0 °C and 4-methylbenzoylchloride (**45**, 1.98 mL, 15.0 mmol, 1.0 eq.) was added dropwise. The reaction mixture was stirred for 3 days at ambient temperature. After diluting the suspension with DCM (8 mL), ice was added and stirring was continued until the ice was thawed. The two layers were separated and the organic layer was filtered. The filtrate was washed with aq. sat.  $\text{NaHCO}_3$  solution (3 × 100 mL) and dried over  $\text{Na}_2\text{SO}_4$ . The solvent was removed *in vacuo* and **44** (2.57 g, 13.0 mmol, 87%) was obtained as a light yellow solid.

Analytics (specific assignment given whenever possible):

**Chemical Formula:**  $\text{C}_{14}\text{H}_{12}\text{O}$ .

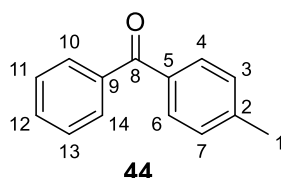
**Molecular Weight [g mol<sup>-1</sup>]:** 196.25.

**<sup>1</sup>H NMR (300 MHz,  $\text{CDCl}_3$ , r.t.):**  $\delta$  [ppm] 7.78 (m, 2H; H-10, H-14), 7.75 – 7.69 (m, 2H; H-4, H-6), 7.61 – 7.52 (m, 1H; H-12), 7.51 – 7.42 (m, 2H; H-11, H-13), 7.27 (d,  $J = 8.2$  Hz, 2H; H-3, H-7), 2.43 (s, 3H, H-1).

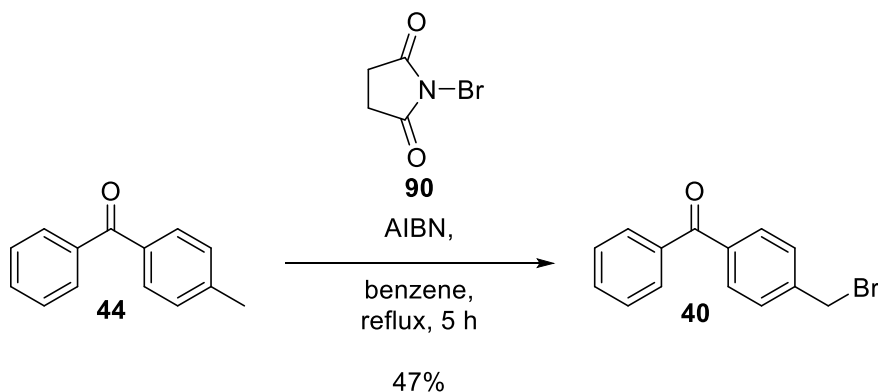
**<sup>13</sup>C NMR (75 MHz,  $\text{CDCl}_3$ , r.t.):**  $\delta$  [ppm] = 196.63 (C-8), 143.35 (C-9), 138.07 (C-5), 135.00 (C-2), 132.27 (C-12), 130.42 (C-10, C-14), 130.04 (C-4, C-6), 129.09 (C-11, C-13), 128.32 (C-3, C-7), 21.77 (C-1).

**GC-MS (EI<sup>+</sup>):** calculated for  $[\text{M}^+]$  ( $\text{C}_{14}\text{H}_{12}\text{O}$ ): 197.26; found:  $m/z = 196.1$ .

**FT-IR:** 3036 (w), 2913 (w), 1935 (w), 1825 (w), 1645 (s,  $\nu(\text{C}=\text{O})$ ), 1601 (m), 1595 (m), 1576 (w), 1449 (w), 1408 (w), 1375 (w), 1319 (m), 1275 (s,  $\delta(\text{-CH}_3)$ ), 1250 (m), 1213 (w), 1180 (m), 1150 (w), 1119 (w), 1072 (w), 1034 (w), 976 (w), 962 (w), 935 (m), 918 (m), 851 (w), 839 (m), 789 (m), 729 (s,  $\delta(\text{C-H})\text{o.o.p.}$ ), 700 (s,  $\delta(\text{C-H})\text{o.o.p.}$ ), 679 (m), 637 (w), 617 (w), 600 (s).



### 5.2.7.24 Synthesis of 4-(bromomethyl)benzophenone (**40**)





### Experimental Procedure:

Synthesis of benzophenone **40** was conducted using an adapted procedure published by Tayama *et al.*<sup>383</sup>

4-Methylbenzophenone (**44**, 2.57 g, 13.0 mmol, 1.0 eq.) was dissolved in benzene (65 mL). *N*-bromosuccinimide (**90**; 2.55 g, 14.3 mmol, 1.1 eq.) and azobis(isobutyronitrile) (0.11 g, 0.65 mmol, 0.05 eq.) were added and the suspension was heated to reflux for 5 h, during which all components dissolved completely. After stirring at room temperature overnight, the solution was cooled to 0 °C, resulting in the formation of a colorless solid. The mixture was filtered and the solvent from the filtrate was removed under reduced pressure giving a yellow oil. Upon the addition of *c*-Hex/EtOAc (10:1, *v/v*) to the residue, a colorless solid formed. The solid was washed with *c*-Hex/EtOAc (10:1, *v/v*) and dried *in vacuo*. 4-(Bromomethyl)benzophenone (**40**, 1.68 g, 6.1 mmol, 47%) was isolated as a colorless solid.

### Analytcs:

**Chemical Formula:** C<sub>14</sub>H<sub>11</sub>OBr.

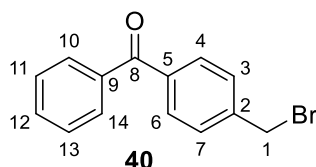
**Molecular Weight [g mol<sup>-1</sup>]:** 275.15.

**<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, r.t.):** δ [ppm] = 7.83 - 7.72 (m, 4H; H-4, H-6, H-10, H-14), 7.65 - 7.54 (m, 1H; H-12), 7.54 - 7.43 (m, 4H; H-3, H-7, H-11, H-13), 4.53 (s, 2H; H-1).

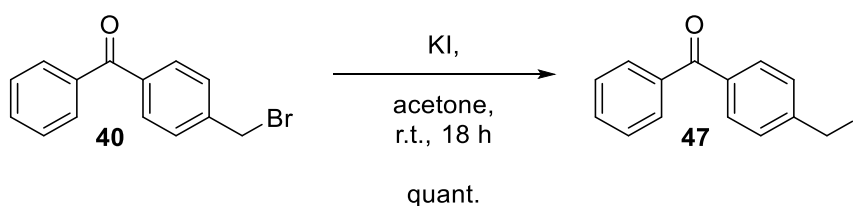
**<sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>, r.t.):** δ [ppm] = 196.09 (C-8), 142.21 (C-9), 137.56 (C-5), 137.49 (C-2), 132.69 (C-12), 130.66 (C-10, C-14), 130.12 (C-4, C-6), 129.06 (C-11, C-13), 128.47 (C-3, C-7), 32.39 (C1).

**GC-MS (EI<sup>+</sup>):** calculated for [M<sup>+</sup>] (C<sub>14</sub>H<sub>11</sub>OBr): 276.15; found: *m/z* = 276.0.

**FT-IR:** 3649 (w), 3024 (w), 1944 (w), 1869 (w), 1773 (w), 1717 (w), 1647 (m, ν(C=O)), 1607 (w), 1591 (w), 1447 (w), 1412 (w), 1310 (w), 1279 (s), 1225 (m), 1202 (m), 1177 (w), 1150 (w), 1099 (w), 1074 (w), 1026 (w), 997 (w), 974 (w), 924 (m), 878 (w), 851 (m), 816 (w), 791 (m), 745 (s, δ(C-H)o.o.p.), 700 (s, δ(C-H)o.o.p.), 677 (s, ν(C-Br)), 617 (s), 600 (s).



### 5.2.7.25 Synthesis of 4-(iodomethyl)benzophenone (47)



### Experimental Procedure:

Compound **47** was synthesized employing an adapted protocol of Ortiz-Trankina *et al.*<sup>384</sup> 4-(Bromomethyl)benzophenone (**40**, 300 mg, 1.09 mmol, 1.0 eq.) was suspended in acetone (15 mL) and KI (724 mg, 4.36 mmol, 4.0 eq) was successively added. The reaction mixture was vigorously stirred overnight at room temperature turning into a deep yellow solution with a colorless precipitate. The precipitate was filtered off and the solvent from the filtrate was removed under reduced pressure. The residue was dissolved in DCM (25 mL) and washed with H<sub>2</sub>O (3 × 25 mL). The aqueous layer was back extracted with DCM (2 × 25 mL). The combined organic layers were dried over MgSO<sub>4</sub> and the solvent was removed *in vacuo*. Compound **47** (380 mg, 1.7 mmol, quant., *c*-Hex/EtOAc, 9/1, *v/v*, *R<sub>f</sub>* = 0.57) was obtained as a brown-yellowish solid in quantitative yield.

### Analytcs:

**Chemical Formula:** C<sub>14</sub>H<sub>11</sub>OI.

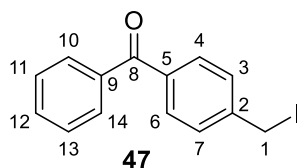
**Molecular Weight [g mol<sup>-1</sup>]:** 322.15.

**<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, r.t.):** δ [ppm] = 7.85 - 7.69 (m, 4H; H-4, H6, H-10, H-13), 7.66 - 7.54 (m, 1H; H-12), 7.53 - 7.45 (m, 4H; H-3, H-7, H-11, H-14), 4.50 (s, 2H, H-1).

**<sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>, r.t.):** δ [ppm] = 196.09 (C-8), 144.06 (C-9), 137.54 (C-5), 132.65 (C-12), 130.78 (C-10, C-14), 130.46 (C-2), 130.12 (C-4, C-6), 128.79 (C-11, C-13), 128.46 (C-3, C-7), 4.10 (C-1).

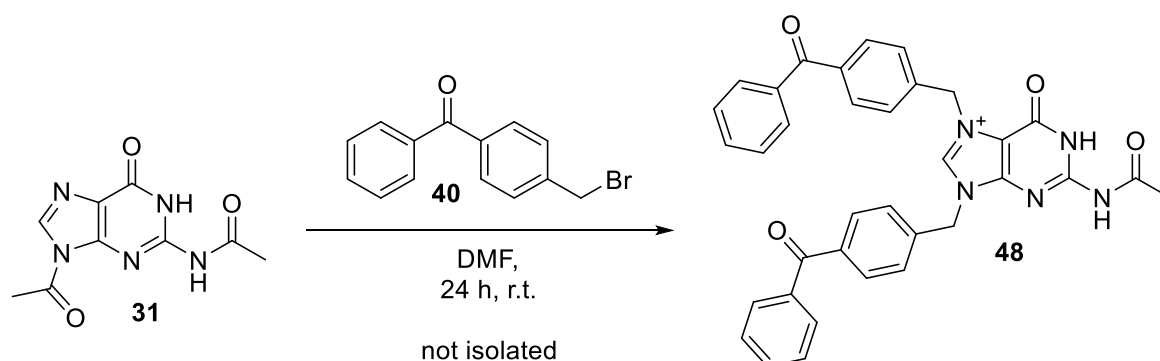
**GC-MS (EI<sup>+</sup>):** calculated for [M<sup>+</sup>] (C<sub>14</sub>H<sub>11</sub>OI): 322.9; found: *m/z* = 322.1.

**FT-IR:** 3026 (w), 2428 (w), 2307 (w), 1701 (w), 1647 (s, ν(C=O)), 1593 (m), 1560 (w), 1447 (m), 1412 (w), 1308 (m), 1279 (s), 1215 (w), 1175 (m), 1150 58 (m), 1082 (m), 1072 (m), 1018 (w), 997 (w), 974 (m), 941 (s), 924 (s), 851 (s), 831 (m), 816 (m), 789 (m), 743 (s, δ(C-H)o.o.p.), 698 (s, δ(C-H)o.o.p.), 677 (s), 635 (m), 613 (s, ν(C-I)), 602 (m), 567 (m).



### 5.2.7.26 Synthesis of 2-acetamido-9-acetyl-7-(4-benzoylbenzyl)-6-oxo-6,9-dihydro-1*H*-purin-7-ium (**48**)

#### Method 1:

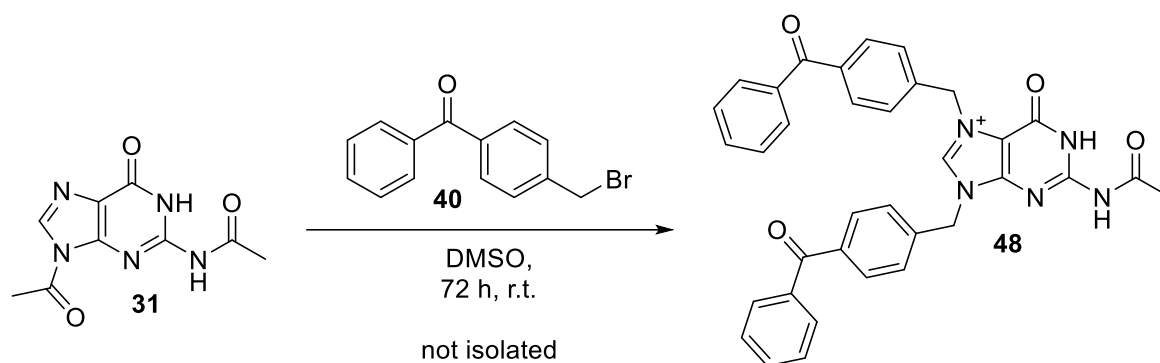


#### Experimental Procedure:

The reaction conditions for the synthesis of **48** were adapted from a procedure published by Pautus *et al.*<sup>385</sup>

*N*<sup>2</sup>,<sup>9</sup>-diacetylated guanine **31** (20 mg, 0.09 mmol, 1.0 eq.) and 4-(bromomethyl)benzophenone (**40**, 117 mg, 0.43 mmol, 5.0 eq.) were dissolved in DMF (0.5 mL) and stirred at r.t. under exclusion of light for 24 h. The molecular weight of molecule **48** was verified by LC-MS. However, product **48** was not isolated due to the small reaction scale (please refer to Chapter 3.2.1 for more information).

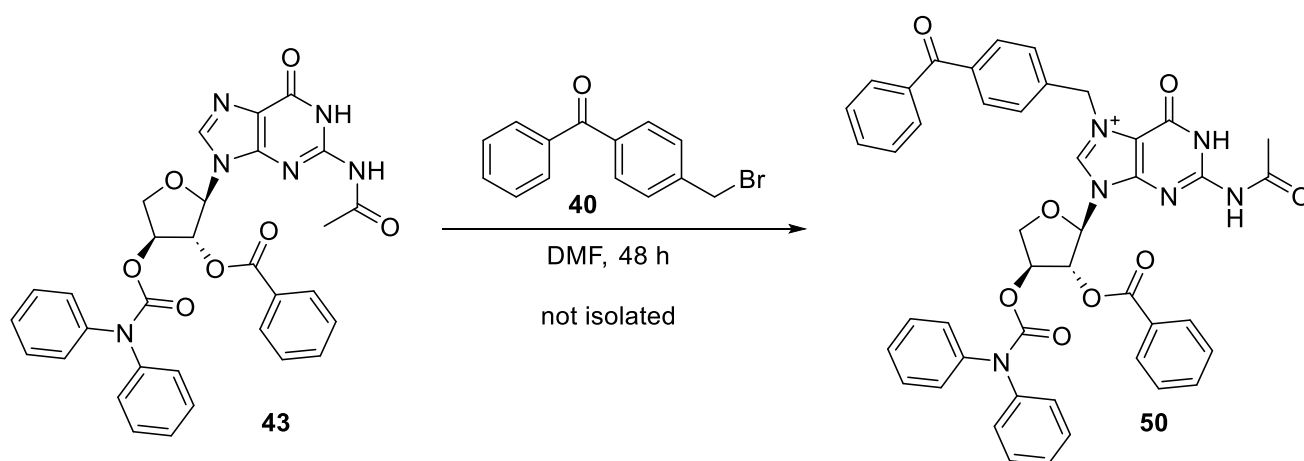
#### Method 2:



#### Experimental Procedure:

The reaction conditions for the synthesis of **48** were adapted from a procedure published by Anhäuser *et al.*<sup>381</sup>

*N*<sup>2</sup>,<sup>9</sup>-diacetylated guanine **31** (20 mg, 0.09 mmol, 1.0 eq.) and 4-(bromomethyl)benzophenone (**40**, 117 mg, 0.43 mmol, 5.0 eq.) were dissolved in DMSO (0.5 mL) and stirred at r.t. under exclusion of light for 72 h. The molecular weight of molecule **48** was verified by LC-MS. However, product **48** was not isolated due to the small reaction scale (please refer to Chapter 3.2.1 for more information).

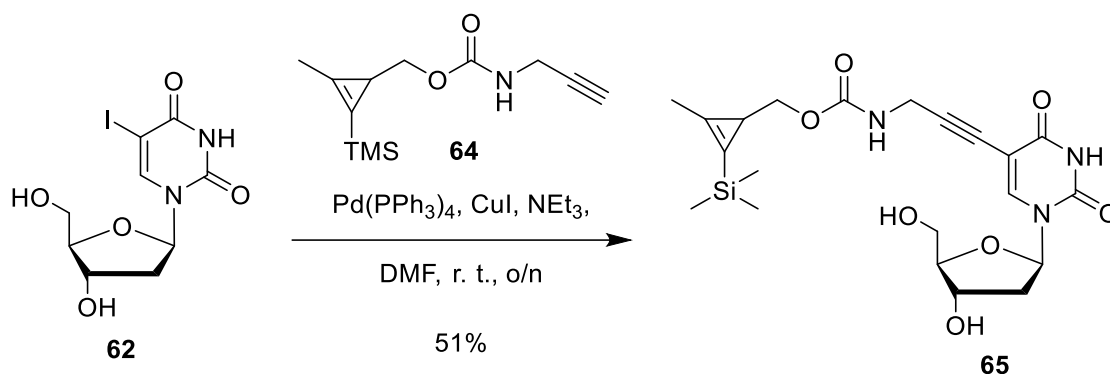
Analytcs:**Chemical Formula:** C<sub>35</sub>H<sub>28</sub>N<sub>5</sub>O<sub>4</sub><sup>+</sup>.**Molecular Weight [g mol<sup>-1</sup>]:** 582.21.**LC-MS (ESI):** calculated for [M - 2H]<sup>-</sup> (C<sub>35</sub>H<sub>26</sub>N<sub>5</sub>O<sub>4</sub>): 580.20; found: *m/z* = 580.18.**5.2.7.27 Synthesis of 2-acetamido-7-(4-benzoylbenzyl)-9-((2*R*,3*R*,4*S*)-3-(benzoyloxy)-4-((diphenylcarbamoyl)oxy) tetrahydrofuran-2-yl)-6-oxo-6,9-dihydro-1*H*-purin-7-ium (50)**Experimental Procedure:

The reaction conditions for the synthesis of **48** were adapted from a procedure published by Pautus *et al.*<sup>385</sup>

Nucleoside **43** (55 mg, 0.09 mmol, 1.0 eq.) and 4-(bromomethyl)benzophenone (**40**; 127 mg, 0.46 mmol, 5.0 eq.) were dissolved in DMF (1.5 mL) and stirred at r.t. under exclusion of light. The reaction was stirred for 24 h and evaluated using analytical HPLC (60 % MeCN in 0.1 % (w/v) aq. ammonium acetate buffer for 10 min followed by a gradient from 60 % → 100 % MeCN in 10 min, flow rate = 1.0 mL min<sup>-1</sup>) and LC-MS. Stirring of the reaction was continued for another 24 h, again followed by analytical HPLC analysis and LC-MS evaluation. The molecular weight of the target molecule **50** was verified by LC-MS. However, product **50** was not isolated due to the small reaction scale (please refer to Chapter 3.2.1 for more information).

Analytcs:**Chemical Formula:** C<sub>45</sub>H<sub>37</sub>N<sub>6</sub>O<sub>8</sub><sup>+</sup>.**Molecular Weight [g mol<sup>-1</sup>]:** 789.82.**LC-MS (ESI):** calculated for [M - 2H]<sup>-</sup> (C<sub>45</sub>H<sub>35</sub>N<sub>6</sub>O<sub>8</sub>): 787.25; found: *m/z* = 787.35.

### 5.2.7.28 Synthesis of {3-[1-(4-Hydroxy-5-hydroxymethyl-tetrahydro-furan-2-yl)-2,4-dioxo-1,2,3,4-tetrahydropyrimidin-5-yl]-prop-2-ynyl}-carbamic acid 2-methyl-3-trimethylsilanyl-cycloprop-2-enylmethyl ester (**65**)



#### Experimental Procedure:

Reaction conditions for Sonogashira coupling were adapted from Eggert *et al.*<sup>214</sup>

In a dry schlenk flask, 5-iodo-2'-deoxyuridine (**62**; 400 mg, 1.12 mmol, 1.0 eq.), CuI (150 mg, 0.78 mmol, 0.7 eq.) and Pd(PPh<sub>3</sub>)<sub>4</sub> (262 mg, 0.22 mmol, 0.2 eq.) were combined under argon atmosphere. The flask was sealed and 3 cycles of evacuation with vacuum and flushing with argon were performed. Subsequently, solids were dissolved in dry DMF (25 mL) and the resulting solution was degassed by a stream of argon. A previously degassed solution of cyclopropene linker **64** (402 mg, 1.7 mmol, 1.5 eq.) in anhydrous DMF (25 mL) and freshly distilled NEt<sub>3</sub> (0.46 mL, 3.38 mmol, 3.0 eq.) was then transferred into the reaction flask while stirring. The resulting reaction mixture was stirred under argon atmosphere at room temperature overnight under exclusion of light and was subsequently concentrated under reduced pressure. The crude product was purified by silica gel column chromatography (MeOH/ DCM, 1/9, v/v, *R<sub>f</sub>* = 0.25) to afford **65** (281 mg, 0.62 mmol, 51%) as a white solid.

#### Analytcs (specific assignment given whenever possible):

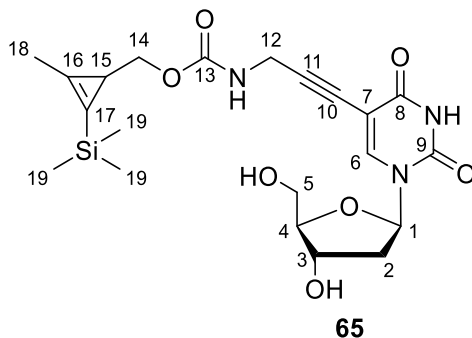
**Chemical Formula:** C<sub>21</sub>H<sub>29</sub>N<sub>3</sub>O<sub>7</sub>Si.

**Molecular Weight [g mol<sup>-1</sup>]:** 463.56.

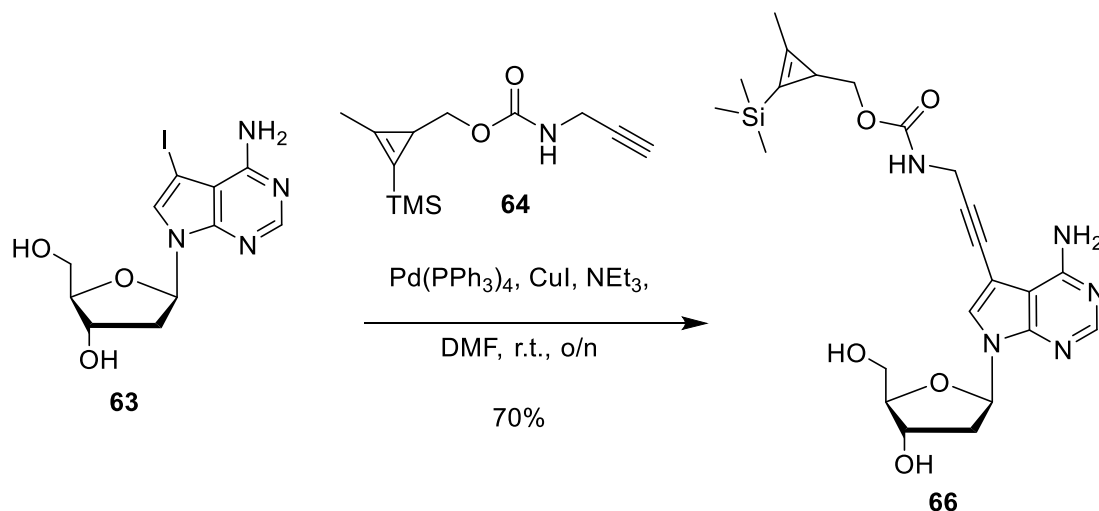
**<sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD, r.t.):** δ [ppm] = 8.28 (s, 1H; H-6), 6.23 (t, <sup>3</sup>*J*<sub>H-1/H-2</sub> = 6.5 Hz, 1H; H-1), 4.39 (dt, *J* = 6.6 Hz, <sup>3</sup>*J*<sub>H-3/H-2</sub> = 3.5 Hz, <sup>3</sup>*J*<sub>H-3/H-4</sub> = 3.5 Hz, 1H; H-3), 4.07 (s, 2H; H-12), 3.93 (q, <sup>3</sup>*J*<sub>H-4/H-3</sub> = 3.4 Hz, <sup>3</sup>*J*<sub>H-4/H-5</sub> = 3.4 Hz, 1H; H-4), 3.87 (dp, <sup>2</sup>*J* = 10.9 Hz, <sup>3</sup>*J*<sub>H-14/H-15</sub> = 5.3 Hz, 2H; H-14), 3.81 (dt, <sup>2</sup>*J*<sub>H-5b/H-5a</sub> = 11.9 Hz, 2.0 Hz, 1H; H-5b/H-5a), 3.73 (dd, <sup>2</sup>*J*<sub>H-5a/H-5b</sub> = 12.0 Hz, <sup>3</sup>*J*<sub>H-5a/H-4</sub> = 3.4 Hz, 1H; H-5a/H-5b), 2.31 (ddd, <sup>2</sup>*J*<sub>H-2b/H-2a</sub> = 13.6 Hz, <sup>3</sup>*J*<sub>H-2/H-1</sub> = 6.3, <sup>3</sup>*J*<sub>H-2b/H-3</sub> = 3.5 Hz, 1H; H-2b/H-2a), 2.24 – 2.18 (m, 4H; H-18, H-2a/H-2b), 1.55 (t, <sup>3</sup>*J*<sub>H-15/H-14</sub> = 5.3 Hz, 1H; H-15), 0.16 (s, 9H; H-19).

**$^{13}\text{C-NMR}$  (126 MHz,  $\text{CD}_3\text{OD}$ , r.t.):**  $\delta$  [ppm] = 164.54 (C-8), 159.05 (C-13) 151.15 (C-9), 145.31 (C-6), 136.03 (C-16), 112.16 (C-17), 100.06 (C-7), 90.79 (C-11), 89.15 (C-4), 87.01 (C-1), 74.88 (C-10), 74.64 (C-14), 72.08 (C-3), 62.64 (C-5), 41.68 (C-2), 31.91 (C-12), 19.84 (C-15), 13.15 (C-18), -1.18 (C-19).

**HR-MS (ESI<sup>+</sup>):** calculated for  $[\text{M} + \text{Na}]^+$  ( $\text{C}_{21}\text{H}_{29}\text{N}_3\text{NaO}_7\text{Si}^+$ ): 484.1667; found:  $m/z = 486.1665$ .



### 5.2.7.29 Synthesis of {3-[4-Amino-7-(4-hydroxy-5-hydroxymethyl-tetrahydro-furan-2-yl)-7H-pyrrolo[2,3-d]pyrimidin-5-yl]-prop-2-ynyl}-carbamic acid 2-methyl-3-trimethylsilyl-cycloprop-2-enylmethyl ester (**66**)



#### Experimental Procedure:

Reaction conditions for Sonogashira coupling were adapted from Eggert *et al.*<sup>214</sup>

In a dry schlenk flask, 5-iodo-2'-deoxyuridine (**63**; 400 mg, 1.06 mmol, 1.0 eq.), CuI (142 mg, 0.74 mmol, 0.7 eq.) and Pd(PPh<sub>3</sub>)<sub>4</sub> (246 mg, 0.21 mmol, 0.2 eq.) were combined under argon atmosphere. The flask was sealed and 3 cycles of evacuation with vacuum and flushing with argon were performed. Subsequently, solids were dissolved in dry DMF (25 mL) and the resulting solution was degassed by a stream of argon. A previously degassed solution of cyclopropene linker **64** (378 mg, 1.6 mmol, 1.5 eq.) in anhydrous DMF (25 mL) and freshly distilled NEt<sub>3</sub> (0.44 mL, 3.2 mmol, 3.0 eq.) was then transferred into the reaction flask while stirring. The resulting reaction mixture was stirred under argon atmosphere at room

temperature overnight under exclusion of light and was subsequently concentrated under reduced pressure. The crude product was purified by silica gel column chromatography (MeOH/ DCM, 1/9, v/v,  $R_f = 0.49$ ) to afford **66** (345 mg, 0.71 mmol, 70%) as a white solid.

Analytics (specific assignment given whenever possible):

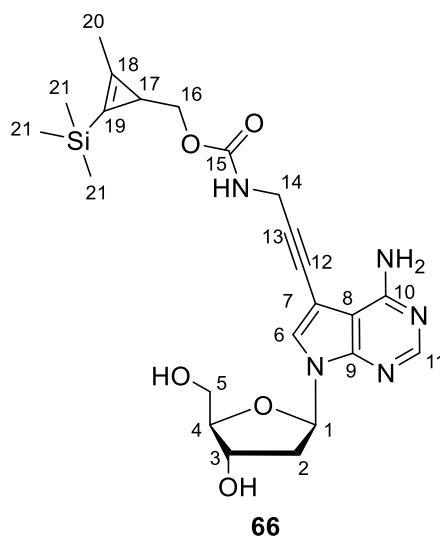
**Chemical Formula:**  $C_{23}H_{31}N_5O_5Si$ .

**Molecular Weight [g mol<sup>-1</sup>]:** 485.62.

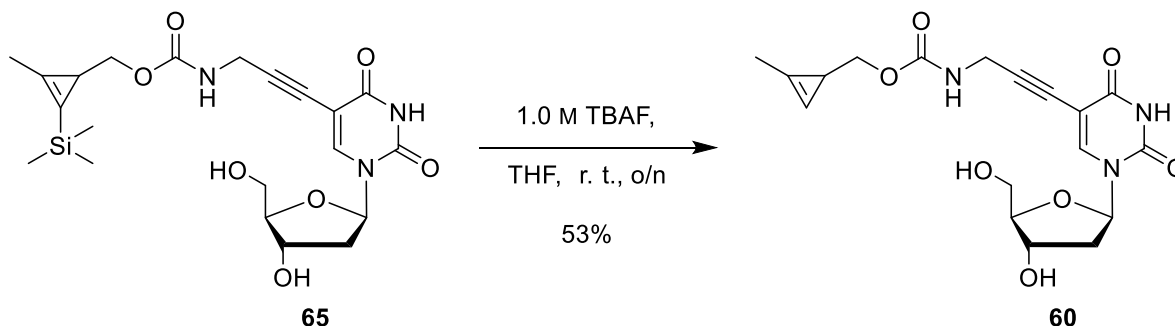
**<sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD, r.t.):**  $\delta$  [ppm] = 8.10 (s, 1H; H-11), 7.53 (s, 1H; H-6), 6.46 (dd,  $^3J_{H-1/H-2b} = 8.0$  Hz,  $^3J_{H-1/H-2a} = 6.0$  Hz, 1H; H-1), 4.51 (dt,  $^3J_{H-3/H-2b} = 5.7$  Hz,  $^3J_{H-3/H-2a} = 2.7$  Hz, 1H; H-3), 4.09 (s, 2H; H-14), 4.00 (q,  $^3J_{H-4/H-5} = 3.2$  Hz, 1H; H-4), 3.91 (d,  $^3J_{H-16/H-17} = 5.2$  Hz, 2H; H-16), 3.79 (dd,  $^2J_{H-5b/H-5a} = 12.1$  Hz,  $^3J_{H-5b/H-4} = 3.3$  Hz, 1H; H-5b/H-5a), 3.72 (dd,  $^2J_{H-5a/H-5b} = 12.1$  Hz, 3.6 Hz, 1H; H-5a/H-5b), 2.61 (ddd,  $^2J_{H-2b/H-2a} = 13.4$  Hz,  $^3J_{H-2b/H-1} = 8.0$  Hz,  $^3J_{H-2b/H-3} = 5.8$  Hz, 1H; H-2b/H-2a), 2.32 (ddd,  $^2J_{H-2a/H-2b} = 13.4$  Hz,  $^3J_{H-2a/H-1} = 6.0$  Hz,  $^3J_{H-2a/H-3} = 2.7$  Hz, 1H; H-2a/H-2b), 2.18 (s, 3H; H-20), 1.56 (t,  $^3J_{H-17/H-16} = 5.2$  Hz, 1H; H-17), 0.14 (s, 9H; H-21).

**<sup>13</sup>C-NMR (126 MHz, CD<sub>3</sub>OD, r.t.):**  $\delta$  [ppm] = 159.36 (C-10), 159.06 (C-15), 153.21 (C-11), 149.77 (C-9), 135.90 (C-18), 127.79 (C-6), 112.17 (C-19), 97.05 (C-8), 90.11 (C-7), 89.15 (C-4), 86.62 (C-1), 75.96 (C-13), 74.70 (C-16), 72.94 (C-3), 71.75 (C-12), 63.59 (C-5), 41.53 (C-2), 32.03 (C-14), 19.83 (C-17), 13.19 (C-20), -1.15 (C-21).

**HR-MS (ESI<sup>+</sup>):** calculated for  $[M + H]^+$  ( $C_{23}H_{32}N_5O_5Si^+$ ): 486.2167; found:  $m/z = 486.2166$ .



### 5.2.7.30 Synthesis of {3-[1-(4-Hydroxy-5-hydroxymethyl-tetrahydro-furan-2-yl)-2,4-dioxo-1,2,3,4-tetrahydropyrimidin-5-yl]-prop-2-ynyl}-carbamic acid 2-methyl-cycloprop-2-enylmethyl ester (**60**)



#### Experimental procedure:

TMS deprotection was performed by adapting a protocol from Yang *et al.*<sup>415</sup>

Under an atmosphere of argon, dUC<sup>CP</sup>-TMS **65** (281 mg, 0.60 mmol, 1.0 eq.) was dissolved in THF (5 mL) and TBAF in THF (1.0 M, 3.0 mL, 3.0 mmol, 5.0 eq.) was added while stirring. Stirring at r.t. was continued overnight and the solvent was removed *in vacuo*. Purification using silica gel column chromatography (MeOH/DCM, 1/9, *v/v*,  $R_f = 0.27$ ) yielded methyl cyclopropene nucleoside **60** (136 mg, 0.35 mmol, 53%) as an off-white foam.

#### Analytcs (specific assignment given whenever possible):

**Chemical Formula:** C<sub>18</sub>H<sub>21</sub>N<sub>3</sub>O<sub>7</sub>.

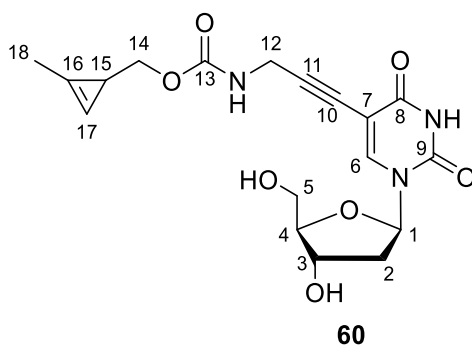
**Molecular Weight [g mol<sup>-1</sup>]:** 391.38.

**<sup>1</sup>H NMR (600 MHz, CD<sub>3</sub>OD, r.t.):**  $\delta$  [ppm] = 8.27 (s, 1H; H-6), 6.65 (s, 1H; H-17), 6.23 (dd,  $^3J_{H-1/H-2a} = 7.0$  Hz,  $^3J_{H-1/H-2b} = 6.1$  Hz, 1H; H-1), 4.40 (dt,  $^3J_{H-3/H-2a} = 6.5$  Hz,  $^3J_{H-3/H-2b} = 3.5$  Hz,  $^3J_{H-3/H-4} = 3.5$  Hz, 1H; H-3), 4.07 (s, 2H; H-12), 3.97 (dd,  $^2J_{H-14b/H-14a} = 11.2$  Hz,  $^3J_{H-14/H-15} = 4.7$  Hz, 1H; H-14b/H-14a), 3.94 (q,  $^3J_{H-4/H-5} = 3.4$  Hz,  $^3J_{H-4/H-3} = 3.4$  Hz, 1H; H-4), 3.86 – 3.79 (m, 2H; H-14a/H-14b, H-5b/H-5a), 3.74 (dd,  $^2J_{H-5a/H-5b} = 12.0$  Hz,  $^3J_{H-5a/H-4} = 3.5$  Hz, 1H; H-5a/H-5b), 2.31 (ddd,  $^2J_{H-2b/H-2a} = 13.6$  Hz,  $^3J_{H-2b/H-1} = 6.1$  Hz,  $^3J_{H-2b/H-3} = 3.6$  Hz, 1H; H-2b/H-2a), 2.22 (ddd,  $^2J_{H-2a/H-2b} = 13.5$  Hz,  $^3J_{H-2a/H-1} = 7.1$  Hz,  $^3J_{H-2a/H-3} = 6.4$  Hz, 1H; H-2a/H-2b), 2.13 (d,  $^4J_{H-18/H-17} = 1.2$  Hz, 3H; H-18), 1.61 (t,  $^3J_{H-15/H-14} = 4.7$  Hz, 1H; H-15).

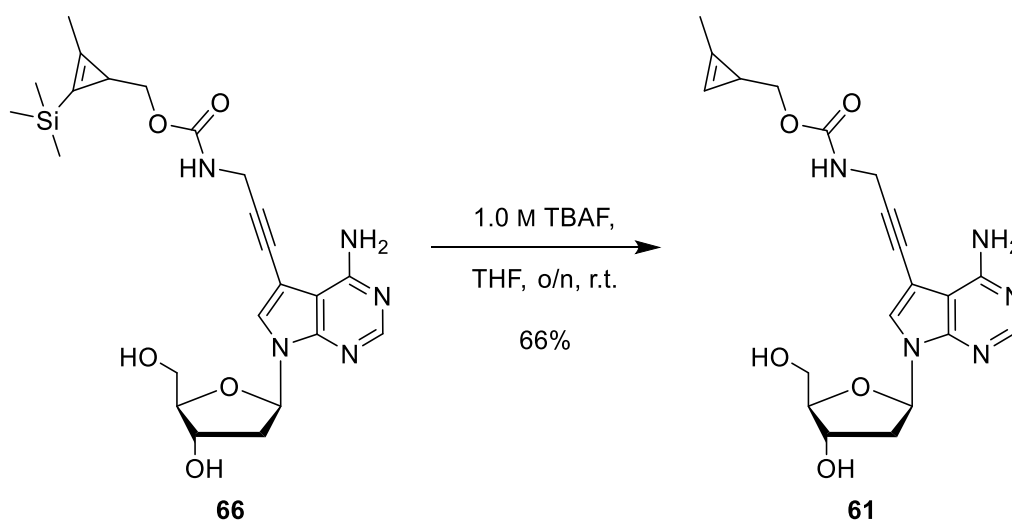
**<sup>13</sup>C-NMR (151 MHz, CD<sub>3</sub>OD, r.t.):**  $\delta$  [ppm] = 164.45 (C-8), 158.91 (C-13), 151.10 (C-9), 145.29 (C-6), 122.11 (C-16), 102.78 (C-17), 99.99 (C-7), 90.74 (C-11), 89.08 (C-4), 86.98 (C-1), 74.90 (C-10), 73.53 (C-14), 72.01 (C-3), 62.59 (C-5), 41.62 (C-2), 31.91 (C-12), 18.22 (C-15), 11.59 (C-18).

**HR-MS (ESI<sup>+</sup>):** calculated for [M + Na]<sup>+</sup> (C<sub>18</sub>H<sub>21</sub>N<sub>3</sub>NaO<sub>7</sub><sup>+</sup>): 414.1271; found:  $m/z = 414.1274$ .





### 5.2.7.31 Synthesis of {3-[4-Amino-7-(4-hydroxy-5-hydroxymethyl-tetrahydro-furan-2-yl)-7H-pyrrolo[2,3-d]pyrimidin-5-yl]-prop-2-ynyl}-carbamic acid 2-methyl-cycloprop-2-enylmethyl ester (**61**)



#### Experimental procedure:

TMS deprotection was performed by adapting a protocol from Yang *et al.*<sup>415</sup>

Under an atmosphere of argon, **dA**<sup>CP</sup>-TMS **66** (345 mg, 0.71 mmol, 1.0 eq.) was dissolved in THF (5 mL) and TBAF in THF (1.0 M, 3.6 mL, 3.6 mmol, 5.0 eq.) was added while stirring. Stirring was continued overnight and the solvent was removed *in vacuo*. Purification using silica gel column chromatography (MeOH/DCM, 1/9, *v/v*,  $R_f = 0.37$ ) yielded methyl cyclopropene nucleoside **61** (195 mg, 0.47 mmol, 66%) as an off-white foam.

#### Analytics (specific assignment given whenever possible):

**Chemical Formula:** C<sub>20</sub>H<sub>23</sub>N<sub>5</sub>O<sub>5</sub>.

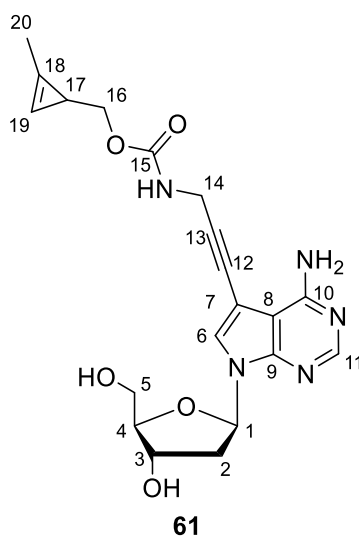
**Molecular Weight [g mol<sup>-1</sup>]:** 413.43.

**<sup>1</sup>H NMR (600 MHz, CD<sub>3</sub>OD, r.t.):**  $\delta$  [ppm] = 8.07 (s, 1H; H-11), 7.53 (s, 1H; H-6), 6.63 (s, 1H; H-19), 6.46 (dd,  $^3J_{H-1/H-2b} = 8.0$  Hz,  $^3J_{H-1/H-2a} = 6.0$  Hz, 1H; H-1), 4.51 (dt,  $^3J_{H-3/H-2b} = 5.7$  Hz,  $^3J_{H-3/H-2a} = 2.7$  Hz, 1H; H-3), 4.09 (s, 2H; H-14), 4.03 – 3.98 (m, 2H; H-4, H-16), 3.87 (dd,  $^2J =$

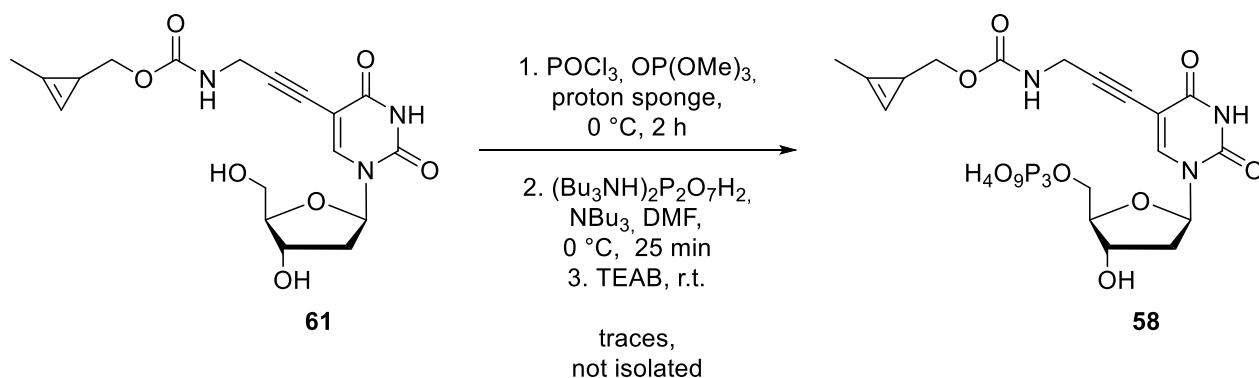
11.0 Hz,  $^3J = 5.4$  Hz, 1H; H-16), 3.79 (dd,  $^2J_{H-5b/H-5a} = 12.1$  Hz,  $^3J_{H-5b/H-4} = 3.3$  Hz, 1H; H-5b/H-5a), 3.72 (dd,  $^2J_{H-5a/H-5b} = 12.1$  Hz, 3.7 Hz, 1H; H-5a/H-5b), 2.61 (ddd,  $^2J_{H-2b/H-2a} = 13.7$  Hz,  $^3J_{H-2b/H-1} = 8.1$  Hz,  $^3J_{H-2b/H-3} = 5.8$  Hz, 1H; H-2b/H-2a), 2.32 (ddd,  $^2J_{H-2a/H-2b} = 13.5$  Hz,  $^3J_{H-2a/H-1} = 6.0$ ,  $^3J_{H-2a/H-3} = 2.8$  Hz, 1H; H-2a/H-2b), 2.11 (d,  $^4J_{H-20/H-19} = 1.3$  Hz, 3H; H-20), 1.65 – 1.60 (m, 1H; H-17).

$^{13}\text{C-NMR}$  (151 MHz,  $\text{CDCl}_3$ , r.t.):  $\delta$  [ppm] = 159.30 (C-10), 159.10 (C-15), 153.18 (C-11), 149.84 (C-9), 127.78 (C-6), 122.05 (C-18), 104.71 (C-8), 102.81 (C-19), 96.97 (C-7), 90.02 (C-13), 89.16 (C-4), 86.60 (C-1), 76.00 (C-12), 73.54 (C-16), 72.93 (C-3), 63.59 (C-5), 41.51 (C-2), 32.03 (C-14), 18.26 (C-17), 11.58 (C-20).

**HR-MS (ESI<sup>+</sup>):** calculated for  $[\text{M} + \text{H}]^+$  ( $\text{C}_{20}\text{H}_{24}\text{N}_5\text{O}_5^+$ ): 414.1772; found:  $m/z = 414.1772$ .



### 5.2.7.32 Synthesis of ((2-methylcycloprop-2-en-1-yl)methyl (3-(1-((2*R*,4*S*,5*R*)-4-hydroxy-5-(((hydroxy((hydroxy (phosphonoxy)phosphoryl)oxy)phosphoryl)oxy)methyl) tetrahydrofuran-2-yl)-2,4-dioxo-1,2,3,4-tetrahydropyrimidin-5-yl)prop-2-yn-1-yl)carbamate (58)



### Experimental Procedure:

Triphosphate synthesis was performed according to protocols published Srivatsan *et al.* and Merkel *et al.*<sup>363,370</sup>

Nucleoside **61** (68 mg, 0.17 mmol, 1.0 eq.) and proton sponge (30 mg, 0.13 mmol, 0.8 eq.) were dried under high vacuum overnight and were dissolved in trimethylphosphate (3.25 mL, freshly distilled and stored over 3 Å molecular sieves and under argon) under an atmosphere of argon and cooled to 0 °C. POCl<sub>3</sub> (40 µL, 0.43 mmol, 2.5 eq., freshly distilled and stored under argon) was added dropwise at 0 °C and the reaction was stirred under ice cold conditions for 2 h. Subsequently, tributylamine (0.43 mL, 1.82 mmol, 10.5 eq., freshly distilled and stored under argon) and a solution of bis-tri-*n*-butylammonium pyrophosphate in dry DMF (0.5 M, 1.88 mL, 0.94 mmol, 5.4 eq.) were added simultaneously in one portion at 0 °C. The mixture was stirred at 0 °C for 25 min and quenched by the addition of TEAB buffer (1.0 M, 18 mL, pH = 8.0). Volatiles were evaporated under high vacuum at room temperature. Coevaporation of the crude product with H<sub>2</sub>O (5 × 30 mL) was performed. The crude product was lyophilized overnight. LC-MS analysis detected only traces of the desired compound **58** in the crude reaction mixture. Micro purification employing HPLC (5 % → 40 % MeCN in 0.1 % (w/v) ammonium acetate buffer in 40 min, flow rate = 1.0 mL min<sup>-1</sup>) was performed. Separation of target compound **58** from side products was not successful, hence product was not isolated (please refer to Chapter 3.3.1 for more information).

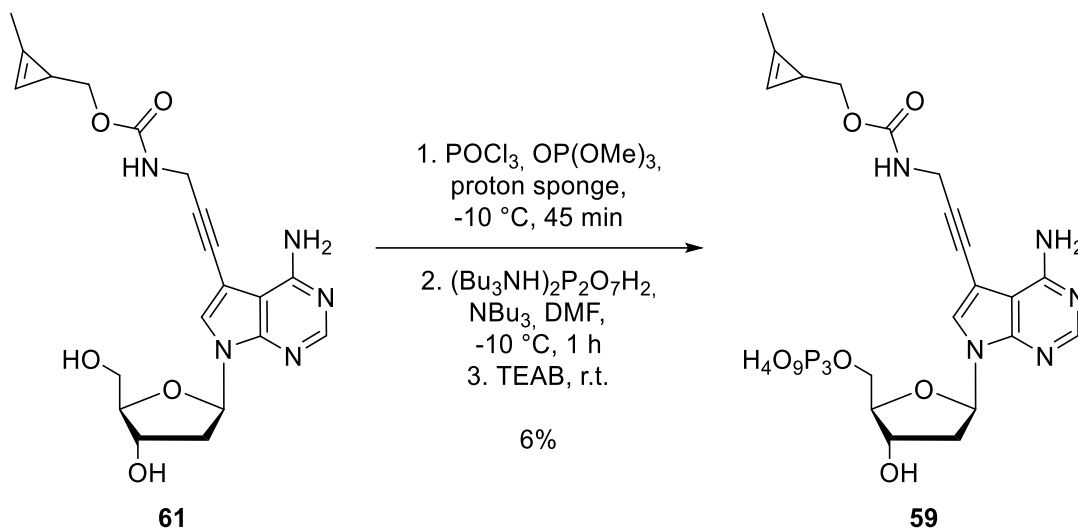
### Analytcs:

**Chemical Formula:** C<sub>18</sub>H<sub>24</sub>N<sub>3</sub>O<sub>16</sub>P<sub>3</sub>.

**Molecular Weight [g mol<sup>-1</sup>]:** 631.32.

**LC-MS (ESI):** calculated for [M - H]<sup>-</sup> (C<sub>18</sub>H<sub>23</sub>N<sub>3</sub>O<sub>16</sub>P<sub>3</sub><sup>-</sup>): 630.03; found: *m/z* = 630.10.

### 5.2.7.33 Synthesis of (2-methylcycloprop-2-en-1-yl)methyl (3-(4-amino-7-((2*R*,4*S*,5*R*)-4-hydroxy-5-(((hydroxy((hydroxy(phosphonoxy)phosphoryl)oxy)phosphoryl)oxy)methyl)tetrahydrofuran-2-yl)-7*H*-pyrrolo[2,3-*d*]pyrimidin-5-yl)prop-2-yn-1-yl)carbamate (**59**)



#### Experimental Procedure:

Triphosphate synthesis was performed according to protocols published Srivatsan *et al.* and Ploschik *et al.*<sup>263,370</sup>

Nucleoside **61** (77 mg, 0.19 mmol, 1.0 eq.) and proton sponge (32 mg, 0.15 mmol, 0.8 eq.) were dried under high vacuum overnight and were dissolved in trimethylphosphate (3.7 mL, freshly distilled and stored over 3 Å molecular sieves and under argon) under an atmosphere of argon and cooled to -10 °C. POCl<sub>3</sub> (20 µL, 0.47 mmol, 2.5 eq., freshly distilled and stored under argon) was added dropwise at -10 °C and the reaction was stirred at -10 °C for 45 min. Subsequently, tributylamine (0.46 mL, 1.96 mmol, 10.5 eq., freshly distilled and stored under argon) and a solution of bis-tri-*n*-butylammonium pyrophosphate in dry DMF (0.5 M, 2.0 mL, 1.00 mmol, 5.4 eq.) were added simultaneously in one portion at -10 °C. The mixture was stirred at -10 °C for 1 h and quenched by the addition of triethylammonium bicarbonate (TEAB) buffer (1.0 M, 20 mL, pH = 8.0). Volatiles were evaporated under high vacuum at room temperature. Coevaporation of the crude product with H<sub>2</sub>O (3 × 20 mL) was performed. The crude product was lyophilized overnight and purification employing HPLC (5 % → 30 % MeCN in 0.1 M aq. TEAB buffer in 50 min, flow rate = 16.0 mL min<sup>-1</sup>) was performed. After HPLC purification, the residue was again coevaporated with H<sub>2</sub>O (3 × 10 mL) and lyophilized overnight. Purified triphosphate **59** (11.2 mg, 0.012 mmol, 6%) was obtained as the 3 × triethylammonium salt as an off-white solid.

Analytics (specific assignment given whenever possible):

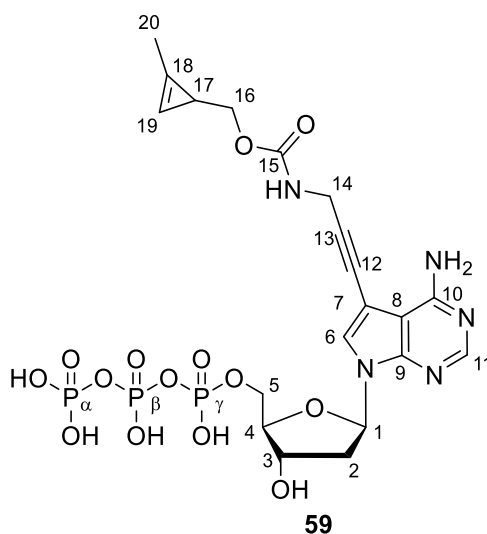
**Chemical Formula:** C<sub>20</sub>H<sub>26</sub>N<sub>5</sub>O<sub>14</sub>P<sub>3</sub>.

**Molecular Weight [g mol<sup>-1</sup>]:** 653.37.

**<sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O, r.t.):** δ [ppm] = 8.15 (b, 1H; H-11), 7.67 (s, 1H; H-6), 6.64 (s, 1H; H-19), 6.57 (dd, *J* = 7.8, <sup>3</sup>*J*<sub>H-1/H-2a</sub> = 6.2 Hz, 1H; H-1), 4.73 (dq, *J* = 6.0, <sup>3</sup>*J*<sub>H-3/H-2a</sub> = 3.2 Hz, 1H; H-3), 4.29 – 3.89 (m, 7H; H-4, H-5, H-14, H-16), 2.65 (dt, <sup>2</sup>*J*<sub>H-2b/H-2a</sub> = 14.1 Hz, 7.0 Hz, 1H; H-2b/H-2a), 2.48 (ddd, <sup>2</sup>*J*<sub>H-2a/H-2b</sub> = 14.0 Hz, <sup>3</sup>*J*<sub>H-2a/H-1</sub> = 6.2, <sup>3</sup>*J*<sub>H-2a/H-3</sub> = 3.2 Hz, 1H; H-2a/H-2b), 2.09 (s, 3H; H-20), 1.64 (s, 1H; H-17).

**<sup>31</sup>P NMR (243 MHz, D<sub>2</sub>O, r.t.):** δ [ppm] = -10.50 (s; P-γ), -11.37 (P-α), -23.21 (P-β).

**LC-MS (ESI):** calculated for [M - H]<sup>-</sup> (C<sub>20</sub>H<sub>25</sub>N<sub>5</sub>O<sub>14</sub>P<sub>3</sub>): 653.07; found: *m/z* = 652.02.



## 5.3 Biochemical methods

When working with nucleic acids, special precautions need to be considered to protect nucleic acid samples from contamination and degradation by ubiquitous nucleases. Disposable nitrile gloves were worn at all times and sterilized with 70 % EtOH prior to sample handling. Disposable reaction tubes (0.2 mL, 0.5 mL, 1.5 mL, and 2.0 mL) were autoclaved and stored in sealed glass containers. Pipette tips were autoclaved and stored in sealed plastic boxes. Conical tubes (15 mL and 50 mL) were purchased from the supplier, classified as sterile, pyrogen-free, DNA-free, DNase-free, and RNase-free, and used without further treatment. Ultrapure water (referred to as ddH<sub>2</sub>O) for direct use or buffer preparation was prepared from deionized water using a *Barnstead MicroPure (ThermoFisher Scientific)* or *Arium Mini (Sartorius)* ultrapure water system, then autoclaved, sterile-filtered and stored in aliquots at -20 °C.

### 5.3.1 Equipment

The equipment that has been used for experimental work in this study is depicted in Table 7.

**Table 7: Equipment used in this study.**

<b>Equipment</b>	<b>Model</b>	<b>Manufacturer</b>
Agarose gel electrophoresis chamber	Mini-Sub Cell GT	<i>Bio-Rad</i>
Autoclave	D65 VX-150	<i>Systec</i>
Bacterial cell culture incubator	BD400	<i>Binder</i>
Block heater	Iso-Block HBT-1-1-131	<i>Benchmark</i> <i>HLC BioTech</i>
Centrifuges	Mikro 200 5425 R 5417 C 5920 R	<i>Hettich</i> <i>Eppendorf</i> <i>Eppendorf</i> <i>Eppendorf</i>
FPLC	Äkta Start	<i>Cytiva</i>
Freeze dryer	Alpha 2-4 LD plus Alpha 1-2 LD plus	<i>Martin Christ</i> <i>Martin Christ</i>
Freezer	Premium NoFrost -20 °C Forma 900	<i>Liebherr</i> <i>Thermo Scientific</i>
Gel documentation system	GelDoc 2000	<i>Bio-Rad</i>

	Genoplex	VWR
	ChemoStar Touch	INTAS
HPLC	Elute SP	Bruker
HPLC columns	Zorbax StableBond (80 Å, 2.1 × 50 mm, 5 µm, C18); Xterra MS (2.1 × 100 mm, 5 µm, C18)	Agilent Technologies Waters
Incubation shaker	Multitron HT Innova 4430	Infors Thermo Fisher
Laminar flow hood	HeraSafe HS18 L miniAir 2448	Heraeus Heraeus
LED transilluminator	LED BluePad Transilluminator	INTAS
Light panel	Kaiser slimlite plano	Kaiser Fototechnik
Mass spectrometer	HTC Esquire, Amazon SL	Bruker Daltonics
Microscales	AT261 Delta Range 1602 MP	Mettler Toledo Satorius
Microscale thermophoresis	Monolith NT.115	Nano Temper Technologies
Microwave	MW 7873	Severin
Mini centrifuges	SU1550 MyFuge	Sunlab Benchmark
PAGE chamber	Mini Protean	Bio-Rad
pH meter	FiveEasyPlus	Mettler Toledo
Pipette controller	accu-jet pro	Brand
Pipettes	2.5 µL, 10 µL, 20 µL, 100 µL, 200 µL, 1000 µL, 5 mL, 10 mL (Research Plus); 2.5 µL, 10 µL, 20 µL, 100 µL, 200 µL, 1000 µL, 5 mL (IKA PETTE)	Eppendorf IKA-Werke
Power supply for gel electrophoresis	Power Pack Universal E132 P25	Bio-Rad Consort Biometra
Scale	PCD	Kern
Shaker	KS 130 basic	IKA-Werke
Spectrophotometer	Nanodrop 2000c DS-11 FX +	Thermo Fisher DeNovix

Thermocycler	LifeEco	<i>Biozym</i>
Thermoshaker	PocketBloc Thermomixer	<i>Bioer</i>
	Thermomixer Comfort	<i>Eppendorf</i>
Ultrasonic homogenizer	Sonoplus	<i>Bandelin</i>
Ultrapure watersystem	Barnstead MicroPure	<i>Thermo Fisher Scientific</i>
	Arium Mini	<i>Sartorius</i>
Vortex mixer	Vortex-Genie 2	<i>Scientific Industries</i>
	RS-VA 10	<i>Phoenix Instruments</i>
Water bath	JB Nova 12	<i>Grant</i>

### 5.3.2 Consumables

The consumables that have been used for experimental work in this study are depicted in Table 8.

**Table 8: Consumables used in this study.**

<b>Consumable</b>	<b>Specification</b>	<b>Manufacturer</b>
Capillaries	Monolith NT.115	<i>Nano Temper Technologies</i>
Centrifugal filters	Amicon Ultra 0.5, 3k	<i>Merck Millipore</i>
	Amicon Ultra 0.5, 10k	
	Amicon Ultra 4, 3k	
	Amicon Ultra 15, 10k	
	Amicon Ultra 4, 30k	
	Amicon Ultra 15, 30k	
Conical tubes	15 mL, 50 mL, PP, DNA-free, DNase-free, RNase-free, sterile	<i>Sarstedt</i>
Cryotubes	2 mL, PP, round bottom, screw top, sterile	<i>Greiner Bio-One</i>
Cuvettes	semi-micro (PS)	<i>VWR</i>
Dialysis membrane	Spectra/Por 7, 10 kDa	<i>SpectrumLabs</i>
DNA LoBind tubes	0.5 mL	<i>Eppendorf</i>
FPLC columns	HisTrap FF crude	<i>GE Healthcare</i>
	HiTrap TALON crude	<i>Cytiva</i>
PCR reaction tubes	200 µL, transparent, PP,	<i>neoLab Migge</i>
	DNA-free, flat lid	
Petri dishes	100x15 mm, PS, sterile	<i>Corning</i>



Pipette tips	10 $\mu$ L, 20 $\mu$ L, 200 $\mu$ L, 1 mL, 5 mL, 10 mL, PP, non-sterile Multiflex 200 $\mu$ L, non-sterile	<i>Sarstedt,</i> <i>Labsolute Th. Geyer</i> <i>Carl Roth</i>
Reaction tubes	1.5 mL and 2 mL, PP, transparent, SafeSeal cap, cap attached	<i>Sarstedt</i>
Plastic syringe	10 mL, 50 mL, sterile	<i>Henke-Sass Wolf</i>
Serological pipettes	5 mL, 10 mL, 25 mL, 50 mL, PS, plugged, sterile, non- cytotoxic	<i>Sarstedt</i>
Sterile filter storage bottles	250 mL, 500 mL, 1000 mL	<i>TPP, VWR</i>
Syringe filters	0.22 $\mu$ m pore size, cellulose acetate filter, sterile	<i>Labsolute Th. Geyer</i>

### 5.3.3 Reagents and buffers

Reagents and buffers that have been used for experimental work in this study are listed in Tables 9 and 10.

**Table 9: Reagents used in this study.**

<b>Reagent</b>	<b>Supplier</b>
5-Bromo-4-chloro-3-indoyl-beta-D-galactopyranoside (X-gal)	<i>Carl Roth</i>
6-Methyl-Tetrazine-BODIPY FL	<i>Jena Bioscience</i>
Acetic acid (glacial)	<i>Fisher Scientific</i>
Acetonitrile	<i>VWR Chemicals</i>
Alexa Fluor 488 NHS-ester	<i>Lumiprobe</i>
Ampicillin	<i>Carl Roth</i>
Agar, bacteriology grade	<i>AppliChem</i>
Agarose (high resolution)	<i>Sigma Aldrich</i>
Ammoniumperoxodisulfate (APS)	<i>Merck</i>
Bovine serum albumin (BSA)	<i>Carl Roth</i>
Boric acid	<i>Alfa Aesar</i>
Bromophenol blue	<i>Carl Roth</i>
Comassie blue	<i>Carl Roth</i>
Dimethylsulfoxide (DMSO)	<i>Sigma Aldrich</i>
Dithiothreitol (DTT)	<i>Sigma Aldrich</i>

---

dNTPs	<i>Jena Bioscience</i>
Ethanol (absolute)	<i>VWR Chemicals</i>
Ethidium bromide	<i>Carl Roth</i>
Ethylenediaminetetraacetic acid (EDTA)	<i>Thermo Fisher</i>
Ficoll 400	<i>Sigma Aldrich</i>
Formamide	<i>Fluka</i>
GeneRuler Ultra Low Range DNA Ladder	<i>Thermo Fisher</i>
GeneRuler 1 kb Plus DNA Ladder	<i>Thermo Fisher</i>
Glycerol	<i>Fisher bioreagents, Sigma Aldrich</i>
Glycine	<i>Carl Roth</i>
HEPES	<i>Thermo Scientific</i>
Hexafluoroisopropanol (HFIP), LC-MS grade	<i>Fluka</i>
Imidazole	<i>Carl Roth, Sigma Aldrich</i>
Isopropyl- $\beta$ -thiogalactosid	<i>Carl Roth</i>
Kanamycinsulfate	<i>Carl Roth</i>
Luria Bertani (LB) broth (Lennox)	<i>Carl Roth</i>
Magnesium chloride ( $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ )	<i>Alfa Aesar</i>
Manganese chloride ( $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ )	<i>Carl Roth</i>
PageRuler prestained protein ladder (10-180 kDa)	<i>Thermo Fisher Scientific</i>
PageRuler unstained protein ladder (10-200 kDa)	<i>Thermo Fisher Scientific</i>
Phenylmethylsulfonyl fluoride (PMSF)	<i>Thermo Fisher Scientific</i>
Rotiphorese sequencing gel concentrate	<i>Carl Roth</i>
Rotiphorese gel 30	<i>Carl Roth</i>
Sodium chloride	<i>Fisher Scientific</i>
Sodium dodecyl sulfate (SDS)	<i>Carl Roth</i>
Sodium hydrogen carbonate	<i>Carl Roth</i>
SYBR safe DNA stain	<i>Life Technologies</i>
Terrific broth (TB)	<i>Carl Roth</i>
Tetramethylenediamine (TEMED)	<i>Merck</i>
Triethylamine, LC-MS grade	<i>Fluka</i>
Tris(hydroxymethyl)aminoethane (Tris)	<i>Carl Roth</i>
Tween-20	<i>Merck</i>
Urea	<i>Merck</i>
Xylene cyanol	<i>Applichem</i>
$\beta$ -Mercaptoethanol	<i>Sigma Aldrich</i>

---

**Table 10: Buffers used in this study.**

<b>Buffer</b>	<b>Composition</b>
10x Antarctic phosphatase reaction buffer	500 mM bis-tris-propane-HCl, 10 mM MgCl <sub>2</sub> , 1 mM ZnCl <sub>2</sub> , pH 6.0
10x CutSmart buffer ( <i>New England Biolabs</i> )	500 mM potassium acetate, 200 mM Tris-acetate, 100 mM magnesium acetate, 1 mg mL <sup>-1</sup> BSA, pH 7.8
10x Lambda exonuclease reaction buffer ( <i>New England Biolabs</i> )	670 mM glycine-KOH, 25 mM MgCl <sub>2</sub> , 50 µg mL <sup>-1</sup> BSA, pH 9.4
10x SDS running buffer*	250 mM Tris, 2 M Glycin, 1% (w/v) SDS
10x T4 ligase buffer ( <i>New England Biolabs</i> )	500 mM Tris-HCl, 100 mM MgCl <sub>2</sub> , 100 mM DTT, 10 mM ATP, pH 7.8
10x T4 ligase buffer ( <i>Thermo Fisher</i> )	400 mM Tris-HCl, 100 mM MgCl <sub>2</sub> , 100 mM DTT, 5 mM ATP, pH 7.8
10x Taq reaction buffer*	100 mM Tris, 500 mM KCl, 1 % (v/v) Triton X-100, pH 9
10x TB*	890 mM Tris, 890 mM boric acid, 250 mM NaCl, pH 8.0
10x TBE*	890 mM Tris, 890 mM boric acid, 20 mM EDTA, pH 8.0
10x Thermopol reaction buffer ( <i>New England Biolabs</i> )	200 mM Tris-HCl, 100 mM (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , 100 mM KCl, 20 mM MgSO <sub>4</sub> , 0.1% Triton X-100, pH 8.8
2x dPAGE loading buffer (colorless)*	95% formamide in 20 mM EDTA, 8.3 M urea
nPAGE loading buffer (colorless)	50 % glycerol in ddH <sub>2</sub> O
4x Laemmli buffer*	250 mM Tris pH 6.8, 8% (w/v) SDS, 40% (v/v) glycerol, 20 % (v/v) β-mercaptoethanol, 0.2% (w/v) bromophenol blue
5x OneTaq reaction buffer ( <i>New England Biolabs</i> )	200 mM Tris-HCl, 220 mM NH <sub>4</sub> Cl, 220 mM KCl, 18 mM MgCl <sub>2</sub> , 0.6% IGEPAL CA-630, 0.5% Tween20, pH 8.9
6x Agarose loading buffer (colorless)*	50 mM Tris, 15% (w/v) Ficoll 400, 6 mM EDTA, pH 8.3

6x TriTrack gel loading dye ( <i>Thermo Fisher Scientific</i> )	10 mM Tris-HCl, 0.03 % bromophenol blue, 0.03 % xylene cyanol FF, 0.15 % orange G, 60 % glycerol, 60 mM EDTA, pH 7.6
Binding buffer*	20 mM HEPES, 300 mM NaCl, 2 mM DTT, 50% glycerol, pH 7.4
Dialysis buffer (high salt)*	20 mM HEPES, 600 mM NaCl, 2 mM DTT, pH 7.4
FPLC buffer A1*	10 mM Tris-HCl, 600 mM NaCl, 10% glycerol, 10 mM imidazole, pH 8.0
FPLC buffer A2*	50 mM Tris-HCl, 1.0 M NaCl, 10 mM imidazole, 10% glycerol, 5 mM $\beta$ -mercaptoethanol, pH 8.0
FPLC buffer B1*	10 mM Tris-HCl, 600 mM NaCl, 10% glycerol, 500 mM imidazole, pH 8.0
FPLC buffer B2*	50 mM Tris-HCl, 300 mM NaCl, 500 mM imidazole, 10% glycerol, 1 mM $\beta$ -mercaptoethanol, pH 8.0
Interaction buffer*	20 mM HEPES, 150 mM NaCl, 2 mM DTT, 0.05 % (v/v) Tween 20, pH 7.4
Lower separating gel buffer*	1.5 M Tris, pH 8.8, 0.4% (w/v) SDS
Lysis buffer 1*	10 mM Tris-HCl, 500 mM NaCl, 10% glycerol, pH 8.0
Lysis buffer 2*	50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, 10% glycerol, 5 mM $\beta$ -mercaptoethanol, pH 8.0
SDS-PAGE staining solution*	0.1% (w/v) Coomassie blue, 40% (v/v) EtOH, 10% (v/v) glacial acetic acid
SDS-PAGE destaining solution*	10% (v/v) EtOH, 7.5% (v/v) glacial acetic acid
TNA polymerase storage buffer*	10 mM Tris-HCl, 250 mM NaCl, 50% glycerol, pH 8.0
Upper stacking gel buffer*	0.5 M Tris pH 6.8, 0.4% (w/v) SDS
ZBP1 storage buffer*	20 mM HEPES, 300 mM NaCl, 2 mM DTT, 50% glycerol, pH 7.4

---

\* made in-house

### 5.3.4 Enzymes

Enzymes that have been used for experimental work in this study are listed in Table 11.

**Table 11: Enzymes used in this study**

<b>Enzyme</b>	<b>Supplier</b>
Antarctic phosphatase	<i>New England Biolabs</i>
DpnI	<i>New England Biolabs</i>
EcoRI-HF	<i>New England Biolabs</i>
Esp3I	<i>New England Biolabs</i>
HindIII-HF	<i>New England Biolabs</i>
Kod-RI polymerase	<i>made in-house</i>
Kod-RSGA polymerase	<i>made in-house</i>
Lambda exonuclease	<i>New England Biolabs</i>
OneTaq DNA polymerase	<i>New England Biolabs</i>
Phusion Flash High Fidelity PCR Master Mix	<i>Thermo Fisher Scientific</i>
SacI-HF	<i>New England Biolabs</i>
T4 DNA ligase	<i>New England Biolabs, Thermo Fisher Scientific</i>
T4 polynucleotide kinase	<i>New England Biolabs</i>
Taq DNA polymerase	<i>made in-house</i>
Therminator polymerase	<i>New England Biolabs</i>

### 5.3.5 Bacterial cells, vectors, and plasmids

Bacterial cells, vectors, genes and plasmids that have been used in this study are listed in Tables 12, 13, and 14.

**Table 12: Bacterial cells used in this study.**

<b>Product</b>	<b>Supplier</b>
NEB10-beta (chemically competent)	<i>New England Biolabs</i>
OneShot TOP10 (chemically competent)	<i>Thermo Fisher Scientific</i>
Stellar (chemically competent)	<i>Takara Bio</i>
T7 Express <i>lysY</i> <sup>q</sup> (chemically competent)	<i>New England Biolabs</i>
NEBExpress® I <sup>q</sup> (chemically competent)	<i>New England Biolabs</i>

**Table 13: Vectors used in this study.**

<b>Vector</b>	<b>Supplier</b>
pQE80HT	<i>kindly provided by the Famulok group (University of Bonn)</i>
pASG-IBA35	<i>IBA Lifescience</i>

**Table 14: Plasmids and gene strings used in this study.**

<b>Plasmid</b>	<b>Supplier</b>
pET28a(+)_mZBP1short <sup>WT</sup>	<i>Twist Bioscience</i>
pET28a(+)_mZBP1short <sup>mZα1</sup>	<i>Twist Bioscience</i>
pET28a(+)_mZBP1short <sup>mZα2</sup>	<i>Twist Bioscience</i>
pET28a(+)_mZBP1short <sup>mZα1-2</sup>	<i>Twist Bioscience</i>
Kod-RI gene string	<i>Thermo Fisher GeneArt</i>
pMA-RQ_Kod-RI	<i>Thermo Fisher GeneArt</i>
pQE80HT_Kod-RI	<i>made in-house</i>
pQE80HT_Kod-RI_2313delA	<i>made in-house</i>
pQE80HT_Kod-RSGA	<i>made in-house</i>

### 5.3.6 Kits

The kits that have been used for experimental work in this study are listed in Table 15.

**Table 15: Kits used in this study.**

<b>Product</b>	<b>Supplier</b>
InFusion HD EcoDry Cloning Kit	<i>Takara Bio</i>
NucleoSpin Gel and PCR clean-up kit	<i>Macherey Nagel</i>
NucleoSpin plasmid	<i>Macherey Nagel</i>
Pierce 660 nm Protein Assay Kit	<i>Thermo Fisher</i>

## 5.3.7 General biochemical methods

### 5.3.7.1 Preparation of DNA templates and primers

The dNaM cyanoethyl phosphoramidite, suitably protected for oligonucleotide synthesis was purchased from *Berry & Associates Inc., USA*. The dTPT3 phosphoramidite (**36**), suitably protected for oligonucleotide synthesis was synthesized according to standard protocols<sup>206,207</sup> (see Chapters 3.1.2, 5.2.7.21 and 5.2.7.22). Both phosphoramidites were provided to *Ella Biotech, Germany*. All modified oligonucleotides were synthesized by *Ella Biotech, Germany*. Primers and unmodified DNA oligonucleotides were synthesized by *Biomers, Germany*. Purchased oligonucleotides were received as lyophilized samples that were dissolved in ddH<sub>2</sub>O upon receipt to obtain 100  $\mu$ M stock solutions. All DNA primers and templates were stored at -20 °C.

### 5.3.7.2 Gel electrophoresis

#### Denaturing polyacrylamide gel electrophoresis

Oligonucleotides were analyzed by denaturing polyacrylamide gel electrophoresis using analytical dPAGE with approximately 5 mL gel volume. For analytical dPAGE separation, solution A (Rotiphorese sequencing gel concentrate), solution B (8.3 M urea, *homemade*) and solution C (8.3 M urea in 10 $\times$  TBE, *homemade*) were combined as depicted in Table 16. Polymerization was initiated with 10 % APS and TEMED and gels were allowed to polymerize for at least 60 min at room temperature.

**Table 16: Composition of denaturing polyacrylamide gels (5 mL total volume) with different polyacrylamide concentrations.** Solution A: 25 % acrylamide/bisacrylamide 19:1 and 50 % urea, solution B: 8.3 M urea, solution C: 8.3 M urea in 10 $\times$  TBE.

Gel type	A [mL]	B [mL]	C [mL]	10 % APS [ $\mu$ L]	TEMED [ $\mu$ L]
20 % dPAGE	4	0.5	0.5	40	4
15 % dPAGE	3	1.5	0.5	40	4

Subsequently, 2 $\times$  dPAGE loading buffer (colorless) was added to the samples in equal ratio followed by denaturation by heating to 95 °C for 5-10 min prior to loading onto the gel. GeneRuler Ultra Low Range DNA Ladder was used as a reference for analytical characterization. 1 $\times$  TBE was employed as running buffer. Gel electrophoresis was performed at 300 V for 45-75 min depending on the sequence length of the oligonucleotide. Fluorescently-labeled oligonucleotides were visualized by fluorescence scanning using a gel documentation system ( $\lambda_{\text{excitation}} = 477 \text{ nm}$ ,  $\lambda_{\text{emission}} = 535 \text{ nm}$ ). Unlabeled oligonucleotides were

stained with SYBR safe DNA stain and visualized by UV illumination using a gel documentation system.

#### Native polyacrylamide gel electrophoresis

Oligonucleotide analysis or electrophoretic mobility shift assays were performed by native polyacrylamide gel electrophoresis using analytical nPAGE with approximately 5 mL gel volume. For nPAGE separation, solution A (Rotiphorese gel 30), solution B (10× TB, *homemade*) and ddH<sub>2</sub>O (component C) were combined as depicted in Table 17. Polymerization was initiated with 10 % APS and TEMED and gels were allowed to polymerize for at least 60 min at room temperature.

**Table 17: Composition of 8 % native polyacrylamide gels (5 mL total volume).** Solution A: 30 % acrylamide/bisacrylamide 37.5:1, solution B: 10× TB, component C: ddH<sub>2</sub>O.

Gel type	A [mL]	B [mL]	C [mL]	10 % APS [μL]	TEMED [μL]
8 % nPAGE	1.3	0.5	3.1	65	4

Subsequently, 2× nPAGE loading buffer was added to the samples in equal ratio and samples were loaded onto the gel. 1× TB was employed as running buffer. Gel electrophoresis was performed at 60 V for 60 min at 4 °C. Fluorescently-labeled oligonucleotides were visualized by fluorescence scanning using a gel documentation system ( $\lambda_{\text{excitation}} = 477 \text{ nm}$ ,  $\lambda_{\text{emission}} = 535 \text{ nm}$ ).

#### SDS polyacrylamide gel electrophoresis

Proteins were analyzed via denaturing SDS polyacrylamide gel electrophoresis. For the preparation of the separating gel (approximately 5 mL gel volume), solution A (Rotiphorese gel 30) was mixed with solution B (lower separating gel buffer) and water as depicted in Table 18. Polymerization was initiated by addition of 10 % APS and TEMED. Subsequently, the gel was casted and overlaid with isopropanol, followed by polymerization for at least 60 min. After polymerization, the isopropanol was discarded.

**Table 18: Composition of denaturing SDS polyacrylamide separating gels (5 mL total volume) with different polyacrylamide concentrations.** Solution A: 30 % acrylamide/ bisacrylamide 37.5:1, solution B: 1.5 M Tris-HCl, 1.4 mM SDS; pH 8.8.

Gel type	A [mL]	B [mL]	H <sub>2</sub> O [mL]	APS [μL]	TEMED [μL]
10 % SDS-PAGE	1.65	1.25	2.1	25	4
15 % SDS-PAGE	2.5	1.25	1.25	25	4

Subsequently, the stacking gel solution was prepared by mixing 7 mL of solution A with 12.5 mL upper stacking gel buffer (0.5 M Tris-HCl, 0.4 % (w/v) SDS; pH 6.5) and 30 μL 2 % bromophenol blue solution. 2 mL of the prepared stacking gel solution were mixed with 10 μL



10 % APS and 8  $\mu$ L TEMED. The stacking gel was poured on top of the separating gel and was allowed to polymerize for 60 min. Then, 4 $\times$  Laemmli buffer was added to the protein samples to a final concentration of 1 $\times$  and samples were denatured at 95  $^{\circ}$ C for 5 min. Either PageRuler prestained protein ladder (10-180 kDa) or PageRuler unstained protein ladder (10-200 kDa) was used as a size reference. Gel electrophoresis was performed at 175 V for 60 min in 1 $\times$  SDS running buffer. For protein visualization, the gel was heated in SDS-PAGE staining solution in the microwave for 1 min, followed by incubation on a shaking plate (~65 rpm) at r.t. for 20 min. Afterwards, destaining was performed by heating the gel in SDS-PAGE destaining solution in the microwave for 1 min, followed by incubation on a shaking plate (~65 rpm) at r.t. for 10 min. The destaining step was repeated one more time. Lastly, the gel was incubated in H<sub>2</sub>O by heating in the microwave for 1 min followed shaking (~65 rpm) at r.t. for 20 min. SDS-PAGE gels were imaged using a gel documentation system in chemiluminescent mode in combination with a light panel.

#### Agarose gel electrophoresis

For agarose gel electrophoresis, a 1 % agarose solution was prepared by dissolving 2 g agarose in 200 mL 0.5  $\times$  TBE, followed by heating to near-boiling temperatures in a microwave. Agarose solutions were stored at 60  $^{\circ}$ C in a heating cabinet. Prior to casting, 50 mL of agarose solution were supplemented with 5  $\mu$ L of SYBR safe DNA gel stain. Subsequently, the gel was casted and solidified for at least 30 min at room temperature. Samples were mixed with 6 $\times$  TriTrack gel loading dye to a final concentration of 1 $\times$ . GeneRuler 1 kb Plus DNA Ladder was used as an internal standard. Gel electrophoresis was performed for 30 min at 130 V in 0.5 $\times$  TBE. Agarose gels were visualized by UV illumination using a gel documentation system.

### **5.3.7.3 Oligonucleotide purification**

#### Centrifugal filters

For purification, buffer exchange, or salt removal from oligonucleotide samples, *Amicon 3K Ultra 0.5 mL* centrifugal filters were employed. Oligonucleotide solutions were applied and centrifuged at 14000  $\times$  g for 15 min at 4  $^{\circ}$ C. 450  $\mu$ L ddH<sub>2</sub>O were added and passed through the membrane by centrifugation at 14000  $\times$  g for 15 min at 4  $^{\circ}$ C. This step was repeated twice. Subsequently, 450  $\mu$ L ddH<sub>2</sub>O were added and passed through the membrane by centrifugation at 14000  $\times$  g for 45 min at 4  $^{\circ}$ C. Elution was carried out by upside-down centrifugation at 1000  $\times$  g for 2 min.

### Silica-membrane spin columns

For the extraction of oligonucleotides from preparative agarose gels or for the purification of PCR reactions, the *Nucleospin Gel and PCR clean-up kit* was used according to the manufacturer's protocol. To perform DNA extraction from agarose gels, SYBR safe-stained nucleic acid fragments were visualized either using a blue LED transilluminator or UV light. The DNA fragments were excised from the gel and the weight of the gel slice was determined. For each 100 mg of agarose gel, 200  $\mu\text{L}$  of Buffer NT1 were added and the samples were incubated for 5-10 min at 50°C while intermittently vortexing until the gel was completely dissolved. For PCR clean up, one volume of sample was mixed with two volumes of Buffer NT1. Subsequently, the prepared oligonucleotide samples from gel extractions or PCR reactions were handled according to the same protocol. To bind DNA to the silica membrane, up to 700  $\mu\text{L}$  of the prepared samples were loaded onto a Nucleospin Gel and PCR Clean-up column. Samples were centrifuged for 30 s at 11000  $\times$  g and the flow-through was discarded. For larger sample volumes, the columns were loaded with remaining sample multiple times and the centrifugation step was repeated. Subsequently, the silica membrane was washed with 700  $\mu\text{L}$  of EtOH-based Buffer NT3 followed by centrifugation at 11000  $\times$  g for 30 s. The flow-through was discarded. This washing step was repeated. To dry the silica membrane and remove residual EtOH, the column was first centrifuged for 1 min at 11000  $\times$  g followed by heating to 70 °C for 5 min. Elution was performed in two steps by adding 15  $\mu\text{L}$  ddH<sub>2</sub>O to the membrane, followed by incubation at 70 °C for 5 minutes. Subsequently oligonucleotides were eluted by centrifugation at 11000  $\times$  g for 1 min. The elution step was repeated one more time.

### Ethanol precipitation

To purify oligonucleotides by precipitation, a 9:1 (v/v) solution of ethanol and aq. sodium acetate (3.0 M) was added to the sample with six-fold excess. The mixture was vortexed thoroughly and incubated overnight at -20 °C. Subsequently, precipitated nucleic acids were centrifuged at 15000  $\times$  g for 15 min at 4 °C. The supernatant was removed and the pellet was washed with ice-cold 90 % EtOH and vortexed thoroughly. Centrifugation at 15000  $\times$  g at 4 °C was again performed for 10 min and the supernatant was discarded. The washing step was repeated. The pellet was then dried at 70 °C for 10 min to remove residual EtOH. The pellet was redissolved in an appropriate amount of ddH<sub>2</sub>O by heating to 65 °C for 10 min while intermittently vortexing.

#### 5.3.7.4 Determination of nucleic acid concentration

Concentrations of nucleic acid samples were determined by measuring the absorption at 260 nm ( $A_{260}$ ) using a spectrophotometer. Concentrations were obtained from the  $A_{260}$  value and software-assisted calculation with OligoCalc<sup>417</sup> employing the native oligonucleotide sequence.

#### 5.3.7.5 $\lambda$ -exonuclease digest

To facilitate HPLC-MS analysis of oligonucleotide products from primer extension assays, 5'-phosphate-modified DNA templates were digested with  $\lambda$ -exonuclease. Therefore, purified primer extension assays were mixed with 5  $\mu$ L of 10 $\times$  lambda exonuclease buffer and 1  $\mu$ L of  $\lambda$ -exonuclease in a total volume of 50  $\mu$ L. The digestion reaction was incubated at 37 °C for 30 min, followed by heat inactivation of the enzyme for 10 min at 75 °C. After digestion, the samples were purified using centrifugal filters (Chapter 5.3.7.3).

#### 5.3.7.6 HPLC-MS analysis of oligonucleotides

HPLC-MS analysis of oligonucleotides was performed to investigate the incorporation of non-canonical nucleotides into a nascent nucleic acid strand. HPLC-ESI mass spectrometry measurements (negative mode) of purified oligonucleotide samples were performed on a *Bruker Daltonics amaZon SL* mass spectrometer in combination with a *Bruker Elute SP* HPLC system.

A *Zorbax StableBond* column (80 Å, 2.1  $\times$  50 mm, 5  $\mu$ m, C18) or an *XTerra MS* column (2.1  $\times$  100 mm, 5  $\mu$ m, C18) was used as the stationary phase. Samples were generally submitted to HPLC-MS using 10 mM triethylamine/100 mM hexafluoroisopropanol as solvent A and MeCN as solvent B. When the *Xterra* column was used as the stationary phase, a gradient of 3% B for 9 min followed by 3%  $\rightarrow$  40% B in 11 min at a flow rate of 0.2 mL min<sup>-1</sup> was applied. When the *Zorbax* column was used as the stationary phase, a gradient of 3%  $\rightarrow$  20% B within 20 min at a flow rate of 0.4 mL min<sup>-1</sup> was used. Crude ESI<sup>-</sup> spectra obtained by integration from a suitable time frame within the respective UV trace were evaluated via *Compass DataAnalysis* using suitable deconvolution parameters (low mass: 400, high mass: 25000, max. charge = 20, max. number of compounds = 10, min. 5 peaks per compound).

### 5.3.8 General bacterial cell culture methods

Bacterial cell culture was used to produce plasmids containing Kod-RI and Kod-RSGA genes by different cloning approaches. Additionally, overexpression of genes to produce recombinant proteins was performed to yield TNA polymerases as well as mZBP1 variants.

#### 5.3.8.1 Preparation of genes, vectors and plasmids

Gene synthesis of the Kod-RI polymerase coding sequence and construction of the pMA-RQ\_Kod-RI plasmid was performed by *ThermoFisher GeneArt (Germany)*. Upon receipt, the freeze-dried aliquots of the Kod-RI gene string and the pMA-RQ\_Kod-RI plasmid were dissolved in ddH<sub>2</sub>O to obtain stock solutions of 10 ng  $\mu\text{L}^{-1}$ . The pASG-IBA 35 vector was purchased as a freeze-dried aliquot from IBA lifesciences (*Germany*) and was dissolved in ddH<sub>2</sub>O to a final concentration of 250 ng  $\mu\text{L}^{-1}$  upon receipt. The pQE80HT vector was generously provided as a 25 ng  $\mu\text{L}^{-1}$  stock by the Famulok group (*University of Bonn*). All vectors were stored at -20 °C. pET28a+ plasmids bearing coding sequences of mZBP1<sup>WT</sup>, mZBP1<sup>mZ $\alpha$ 1</sup>, mZBP1<sup>mZ $\alpha$ 2</sup>, and mZBP<sup>mZ $\alpha$ 1-2</sup> were purchased from *Twist Bioscience (USA)* and were kindly provided as glycerol stocks in T7 Express *lysY/l<sup>q</sup>* *E.coli* cells by the Pasparakis group (*University of Cologne*). Bacterial glycerol stocks were stored at -80 °C.

#### 5.3.8.2 Antibiotics

When working with bacterial cell culture, plasmids carry antibiotic resistance genes, hence media and agar plates were supplemented with antibiotics to grow only plasmid-containing bacteria. The antibiotic concentrations used in this study are depicted in Table 19.

**Table 19: Overview of working concentrations for different antibiotics for bacterial cultivation used in this study.**

Antibiotic	Working concentration
Kanamycin	50 $\mu\text{g mL}^{-1}$
Ampicillin	100 $\mu\text{g mL}^{-1}$

*E.coli* cells harboring the pQE80HT vector were generally cultivated on agar plates and in liquid LB medium supplemented with kanamycin. *E.coli* cells harboring the pET28a(+) vector were generally cultivated in liquid TB medium supplemented with kanamycin. *E.coli* cells harboring the pASG-IBA 35 vector were cultivated on agar plates and in LB medium supplemented with ampicillin.

### 5.3.8.3 Transformation

#### Transformation of NEB 10-beta competent *E.coli*

For transformation of plasmid DNA, a 50  $\mu\text{L}$  aliquot of NEB 10-beta chemically competent cells was thawed on ice for 10 min. Cells were carefully mixed with an appropriate amount of plasmid or ligation mixture and incubated on ice for 30 min. Heat shock was performed for 30 s at 42 °C, followed by incubation on ice for 5 min. Subsequently, 950  $\mu\text{L}$  of room temperature NEB 10-beta stable outgrowth medium (supplied with the cells) was added and cells were incubated for 120 min at 37 °C and 300 rpm shaking. 20  $\mu\text{L}$ , 50  $\mu\text{L}$  and 100  $\mu\text{L}$  of cell suspension were spread on pre-warmed selection plates supplemented with the appropriate antibiotic (see Chapter 5.3.8.2). Furthermore, the remaining cells were centrifuged at 6000  $\times$  g for 5 min, 900  $\mu\text{L}$  of the supernatant were discarded and the cell pellet was resuspended and also spread onto an agar plate. Agar plates with bacteria were incubated at 37 °C overnight. The next day, bacterial liquid cultures were prepared (Chapter 5.3.8.4).

#### Transformation of Stellar competent *E.coli*

For transformation of plasmid DNA, a 50  $\mu\text{L}$  aliquot of Stellar chemically competent cells was thawed on ice for 10 min. Cells were carefully mixed with an appropriate amount of plasmid or ligation mixture and incubated on ice for 30 min. Heat shock was performed for 30 s at 45 °C, followed by incubation on ice for 2 min. Subsequently, 450  $\mu\text{L}$  of room temperature stable outgrowth medium (supplied with the cells) was added and cells were incubated for 60 min at 37 °C and 300 rpm shaking. 20  $\mu\text{L}$ , 50  $\mu\text{L}$  and 100  $\mu\text{L}$  of cell suspension were spread on pre-warmed selection plates supplemented with the appropriate antibiotic (see Chapter 5.3.8.2). Furthermore, the remaining cells were centrifuged at 6000  $\times$  g for 5 min, 350  $\mu\text{L}$  of supernatant were discarded and the cell pellet was resuspended and also spread onto an agar plate. Agar plates with bacteria were incubated at 37 °C overnight. The next day, bacterial liquid cultures were prepared (Chapter 5.3.8.4).

#### Transformation of OneShot TOP10 competent *E.coli*

For transformation of plasmid DNA, a 50  $\mu\text{L}$  aliquot of OneShot TOP10 chemically competent cells was thawed on ice for 10 min. Cells were carefully mixed with an appropriate amount of plasmid or ligation mixture and incubated on ice for 30 min. Heat shock was performed for 30 s at 42 °C, followed by incubation on ice for 5 min. Subsequently, 900  $\mu\text{L}$  of room temperature LB medium was added and cells were incubated for 60 min at 37 °C and 300 rpm shaking. Agar plates were treated with X-gal before spreading the bacteria, therefore 15  $\mu\text{L}$  X-gal (100 mg  $\text{mL}^{-1}$ ) were mixed with 35  $\mu\text{L}$  ddH<sub>2</sub>O and spread on top of the agar. X-gal was allowed to diffuse into the agar for 1 h. Subsequently, 20  $\mu\text{L}$ , 50  $\mu\text{L}$  and 100  $\mu\text{L}$  of cell suspension were spread on pre-warmed selection plates supplemented with the appropriate antibiotic (see

Chapter 5.3.8.2). Furthermore, the remaining cells were centrifuged at 6000 × g for 5 min, 850 µL of the supernatant were discarded and the cell pellet was resuspended and also spread onto an agar plate. Agar plates with bacteria were incubated at 37 °C overnight. The next day, bacterial liquid cultures were prepared from white colonies (Chapter 5.3.8.4).

#### Transformation of NEBExpress® I<sup>q</sup> competent *E.coli*

For transformation of plasmid DNA, a 50 µL aliquot of NEBExpress® I<sup>q</sup> chemically competent cells was thawed on ice for 10 min. Cells were carefully mixed with an appropriate amount of plasmid and incubated on ice for 30 min. Heat shock was performed for 20 s at 42 °C, followed by incubation on ice for 5 min. Subsequently, 950 µL of room temperature stable outgrowth medium (supplied with the cells) was added and cells were incubated for 60 min at 37 °C and 300 rpm shaking. 50 µL and 100 µL of cell suspension were spread on pre-warmed selection plates supplemented with the appropriate antibiotic (see Chapter 5.3.8.2). Furthermore, the remaining cells were centrifuged at 6000 × g for 5 min, 900 µL of the supernatant were discarded and the cell pellet was resuspended and also spread onto an agar plate. Agar plates with bacteria were incubated at 37 °C overnight. The next day, bacterial liquid cultures were prepared (Chapter 5.3.8.4).

### **5.3.8.4 Bacterial cultivation for plasmid preparation**

For bacterial cultivation, LB medium was inoculated with single clones picked from previously prepared agar plates (Chapter 5.3.8.3). Clones were transferred into 10 mL LB medium supplemented with an appropriate antibiotic (Chapter 5.3.8.2). Subsequently, cultures were incubated overnight at 37 °C and 140 rpm shaking. The next day, bacterial glycerol stocks were prepared (Chapter 5.3.8.5) and plasmid preparation was performed (Chapter 5.3.8.4).

### **5.3.8.5 Bacterial glycerol stock preparation**

For the preparation of bacterial glycerol stocks, 750 µL of 50 % glycerol were cooled on ice and then mixed with 750 µL saturated *E.coli* overnight culture (Chapter 5.3.8.4). Bacterial glycerol stocks were stored at -80 °C.

### **5.3.8.6 Plasmid preparation**

Isolation of plasmid DNA was performed using *Nucleospin plasmid kit* according to the manufacturer's protocol. Therefore, *E.coli* cells were harvested by centrifugation of a saturated

10 mL *E.coli* culture (Chapter 5.3.8.4) at 11000 × g for 1 min at r.t. The supernatant was removed and the pellet was resuspended in 500 µL of Resuspension Buffer A1 containing RNase A by vigorous vortexing. To induce cell lysis 500 µL of Lysis Buffer A2 were added and samples were inverted 8 times and incubated at r.t. for 5 min. Subsequently, Neutralization Buffer A3 was added and tubes were inverted 8 times to precipitate all proteins and chromosomal DNA. Samples were centrifuged 10 min at 11000 × g at r.t. to clarify the lysate. To bind plasmid DNA to the silica membrane, up to 700 µL of the clarified supernatant were loaded onto a Nucleospin plasmid isolation column. Samples were centrifuged for 1 min at 11000 × g and the flow-through was discarded. For larger sample volumes, the columns were loaded with the remaining sample multiple times and the centrifugation step was repeated. Subsequently, the silica membrane was washed with 500 µL Wash Buffer AW (preheated to 50 °C) followed by centrifugation at 11000 × g for 1 min. The flow-through was discarded. A second membrane washing step was performed by adding 600 µL of EtOH-based Wash Buffer A4 to the membrane followed by centrifugation at 11000 × g for 1 min. The flow-through was discarded. To dry the silica membrane and remove residual EtOH, the column was first centrifuged for 2 min at 11000 × g followed by heating to 70 °C for 5 min. Elution was performed in two steps by adding 25 µL ddH<sub>2</sub>O to the membrane and incubation at 70 °C for 5 minutes, followed by centrifugation at 11000 × g for 1 min. The elution step was repeated one more time.

### 5.3.8.7 Colony PCR

Colony PCR was performed to verify that single clones contained the Kod-RI polymerase coding sequence. Therefore, the isolated plasmid from the plasmid preparation (Chapter 5.3.8.6) was diluted with ddH<sub>2</sub>O to a final concentration of 2 ng µL<sup>-1</sup> and used as a template in a PCR reaction. Primers for colony PCR were designed to flank the insert of interest at its 5'- and 3'-end. Colony PCR was performed in 20 µL scales with final concentrations of 1 × Taq buffer, 2 mM MgCl<sub>2</sub>, 500 µM of each dNTP, 1 ng of template, 0.5 µM of Kod-RI\_ColonyPCR\_fw primer and 0.5 µM Kod-RI\_ColonyPCR\_rv primer, and 0.1 U µL<sup>-1</sup> of Taq polymerase. Table 20 provides a detailed pipetting scheme.

**Table 20: General pipetting scheme for colony PCRs.** The stock concentrations of the isolated plasmids and the used volumes vary between the samples.

Component	Stock concentration	Final concentration	V [ $\mu\text{L}$ ]
Isolated plasmid	varied	1.0 ng	varied
Kod-RI_ColonyPCR_fw	100 $\mu\text{M}$	0.5 $\mu\text{M}$	0.1
Kod-RI_ColonyPCR_rv	100 $\mu\text{M}$	0.5 $\mu\text{M}$	0.1
dNTPs	25 mM each	500 $\mu\text{M}$	0.4
MgCl <sub>2</sub>	100 mM	2 mM	0.4
Taq Reaction Buffer	10 $\times$	1 $\times$	2
Taq polymerase	5 U $\mu\text{L}^{-1}$	0.1 U $\mu\text{L}^{-1}$	0.4
ddH <sub>2</sub> O	-	-	ad 20

PCR cycling was performed with an initial denaturing step at 94 °C for 2 min, followed by 30 cycles of denaturing for 30 s at 94 °C, annealing at 54 °C for 40 s, and elongation at 72 °C for 2.5 min. A final elongation step of 5 min was performed before cooling to 4 °C at the end.

Table 21 provides an overview of the general cycling conditions.

**Table 21: General cycling conditions for colony PCR employed in this study.**

	T [°C]	t [s]
Initial denaturation	94	120
<u>30 cycles:</u>		
<b>Denaturation</b>	94	30
<b>Annealing</b>	54	40
<b>Elongation</b>	72	150
Final Elongation	72	300
Hold	4	$\infty$

Colony PCR reactions were analyzed regarding the presence of the insert and its size using 1 % agarose gel electrophoresis (Chapter 5.3.7.2) and sent for sequencing (Chapter 5.3.8.9).

### 5.3.8.8 Diagnostic restriction enzyme digest

Diagnostic restriction enzyme digests of pQE80HT\_Kod-RIS, pQE80HT\_Kod-RISG, pQE80HT\_Kod-RSG and pQE80HT\_Kod-RSGA plasmids were performed to determine if the clones contained the insert of interest. Therefore, EcoRI was selected as a restriction enzyme as it is a dual cutter for these plasmids, with one cleavage site located in the insert and one in the vector backbone, producing a distinct pattern of two easily resolved bands of 1767 bp and 5114 bp. Diagnostic restriction enzyme digests were performed in 10  $\mu\text{L}$  scale. ~400 ng of



plasmid were digested in 1× CutSmart buffer utilizing 1 U  $\mu\text{L}^{-1}$  of EcoRI-HF. Restriction enzyme digests were incubated for 2 h at 37 °C and then analyzed by 1 % agarose gel electrophoresis (Chapter 5.3.7.2) and sent for sequencing (Chapter 5.3.8.9).

### 5.3.8.9 Sanger sequencing

To validate the sequence integrity of the cloned genes, isolated plasmids that were positive clones according to colony PCRs or diagnostic restriction enzyme digests (Chapters 5.3.8.7 and 5.3.8.8) were sent for sequencing. Sanger sequencing was performed by *GATC Services (Eurofins Genomics, Germany)*.

### 5.3.8.10 Pierce 660 nm Protein-Assay-Kit

To determine protein concentrations of expressed and purified proteins, the *Pierce 660 nm Protein Assay Kit* was used according to the manufacturer's protocol. This protein assay is compatible with high levels of glycerol as well as DTT in storage buffers. As a first step, a BSA standard curve with a working range of 25 – 2000  $\mu\text{g mL}^{-1}$  was prepared using the pre-diluted BSA standards supplied with the kit. Therefore 100  $\mu\text{L}$  of each standard were mixed with 1.5 mL of protein assay reagent, vortexed thoroughly and incubated for 5 minutes at ambient temperatures. Subsequently, the spectrophotometer was set to a wavelength of 660 nm and was blanked with a cuvette filled with water. Then, the samples containing the protein standards were measured. The acquired absorbance at 660 nm for each BSA standard was plotted versus its concentration in  $\mu\text{g mL}^{-1}$  in a coordinate system. The standard curve and its corresponding equation was obtained by linear regression.

Subsequently, the protein concentration of the unknown protein samples was determined. Therefore, 100  $\mu\text{L}$  of unknown protein sample were mixed with 1.5 mL of assay reagent, followed by thorough vortexing and incubation at ambient temperatures for 5 min. After incubation, the samples were also measured at 660 nm. Then, the previously prepared standard curve was used to determine the protein concentration of the unknown sample by interpolation.

### 5.3.9 PCR-based cloning of Kod-RI TNA polymerase

For cloning of the Kod-RI gene into an expression vector, the Kod-RI coding sequence was designed based on the published protein sequence of the Kod-RI polymerase.<sup>151</sup> The sequence was then codon optimized for expression in *E.coli* and purchased from *Thermo*

*Fisher GeneArt* as a gene string and as a plasmid in which the gene was inserted into the pMA-RQ vector. Different cloning approaches were performed in this study including AQUA cloning, In-Fusion cloning, restriction enzyme cloning and Stargate cloning. To facilitate these cloning strategies, the gene of interest needs to meet certain criteria. For AQUA cloning, the gene of interest is designed to have short overhangs consisting of regions of homology to the 3'- and 5'-ends of a linearized vector. 25 bp overhangs are recommended for AQUA cloning.<sup>375</sup> Therefore, the purchased Kod-RI gene string was designed to contain flanking extensions of 25 bp of homologous sequence overlaps to the PCR linearized pQE80HT vector specifically provided for AQUA cloning. However, for In-Fusion cloning where shorter overhangs of 15 bp are recommended, or when AQUA cloning and In-Fusion cloning is performed into a different vector, the overhangs must be modified accordingly. In addition, restriction enzyme cloning as well as Stargate cloning require the insertion of restriction enzyme binding sites at the 3'- and 5'-ends of the gene that match those present in the multiple cloning site of the vector. Hence, the purchased Kod-RI gene was amplified by PCR using primers specifically designed to add the desired flanking extensions for the particular cloning approach, respectively.

### **5.3.9.1 Retransformation of pQE80HT and pASG-IBA vectors**

To amplify the pQE80HT and pASG-IBA35 vectors for subsequent use in cloning approaches, the vectors were transformed into competent *E.coli* cells. Therefore, 50 ng of pQE80HT vector were transformed into chemically competent Stellar *E.coli* cells and 50 ng of pASG-IBA35 vector were transformed into chemically competent NEB10-beta *E.coli* cells according to the protocols described in Chapter 5.3.8.3. Then, single colonies were picked and bacteria were cultivated for plasmid preparation (Chapter 5.3.8.4). Bacterial glycerol stocks were prepared (Chapter 5.3.8.5) before plasmid isolation (Chapter 5.3.8.4). The concentrations of the isolated plasmids were determined using a spectrophotometer (Chapter 5.3.7.4). Clones were sent for sequencing (Chapter 5.3.8.9) using Seq\_pQE80HT\_1\_fw and Seq\_pQE80HT\_4\_rv primers for sequencing of the multiple cloning site of the pQE80HT vector and using Seq\_pASG-IBA35\_1\_fw and Seq\_pASG-IBA35\_4\_rv primers for sequencing of the multiple cloning site of the pASG-IBA35 vector.

### **5.3.9.2 PCR amplification of the Kod-RI gene for cloning**

Kod-RI gene amplification was performed with Phusion Flash High Fidelity PCR Master Mix. PCR was performed in 20  $\mu$ L scale with final concentrations of 1  $\times$  Phusion Flash High Fidelity PCR Master Mix, 1 ng of either Kod-RI gene or pMA-RQ\_Kod-RI as template and 0.5  $\mu$ M forward and reverse primers, respectively. A detailed pipetting scheme is depicted in Table 22.

**Table 22: General pipetting scheme for the amplification of the Kod-RI gene with Phusion Flash High Fidelity PCR Master Mix.** Either the Kod-RI gene or the pMA-RQ\_Kod-RI plasmid was used as a template. Different primers were used in order to introduce the desired overhangs for the different cloning approaches.

Component	Stock concentration	Final concentration	V [ $\mu$ L]
Template	10 ng $\mu$ L <sup>-1</sup>	1.0 ng	0.1
Forward primer	100 $\mu$ M	0.5 $\mu$ M	0.1
Reverse primer	100 $\mu$ M	0.5 $\mu$ M	0.1
Phusion Flash PCR Master Mix	2 x	1 x	10
ddH <sub>2</sub> O	-	-	9.7

For amplification of the Kod-RI gene containing 25 bp overhangs for AQUA cloning into the pQE80HT vector or for amplification of Kod-RI gene containing 15 bp overhangs for In-Fusion cloning into the pQE80HT vector, PCR cycling was performed with an initial denaturing step at 98 °C for 30 s, followed by 25 cycles of denaturing for 10 s at 98 °C, annealing at the individual temperature for each primer/template set for 5 s and elongation at 72 °C for 48 s. Eventually, a final elongation step was performed at 72 °C for 48 s before cooling to 4 °C at the end. Table 23 provides the detailed PCR cycling conditions for the primer sets used during PCR. For amplification of the template for AQUA cloning, the Kod-RI gene was used as a template and the Kod-RI\_ampli\_25bp\_pQE80HT\_fw and Kod-RI\_ampli\_25bp\_pQE80HT\_rv were used as primers. To shorten the overhangs to 15 bp for in-fusion cloning, the Kod-RI gene was used as a template and the Kod-RI\_ampli\_15bp\_pQE80HT\_fw and Kod-RI\_ampli\_15bp\_pQE80HT\_rv\_ were used as primers.

**Table 23: PCR cycling conditions using Phusion Flash High Fidelity PCR Master Mix for the amplification of the Kod-RI gene with 25 bp overhangs for AQUA cloning and 15 bp overhangs for In-Fusion cloning into the pQE80HT vector.**

Overhangs	T [°C]	t [s]
Initial denaturation	98	30
<u>25 cycles:</u>		
<b>Denaturation</b>	98	10
<b>Annealing</b>	25 bp - pQE80HT	5
	15 bp - pQE80HT	
<b>Elongation</b>	72	48
Final Elongation	72	48
Hold	4	$\infty$

When new sequences were added to the Kod-RI coding sequence as overhangs, the total number of 25 cycles was divided into five initial cycles at lower annealing temperatures corresponding to the melting temperature of the primer without considering the introduced

flanking overhangs, followed by 20 cycles at higher annealing temperatures corresponding to the melting temperature of the primer including the overhangs. Table 24 provides the detailed cycling conditions for the respective primer sets used. To introduce restriction enzyme recognition sequences (RE overhangs) for *SacI*-HF and *HindIII*-HF restriction enzymes at the 3'- and 5'-ends of the *Kod-RI* gene, the purchased *Kod-RI* gene string was used as a template using *Kod-RI\_ampli\_REcloning\_pQE80HT\_fw* and *Kod-RI\_ampli\_REcloning\_pQE80HT\_rv* as primers. To introduce overhangs compatible with Stargate cloning (Stargate overhangs), the *pMA-RQ\_Kod-RI* plasmid was used as a template employing *Kod-RI\_ampli\_Stargate\_pASG-IBA35\_fw* and *Kod-RI\_ampli\_Stargate\_pASG-IBA35\_rv* as primers. To allow AQUA cloning of the *Kod-RI* gene into the *pASG-IBA35* vector (25 bp – *pASG-IBA35*), *pMA-RQ\_Kod-RI* vector was used as a template employing *Kod-RI\_ampli\_25bp\_pASG-IBA35\_fw* and *Kod-RI\_ampli\_25bp\_pASG-IBA35\_rv* as primers. To introduce overhangs compatible with In-Fusion cloning (15 bp - *pASG-IBA35*), the *pMA-RQ\_Kod-RI* plasmid was used as a template employing *Kod-RI\_ampli\_15bp\_pASG-IBA35\_fw* and *Kod-RI\_ampli\_15bp\_pASG-IBA35\_rv* as primers.

**Table 24: PCR cycling conditions using Phusion Flash High Fidelity PCR Master Mix for the amplification of the *Kod-RI* gene and simultaneous introduction of overhangs compatible with either restriction enzyme cloning into the *pQE80HT* vector or Stargate cloning, AQUA cloning and In-Fusion cloning into the *pASG-IBA 35* vector.**

Added overhangs		T [°C]	t [s]
Initial denaturation		98	30
<u>5 cycles</u>			
<b>Denaturation</b>		98	10
<b>Annealing</b>	RE overhangs	56	5
	Stargate overhangs	56	
	25 bp - <i>pASG-IBA35</i>	56	
	15 bp - <i>pASG-IBA35</i>	53.5	
<b>Elongation</b>		72	48
<u>20 cycles:</u>			
<b>Denaturation</b>		98	10
<b>Annealing</b>	RE overhangs	67	5
	Stargate overhangs	68.5	
	25 bp - <i>pASG-IBA35</i>	72	
	15 bp - <i>pASG-IBA35</i>	72	
<b>Elongation</b>		72	48
Final Elongation		72	48
Hold		4	∞

After PCR, the sequence length was analyzed by 1 % analytical agarose gel (Chapter 5.3.7.2). Subsequently, for PCRs using the pMA-RQ\_Kod-RI vector as a template, the vector was digested with DpnI by adding 10 units of DpnI in 1 × CutSmart buffer, followed by incubation of the sample at 37 °C overnight. Then, the PCR product was purified via 1 % agarose gel (Chapter 5.3.7.2) followed by excision from the agarose gel and isolation using *Nucleospin Gel and PCR clean-up kit* (Chapter 5.3.7.3). For PCRs using the Kod-RI gene as a template, reactions were directly purified using *Nucleospin Gel and PCR clean-up kit* (Chapter 5.3.7.2). Lastly, the concentration of the samples was determined using a spectrophotometer (Chapter 5.3.7.7).

### 5.3.9.3 PCR linearization of pQE80HT and pASG-IBA35 vectors for AQUA and In-Fusion cloning

pQE80HT and pASG-IBA35 expression vectors were PCR linearized with Phusion Flash High Fidelity PCR Master Mix according to manufacturer's protocol. For linearization of the pQE80HT vector, pQE80HT\_PCRIin\_fw and pQE80HT\_PCRIin\_rv were used as primers. For linearization of the pASG-IBA35 vector, pASG-IBA35\_PCRIin\_fw and pASG-IBA35\_PCRIin\_rv were used as primers. Primers were designed to bind in the multiple cloning site downstream of the His-tag coding sequence. PCR was performed in a 20 µL scale with final concentrations of 1 × Phusion Flash High Fidelity PCR Master Mix, 1 ng of template and 0.5 µM forward and 0.5 µM reverse primers. A detailed pipetting scheme is provided in Table 25.

**Table 25: General pipetting scheme for the linearization of pQE80HT and pASG-IBA35 vectors with Phusion Flash High Fidelity PCR Master Mix.**

Component	Stock concentration	Final concentration	V [µL]
Template	10 ng µL <sup>-1</sup>	1.0 ng	0.1
Forward primer	100 µM	0.5 µM	0.1
Reverse primer	100 µM	0.5 µM	0.1
Phusion Flash PCR Master Mix	2 ×	1 ×	10
ddH <sub>2</sub> O	-	-	9.7

PCR cycling was performed with an initial denaturing step at 98 °C for 30 s, followed by 25 cycles of denaturing for 10 s at 98 °C, annealing at the individual temperature for each primer/template set for 5 s and elongation at 72 °C for 20 s per kb. A final elongation step was performed before cooling to 4 °C at the end. Table 26 provides an overview of the cycling conditions for each vector.

**Table 26: PCR cycling conditions for high fidelity PCR linearization of pQE80HT and pASG-IBA35 vectors using Phusion Flash High Fidelity PCR Master Mix.**

	Template	T [°C]	t [s]
Initial denaturation		98	30
<u>25 cycles:</u>			
<b>Denaturation</b>		98	10
<b>Annealing</b>	pQE80HT	58.5	5
	pASG-IBA35	59.5	
<b>Elongation</b>	pQE80HT	72	92
	pASG-IBA35		78
Final Elongation	pQE80HT	72	92
	pASG-IBA35		78
Hold		4	∞

After the PCR reaction, the sequence length was analyzed by 1 % analytical agarose gel (Chapter 5.3.7.2). Then, the template vector was DpnI digested by addition of 10 units of DpnI in 1× CutSmart buffer, followed by incubation of the sample at 37 °C overnight. Then, the linearized plasmid was purified via 1 % agarose gel (Chapter 5.3.7.2) followed by excision from the agarose gel and isolation using *Nucleospin Gel and PCR clean-up kit* (Chapter 5.3.7.3). Lastly, the oligonucleotide concentration was determined (Chapter 5.3.7.4).

#### 5.3.9.4 AQUA Cloning

The purified, linearized vector and the Kod-RI gene with 25 bp overhangs consisting of regions of homology to the 3'- and 5'-ends of the linearized vector were mixed and assembled by AQUA cloning.<sup>375</sup> Therefore, different vector:insert molar ratios were used with a defined vector DNA mass of 100 ng. The pQE80HT vector has a sequence length of 4558 bp and the pASG-IBA35 vector has a sequence length of 3883 bp. Based on the vector length, the required insert DNA mass of the 2375 bp Kod-RI coding sequence was calculated using the *QIAGEN Ligation Calculator*. The required insert DNA mass for AQUA cloning into both vectors is depicted in Tables 27 and 28.

**Table 27: Calculated mass of DNA insert for the corresponding vector:insert ratios used in AQUA cloning with the pQE80HT vector.** The vector DNA mass amounts to 100 ng.

Vector:insert ratio	Required insert DNA mass [ng]
1:5	260.6
1:7	375.3
1:10	521.3
1:20	1042.6

**Table 28: Calculated mass of DNA insert for the corresponding vector:insert ratios used in AQUA cloning with the pASG-IBA35 vector.** The vector DNA mass amounts to 100 ng.

Vector:insert ratio	Required insert DNA mass [ng]
1:10	611.6
1:20	1223

The calculated amounts of vector and insert were mixed together in a total volume of 10  $\mu$ L and the AQUA cloning reactions were incubated at 25 °C for 1 h. Subsequently, 5  $\mu$ L of the ligation mixture were transformed into competent NEB 10-beta or Stellar *E.coli* cells (Chapter 5.3.8.3). Then, single colonies were picked and bacteria were cultivated for plasmid preparation (Chapters 5.3.8.4 and 5.3.8.6). Colony PCR was performed to check if the vector contains the desired Kod-RI insert (Chapter 5.3.8.7). Positive clones were sent for sequencing (Chapter 5.3.8.9) using Seq\_Kod-RI\_2\_fw and Seq\_Kod-RI\_3\_rv primers that are located within the coding sequence of the Kod-RI gene and either the pQE80HT vector specific primers Seq\_pQE80HT\_1\_fw and Seq\_pQE80HT\_4\_rv or the pASG-IBA35 specific primers Seq\_pASG-IBA35\_1\_fw and Seq\_pASG-IBA35\_4\_rv.

### 5.3.9.5 In-Fusion cloning

The purified, linearized vector and Kod-RI gene with 15 bp extensions complementary to the ends of the linearized vector were mixed and assembled via In-Fusion cloning. Therefore, different vector:insert molar ratios were used with a defined vector DNA mass of 100 ng. The pQE80HT vector has a sequence length of 4558 bp and pASG-IBA35 vector has a sequence length of 3883 bp. Based on the vector length, the required insert DNA mass of the 2355 bp Kod-RI coding sequence was calculated using *QIAGEN Ligation Calculator*. The required insert DNA mass for In-Fusion cloning into both vectors is depicted in Tables 29 and 30.

**Table 29: Calculated mass of DNA insert for the corresponding vector:insert ratios used in In-Fusion cloning with the pQE80HT vector.** The vector DNA mass amounts to 100 ng.

Vector:insert ratio	Required insert DNA mass [ng]
1:2	103.4

**Table 30: Calculated mass of DNA insert for the corresponding vector:insert ratios used in In-Fusion cloning with the pASG-IBA35 vector.** The vector DNA mass amounts to 100 ng.

Vector:insert ratio	Required insert DNA mass [ng]
1:1	60.7
1:2	121.3
1:5	303.2

The calculated amounts of vector and insert were mixed together in a total volume of 10  $\mu$ L and were added to the *In-Fusion HD EcoDry Cloning Kit*. The components were mixed by pipetting up and down. The In-Fusion cloning reaction was first incubated at 37 °C for 15 min and then at 50 °C for 15 min. 2.5  $\mu$ L of the ligation mixture were then transformed into competent NEB 10-beta or Stellar *E.coli* cells (Chapter 5.3.8.3). Then, single colonies were picked and bacteria were cultivated for plasmid preparation (Chapters 5.3.8.4 and 5.3.8.6). Colony PCR was performed to check if the vector contained the desired Kod-RI insert (Chapter 5.3.8.7). Positive clones were sent for sequencing (Chapter 5.3.8.9) using Seq\_Kod-RI\_2\_fw and Seq\_Kod-RI\_3\_rv primers that are located within the coding sequence of Kod-RI gene and either the pQE80HT vector specific primers Seq\_pQE80HT\_1\_fw and Seq\_pQE80HT\_4\_rv or the pASG-IBA35 specific primers Seq\_pASG-IBA35\_1\_fw and Seq\_pASG-IBA35\_4\_rv.

### 5.3.9.6 Stargate cloning

The pASG-IBA35 vector was mixed with the Kod-RI gene, which was previously modified by PCR with combinatorial Esp3I recognition sites at both termini (Chapter 5.3.9.2). The Stargate experiment was set up in a 25  $\mu$ L scale containing 5 ng of vector and 36 ng of Kod-RI gene at final concentrations of 1 $\times$  CutSmart buffer, 10 mM DTT, 0.5 mM ATP, 2.5 U T4 DNA ligase and 5 U Esp3I. A detailed pipetting scheme for the Stargate digestion and ligation experiment is depicted in Table 31.



**Table 31: General pipetting scheme for the digestion and ligation reaction during Stargate Cloning.**

<b>Component</b>	<b>Stock concentration</b>	<b>Final concentration</b>	<b>Volume [<math>\mu</math>]</b>
pASG-IBA35	2.5 ng $\mu\text{L}^{-1}$	5 ng	2
Kod-RI gene	113 ng $\mu\text{L}^{-1}$	36 ng	0.32
CutSmart Buffer	10 $\times$	1 $\times$	2.5
DTT/ATP-Mix	50 mM/ 2.5 mM	10 mM / 0.5 mM	5
T4 DNA Ligase	5 U $\mu\text{L}^{-1}$	2.5 U	0.5
Esp3I	10 U $\mu\text{L}^{-1}$	5 U	0.5
ddH <sub>2</sub> O	-	-	14.18

The sample was mixed thoroughly and incubated for 1 h at 30 °C. Subsequently, 10  $\mu\text{L}$  of the ligation mix were transformed into OneShot TOP10 or NEB10-beta chemically competent *E.coli* cells (Chapter 5.3.8.3). Then, single colonies were picked and bacteria were cultivated for plasmid preparation (Chapters 5.3.8.4 and 5.3.8.6). Colony PCR was performed to check if the vector contained the desired Kod-RI insert (Chapter 5.3.8.7). Positive clones were sent for sequencing (Chapter 5.3.8.9) using Seq\_Kod-RI\_2\_fw and Seq\_Kod-RI\_3\_rv primers that are located within the coding sequence of the Kod-RI gene and the pASG-IBA35 specific primers Seq\_pASG-IBA35\_1\_fw and Seq\_pASG-IBA35\_Stargate\_rv4.

### **5.3.9.7 Restriction enzyme cloning**

#### Restriction enzyme digest

For restriction enzyme cloning of the Kod-RI gene into the pQE80HT vector, the Kod-RI gene was extended with a SacI restriction enzyme recognition site at the 5-terminus and a HindIII restriction enzyme recognition site at the 3-terminus by PCR (Chapter 5.3.9.2). Then, both vector and gene were digested separately using SacI-HF and HindIII-HF restriction enzymes in a 20  $\mu\text{L}$  scale reaction containing either 1  $\mu\text{g}$  of vector or 1  $\mu\text{g}$  of the gene in a final concentration of 1 $\times$  CutSmart buffer and 20 units of both restriction enzymes. A detailed pipetting scheme for the restriction enzyme digest of pQE80HT is depicted in Table 32 and for the Kod-RI gene in Table 33.

**Table 32: Detailed pipetting scheme for the restriction enzyme digest of the pQE80HT vector with SacI-HF and HindIII-HF restriction enzymes.**

Component	Stock concentration	Final concentration	V [ $\mu\text{L}$ ]
pQE80HT	71.6 ng $\mu\text{L}^{-1}$	1 $\mu\text{g}$	14
CutSmart Buffer	10x	1x	2
SacI-HF	20 U $\mu\text{L}^{-1}$	20 U	1
HindIII-HF	20 U $\mu\text{L}^{-1}$	20 U	1
ddH <sub>2</sub> O	-	-	2

**Table 33: Detailed pipetting scheme for the restriction enzyme digest of the Kod-RI gene with SacI-HF and HindIII-HF restriction enzymes.**

Component	Stock concentration	Final concentration	V [ $\mu\text{L}$ ]
Kod-RI gene	115.4 ng $\mu\text{L}^{-1}$	1 $\mu\text{g}$	8.7
CutSmart Buffer	10x	1x	2
SacI-HF	20 U $\mu\text{L}^{-1}$	20 U	1
HindIII-HF	20 U $\mu\text{L}^{-1}$	20 U	1
ddH <sub>2</sub> O	-	-	7.3

The restriction enzyme digests were incubated at 37 °C for 1 h, followed by heat inactivation of the enzymes at 80 °C for 20 min. Subsequently, the digested vector was treated with 1  $\mu\text{L}$  Antarctic phosphatase and 2  $\mu\text{L}$  10x Antarctic phosphatase reaction buffer and was again incubated at 37 °C for 1 h. Heat inactivation of the Antarctic phosphatase was performed at 80 °C for 20 min. Then, both samples were purified using preparative 1 % agarose gel electrophoresis (Chapter 5.3.7.2), followed by excision from the agarose gel and isolation using *Nucleospin Gel and PCR clean-up kit* (Chapter 5.3.7.3). The concentration of both samples was determined (Chapter 5.3.7.4) for the subsequent ligation reactions.

### Ligation

The SacI- and HindIII-digested pQE80HT vector and Kod-RI gene were ligated using T4 DNA ligase. Therefore, a vector:insert molar ratio of 1:2 was used with a defined vector DNA mass of 50 ng. The digested pQE80HT vector has a sequence length of 4524 bp. Based on the vector length, the required insert DNA mass of the 2327 bp Kod-RI sequence was calculated using *QIAGEN Ligation Calculator* and amounted to 51.4 ng. Then vector, insert, T4 DNA ligase, and T4 ligase buffer were mixed on ice as indicated in Table 34.

**Table 34: General pipetting scheme for the ligation reaction during restriction enzyme cloning.**

Component	Stock concentration	Final concentration	V [ $\mu$ L]
pQE80HT	24.3 ng $\mu$ L <sup>-1</sup>	50 ng	2.06
Kod-RI gene	16.4 ng $\mu$ L <sup>-1</sup>	51.4 ng	2.12
T4 DNA ligase	400 U $\mu$ L <sup>-1</sup>	400 U	1
T4 ligase buffer	10 $\times$	1 $\times$	1
ddH <sub>2</sub> O	-	-	3.82

The ligation reaction was incubated at 16 °C overnight, followed by heat inactivation at 65 °C for 10 min. Subsequently, the ligation reaction was cooled on ice and 2  $\mu$ L were transformed either into Stellar or NEB10-beta chemically competent *E.coli* cells (Chapter 5.3.8.3). Then, single colonies were picked and bacteria were cultivated for plasmid preparation (Chapters 5.3.8.4 and 5.3.8.6). Colony PCR was performed to check if the vector contained the desired Kod-RI insert (Chapter 5.3.8.7). Positive clones were sent for sequencing (Chapter 5.3.8.9) using the Seq\_Kod-RI\_2\_fw and Seq\_Kod-RI\_3\_rv primers that are located within the coding sequence of the Kod-RI gene and the pQE80HT specific primers Seq\_pQE80HT\_1\_fw and Seq\_pQE80HT\_4\_rv.

### 5.3.9.8 Site-directed mutagenesis of a deletion mutant of pQE80HT\_Kod-RI

The plasmid pQE80HT\_Kod-RI\_2313delA was previously isolated during AQUA cloning of the Kod-RI gene into the pQE80HT vector. The pQE80HT\_Kod-RI\_2313delA carries a single deletion of A at position 2313 in the coding sequence of the Kod-RI gene (Chapter 7.5.4).

This single point mutation was specifically changed by site-directed mutagenesis to re-introduce the deleted nucleotide back into the coding sequence. Therefore, the plasmid was linearized by PCR using one regular primer and a primer with an additional unpaired adenosine, which was specifically designed to insert the desired adenosine into the plasmid in a targeted manner.

#### Linearization of pQE80HT\_Kod-RI\_2313delA

The pQE80HT\_Kod-RI\_2313delA was PCR linearized with Phusion Flash High Fidelity PCR Master Mix according to the manufacturer's protocol. Therefore, the pQE80HT\_Kod-RI\_2313delA\_fw and pQE80HT\_Kod-RI\_2313delA\_rv primers were used. PCR was performed in a 20  $\mu$ L scale with final concentrations of 1 $\times$  Phusion Flash High Fidelity PCR Master Mix, 1 ng of template and 0.5  $\mu$ M forward and 0.5  $\mu$ M reverse primer, respectively. Three reactions were set up as depicted in Table 35.

**Table 35: General pipetting scheme for the linearization and site-directed mutagenesis of pQE80HT\_Kod-RI\_2313delA with Phusion Flash High Fidelity PCR Master Mix.**

<b>Component</b>	<b>Stock concentration</b>	<b>Final concentration</b>	<b>V [<math>\mu</math>L]</b>
pQE80HT_Kod-RI_2313delA	10 ng $\mu$ L <sup>-1</sup>	1.0 ng	0.1
pQE80HT_Kod-RI_2313delA_fw	100 $\mu$ M	0.5 $\mu$ M	0.1
pQE80HT_Kod-RI_2313delA_rv	100 $\mu$ M	0.5 $\mu$ M	0.1
Phusion Flash PCR Mastermix	2 $\times$	1 $\times$	10
ddH <sub>2</sub> O	-	-	9.7 $\mu$ L

PCR cycling was performed with an initial denaturing step at 98 °C for 30 s, followed by 25 cycles of denaturing for 10 s at 98 °C, annealing at 59 °C for 5 s and elongation at 72 °C for 138 s. A final elongation step was performed before cooling to 4 °C at the end. Table 36 provides an overview of the cycling conditions for pQE80HT\_Kod-RI\_2313delA linearization.

**Table 36: PCR cycling conditions for the PCR linearization and site-directed mutagenesis of pQE80HT\_Kod-RI\_2313delA with Phusion Flash High Fidelity PCR Master Mix.**

	<b>T [°C]</b>	<b>t [s]</b>
Initial denaturation	98	30
<u>25 cycles:</u>		
<b>Denaturation</b>	98	10
<b>Annealing</b>	59	5
<b>Elongation</b>	72	138
Final Elongation	72	138
Hold	4	$\infty$

After the PCR reaction, the sequence length was analyzed by 1 % analytical agarose gel (Chapter 5.3.7.2). Then, the PCR reactions were pooled and the template vector was DpnI digested by addition of 20 units of DpnI in 1 $\times$  CutSmart buffer, followed by incubation of the sample at 37 °C overnight. Then, the linearized plasmid was purified using *Nucleospin Gel and PCR clean-up kit* (Chapter 5.3.7.3) with only a single elution step with 20  $\mu$ L.

#### 5'-phosphorylation of the linearized pQE80HT\_Kod-RI\_2313delA

The linearized pQE80HT\_Kod-RI\_2313delA was phosphorylated at the 5'-hydroxyl termini. Therefore, the 20  $\mu$ L sample originating from the PCR-based linearization of pQE80HT\_Kod-RI\_2313delA were mixed with 10 U of T4 polynucleotide kinase in 1 $\times$  T4 ligase buffer. The sample was incubated for 1 h at 37 °C followed by heat inactivation at 65 °C for 20 min. Subsequently, purification using 1 % preparative agarose gel electrophoresis was

performed (Chapter 5.3.7.2), followed by excision from the agarose gel and isolation using *Nucleospin Gel and PCR clean-up kit* (Chapter 5.3.7.3). The concentration of the sample was determined using a spectrophotometer (Chapter 5.3.7.4).

#### Ligation of the linearized pQE80HT\_Kod-RI\_2313delA

The linearized and 5'-phosphorylated pQE80HT\_Kod-RI\_2313delA was ligated using T4 DNA ligase. The ligation reaction was performed in a 20  $\mu$ L scale, using 75 ng of linearized plasmid in 1 $\times$  T4 ligase buffer. A detailed pipetting scheme is shown in Figure 37.

**Table 37: General pipetting scheme for the ligation reaction of the linearized and 5'-phosphorylated pQE80HT\_Kod-RI\_2313delA.**

Component	Stock concentration	Final concentration	V [ $\mu$ L]
pQE80HT_Kod-RI_2313delA	23 ng $\mu$ L <sup>-1</sup>	75 ng	3.26
T4 DNA ligase	400 U $\mu$ L <sup>-1</sup>	800 U	2
T4 ligase buffer	10 $\times$	1 $\times$	2
ddH <sub>2</sub> O	-	-	3.82

The ligation reaction was incubated at 16 °C overnight followed by heat inactivation at 65 °C for 10 min. Subsequently, 10  $\mu$ L of the ligation reaction were transformed into NEB10-beta chemically competent *E.coli* cells (Chapter 5.3.8.3). Then, single colonies were picked and bacteria were cultivated for plasmid preparation (Chapter 5.3.8.4). Bacterial glycerol stocks of pQE80HT\_Kod-RI were prepared (Chapter 5.3.8.5) before plasmid isolation was performed (Chapter 5.3.8.6). The concentration of the samples was determined using a spectrophotometer (Chapter 5.3.7.4). Positive clones were sent for sequencing (Chapter 5.3.8.9) using the Seq\_Kod-RI\_2\_fw and Seq\_Kod-RI\_3\_rv primers that are located within the coding sequence of Kod-RI gene and the pQE80HT specific primers Seq\_pQE80HT\_1\_fw and Seq\_pQE80HT\_4\_rv.

#### Transformation of pQE80HT\_Kod-RI into *E.coli* expression strain

For protein expression, the pQE80HT\_Kod-RI plasmid was transformed into an *E.coli* expression strain. Therefore, 55 ng of pQE80HT\_Kod-RI were transformed into chemically competent NEBExpress® I<sup>q</sup> *E.coli* cells. Then, single colonies were picked and bacteria were cultivated for plasmid preparation (Chapter 5.3.8.4). Bacterial glycerol stocks were prepared (Chapter 5.3.8.5) before plasmid isolation (Chapter 5.3.8.6). The concentration of the samples was determined using a spectrophotometer (Chapter 5.3.7.4). Clones were sent for sequencing (Chapter 5.3.8.9) using the Seq\_Kod-RI\_2\_fw and Seq\_Kod-RI\_3\_rv primers that are located within the coding sequence of the Kod-RI gene and the pQE80HT specific primers Seq\_pQE80HT\_1\_fw and Seq\_pQE80HT\_4\_rv.

### 5.3.10 Site-directed mutagenesis of the Kod-RI coding sequence to the Kod-RSGA coding sequence

The coding sequence of Kod-RSGA polymerase<sup>152</sup> in the pQE80HT expression vector was constructed from the pQE80HT\_Kod-RI plasmid by site-directed mutagenesis. Using specially designed primers with mismatches introducing the four desired mutations<sup>152</sup>, the pQE80HT\_Kod-RSGA expression plasmid was prepared in four consecutive site-directed mutagenesis rounds, each consisting of linearization and targeted mutation of the plasmid, 5'-phosphorylation and ligation of the plasmid, followed by transformation into competent *E.coli* cells and subsequent isolation of the mutant plasmid.

#### Plasmid linearization and targeted mutation

pQE80HT\_Kod-RI was linearized by PCR with Phusion Flash High Fidelity PCR Master Mix) according to manufacturer's protocol. Primer set 1 (Mut1S\_RSGA\_fw and Mut1\_RSGA\_rv) was used in the first round of site-directed mutagenesis. PCR was performed in a 20  $\mu$ L scale with final concentrations of 1 $\times$  Phusion Flash High Fidelity PCR Master Mix, 1 ng of template and 0.5  $\mu$ M forward primer and 0.5  $\mu$ M reverse primer. A detailed pipetting scheme is depicted in Table 38.

**Table 38: General pipetting scheme for the linearization and targeted mutation of pQE80HT\_Kod-RI with Phusion Flash High Fidelity PCR Master Mix.**

<b>Component</b>	<b>Stock concentration</b>	<b>Final concentration</b>	<b>V [<math>\mu</math>L]</b>
pQE80HT_Kod-RI	10 ng $\mu$ L <sup>-1</sup>	1.0 ng	0.1
Mut1S_RSGA_fw	100 $\mu$ M	0.5 $\mu$ M	0.1
Mut1_RSGA_rv	100 $\mu$ M	0.5 $\mu$ M	0.1
Phusion Flash PCR Mastermix	2 $\times$	1 $\times$	10
ddH <sub>2</sub> O	-	-	9.7 $\mu$ L

PCR cycling was performed with an initial denaturing step at 98 °C for 30 s, followed by 25 cycles of denaturing for 10 s at 98 °C, annealing at 61.5 °C for 5 s and elongation at 72 °C for 138 s. A final elongation step was performed before cooling to 4 °C at the end. Table 39 provides an overview of the cycling conditions for linearization and mutation of pQE80HT\_Kod-RI.

**Table 39: PCR cycling conditions for the PCR-based linearization and targeted mutation of pQE80HT\_Kod-RI using Phusion Flash High Fidelity PCR Master Mix.** The utilized annealing temperature is the melting temperature of the primer set 1: Mut1S\_RSGA\_fw and Mut1\_RSGA\_rv.

	<b>T [°C]</b>	<b>t [s]</b>
Initial denaturation	98	30
<u>30 cycles:</u>		
<b>Denaturation</b>	98	10
<b>Annealing</b>	58.5	5
<b>Elongation</b>	72	138
Final Elongation	72	138
Hold	4	∞

After the PCR reaction, the sequence length was analyzed by 1 % analytical agarose gel (Chapter 5.3.7.2). Then, the template vector was DpnI digested by addition of 10 units of DpnI in 1×CutSmart buffer, followed by incubation of the sample at 37 °C overnight. Then, the mutated and linearized pQE80HT\_RIS plasmid was purified using *Nucleospin Gel and PCR clean-up kit* (Chapter 5.3.7.3) with only a single elution step with 20 µL.

#### 5'-phosphorylation of the linearized plasmid

The linearized pQE80HT\_Kod-RIS plasmid was phosphorylated at the 5'-hydroxyl termini. Therefore, the 20 µL sample originating from PCR-based linearization and mutation of pQE80HT\_Kod-RI plasmid was mixed with 2 U of T4 polynucleotide kinase in 1× T4 ligase buffer. The sample was incubated for 1 h at 37 °C, followed by heat inactivation at 65 °C for 20 min. Subsequently, purification using 1 % preparative agarose gel electrophoresis was performed (Chapter 5.3.7.2), followed by excision from the agarose gel and isolation using *Nucleospin Gel and PCR clean-up kit* (Chapter 5.3.7.3). The concentration was determined using a spectrophotometer (Chapter 5.3.7.4).

#### Ligation of the linearized plasmid DNA

The linearized and 5'-phosphorylated plasmid was ligated using T4 DNA ligase. The ligation reaction was performed in a 20 µL scale using 75 ng of linearized plasmid in 1× ligase buffer. A detailed pipetting scheme is provided in Table 40.

**Table 40: General pipetting scheme for the ligation reaction of linearized and 5'-phosphorylated plasmid.**

Components	Stock concentration	Final concentration	V [ $\mu\text{L}$ ]
pQE80HT_Kod-RIS (linearized)	129.5 ng $\mu\text{L}^{-1}$	75 ng	0.56
T4 DNA Ligase	1 U $\mu\text{L}^{-1}$	2 U	2
T4 ligase buffer	5 x	1 x	4
ddH <sub>2</sub> O	-	-	13.44

The ligation reaction was incubated at 24 °C for 1 h and was then put on ice. Subsequently, 2.5  $\mu\text{L}$  of the ligation reaction were transformed into NEB10-beta chemically competent *E.coli* cells (Chapter 5.3.8.3). Then, single colonies were picked and bacteria were cultivated for plasmid preparation (Chapters 5.3.8.4 and 5.3.8.6). The concentration of samples was determined using a spectrophotometer (Chapter 3.5.7.4) and diagnostic restriction enzyme digest was performed (Chapter 5.3.8.8). Positive clones were sent for sequencing (Chapter 5.3.8.9) using the Seq\_RSGA\_2\_fw and Seq\_RSGA\_3\_rv primers that are located within the coding sequence and the pQE80HT specific primers Seq\_pQE80HT\_1\_fw and Seq\_pQE80HT\_4\_rv.

#### Three additional rounds of site-directed mutagenesis

The site-directed mutagenesis procedure, including PCR-based linearization and simultaneous introduction of mutations, 5'-phosphorylation, ligation and transformation, was repeated three times in a row. Therefore, three different primer sets (2: Mut2\_RSGA\_fw and Mut2G\_RSGA\_rv, 3: Mut3\_RSGA\_fw and Mut3E\_RSGA\_rv, 4: Mut4\_RSGA\_fw and Mut4A\_RSGA\_rv) were used sequentially to introduce all four desired mutations into the coding sequence. The applied annealing temperatures for these primer sets, which were used during the PCR-based linearization and targeted mutation of the plasmid from the previous round, are depicted in Table 41. The remaining steps of the protocol remained unchanged.

**Table 41: Annealing temperatures for the primer sets 2-4 for PCR-based linearization and targeted mutation of the plasmid from the previous round using Phusion Flash High Fidelity PCR Master Mix.**

Primer set	Annealing	
	temperature [°C]	Resulting plasmid
2	59	pQE80HT_Kod-RISG
3	59.3	pQE80HT_Kod-RSG
4	59.5	pQE80HT_Kod-RSGA



### Retransformation of pQE80HT\_Kod-RSGA into *E.coli* expression strain

For further protein expression, pQE80HT\_Kod-RSGA plasmid was transformed into an *E.coli* expression strain. Therefore, 55 ng of pQE80HT\_Kod-RSGA were transformed into chemically competent NEBExpress® I<sup>q</sup> *E.coli* cells (Chapter 5.3.8.3). Then, single colonies were picked and bacteria were cultivated for plasmid preparation (Chapter 5.3.8.4). Bacterial glycerol stocks were prepared (Chapter 5.3.8.5) before plasmid isolation (Chapter 5.3.8.6). The concentration of the samples were determined using a spectrophotometer (Chapter 5.3.7.4). The integrity of the coding sequence was verified by Sanger sequencing (Chapter 5.3.8.9) using the Seq\_RSGA\_2\_fw and Seq\_RSGA\_3\_rv primers that are located within the coding sequence of the Kod-RSGA gene and the pQE80HT specific primers Seq\_pQE80HT\_1\_fw and Seq\_pQE80HT\_4\_rv.

## **5.3.11 Expression and purification of TNA polymerases**

The two TNA polymerases, Kod-RI and Kod-RSGA, were overexpressed and purified by affinity chromatography to investigate their capability for the incorporation of tTPT3TP and tNaMTP into nascent TNA strands.

### **5.3.11.1 Expression of TNA polymerases**

Protein expression and purification were conducted using an adapted protocol published by Nikoomanzar *et al.*<sup>376</sup>

A starter culture was prepared by inoculating a 100 mL culture of liquid LB containing 50 µg mL<sup>-1</sup> kanamycin with a glycerol stock of NEBExpress® I<sup>q</sup> cells harboring either the pQE80HT\_Kod-RI or the pQE80HT\_Kod-RSGA expression plasmid. The *E.coli* cells were grown aerobically at 37 °C and 120 rpm overnight. The OD<sub>600</sub> of a 1:10 dilution of the starter culture was measured and the needed volume of the starter culture to prepare a 1000 mL expression culture with an OD<sub>600</sub> of 0.1 was calculated. The expression culture was prepared accordingly and was incubated at 37 °C and 120 rpm shaking. At an OD<sub>600</sub> of 0.6, protein expression was induced by the addition of 0.5 mL 1.0 M isopropyl-β-D-thiogalactopyranoside to a final concentration of 0.5 mM. The induced expression culture was incubated at 15 °C for 18 h and 120 rpm. Subsequently, the cells were harvested by centrifugation at 5000 × g for 30 min at 4 °C and the supernatant was discarded. The cell pellet was washed by resuspension in 40 mL lysis buffer 1 (10 mM Tris-HCl, 500 mM NaCl, 10 % glycerol, 20 mM imidazole, pH 8.0), followed by centrifugation at 5000 × g for 30 min at 4 °C. The supernatant was discarded and the cell pellet was resuspended in 40 mL lysis buffer 1 supplemented with 80 µL of 100 mM phenylmethylsulfonyl fluoride to a final concentration of 0.2 mM. The cells

were then lysed by sonication on ice using a 2 sec on/ 4 sec off protocol for 2 min working time. The sonication step was performed twice. The cell debris was precipitated by centrifugation at 20000 × g for 30 min at 4 °C and the clarified supernatant was collected. Subsequently, the supernatant was heat treated at 80 °C for 1 h to denature endogenous *E.coli* proteins. The sample was cooled on ice for 45 min to aggregate denatured proteins and was then centrifuged at 4000 × g for 60 min at 4 °C. Finally, the supernatant was collected and subjected to purification (Chapter 5.3.11.2).

### 5.3.11.2 IMAC purification of TNA polymerases

#### Purification protocol

The supernatant from the previous step containing either Kod-RI or Kod-RSGA TNA polymerase was purified using an *ÅKTA start* FPLC chromatography system equipped with a 1 mL *HisTrap FF crude* column. Prior to sample application, the column was equilibrated with 20 column volumes (CV) of 100 % Buffer B1 (10 mM Tris-HCl, 600 mM NaCl, 10 % glycerol, 500 mM imidazole, pH 8.0), followed by equilibration with 10 CV of 100 % Buffer A1 (10 mM Tris-HCl, 600 mM NaCl, 10 % glycerol, 20 mM imidazole, pH 8.0) at a flow rate of 1 mL min<sup>-1</sup>. Then, 15 mL of the supernatant were applied to the column. After sample application, the column was washed with 10 column volumes of buffer A1. The TNA polymerase was then eluted using a linear gradient 0 % → 100 % of buffer B1 at a flow rate of 1 mL min<sup>-1</sup> for 10 column volumes. The eluted fractions containing the TNA polymerase were visualized using 10 % SDS-PAGE (Chapter 5.3.7.2) and were subsequently pooled.

#### Buffer exchange

Buffer exchange of the pooled polymerase fractions into TNA polymerase storage buffer (10 mM Tris-HCl, 250 mM NaCl, 50 % glycerol, pH 8.0) and concentration was performed using *Amicon Ultra-15* centrifugal filters with a molecular weight cut-off of 30 kDa. Therefore, the pooled fractions were loaded onto the centrifugal filter and centrifugation at 4000 × g was performed for 15 min at 4 °C. The flow-through was discarded and 7 mL of TNA polymerase storage buffer were added to the protein sample. Centrifugation at 4000 × g was performed for 15 min at 4 °C. Then, the membrane was unclogged by pipetting up and down and centrifugation at 4000 × g was continued for 15 min at 4 °C. The flow-through was discarded. To ensure sufficient buffer exchange, the addition of TNA polymerase storage buffer and the centrifugation steps were repeated three times. Lastly, the membrane was again unclogged and the sample was mixed by pipetting up and down. Then, the protein-containing sample was removed from the filter device using a pipette and was stored in 1 mL aliquots at -20 °C. The

protein concentration of the TNA polymerase aliquots was determined using the *Pierce 660 nm Protein Assay Kit* (Chapter 5.3.8.10).

### 5.3.11.3 TNA polymerase activity assay

A TNA polymerase activity assay was conducted to ascertain the optimal working concentration of the TNA polymerases.<sup>376</sup> Therefore, a two-fold dilution series with six different concentrations was prepared using either the Kod-RI polymerase stock (1544 ng  $\mu\text{L}^{-1}$ ) or the Kod-RSGA polymerase stock (2253 ng  $\mu\text{L}^{-1}$ ). For the Kod-RI polymerase, the stock solution was diluted six times two-fold within the range of 483 ng  $\mu\text{L}^{-1}$  to 1544 ng  $\mu\text{L}^{-1}$ . For the Kod-RSGA polymerase, the stock solution was subjected to six two-fold dilutions ranging from 704 ng  $\mu\text{L}^{-1}$  to 2253 ng  $\mu\text{L}^{-1}$ . *In vitro* primer extension assays were performed in a 10  $\mu\text{L}$  scale containing 1  $\mu\text{M}$  final concentration of 6-FAM-labeled DNA primer and an equimolar amount of DNA template (DNA\_Temp\_longATGC). The following reaction conditions were applied: 1 $\times$  Thermopol reaction buffer, 1 mM tNTPs, 1 mM  $\text{MnCl}_2$  and varying quantities of TNA polymerase. For Kod-RI polymerase, the final assay concentrations ranged from 48.3 ng  $\mu\text{L}^{-1}$  to 154.4 ng  $\mu\text{L}^{-1}$  and for Kod-RSGA polymerase from 70.4 ng  $\mu\text{L}^{-1}$  to 225.3 ng  $\mu\text{L}^{-1}$ . A comprehensive pipetting scheme of the TNA activity assay is depicted in Table 42.

**Table 42: General pipetting scheme for the TNA polymerase activity assay.** As TNA polymerase either Kod-RI or Kod-RSGA TNA polymerase was employed at different concentrations.

Component	Stock concentration	Final concentration	V [ $\mu\text{L}$ ]
DNA_Temp_longATGC	100 $\mu\text{M}$	1 $\mu\text{M}$	0.1
6-FAM_PEx_primer	100 $\mu\text{M}$	1 $\mu\text{M}$	0.1
Thermopol reaction buffer	10 $\times$	1 $\times$	1
tNTPs	1 mM each	0.1 mM each	1
$\text{MnCl}_2$	10 mM	1 mM	1
TNA polymerase	varied	varied	1
ddH <sub>2</sub> O	-	-	4.8

As a positive control, a primer extension assay using commercially available Terminator polymerase with a final concentration of 1 U and similar reaction conditions was included.

First, primer and template were annealed in 1 $\times$  Thermopol reaction buffer containing tNTPs by heating to 95  $^{\circ}\text{C}$  for 5 min, followed by cooling to 4  $^{\circ}\text{C}$  for 10 min. Then, the premixed polymerase dilutions supplemented with  $\text{MnCl}_2$  were added and reactions were incubated at 55  $^{\circ}\text{C}$  for 3 h. After 3 h of incubation, the primer extension was cooled down to 4  $^{\circ}\text{C}$ . The employed temperature program for TNA polymerase activity assays is depicted in Table 43.

**Table 43: Temperature profile for TNA polymerase activity assays.**

	T [°C]	T [min]
Denaturation	95	5
Annealing	4	10
Elongation	55	180
Hold	4	∞

Subsequently, primer extensions were quenched with 10  $\mu\text{L}$  2 $\times$  dPAGE loading buffer. The results were evaluated by 15 % polyacrylamide gel (Chapter 5.3.7.2). Based on the results of the activity assays, working concentrations of the TNA polymerases were selected.

### 5.3.12 TNA primer extension assays

TNA primer extensions assays were performed to investigate incorporation of highly modified TNA triphosphates bearing the unnatural nucleobases, **TPT3** and **NaM**. Primer extensions were either performed in a small 10  $\mu\text{L}$  scale for dPAGE analysis or in a large 50  $\mu\text{L}$  scale for subsequent LC-MS analysis.

#### 5.3.12.1 Protocols for small-scale primer extension assays

##### Small-scale PEx at 55 °C and 70 °C

Small-scale PEx were performed in a total volume of 10  $\mu\text{L}$  with final concentrations of 1 $\times$  Thermopol reaction buffer, 1  $\mu\text{M}$  6-FAM\_PEx\_primer, 1  $\mu\text{M}$  DNA template bearing d**NaM** (DNA\_ATGC\_shortTemp<sup>NaM</sup>), 1 mM MnCl<sub>2</sub>, 0.1 mM tNTPs, 0.1 mM t**TPT3TP** and 0.1 U  $\mu\text{L}^{-1}$  Therminator polymerase. A general pipetting scheme for the TNA primer extension assays at different temperatures is shown in Table 44.

**Table 44: Pipetting scheme for a small-scale TNA primer extension assay using Therminator polymerase with tTPT3TP at 50 °C and 70 °C.**

Components	Stock concentration	Final concentration	Volume [ $\mu\text{L}$ ]
6-FAM_PEx_primer	100 $\mu\text{M}$	1 $\mu\text{M}$	0.1
DNA_ATGC_shortTemp <sup>NaM</sup>	100 $\mu\text{M}$	1 $\mu\text{M}$	0.1
Thermopol reaction buffer	10 $\times$	1 $\times$	1
ddH <sub>2</sub> O	-	-	6.3
MnCl <sub>2</sub>	10 mM	1 mM	1
Therminator polymerase	2 U $\mu\text{L}^{-1}$	0.1 U $\mu\text{L}^{-1}$	0.5
tNTPs	5 mM each	0.1 mM each	0.8
t <b>TPT3TP</b>	5 mM	0.1 mM	0.2

In addition, a positive control using the same reaction setup with an unmodified template (DNA\_ATGC\_shortTemp) and without tTPT3TP addition was included.

In the annealing step, primer and template were hybridized in 1× Thermopol reaction buffer by heating to 95 °C for 5 min followed by cooling to 4 °C for 10 min. Then, a premixed solution of MnCl<sub>2</sub> and Therminator polymerase was added to the annealed primer/template complex. After adding the other components, the reaction was incubated at 55 °C or 70 °C for 3 h. The temperature profile is depicted in Table 45.

**Table 45: General temperature profile for a small-scale TNA primer extension assay using Therminator polymerase with tTPT3TP at 50 °C and 70 °C.**

	T [°C]	t [min]
Denaturation	95	5
Annealing	4	10
Elongation	55 <u>or</u> 70 °C	180
Hold	4	∞

Subsequently, primer extensions were quenched by the addition of 10 µL 2× dPAGE loading buffer and analyzed by 20% dPAGE (Chapter 5.3.7.2).

#### Small-scale PEx with additives

Small-scale PEx were performed in a total volume of 10 µL with final concentrations of 1× Thermopol reaction buffer, 1 µM 6-FAM-labeled DNA primer, 1 µM DNA template bearing dNaM (DNA\_ATGC\_midTemp<sup>NaM</sup>), 1mM MnCl<sub>2</sub>, 0.25 mM tNTPs, 0.25 mM tTPT3TP, and 0.1 U µL<sup>-1</sup> Therminator polymerase. Five PEx were set up with different combinations of additives including spermidine, DMSO, BSA, and DTT. A general pipetting scheme for TNA primer extension assays with additives is shown in Table 46 and the additives are defined in Table 47.

**Table 46: Pipetting scheme for small-scale TNA primer extension assays using Therminator polymerase with tTPT3TP and different additives.** The used additives are depicted in Table 47.

Component	Stock concentration	Final concentration	Volume [ $\mu\text{L}$ ]
6-FAM_PEx_primer	100 $\mu\text{M}$	1 $\mu\text{M}$	0.1
DNA_ATGC_midTemp <sup>NaM</sup>	100 $\mu\text{M}$	1 $\mu\text{M}$	0.1
Thermopol reaction buffer	10x	1x	1
ddH <sub>2</sub> O	-	-	ad 10
MnCl <sub>2</sub>	10 mM	1 mM	1
Therminator polymerase	2 U $\mu\text{L}^{-1}$	0.1 U $\mu\text{L}^{-1}$	0.5
Additives*	varied	varied	varied
tNTPs	5 mM each	0.25 mM each	2
tTPT3TP	5 mM	0.25 mM	0.5

\* Five different PEx were set up with different combinations of additives including spermidine, DMSO, BSA, and DTT. Please refer to Table 47 for a detailed list of the additives.

**Table 47: Combination of additives used in small-scale primer extension assays.** The depicted amounts are supplemented in the PEx shown in Table 46.

PEx	Additive	Stock concentration	Final concentration	Volume [ $\mu\text{L}$ ]
1	Spermidine	100 mM	2 mM	0.2
	BSA	1 mg $\text{mL}^{-1}$	100 $\mu\text{g mL}^{-1}$	1.0
2	Spermidine	100 mM	2 mM	0.2
	DMSO	$\geq 99.9\%$	2 %	0.2
3	BSA	1 mg $\text{mL}^{-1}$	100 $\mu\text{g mL}^{-1}$	1.0
	DTT	100 mM	1 mM	0.1
4	Spermidine	100 mM	2 mM	0.2
	DMSO	$\geq 99.9\%$	2 %	0.2
	BSA	1 mg $\text{mL}^{-1}$	100 $\mu\text{g mL}^{-1}$	1.0
	DTT	100 mM	1 mM	0.1
5	Spermidine	100 mM	2 mM	0.2
	DTT	100 mM	1 mM	1.0
	BSA	1 mg $\text{mL}^{-1}$	100 $\mu\text{g mL}^{-1}$	0.1

In the annealing step, primer and template were hybridized in 1x Thermopol reaction buffer by heating to 95 °C for 5 min, followed by cooling to 4 °C for 10 min. Then, a premixed solution of MnCl<sub>2</sub> and Therminator polymerase was added to the annealed primer/template complex. After adding the other components, the reaction was incubated at 55 °C for 3 h. The temperature profile is depicted in Table 48.

**Table 48: General temperature profile for a small-scale TNA primer extension assay using Therminator polymerase with tTPT3TP and different combinations of additives.**

	T [°C]	t [min]
Denaturation	95	5
Annealing	4	10
Elongation	55	180
Hold	4	∞

Subsequently, primer extensions were quenched by the addition of 10  $\mu\text{L}$  2 $\times$  dPAGE loading buffer and analyzed by 20% dPAGE (Chapter 5.3.7.2).

#### Small-scale PEx with varying tTPT3TP concentrations

Small-scale PEx were performed in a total volume of 10  $\mu\text{L}$  with final concentrations of 1 $\times$  Thermopol reaction buffer, 1  $\mu\text{M}$  6-FAM-labeled DNA primer, 1  $\mu\text{M}$  DNA template bearing dNaM (DNA\_ATGC\_midTemp<sup>NaM</sup>), 1 mM MnCl<sub>2</sub>, 0.25 mM tNTPs, 2 % DMSO, 2 mM spermidine and 0.1 U  $\mu\text{L}^{-1}$  Therminator polymerase. Four different PEx were set up with these conditions, containing varying final concentrations of tTPT3TP (*i.e.* 0.25 mM, 0.5 mM, 0.75 mM, 1.0 mM). A general pipetting scheme for TNA primer extension assays with different tTPT3TP concentrations is shown in Table 49.

**Table 49: Pipetting scheme for small-scale TNA primer extension assays using Therminator polymerase with varying tTPT3TP concentrations.**

Component	Stock concentration	Final concentration	Volume [ $\mu\text{L}$ ]
6-FAM_PEx_primer	100 $\mu\text{M}$	1 $\mu\text{M}$	0.1
DNA_ATGC_midTemp <sup>NaM</sup>	100 $\mu\text{M}$	1 $\mu\text{M}$	0.1
Thermopol reaction buffer	10 $\times$	1 $\times$	1
ddH <sub>2</sub> O	-	-	ad 10
MnCl <sub>2</sub>	10 mM	1 mM	1
Therminator polymerase	2 U $\mu\text{L}^{-1}$	0.1 U $\mu\text{L}^{-1}$	0.5
DMSO	$\geq 99.9$ %	2 %	0.2
Spermidine	100 mM	2 mM	0.2
tNTPs	5 mM each	0.25 mM each	2
tTPT3TP*	5 mM	0.25 mM	0.5
		0.5 mM	1.0
		0.75 mM	1.5
		1.0 mM	2.0

\* Four different PEx were set up with the depicted concentrations

In addition, a negative control using the same reaction setup with water instead of **tTPT3TP** was included.

In the annealing step, primer and template were hybridized in 1× Thermopol reaction buffer by heating to 95 °C for 5 min, followed by cooling to 4 °C for 10 min. Then, a premixed solution of MnCl<sub>2</sub> and Therminator polymerase was added to the annealed primer/template complex. After adding the other components, the reaction was incubated at 55 °C for 16 h. The temperature profile is depicted in Table 50.

**Table 50: General temperature profile used for a small-scale TNA primer extension assay using Therminator polymerase with varying tTPT3TP concentrations.**

	T [°C]	t [min]
Denaturation	95	5
Annealing	4	10
Elongation	55	960
Hold	4	∞

Subsequently, primer extensions were quenched by the addition of 10 µL 2× dPAGE loading buffer and analyzed by 20% dPAGE (Chapter 5.3.7.2).

### 5.3.12.2 Protocols for large-scale primer extension assays

A series of large-scale primer extension assays were conducted with Therminator, Kod-RI, and Kod-RSGA polymerases to investigate the incorporation of either **tTPT3TP** or **tNaMTP** into a nascent TNA strand using LC-MS analysis.

Large scale *in vitro* primer extension assays were performed in a 50 µL scale, containing 1 µM final concentration of 6-FAM-labeled DNA primer and an equimolar amount of 5'-monophosphate-modified DNA template bearing either **dTPT3** or **dNaM** (DNA\_pTemp<sup>TPT3</sup>, DNA\_pTemp<sup>NaM</sup>). The sequences of the DNA templates varied in the individual PEx experiments and are provided for each experiment hereafter. The following reaction conditions were applied in all large scale PEx: 1× Thermopol reaction buffer, TNA polymerase, 0.25 mM of the tNTPs needed for the respective DNA template, and either 0.5 mM **tTPT3TP** or 0.5 mM **tNaMTP**. The addition of MnCl<sub>2</sub>, DMSO, and spermidine varied in the individual PEx experiments. Detailed pipetting schemes for the different TNA primer extension assays with Therminator, Kod-RI, and Kod-RSGA polymerases and **tTPT3TP** or **tNaMTP** are shown in Tables 51, 52, 53, 54, 55 and 56.



**Table 51: Pipetting scheme for a large-scale TNA primer extension assay with tTPT3TP using Therminator polymerase.**

Component	Stock concentration	Final concentration	Volume [ $\mu\text{L}$ ]
6-FAM_PEx_primer	100 $\mu\text{M}$	1 $\mu\text{M}$	0.5
DNA_pTemp <sup>NaM*</sup>	100 $\mu\text{M}$	1 $\mu\text{M}$	0.5
Thermopol reaction buffer	10x	1x	5
ddH <sub>2</sub> O	-	-	19.5
MnCl <sub>2</sub>	10 mM	1 mM	5
Therminator polymerase	2 U $\mu\text{L}^{-1}$	0.1 U $\mu\text{L}^{-1}$	2.5
DMSO	$\geq 99.9\%$	2 %	1
Spermidine	100 mM	2 mM	1
tNTPs	5 mM each	0.25 mM each	10
tTPT3TP	5 mM	0.5 mM	5

\* PEx were performed with two templates: DNA\_pTemp\_midATC<sup>NaM</sup> or DNA\_pTemp\_midATG<sup>NaM</sup>

**Table 52: Pipetting scheme for a large-scale TNA primer extension assay with tNaMTP using Therminator polymerase.**

Component	Stock concentration	Final concentration	Volume [ $\mu\text{L}$ ]
6-FAM_PEx_primer	100 $\mu\text{M}$	1 $\mu\text{M}$	0.5
DNA_pTemp <sup>TPT3*</sup>	100 $\mu\text{M}$	1 $\mu\text{M}$	0.5
Thermopol reaction buffer	10x	1x	5
ddH <sub>2</sub> O	-	-	21.5
MnCl <sub>2</sub>	10 mM	1 mM	5
Therminator polymerase	2 U $\mu\text{L}^{-1}$	0.1 U $\mu\text{L}^{-1}$	2.5
tNTPs	5 mM each	0.25 mM each	10
tNaMTP	5 mM	0.5 mM	5

\* PEx were performed with two templates: DNA\_pTemp\_midATC<sup>TPT3</sup> or DNA\_pTemp\_midATG<sup>TPT3</sup>

**Table 53: Pipetting scheme for a large-scale TNA primer extension assay with tTPT3TP using Kod-RI polymerase.**

Component	Stock concentration	Final concentration	Volume [ $\mu\text{L}$ ]
6-FAM_PEx_primer	100 $\mu\text{M}$	1 $\mu\text{M}$	0.5
DNA_pTemp <sup>NaM*</sup>	100 $\mu\text{M}$	1 $\mu\text{M}$	0.5
Thermopol reaction buffer	10x	1x	5
ddH <sub>2</sub> O	-	-	19.5
MnCl <sub>2</sub>	10 mM	1 mM	5
Kod-RI polymerase	1544.4 ng $\mu\text{L}^{-1}$	77.2 ng $\mu\text{L}^{-1}$	2.5
DMSO	$\geq 99.9\%$	2 %	1
Spermidine	100 mM	2 mM	1
tNTPs	5 mM each	0.25 mM each	10
<b>tTPT3TP</b>	5 mM	0.5 mM	5

\* PEx were performed with two templates: DNA\_pTemp\_midATC<sup>NaM</sup> or DNA\_pTemp\_midATG<sup>NaM</sup>

**Table 54: Pipetting scheme for a large-scale TNA primer extension assay with tNaMTP using Kod-RI polymerase.**

Component	Stock concentration	Final concentration	Volume [ $\mu\text{L}$ ]
6-FAM_PEx_primer	100 $\mu\text{M}$	1 $\mu\text{M}$	0.5
DNA_pTemp <sup>TPT3*</sup>	100 $\mu\text{M}$	1 $\mu\text{M}$	0.5
Thermopol reaction buffer	10x	1x	5
ddH <sub>2</sub> O	-	-	21.5
MnCl <sub>2</sub>	10 mM	1 mM	5
Kod-RI polymerase	1544.4 ng $\mu\text{L}^{-1}$	77.2 ng $\mu\text{L}^{-1}$	2.5
tNTPs	5 mM each	0.25 mM each	10
<b>tNaMTP</b>	5 mM	0.5 mM	5

\* PEx were performed with two templates: DNA\_pTemp\_midATC<sup>TPT3</sup> or DNA\_pTemp\_midATG<sup>TPT3</sup>

**Table 55: Pipetting scheme for a large-scale TNA primer extension assay with tTPT3TP using Kod-RSGA polymerase.**

Component	Stock concentration	Final concentration	Volume [ $\mu\text{L}$ ]
6-FAM_PEx_primer	100 $\mu\text{M}$	1 $\mu\text{M}$	0.5
DNA_pTemp <sup>NaM*</sup>	100 $\mu\text{M}$	1 $\mu\text{M}$	0.5
Thermopol reaction buffer	10 $\times$	1 $\times$	5
ddH <sub>2</sub> O	-	-	19.5
MnCl <sub>2</sub> **	10 mM	1 mM	5
Kod-RSGA polymerase	2253.2 ng $\mu\text{L}^{-1}$	112.7 ng $\mu\text{L}^{-1}$	2.5
DMSO	$\geq 99.9\%$	2 %	1
Spermidine	100 mM	2 mM	1
tNTPs	5 mM each	0.25 mM each	10
<b>tTPT3TP</b>	5 mM	0.5 mM	5

\* PEx were performed with two templates: DNA\_pTemp\_midATG<sup>NaM</sup> or DNA\_pTemp\_mid2ATGC<sup>NaM</sup>

\*\* PEx were also performed without 1 mM MnCl<sub>2</sub> and ddH<sub>2</sub>O was added instead

**Table 56: Pipetting scheme for a large-scale TNA primer extension assay with tNaMTP using Kod-RSGA polymerase.**

Component	Stock concentration	Final concentration	Volume [ $\mu\text{L}$ ]
6-FAM_PEx_primer	100 $\mu\text{M}$	1 $\mu\text{M}$	0.5
DNA_pTemp_mid2ATGC <sup>TPT3</sup>	100 $\mu\text{M}$	1 $\mu\text{M}$	0.5
Thermopol reaction buffer	10 $\times$	1 $\times$	5
ddH <sub>2</sub> O	-	-	19.5
MnCl <sub>2</sub> *	10 mM	1 mM	5
Kod-RSGA polymerase	2253.2 ng $\mu\text{L}^{-1}$	112.7 ng $\mu\text{L}^{-1}$	2.5
DMSO	$\geq 99.9\%$	2 %	1
Spermidine	100 mM	2 mM	1
tNTPs	5 mM each	0.25 mM each	10
<b>tNaMTP</b>	5 mM	0.5 mM	5

\* PEx were also performed without 1 mM MnCl<sub>2</sub> and ddH<sub>2</sub>O was added instead

In addition, positive and negative controls were included to serve as a size markers for full-length products and fragments where truncation occurred at the position of the unnatural nucleobase. In negative controls, ddH<sub>2</sub>O was added to the primer extension assays instead of tNaMTP or tTPT3TP. In positive controls, the respective unmodified DNA templates were

employed and ddH<sub>2</sub>O was added to the primer extension assays instead of tNaMTP or tTPT3TP.

In the annealing step, primer and template were hybridized in 1× Thermopol reaction buffer by heating to 95 °C for 5 min, followed by cooling to 4 °C for 12 min. Then, a premixed solution of MnCl<sub>2</sub> and TNA polymerase was added to the annealed primer/template complex. After adding the other components, the reaction was incubated at 55 °C for 16 h. The general temperature profile used for large scale primer extension assays is depicted in Table 57.

**Table 57: General temperature profile for large scale TNA primer extension assay.**

	T [°C]	t [min]
Denaturation	95	5
Annealing	4	12
Elongation	55	960
Hold	4	∞

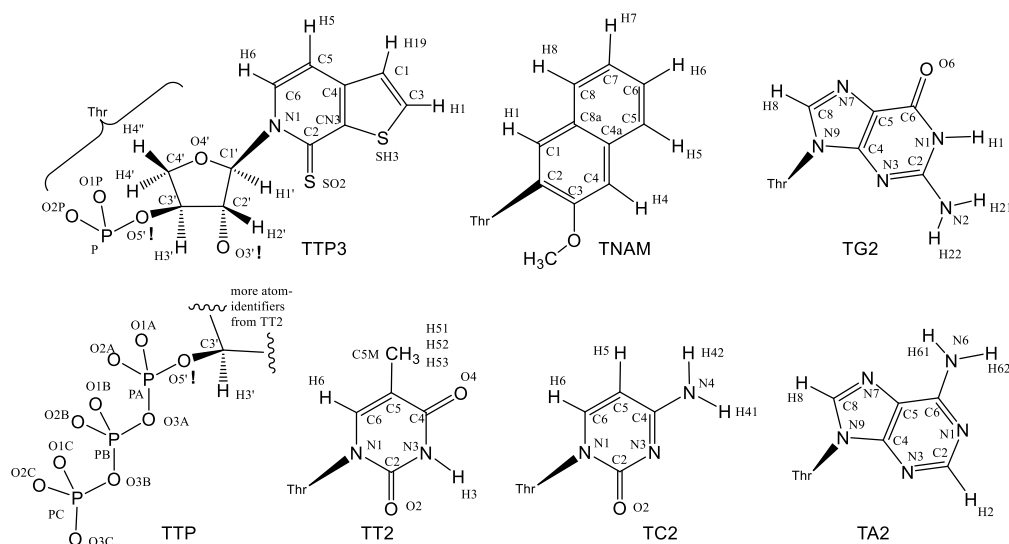
Subsequently, the primer extensions were purified by centrifugal filters (Chapter 5.3.7.3) and the 5'-phosphate-modified DNA template was digested with λ-exonuclease (Chapter 5.3.7.5). Subsequently, the sample was again purified with centrifugal filters (Chapter 5.3.7.3). Next, 7 μL of the sample were analyzed by 20% dPAGE (Chapter 5.3.7.2) and 20 μL of the sample were subjected to LC-MS analysis (Chapter 5.3.7.6).

### 5.3.13 MD simulations of tTPT3 and tNaM elongation in a nascent TNA strand

#### 5.3.13.1 Parametrization of threose nucleic acid building blocks

Threose nucleic acid building blocks were parametrized as new residues TA2, TT2, TG2, TC2, TTP3, TNAM, and TTP (= triphosphate) in the CHARMM force field. The force field generator Cgenff<sup>418,419</sup> and 6-31G\* *ab initio* or svp-b3lyp density functional calculations were used to derive the parameters. Atomic charges, bond lengths, angles and dihedral angles suggested by Cgenff and the quantum calculations were in addition modified using parameters for similar bonding situations already noted in the Charmm<sup>379</sup> force field.

The used atom descriptors are listed in Figure 62. Atom types, charges and connectivities are given in Table 58, which corresponds to the residue topology files in Gromacs (extension .rtp). New parameters for bonds, bending and torsions are noted in Table 59.



**Figure 62: Structure of the L-threose nucleic acid residues named “TTP3, TNAM, TG2, TA2, TC2, TT2 and TTP” in the modified Gromacs<sup>378</sup>– Charmm36<sup>379</sup> force field (see also Table 58).** Please note the unusual (not IUPAC) names of the oxygen atoms at C2' and C3'. The names were useful during the parametrization process, creating and comparing bonding parameters similar to bonding situations already noted in the force field.

**Table 58: Residue topology entries TTP3, TNAM, TG2, TA2, TC2, TT2 and TTP from the modified “merged.rtp” of the Gromacs – Charmm 36 force field.** The partial charges were suggested by CGenFF<sup>418,419</sup> (interface 2.2.0, force field 4.0) and adjusted using AM1 and 6-31G\* QM calculations. The atom definitions are given in Figure 62.

```
[ TA2 ]
; Threose-Adenine
[ atoms ]
; atom atom-type charge
      P      P5      1.500  0
      O1P     ON3     -0.780  1
      O2P     ON3     -0.780  2
      O5'     ON2     -0.570  3 ; O3' !
      C4'     CN8C     0.070  4
      H4'     HN8      0.090  5
      H4''    HN8      0.090  6
      O4'     ON6B    -0.500  7
      C1'     CN7B     0.160  8
      H1'     HN7      0.090  9
      N9      NN2     -0.050 10
      C5      CN5      0.280 11
      N7      NN4     -0.710 12
      C8      CN4      0.340 13
      H8      HN3      0.120 14
      N1      NN3A    -0.740 15
```

C2	CN4	0.500	16
H2	HN3	0.130	17
N3	NN3A	-0.750	18
C4	CN5	0.430	19
C6	CN2	0.460	20
N6	NN1	-0.770	21
H61	HN1	0.380	22
H62	HN1	0.380	23
C2'	CN7B	0.010	24
H2'	HN7	0.090	25
O3'	ON2	-0.570	26 ; O2' !
C3'	CN7	0.010	27
H3'	HN7	0.090	28
[ bonds ]			
-O3'	P		
P	O1P		
P	O2P		
P	O5'		
O5'	C3'		
C4'	O4'		
C4'	C3'		
O4'	C1'		
C1'	N9		
C1'	C2'		
N9	C4		
N9	C8		
C4	N3		
C2	N1		
C6	N6		
N6	H61		
N6	H62		
C6	C5		
C5	N7		
C2'	C3'		
O3'	+P		
C2'	O3'		
C1'	H1'		
C2'	H2'		
C3'	H3'		
C4'	H4'		
C4'	H4''		

C8	H8		
C2	H2		
N1	C6		
C2	N3		
C4	C5		
N7	C8		
[ impropers ]			
N6	C6	H61	H62
C6	N1	C5	N6
[ TT2 ]			
; Threose-Thymidine			
[ atoms ]			
P	P5	1.500	0
O1P	ON3	-0.780	1
O2P	ON3	-0.780	2
O5'	ON2	-0.570	3 ; O3' !!
C4'	CN8C	0.070	4
H4'	HN8	0.090	5
H4''	HN8	0.090	6
O4'	ON6B	-0.500	7
C1'	CN7B	0.160	8
H1'	HN7	0.090	9
N1	NN2B	-0.34	10
C6	CN3	0.17	11
H6	HN3	0.17	12
C2	CN1T	0.51	13
O2	ON1	-0.41	14
N3	NN2U	-0.46	15
H3	HN2	0.36	16
C4	CN1	0.50	17
O4	ON1	-0.45	18
C5	CN3T	-0.15	19
C5M	CN9	-0.11	20
H51	HN9	0.07	21
H52	HN9	0.07	22
H53	HN9	0.07	23
C2'	CN7B	0.010	24
H2'	HN7	0.090	25
O3'	ON2	-0.570	26 ; O2' !
C3'	CN7	0.010	27
H3'	HN7	0.090	28

[ bonds ]

```

      -O3'      P
      P      O1P
      P      O2P
      P      O5'
O5'      C3'
C4'      O4'
C4'      C3'
O4'      C1'
C1'      N1
C1'      C2'
N1      C2
N1      C6
C2      N3
N3      H3
N3      C4
C4      C5
C5      C5M
C2'      C3'
O3'      +P
C2'      O3'
C1'      H1'
C2'      H2'
C3'      H3'
C4'      H4'
C4'      H4' '
C6      H6
C5M      H51
C5M      H52
C5M      H53
C2      O2
C4      O4
C5      C6

```

[ impropers ]

```

C2      N1      N3      O2
C4      N3      C5      O4
C5      C4      C6      C5M

```

[ TG2 ]

; Threose-Guanine

[ atoms ]



P	P5	1.500	0
O1P	ON3	-0.780	1
O2P	ON3	-0.780	2
O5'	ON2	-0.570	3 ; O3' !
C4'	CN8C	0.070	4
H4'	HN8	0.090	5
H4''	HN8	0.090	6
O4'	ON6B	-0.500	7
C1'	CN7B	0.160	8
H1'	HN7	0.090	9
N9	NN2B	-0.02	10
C4	CN5	0.26	11
N2	NN1	-0.68	12
H21	HN1	0.32	13
H22	HN1	0.35	14
N3	NN3G	-0.74	15
C2	CN2	0.75	16
N1	NN2G	-0.34	17
H1	HN2	0.26	18
C6	CN1	0.54	19
O6	ON1	-0.51	20
C5	CN5G	0.00	21
N7	NN4	-0.60	22
C8	CN4	0.25	23
H8	HN3	0.16	24
C2'	CN7B	0.010	25
H2'	HN7	0.090	26
O3'	ON2	-0.570	27 ; O2' !
C3'	CN7	0.010	28
H3'	HN7	0.090	29
[ bonds ]			
-O3'	P		
P	O1P		
P	O2P		
P	O5'		
O5'	C3'		
C4'	O4'		
C4'	C3'		
O4'	C1'		
C1'	N9		
C1'	C2'		

```

N9      C4
N9      C8
C4      N3
C2      N2
C2      N1
N2      H21
N2      H22
N1      H1
N1      C6
C6      C5
C5      N7
  C2'   C3'
  O3'   +P
  C2'   O3'
  C1'   H1'
  C2'   H2'
  C3'   H3'
  C4'   H4'
  C4'   H4''
  C8     H8
  C2     N3
  C4     C5
  N7     C8
  C6     O6

```

[ impropers ]

```

C2      N3      N1      N2
C6      N1      C5      O6
N2      H21 C2      H22

```

[ TTP3 ]

[ atoms ]

```

  P      P5      1.500  0
  O1P    ON3     -0.780  1
  O2P    ON3     -0.780  2
  O5'    ON2     -0.570  3 ; O3' !
  H4'    HN8      0.090  5
  H4''   HN8      0.090  6
  O4'    ON6B    -0.500  7
  C1'    CN7B     0.160  8
  H1'    HN7      0.090  9
N1      NN2B    -0.300 59

```

C6	CG2R61	0.025	60
H6	HGR62	0.185	61
C5	CG2R61	0.005	62
H5	HGR62	0.175	63
C4	CG2RC0	-0.095	64
CN3	CG2RC0	0.055	65
SH3	SG2R50	-0.050	66
C2	CG2R63	0.130	67
SO2	SG2D1	-0.205	68
C1	CG2R51	-0.250	76
C3	CG2R51	-0.069	77
H1	HGR52	0.185	80
H19	HGR51	0.209	83
C2'	CN7B	0.010	24
H2'	HN7	0.090	25
O3'	ON2	-0.570	26 ; O2' !
C3'	CN7	0.010	27
H3'	HN7	0.090	28

[ bonds ]

-O3'	P
P	O1P
P	O2P
P	O5'
O5'	C3'
C4'	O4'
C4'	C3'
O4'	C1'
C1'	N1
C1'	C2'
N1	C6
C6	H6
C6	C5
C5	H5
C5	C4
C4	CN3
CN3	SH3
N1	C2
CN3	C2
C2	SO2
C4	C1
SH3	C3

```

C1  C3
C3  H1
C1  H19
  C2'  C3'
  O3'  +P
  C2'  O3'
  C1'  H1'
  C2'  H2'
  C3'  H3'
  C4'  H4'
  C4'  H4''
[ impropers ]
  C2    CN3    N1    SO2

[ TNAM ]
; Threose-NAM
[ atoms ]
  P    P5    1.500  0
  O1P  ON3   -0.780  1
  O2P  ON3   -0.780  2
  O5'  ON2   -0.570  3 ; O3' !
  C4'  CN8C   0.070  4
  H4'  HN8    0.090  5
  H4'' HN8    0.090  6
  O4'  ON6B  -0.500  7
  C1'  CN7B   0.160  8
  H1'  HN7    0.090  9
  C2  CG2R61 -0.060 12
  C1  CG2R61 -0.125 13
  H1  HGR61   0.115 14
  C3  CG2R61  0.188 15
  O3  OG301  -0.400 16
  C4  CG2R61 -0.131 17
  H4  HGR61   0.115 18
  C4a CG2R61  0.003 19
  C5  CG2R61 -0.137 20
  H5  HGR61   0.105 21
  C6  CG2R61 -0.115 22
  H6  HGR61   0.110 23
  C7  CG2R61 -0.115 24
  H7  HGR61   0.110 25

```

C8	CG2R61	-0.127	26
H8	HGR61	0.105	27
C8a	CG2R61	-0.011	28
CO3	CG331	0.100	29
H9	HGA3	0.090	30
H10	HGA3	0.090	31
H11	HGA3	0.090	32
C2'	CN7B	0.010	24
H2'	HN7	0.090	25
O3'	ON2	-0.570	26 ; O2' !
C3'	CN7	0.010	27
H3'	HN7	0.090	28

[ bonds ]

-O3'	P
P	O1P
P	O2P
P	O5'
O5'	C3'
C4'	O4'
C4'	C3'
O4'	C1'
C1'	C2
C1'	C2'
C2	C1
C2	C3
C3	C4
C3	O3
C4	C4a
C1	C8a
C4a	C8a
C4a	C5
C5	C6
C6	C7
C7	C8
C8	C8a
O3	CO3
C1	H1
C4	H4
C5	H5
C6	H6
C7	H7

```

C8      H8
CO3     H9
CO3     H10
CO3     H11
  C2'   C3'
  O3'   +P
  C2'   O3'
  C1'   H1'
  C2'   H2'
  C3'   H3'
  C4'   H4'
  C4'   H4''

```

```
[ impropers ]
```

```
[ TTP ]
```

```
; TT2 with triphosphate parameters from ATP
```

```
[ atoms ]
```

```

;      P      P5      1.500  0
;      O1P     ON3     -0.780  1
;      O2P     ON3     -0.780  2
      O5'     ON2     -0.600  3 ; O3' !
      C4'     CN8C     0.070  4
      H4'     HN8      0.090  5
H4''      HN8      0.090  6
      O4'     ON6B     -0.500  7
      C1'     CN7B     0.160  8
      H1'     HN7      0.090  9
      N1      NN2B     -0.34   10
      C6      CN3      0.17   11
      H6      HN3      0.17   12
      C2      CN1T     0.51   13
      O2      ON1     -0.41   14
      N3      NN2U     -0.46   15
      H3      HN2      0.36   16
      C4      CN1      0.50   17
      O4      ON1     -0.45   18
      C5      CN3T     -0.15   19
      C5M     CN9     -0.11   20
      H51     HN9      0.07   21
      H52     HN9      0.07   22

```

H53	HN9	0.07	23
C2'	CN7B	0.140	24
H2'	HN7	0.090	25
O3'	ON5	-0.660	27 ; O2' !
C3'	CN7	0.010	28
H3'	HN7	0.090	29
H3T	HN5	0.430	30
PA	P	1.480	31
O1A	ON3	-0.820	32
O2A	ON3	-0.820	33
O3A	ON2	-0.740	34
PB	P3	1.500	35
O1B	ON3	-0.820	36
O2B	ON3	-0.820	37
O3B	ON2	-0.860	38
PG	P3	1.100	39
O1G	ON3	-0.900	40
O2G	ON3	-0.900	41
O3G	ON3	-0.900	42

[ bonds ]

O5'	PA
PA	O1A
PA	O2A
PA	O3A
O3A	PB
PB	O1B
PB	O2B
PB	O3B
O3B	PG
PG	O1G
PG	O2G
PG	O3G
O5'	C3'
C4'	O4'
C4'	C3'
O4'	C1'
C1'	N1
C1'	C2'
N1	C2
N1	C6
C2	N3

```

N3   H3
N3   C4
C4   C5
C5   C5M
  C2'  C3'
  O3'  H3T
  C2'  O3'
  C1'  H1'
  C2'  H2'
  C3'  H3'
  C4'  H4'
  C4'  H4''
  C6   H6
  C5M  H51
  C5M  H52
  C5M  H53
   C2   O2
   C4   O4
   C5   C6

```

[ impropers ]

```

  C2   N1   N3   O2
  C4   N3   C5   O4
  C5   C4   C6   C5M

```

[ TC2 ]

; Threose-Cytidine

[ atoms ]

```

   P   P5   1.500  0
  O1P  ON3  -0.780  1
  O2P  ON3  -0.780  2
  O5'  ON2  -0.570  3 ; O3'
  C4'  CN8C  0.070  4
  H4'  HN8   0.090  5
  H4'' HN8   0.090  6
  O4'  ON6B -0.500  7
  C1'  CN7B  0.160  8
  H1'  HN7   0.090  9

  N1   NN2  -0.13  12
  C6   CN3   0.05  13
  H6   HN3   0.17  14
  C5   CN3  -0.13  15

```



H5	HN3	0.07	16
C2	CN1	0.52	17
O2	ON1C	-0.49	18
N3	NN3	-0.66	19
C4	CN2	0.65	20
N4	NN1	-0.75	21
H41	HN1	0.37	22
H42	HN1	0.33	23
C2'	CN7B	0.010	24
H2'	HN7	0.090	25
O3'	ON2	-0.570	26 ; O2' !
C3'	CN7	0.010	27
H3'	HN7	0.090	28
[ bonds ]			
-O3'	P		
P	O1P		
P	O2P		
P	O5'		
O5'	C3'		
C4'	O4'		
C4'	C3'		
O4'	C1'		
C1'	N1		
C1'	C2'		
N1	C2		
N1	C6		
C2	N3		
C4	N4		
N4	H41		
N4	H42		
C4	C5		
C2'	C3'		
O3'	+P		
C2'	O3'		
C1'	H1'		
C2'	H2'		
C3'	H3'		
C4'	H4'		
C4'	H4''		
C5	H5		
C6	H6		

C2	O2		
C5	C6		
N3	C4		
[ impropers ]			
C2	N1	N3	O2
C4	N3	C5	N4
N4	C4	H41	H42

Table 59: Additional force field parameters used for threose-nucleic acid building blocks.

## Bonds :

CG2DC1	CG3RC1	1	0.14850000	305432.00
NG2D1	NG2D1	1	0.13660000	267776.00
CN8C	ON6B	1	0.14318000	217568.00
CN7	CN8C	1	0.15228000	186188.00
CN8C	HN8	1	0.11020000	258571.20
ON2	P5	1	0.16000000	225936.00
ON3	P5	1	0.14800000	485344.00
CN7B	CG2R61	1	0.15120000	192464.00

## Bending.

ON6B	CN8C	HN8	5	109.880000	378.233600	0.00000000	0.00
HN7	CN7	CN8C	5	112.780000	334.720000	0.00000000	0.00
CN8C	ON6B	CN7B	5	110.800000	920.480000	0.00000000	0.00
CN7	CN8C	ON6B	5	105.350000	836.800000	0.00000000	0.00
CN7B	CN7	CN8C	5	100.820000	502.080000	0.25610000	6694.40
CN8C	CN7	ON2	5	108.390000	962.320000	0.00000000	0.00
HN8	CN8C	HN8	5	109.960000	297.064000	0.18020000	4518.72
HN8	CN8C	CN7	5	111.320000	288.947040	0.21790000	18853.10
CN8C	CN7	ON5	5	108.390000	962.320000	0.00000000	0.00
CN7	ON2	P5	5	120.000000	167.360000	0.23300000	29288.00
CN7B	ON2	P5	5	120.000000	167.360000	0.23300000	29288.00
ON2	P5	ON2	5	104.300000	669.440000	0.00000000	0.00
ON2	P5	ON3	5	111.600000	827.595200	0.00000000	0.00
ON3	P5	ON3	5	120.000000	1004.160000	0.00000000	0.00
ON6	CN7B	G2R61	5	110.500000	920.480000	0.00000000	0.00
HN7	CN7B	CG2R61	5	108.740000	418.400000	0.00000000	0.00
CN8	CN7B	CG2R61	5	115.800000	453.100000	0.00000000	0.00

---

CN7B	CG2R61	CG2R61	5	120.000000	383.400000	0.00000000	0.00
ON6B	CN7B	CG2R61	5	110.500000	920.480000	0.00000000	0.00
CN7B	CN7B	CG2R61	5	115.800000	453.100000	0.00000000	0.00

---

## Torsions:

ON2	CN7B	CN7	ON2	9	0.000000	0.000000	3 ;
ON5	CN7B	CN7	ON2	9	0.000000	20.000000	1 ;
ON5	CN7	CN7B	ON2	9	0.000000	0.000000	3 ;
ON5	CN7	CN7B	ON2	9	0.000000	20.000000	1 ;
ON2	P2	ON2	CN7B	9	180.000000	3.765600	1 ;
ON2	P2	ON2	CN7B	9	180.000000	1.673600	2 ;
ON2	P2	ON2	CN7B	9	180.000000	0.836800	3 ;
ON3	P2	ON2	CN7B	9	0.000000	0.418400	3 ;
P2	ON2	CN7B	HN7	9	0.000000	0.000000	3 ;
CN8C	ON6B	CN7B	NN2	9	0.000000	0.000000	3 ;
CN8C	ON6B	CN7B	NN2B	9	0.000000	0.000000	3 ;
ON2	CN7	CN8C	ON6B	9	0.000000	1.673600	6 ;
ON2	CN7	CN8C	ON6B	9	0.000000	0.000000	5 ;
ON2	CN7	CN8C	ON6B	9	180.000000	0.000000	4 ;
ON2	CN7	CN8C	ON6B	9	0.000000	6.694400	3 ;
HN8	CN8C	CN7	ON2	9	0.000000	0.815880	3 ;
CN8C	CN7	ON2	P2	9	0.000000	2.510400	5 ;
CN8C	CN7	ON2	P2	9	0.000000	0.836800	4 ;
CN8C	CN7	ON2	P2	9	180.000000	0.000000	3 ;
CN8C	CN7	ON2	P2	9	0.000000	1.673600	2 ;
CN8C	CN7	ON2	P2	9	180.000000	7.949600	1 ;
CN7B	CN7	ON2	P2	9	0.000000	2.510400	5 ;
CN7B	CN7	ON2	P2	9	0.000000	0.836800	4 ;
CN7B	CN7	ON2	P2	9	180.000000	0.000000	3 ;
CN7B	CN7	ON2	P2	9	0.000000	1.673600	2 ;
CN7B	CN7	ON2	P2	9	180.000000	7.949600	1 ;
CN8C	CN7	ON2	P	9	0.000000	2.510400	5 ;
CN8C	CN7	ON2	P	9	0.000000	0.836800	4 ;
CN8C	CN7	ON2	P	9	180.000000	0.000000	3 ;
CN8C	CN7	ON2	P	9	0.000000	1.673600	2 ;
CN8C	CN7	ON2	P	9	180.000000	7.949600	1 ;
CN7B	CN7B	ON2	P2	9	0.000000	2.510400	5 ;
CN7B	CN7B	ON2	P2	9	0.000000	0.836800	4 ;
CN7B	CN7B	ON2	P2	9	180.000000	0.000000	3 ;
CN7B	CN7B	ON2	P2	9	0.000000	1.673600	2 ;

---

---

CN7B	CN7B	ON2	P2	9	180.000000	7.949600	1 ;
CN7	CN7B	ON2	P2	9	0.000000	2.510400	5 ;
CN7	CN7B	ON2	P2	9	0.000000	0.836800	4 ;
CN7	CN7B	ON2	P2	9	180.000000	0.000000	3 ;
CN7	CN7B	ON2	P2	9	0.000000	1.673600	2 ;
CN7	CN7B	ON2	P2	9	180.000000	7.949600	1 ;
CN7B	CN7B	ON6B	CN8C	9	0.000000	0.000000	6 ;
CN7B	CN7	CN8C	ON6B	9	0.000000	0.000000	3 ;
CN7B	CN7B	CN7	CN8C	9	0.000000	0.000000	6 ;
CN7	CN8C	ON6B	CN7B	9	180.000000	0.000000	6 ;
HN8	CN8C	CN7	CN7B	9	0.000000	0.000000	3 ;
HN8	CN8C	CN7	HN7	9	0.000000	0.815880	3 ;
HN7	CN7	CN8C	ON6B	9	180.000000	0.815880	3 ;
HN7	CN7B	ON6B	CN8C	9	0.000000	0.000000	3 ;
HN8	CN8C	ON6B	CN7B	9	0.000000	0.815880	3 ;
CN8C	CN7	CN7B	ON2	9	0.000000	0.000000	3 ;
CN8C	CN7	CN7B	ON5	9	0.000000	0.000000	3 ;
HN7	CN7B	CN7	CN8C	9	0.000000	0.815880	3 ;
HN8	CN8C	CN7	ON5	9	0.000000	0.815880	3 ;
ON5	CN7	CN8C	ON6B	9	0.000000	1.673600	6 ;
CN8C	CN7	ON5	HN5	9	0.000000	2.092000	3 ;
CN8C	CN7	ON5	HN5	9	180.000000	1.255200	2 ;
CN8C	CN7	ON5	HN5	9	0.000000	6.276000	1 ;
ON2	P5	ON2	CN7	9	180.000000	3.765600	1 ;
ON2	P5	ON2	CN7	9	180.000000	1.673600	2 ;
ON2	P5	ON2	CN7	9	180.000000	0.836800	3 ;
N3	P5	ON2	CN7	9	0.000000	0.418400	3 ;
P5	ON2	CN7	HN7	9	0.000000	0.000000	3 ;
ON2	P5	ON2	CN7B	9	180.000000	3.765600	1 ;
ON2	P5	ON2	CN7B	9	180.000000	1.673600	2 ;
ON2	P5	ON2	CN7B	9	180.000000	0.836800	3 ;
ON3	P5	ON2	CN7B	9	0.000000	0.418400	3 ;
P5	ON2	CN7B	HN7	9	0.000000	0.000000	3 ;
CN8C	CN7	ON2	P5	9	180.000000	7.949600	1 ;
CN7B	CN7	ON2	P5	9	0.000000	2.510400	5 ;
CN7B	CN7	ON2	P5	9	0.000000	0.836800	4 ;
CN7B	CN7	ON2	P5	9	180.000000	0.000000	3 ;
CN7B	CN7	ON2	P5	9	0.000000	1.673600	2 ;
CN7B	CN7	ON2	P5	9	180.000000	7.949600	1 ;
CN7B	CN7B	ON2	P5	9	0.000000	2.510400	5 ;
CN7B	CN7B	ON2	P5	9	0.000000	0.836800	4 ;

---

---

CN7B	CN7B	ON2	P5	9	180.000000	0.000000	3 ;
CN7B	CN7B	ON2	P5	9	0.000000	1.673600	2 ;
CN7B	CN7B	ON2	P5	9	180.000000	7.949600	1 ;
CN7	CN7B	ON2	P5	9	0.000000	2.510400	5 ;
CN7	CN7B	ON2	P5	9	0.000000	0.836800	4 ;
CN7	CN7B	ON2	P5	9	180.000000	0.000000	3 ;
CN7	CN7B	ON2	P5	9	0.000000	1.673600	2 ;
CN7	CN7B	ON2	P5	9	180.000000	7.949600	1 ;
CN7	CN7	ON2	P5	9	0.000000	2.510400	5 ;
CN7	CN7	ON2	P5	9	0.000000	0.836800	4 ;
CN7	CN7	ON2	P5	9	180.000000	0.000000	3 ;
CN7	CN7	ON2	P5	9	0.000000	1.673600	2 ;
CN7	CN7	ON2	P5	9	180.000000	7.949600	1 ;
CN8	CN7	ON2	P5	9	180.000000	7.949600	1 ;
NN2B	CG2R61	CG2R61	HGR62	9	180.000000	3.347200	2 ;
CN7	ON6	CN7B	CG2R61	9	0.000000	0.000000	3 ;
ON6	CN7B	CG2R61	CG2R61	9	0.000000	3.765600	1 ;
ON6	CN7B	NN2B	CG2R61	9	0.000000	3.765600	1 ;
ON6	CN7B	NN2B	CG2R63	9	180.000000	3.347200	1 ;
CN8	CN7B	NN2B	CG2R63	9	180.000000	0.836800	3 ;
CN8	CN7B	NN2B	CG2R61	9	180.000000	0.000000	3 ;
CN8C	ON6B	CN7B	CG2R61	9	0.000000	0.000000	3 ;
ON6B	CN7B	CG2R61	CG2R61	9	0.000000	3.765600	1 ;
CN7B	CN7B	CG2R61	CG2R61	9	180.000000	0.000000	3 ;
CG2R61	CN7B	CN7B	ON5	9	0.000000	0.000000	3 ;
CG2R61	CN7B	CN7B	CN7	9	0.000000	0.000000	3 ;
CG2R61	CN7B	CN7B	HN7	9	0.000000	0.000000	3 ;
HN7	CN7B	CG2R61	CG2R61	9	0.000000	0.815880	3 ;
CN8	CN7B	CG2R61	CG2R61	9	180.000000	0.000000	3 ;
HN8	CN8	CN7B	CG2R61	9	0.000000	0.000000	3 ;
CN7	CN8	CN7B	CG2R61	9	0.000000	0.000000	3 ;
CG2R61	CG2R61	CG2R61	CN7B	9	180.000000	3.100000	2 ;
CN7B	CG2R61	CG2R61	OG301	9	180.000000	2.400000	2 ;
CN7B	CG2R61	CG2R61	HGR61	9	180.000000	2.400000	2 ;

---

### 5.3.13.2 Preparation of the starting models

To build the starting structure of the Kod-RSGA polymerase in complex with a DNA-TNA hybrid and an incoming TNA triphosphate, the Kod-RI crystal structure (PDB code: 5VU8<sup>151</sup>) was converted to the Kod-RSGA sequence using molecular modeling and local energy minimization via HyperChem (Release 7.01, *Hypercube, Inc.*). In detail, Asn491 was mutated to Ser491, Arg606 to Gly606, Ile664 to Glu664 and Thr723 to Ala723. The DNA template sequence was converted to DNA\_Temp\_mid2ATGC bearing either **dNaM** or **dTPT3** at the -1 templating position. The nascent TNA oligonucleotide was constructed first as DNA, matching the respective nucleobase positions of the original DNA primer in the crystal structure. The carbohydrate part of the DNA was then converted to the TNA structure and the carbohydrate geometry locally relaxed within the HyperChem force field.

Three Mg<sup>2+</sup> ions and the incoming triphosphate, which are resolved in the X-ray structure of Kod-RI, were also included in the starting geometry. The starting structures were placed in a cubic box and solvated with TIP3P water molecules. Sodium and chloride ions were added to get a neutral system with an ionic concentration of 0.15 mol L<sup>-1</sup>. The procedure results in a 12.1 nm<sup>3</sup> cubic box filled with 56212 TIP3P water molecules, 179 Na<sup>+</sup> and 167 Cl<sup>-</sup> ions for both systems, containing the **dNaM-tTPT3** duplex or the **dTPT3-tNaM** duplex.

### 5.3.13.3 MD simulations

The starting model systems were energy minimized switching alternatively between runs using steepest descent gradients or Polak-Ribiere conjugate gradients until convergence to machine precision. Subsequently, 480 ps MD calculations at constant temperature (300 K, NVT) followed by 480 ps MD calculations at constant pressure (1 bar, NPT) equilibrate solvent and ions. Finally, MD trajectories of 200 ns were calculated without restraints or constraints (except bond lengths to hydrogens) at 300 K.

### 5.3.14 Preparation of cyclopropene-modified DNA

Synthesized  $\text{dA}^{\text{CPTP}}$  was employed in DNA primer extension assays to analyze the incorporation of the modified adenosine triphosphate into DNA.

#### 5.3.14.1 DNA primer extension assays with $\text{dA}^{\text{CPTP}}$

DNA primer extensions were performed with complete substitution of  $\text{dATP}$  with  $\text{dA}^{\text{CPTP}}$ . DNA primer extensions were performed in a 25  $\mu\text{L}$  scale with final concentrations of 1 $\times$  Taq reaction buffer, 2.5  $\mu\text{M}$  DNA primer, 2.0  $\mu\text{M}$  5'-phosphate modified DNA template, 2 mM  $\text{MgCl}_2$ , 400  $\mu\text{M}$  of each dNTP except  $\text{dATP}$ , 800  $\mu\text{M}$  of  $\text{dA}^{\text{CPTP}}$  and 0.1  $\text{U } \mu\text{L}^{-1}$  Taq DNA polymerase. A positive control with canonical  $\text{dATP}$  instead of  $\text{dA}^{\text{CPTP}}$  was included to serve as a size marker for the full-length DNA product. The pipetting schemes are provided in Tables 60 and 61.

**Table 60: General pipetting scheme for DNA primer extension assays with the cyclopropene-modified  $\text{dA}^{\text{CPTP}}$ .**

Component	Stock concentration	Final concentration	Volume [ $\mu\text{L}$ ]
DNA_Temp_mid2ATGC	100 $\mu\text{M}$	2 $\mu\text{M}$	0.5
PEX_primer	100 $\mu\text{M}$	2.5 $\mu\text{M}$	0.75
Taq reaction buffer	10 $\times$	1 $\times$	2.5
$\text{MgCl}_2$	100 mM	2 mM	0.5
ddH <sub>2</sub> O	-	-	14.35
Taq polymerase	2.5 $\text{U } \mu\text{L}^{-1}$	0.1 $\text{U } \mu\text{L}^{-1}$	2
d(T, G, C)TP	25 mM each	400 $\mu\text{M}$ each	0.4
$\text{dA}^{\text{CPTP}}$	5 mM	800 $\mu\text{M}$	4

**Table 61: General pipetting scheme for the positive control of DNA primer extension assays using only canonical nucleotides.**

Component	Stock concentration	Final concentration	Volume [ $\mu\text{L}$ ]
DNA_Temp_mid2ATGC	100 $\mu\text{M}$	2 $\mu\text{M}$	0.5
PEX_primer	100 $\mu\text{M}$	2.5 $\mu\text{M}$	0.75
Taq reaction buffer	10 $\times$	1 $\times$	2.5
$\text{MgCl}_2$	100 mM	2 mM	0.5
ddH <sub>2</sub> O	-	-	18.35
Taq polymerase	2.5 $\text{U } \mu\text{L}^{-1}$	0.1 $\text{U } \mu\text{L}^{-1}$	2
dNTPs	25 mM each	400 $\mu\text{M}$ each	0.4

In a first step, Taq reaction buffer, DNA primer and template, MgCl<sub>2</sub> and ddH<sub>2</sub>O were mixed in a tube and a denaturing step at 95 °C for 5 min was performed. Subsequently, primer and template were annealed at 4 °C for 10 min. Then, Taq polymerase and the nucleotides were added and the primer extension was incubated for 30 min at 68 °C. The conditions used for the DNA primer extension assays are depicted in Table 62.

**Table 62: Temperature profile for DNA primer extension assays with dA<sup>CP</sup>TP.**

	T [°C]	t [min]
Denaturation	95	10
Annealing	4	10
Elongation	68	30
Hold	4	∞

Subsequently, the primer extension assays were purified using centrifugal filters (Chapter 5.3.7.3) and the templates were digested using λ.exonuclease (Chapter 5.3.7.5), followed by another purification step with centrifugal filters (Chapter 5.3.7.3). The concentration of the oligonucleotide samples was determined using a spectrophotometer (Chapter 5.3.7.4). Next, the samples were submitted to iEDDA click reaction with a tetrazine-fluorophore conjugate (Chapter 5.3.14.2).

### 5.3.14.2 iEDDA click reaction of DNA<sup>CP</sup> with a tetrazine-fluorophore conjugate

To evidence if dA<sup>CP</sup>TP was incorporated into DNA during the primer extension assay, the prepared oligonucleotides were subjected to an iEDDA click reaction with a tetrazine-BODIPY conjugate. Therefore, 100 pmol of cyclopropene-modified DNA<sup>CP</sup> were mixed with a 50 fold excess of 6-methyl-tetrazine-BODIPY (500 μM in DMSO) in a total volume of 15 μL. Click reactions were incubated at 25 °C for 2 h under the exclusion of light. As controls, the same reactions were concomitantly prepared containing unmodified DNA<sup>WT</sup> or ddH<sub>2</sub>O instead of DNA<sup>CP</sup>. The samples originating from the click reactions were freeze-dried and redissolved in 6 μL ddH<sub>2</sub>O. In addition, 100 pmol of untreated DNA<sup>WT</sup> and DNA<sup>CP</sup> were freeze-dried and redissolved in 6 μL of ddH<sub>2</sub>O serving as unlabeled, non-fluorescent controls. Subsequently, 20 % dPAGE analysis of all samples was performed as described in Chapter 5.3.7.2.



### 5.3.15 Expression and purification of mZBP1 variants

The murine Z-RNA binding protein 1 consists of two different Z-RNA binding domains, Z $\alpha$ 1 and Z $\alpha$ 2. To evaluate the role of these protein domains in Z-RNA binding, mZBP1<sup>WT</sup> and three loss-of-function mutants (mZBP1<sup>mZ $\alpha$ 1</sup>, mZBP1<sup>mZ $\alpha$ 2</sup>, mZBP1<sup>mZ $\alpha$ 1-2</sup>) were overexpressed and purified by affinity chromatography to allow assays to investigate the binding of the different mutants to Z-RNA and A-RNA. The loss-of-function mutations of the Z $\alpha$  domains were adapted from Ha *et al.*<sup>408</sup>

#### 5.3.15.1 Expression of mZBP1 variants

The expression and purification of the mZBP1 variants were conducted using a protocol adapted from the published procedures from Nichols *et al.* and Placido *et al.*<sup>409,410</sup>

A starter culture was prepared, by inoculating a 100 mL culture of liquid TB containing 50  $\mu\text{g mL}^{-1}$  kanamycin with a glycerol stock of NEB T7 Express *lysY/l<sup>q</sup>* *E.coli* cells harboring either the pET28a(+)\_mZBP1\_WT, pET28a(+)\_mZBP1-Z $\alpha$ 1, pET28a(+)\_mZBP1-Z $\alpha$ 2 or pET28a(+)\_mZBP1-Z $\alpha$ 1-2 expression plasmid. The *E.coli* cells were grown aerobically at 37 °C and 140 rpm overnight. The OD<sub>600</sub> of a 1:10 dilution of the starter culture was measured and the required volume of starter culture to prepare a 1000 mL expression culture with an OD<sub>600</sub> of 0.1 was calculated. The expression culture was prepared accordingly and was incubated at 37 °C and 120 rpm shaking. At an OD<sub>600</sub> of 0.6, protein expression was induced by addition of 0.5 mL 1.0 M isopropyl- $\beta$ -D-thiogalactopyranoside to a final concentration of 0.5 mM. The induced expression culture was incubated at 15 °C for 18 h and 120 rpm. Subsequently, cells were harvested by centrifugation at 4250  $\times$  g for 35 min at 4 °C and the supernatant was discarded. The cell pellet was washed by resuspension in 40 mL lysis buffer 2 (50 mM Tris-HCl, 300 mM NaCl, 10 % glycerol, 10 mM imidazole, 5 mM  $\beta$ -mercaptoethanol, pH 8.0) and centrifugation at 4250  $\times$  g for 35 min at 4 °C. The supernatant was discarded and the cell pellet was resuspended in 40 mL lysis buffer 2. Cells were then lysed by sonication on ice using a 2 sec on/ 4 sec off protocol for 2 min working time. The sonication step was performed three times. In between sonication steps, the sample was incubated on ice for 5 min. Subsequently, the cell debris was precipitated by centrifugation at 20000  $\times$  g for 30 min at 4 °C and the clarified supernatant was collected and subjected to purification by affinity chromatography (Chapter 5.3.15.2).

### 5.3.15.2 IMAC purification of mZBP1 variants

#### Purification protocol

The supernatant from the previous step containing either mZBP1<sup>WT</sup>, mZBP1<sup>mZα1</sup>, mZBP1<sup>mZα2</sup>, or mZBP1<sup>mZα1-2</sup> was purified using an *Äkta start* FPLC chromatography system equipped with a 1 mL *HiTrap TALON crude* column. Prior to sample application, the column was equilibrated with 5 CV of 100 % Buffer B2 (50 mM Tris-HCl, 300 mM NaCl, 10 % glycerol, 500 mM imidazole, 1 mM β-mercaptoethanol, pH 8.0), followed by equilibration with 5 CV of 100 % Buffer A2 (50 mM Tris-HCl, 1.0 M NaCl, 10 % glycerol, 10 mM imidazole, 5 mM β-mercaptoethanol, pH 8.0) at a flow rate of 1 mL min<sup>-1</sup>. Then, 15 mL of the supernatant were applied to the column. After sample application, the column was washed with 15 column volumes of Buffer A2. His-tagged protein was then eluted using a stepwise gradient from 0 % → 100 % of Buffer B2 at a flow rate of 1 mL min<sup>-1</sup> as depicted in Table 63. The eluted fractions containing the protein of interest were visualized using 15 % SDS-PAGE (Chapter 5.3.7.2) and pooled.

**Table 63: FPLC gradient of the IMAC purification of the mZBP1 variants.** A stepwise gradient from 0 % → 100 % of Buffer B2 was used.

Buffer B2 [%]	CV
1	2
2	2
3	1
5	2
10	3
20	2
30	3
40	1
50	1
100	10

#### Buffer exchange

Buffer exchange of the pooled polymerase fractions into ZBP1 storage buffer (20 mM HEPES, 300 mM NaCl, 2 mM DTT, 50 % glycerol, pH 7.4) was performed by dialysis using a dialysis membrane with a molecular weight cut-off of 10 kDa. Therefore, the pooled fractions were transferred into the dialysis tubing that was closed on one end. The tubing was sealed tightly and placed in a beaker containing a volume of high salt dialysis buffer (20 mM HEPES, 600 mM NaCl, 2 mM DTT, pH 7.4) equal to 100 times the volume of the sample. Dialysis against the high salt buffer was performed 3 h while stirring at 4 °C. Subsequently, the protein-containing dialysis tubing was transferred into a beaker with ZBP1 storage buffer. Again the volume of the dialysis buffer was 100 times the volume of the sample. Dialysis against the ZBP1 storage

buffer was performed for 24 h at 4 °C while stirring. Then, the protein-containing sample was removed from the dialysis tubing using a pipette and was stored in 1 mL aliquots at -20 °C. The protein concentration of the mZBP1 variants was determined using the *Pierce 660 nm Protein Assay Kit* (Chapter 5.3.8.10).

### 5.3.16 Interaction studies between mZBP1 variants and fluorescently-labeled RNA in A- and Z-conformation

Purified mZBP1 variants were investigated regarding their Z-RNA binding capability using an electrophoretic mobility shift assay. To stabilize the Z-form of RNA, the utilized **GC**-rich RNA hairpin was modified with the m<sup>8</sup>G Z-RNA stabilizer (for the exact sequence, please refer to Chapter 7.5.2 – Z-RNA<sup>m8G</sup>). In addition, for later visualization of the EMSA, the Z-RNA was modified with a 5'-amino group to facilitate click labeling with an NHS-fluorophore conjugate.

#### 5.3.16.1 NHS-ester click reaction for fluorescent labeling of amino-modified Z-RNA

To allow visualization of RNA after EMSA, 5'-amino modified Z-RNA<sup>m8G</sup> and 5'-amino modified A-RNA<sup>WT</sup> were fluorescently labeled with AF488 NHS-ester (in DMSO). Therefore, a click reaction was prepared in a 25 µL scale containing 2 nmol of either Z-RNA<sup>m8G</sup> or A-RNA<sup>WT</sup>, 20 nmol of AF488 NHS-ester in a final concentration of 80 mM NaHCO<sub>3</sub> buffer. A detailed pipetting scheme is depicted in Table 64.

**Table 64: General pipetting scheme for the NHS-ester click reaction of 5'-amino modified RNA with AF488 NHS-ester.** As RNA, either RNA<sup>m8G</sup> or RNA<sup>WT</sup> was employed.

Component	Stock concentration	Final concentration	Volume [µL]
RNA	100 µM	80 pmol µL <sup>-1</sup>	20
AF488 NHS-ester	10 mM	0.8 mM	2
NaHCO <sub>3</sub> buffer	1.0 M	80 mM	2
ddH <sub>2</sub> O	-	-	1

The click reaction was then incubated at 25 °C for 2 h in the dark, followed by purification using ethanol precipitation (Chapter 5.3.7.3). Concentration of AF-488 labeled RNA was determined using NanoDrop (Chapter 5.3.7.4).

### 5.3.16.2 Electrophoretic mobility shift assay

Initially, the minimal detectable amount of fluorescently-labeled RNA was determined by titrating experiments. Therefore, different amounts of fluorescently-labeled Z-RNA<sup>m8G</sup> and A-RNA<sup>WT</sup> (*i.e.* 5 pmol, 15 pmol, 25 pmol, and 50 pmol) were combined with equal ratios of nPAGE loading buffer and subjected to 8 % nPAGE analysis (Chapter 5.3.7.2), followed by visualization by fluorescence scanning. 25 pmol were identified as a sufficient amount for the detection of a fluorescent signal on nPAGE.

Electrophoretic mobility shift assays were performed with mZBP1<sup>WT</sup> and RNA in its Z- or A-conformation in a 10  $\mu$ L scale. Therefore, titrating amounts of mZBP1<sup>WT</sup> (*i.e.* 15 pmol, 25 pmol, 50 pmol, 100 pmol) were incubated with 25 pmol of AF488-labeled Z-RNA<sup>m8G</sup> or A-RNA<sup>WT</sup>, respectively. Detailed pipetting schemes are depicted in Tables 65 and 66.

**Table 65: General pipetting scheme for EMSA with Z-RNA<sup>m8G</sup> and mZBP1<sup>WT</sup>.**

Component	Stock concentrations	Final concentrations	V [ $\mu$ L]
Z-RNA <sup>m8G</sup>	33 $\mu$ M	2.5 $\mu$ M	0.75
			<i>sample 1:</i> 5.0
			<i>sample 2:</i> 4.4
			<i>sample 3:</i> 2.94
			<i>sample 4:</i> 0
		1.5 $\mu$ M	<i>sample 1:</i> 0.88
		2.5 $\mu$ M	<i>sample 2:</i> 1.47
		5.0 $\mu$ M	<i>sample 3:</i> 2.94
		10 $\mu$ M	<i>sample 4:</i> 5.88
ddH <sub>2</sub> O	-	-	4.37

\* binding buffer equals mZBP1 storage buffer

**Table 66: General pipetting scheme for EMSA with A-RNA<sup>WT</sup> and mZBP1<sup>WT</sup>.**

Components	Stock concentrations	Final concentrations	V [ $\mu$ L]
A-RNA <sup>WT</sup>	56 $\mu$ M	2.5 $\mu$ M	0.45
			<i>sample 1:</i> 5.0
			<i>sample 2:</i> 4.4
Binding buffer*	1x	0.6x	<i>sample 3:</i> 2.94
			<i>sample 4:</i> 0
		1.5 $\mu$ M	<i>sample 1:</i> 0.88
		2.5 $\mu$ M	<i>sample 2:</i> 1.47
mZBP1 <sup>WT</sup>	17 $\mu$ M	5.0 $\mu$ M	<i>sample 3:</i> 2.94
		10 $\mu$ M	<i>sample 4:</i> 5.88
ddH <sub>2</sub> O	-	-	4.67

\* binding buffer equals mZBP1 storage buffer

Incubation of the samples was performed in binding buffer at final concentrations of 6 mM HEPES, 30 % glycerol, 175 mM NaCl, and 1.2 mM DTT (pH 7.4) for 30 min at 24 °C in the dark. Subsequently, the samples were loaded onto 8 % nPAGE (Chapter 5.3.7.2) and were visualized by fluorescence scanning for the evaluation of the results.

### 5.3.16.3 Microscale thermophoresis

Microscale thermophoresis was performed to determine the dissociation constant ( $K_d$ ) between AF488-labeled Z-RNA<sup>m8G</sup> and mZBP1<sup>WT</sup>. Binding titrations were performed by titrating varying amounts of mZBP1<sup>WT</sup> into 20 nM AF488-labeled Z-RNA<sup>m8G</sup> in interaction buffer (20 mM HEPES, 150 mM NaCl, 2 mM DTT, 0.05 % (v/v) Tween 20, pH 7.4). Titrations included 16 concentration steps ranging from approximately 0.15 nM to 500 nM (please refer to Table 67 for detailed mZBP1<sup>WT</sup> concentrations). Each solution was filled into Monolith capillaries and thermophoresis was measured at 22 °C with 40 % excitation power and 60 % MST power.

**Table 67: mZBP1<sup>WT</sup> ligand concentrations used for MST measurements with Z-RNA<sup>m8G</sup>.**

<b>Sample</b>	<b>mZBP1 concentration [nM]</b>
<b>1</b>	500
<b>2</b>	250
<b>3</b>	125
<b>4</b>	62.5
<b>5</b>	31.25
<b>6</b>	15.63
<b>7</b>	7.813
<b>8</b>	3.906
<b>9</b>	1.953
<b>10</b>	0.977
<b>11</b>	0.488
<b>12</b>	0.244
<b>13</b>	0.122
<b>14</b>	0.061
<b>15</b>	0.052
<b>16</b>	0.153

## 6 References

- (1) Crick, F. On protein synthesis. *Symposia of the Society for Experimental Biology* **1958**, *12*, 138–163.
- (2) Crick, F. *Ideas on protein synthesis*. Unpublished Note, 1956. <https://wellcomecollection.org/works/xmscu3g4> (accessed 2024-02-13).
- (3) Thieffry, D.; Sarkar, S. Forty years under the central dogma. *Trends in biochemical sciences* **1998**, *23* (8), 312–316. DOI: 10.1016/s0968-0004(98)01244-4.
- (4) Morris, R.; Black, K. A.; Stollar, E. J. Uncovering protein function: from classification to complexes. *Essays in biochemistry* **2022**, *66* (3), 255–285. DOI: 10.1042/EBC20200108.
- (5) Garcia, S.; Acosta-Reyes, F. J.; Saperas, N.; Campos, J. L. *Crystal Structure of rich-AT DNA 20mer*, 2016. DOI: 10.2210/pdb5f9i/pdb.
- (6) Holland, J. A.; Hansen, M. R.; Du, Z.; Hoffman, D. W. *An Investigation of the Structure of the Pseudoknot within the Gene 32 Messenger RNA of Bacteriophage T2 using Heteronuclear NMR Methods*, 1998. DOI: 10.2210/pdb2tpk/pdb.
- (7) Bergen, K.; Betz, K.; Welte, W.; Diederichs, K.; Marx, A. *KOD Polymerase in binary complex with dsDNA*, 2013. DOI: 10.2210/pdb4k8z/pdb.
- (8) Crick, F. Central dogma of molecular biology. *Nature* **1970**, *227* (5258), 561–563. DOI: 10.1038/227561a0.
- (9) Central dogma reversed. *Nature* **1970**, *226* (5252), 1198–1199. DOI: 10.1038/2261198a0.
- (10) Koonin, E. V. Does the central dogma still stand? *Biology direct* **2012**, *7*, 27. DOI: 10.1186/1745-6150-7-27. Published Online: Aug. 23, 2012.
- (11) Prusiner, S. B. Novel proteinaceous infectious particles cause scrapie. *Science (New York, N.Y.)* **1982**, *216* (4542), 136–144. DOI: 10.1126/science.6801762.
- (12) Tuite, M. F.; Serio, T. R. The prion hypothesis: from biological anomaly to basic regulatory mechanism. *Nature reviews. Molecular cell biology* **2010**, *11* (12), 823–833. DOI: 10.1038/nrm3007. Published Online: Nov. 17, 2010.
- (13) Lingner, J.; Hughes, T. R.; Shevchenko, A.; Mann, M.; Lundblad, V.; Cech, T. R. Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science (New York, N.Y.)* **1997**, *276* (5312), 561–567. DOI: 10.1126/science.276.5312.561.
- (14) Finnegan, D. J. Eukaryotic transposable elements and genome evolution. *Trends in genetics : TIG* **1989**, *5* (4), 103–107. DOI: 10.1016/0168-9525(89)90039-5.
- (15) Inouye, S.; Hsu, M. Y.; Eagle, S.; Inouye, M. Reverse transcriptase associated with the biosynthesis of the branched RNA-linked msDNA in *Myxococcus xanthus*. *Cell* **1989**, *56* (4), 709–717. DOI: 10.1016/0092-8674(89)90593-X.
- (16) Inouye, S.; Inouye, M. Structure, function, and evolution of bacterial reverse transcriptase. *Virus genes* **1995**, *11* (2-3), 81–94. DOI: 10.1007/BF01728650.
- (17) Mattick, J. S. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO reports* **2001**, *2* (11), 986–991. DOI: 10.1093/embo-reports/kve230.
- (18) Geisler, S.; Collier, J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nature reviews. Molecular cell biology* **2013**, *14* (11), 699–712. DOI: 10.1038/nrm3679. Published Online: Oct. 9, 2013.
- (19) Mercer, T. R.; Dinger, M. E.; Mattick, J. S. Long non-coding RNAs: insights into functions. *Nature reviews. Genetics* **2009**, *10* (3), 155–159. DOI: 10.1038/nrg2521.
- (20) Mattick, J. S.; Amaral, P. P.; Carninci, P.; Carpenter, S.; Chang, H. Y.; Chen, L.-L.; Chen, R.; Dean, C.; Dinger, M. E.; Fitzgerald, K. A.; Gingeras, T. R.; Guttman, M.; Hirose, T.; Huarte, M.; Johnson, R.; Kanduri, C.; Kapranov, P.; Lawrence, J. B.; Lee, J. T.; Mendell, J. T.; Mercer, T. R.; Moore, K. J.; Nakagawa, S.; Rinn, J. L.; Spector, D. L.; Ulitsky, I.; Wan, Y.; Wilusz, J. E.; Wu, M. Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nature reviews. Molecular cell biology* **2023**, *24* (6), 430–447. DOI: 10.1038/s41580-022-00566-8. Published Online: Jan. 3, 2023.

- (21) Gilbert, W. Origin of life: The RNA world. *Nature* **1986**, 319 (6055), 618. DOI: 10.1038/319618a0.
- (22) Di Giulio, M. On the RNA world: evidence in favor of an early ribonucleopeptide world. *Journal of molecular evolution* **1997**, 45 (6), 571–578. DOI: 10.1007/pl00006261.
- (23) Kumar, D.; Steele, E. J.; Wickramasinghe, N. C. Preface: The origin of life and astrobiology. *Advances in Genetics* **2020**, 106, xv–xviii. DOI: 10.1016/S0065-2660(20)30037-7.
- (24) Weiss, M. C.; Preiner, M.; Xavier, J. C.; Zimorski, V.; Martin, W. F. The last universal common ancestor between ancient Earth chemistry and the onset of genetics. *PLoS Genetics* **2018**, 14 (8). DOI: 10.1371/journal.pgen.1007518.
- (25) Lazcano, A.; Miller, S. L. The origin and early evolution of life: prebiotic chemistry, the pre-RNA world, and time. *Cell* **1996**, 85 (6), 793–798. DOI: 10.1016/S0092-8674(00)81263-5.
- (26) Johnston, W. K.; Unrau, P. J.; Lawrence, M. S.; Glasner, M. E.; Bartel, D. P. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science (New York, N.Y.)* **2001**, 292 (5520), 1319–1325. DOI: 10.1126/science.1060786.
- (27) Cech, T. R. A model for the RNA-catalyzed replication of RNA. *Proceedings of the National Academy of Sciences of the United States of America* **1986**, 83 (12), 4360–4363. DOI: 10.1073/pnas.83.12.4360.
- (28) Moore, P. B.; Steitz, T. A. The involvement of RNA in ribosome function. *Nature* **2002**, 418 (6894), 229–235. DOI: 10.1038/418229a.
- (29) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science (New York, N.Y.)* **2000**, 289 (5481), 905–920. DOI: 10.1126/science.289.5481.905.
- (30) Wimberly, B. T.; Brodersen, D. E.; Clemons, W. M.; Morgan-Warren, R. J.; Carter, A. P.; Vonrhein, C.; Hartsch, T.; Ramakrishnan, V. Structure of the 30S ribosomal subunit. *Nature* **2000**, 407 (6802), 327–339. DOI: 10.1038/35030006.
- (31) Yusupov, M. M.; Yusupova, G. Z.; Baucom, A.; Lieberman, K.; Earnest, T. N.; Cate, J. H.; Noller, H. F. Crystal structure of the ribosome at 5.5 Å resolution. *Science (New York, N.Y.)* **2001**, 292 (5518), 883–896. DOI: 10.1126/science.1060089. Published Online: Mar. 29, 2001.
- (32) Ferus, M.; Pietrucci, F.; Saitta, A. M.; Knížek, A.; Kubelík, P.; Ivanek, O.; Shestivska, V.; Civiš, S. Formation of nucleobases in a Miller-Urey reducing atmosphere. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, 114 (17), 4306–4311. DOI: 10.1073/pnas.1700010114. Published Online: Apr. 10, 2017.
- (33) Benner, S. A.; Kim, H.-J.; Carrigan, M. A. Asphalt, water, and the prebiotic synthesis of ribose, ribonucleosides, and RNA. *Accounts of chemical research* **2012**, 45 (12), 2025–2034. DOI: 10.1021/ar200332w. Published Online: Mar. 28, 2012.
- (34) Powner, M. W.; Gerland, B.; Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **2009**, 459 (7244), 239–242. DOI: 10.1038/nature08013.
- (35) Pearce, B. K. D.; Pudritz, R. E.; Semenov, D. A.; Henning, T. K. Origin of the RNA world: The fate of nucleobases in warm little ponds. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, 114 (43), 11327–11332. DOI: 10.1073/pnas.1710339114. Published Online: Oct. 2, 2017.
- (36) Becker, S.; Feldmann, J.; Wiedemann, S.; Okamura, H.; Schneider, C.; Iwan, K.; Crisp, A.; Rossa, M.; Amatov, T.; Carell, T. Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides. *Science (New York, N.Y.)* **2019**, 366 (6461), 76–82. DOI: 10.1126/science.aax2747.
- (37) Saladino, R.; Carota, E.; Botta, G.; Kapralov, M.; Timoshenko, G. N.; Rozanov, A. Y.; Krasavin, E.; Di Mauro, E. Meteorite-catalyzed syntheses of nucleosides and of other prebiotic compounds from formamide under proton irradiation. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, 112 (21), E2746-55. DOI: 10.1073/pnas.1422225112.



- (38) Ferris, J. P.; Hill, A. R.; Liu, R.; Orgel, L. E. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **1996**, *381* (6577), 59–61. DOI: 10.1038/381059a0.
- (39) Higgs, P. G.; Lehman, N. The RNA World: molecular cooperation at the origins of life. *Nature reviews. Genetics* **2015**, *16* (1), 7–17. DOI: 10.1038/nrg3841. Published Online: Nov. 11, 2014.
- (40) Bernhardt, H. S. The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others)(a). *Biology direct* **2012**, *7*, 23. DOI: 10.1186/1745-6150-7-23. Published Online: Jul. 13, 2012.
- (41) Joyce, G. F.; Schwartz, A. W.; Miller, S. L.; Orgel, L. E. The case for an ancestral genetic system involving simple analogues of the nucleotides. *Proceedings of the National Academy of Sciences of the United States of America* **1987**, *84* (13), 4398–4402. DOI: 10.1073/pnas.84.13.4398.
- (42) Joyce, G. F. RNA evolution and the origins of life. *Nature* **1989**, *338* (6212), 217–224. DOI: 10.1038/338217a0.
- (43) Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **2002**, *418* (6894), 214–221. DOI: 10.1038/418214a.
- (44) Eschenmoser, A. Chemical etiology of nucleic acid structure. *Science (New York, N.Y.)* **1999**, *284* (5423), 2118–2124. DOI: 10.1126/science.284.5423.2118.
- (45) Stadler B. M. R.; Stadler P. F. Molecular Replicator Dynamics. *Adv. Complex Syst.* **2003**, *06* (01), 47–77. DOI: 10.1142/S0219525903000724.
- (46) Wittung, P.; Nielsen, P. E.; Buchardt, O.; Egholm, M.; Nordén, B. DNA-like double helix formed by peptide nucleic acid. *Nature* **1994**, *368* (6471), 561–563. DOI: 10.1038/368561a0.
- (47) Gait, M. J.; Agrawal, S. Introduction and History of the Chemistry of Nucleic Acids Therapeutics. *Methods in molecular biology (Clifton, N.J.)* **2022**, *2434*, 3–31. DOI: 10.1007/978-1-0716-2010-6\_1.
- (48) Anosova, I.; Kowal, E. A.; Dunn, M. R.; Chaput, J. C.; van Horn, W. D.; Egli, M. The structural diversity of artificial genetic polymers. *Nucleic acids research* **2016**, *44* (3), 1007–1021. DOI: 10.1093/nar/gkv1472. Published Online: Dec. 15, 2015.
- (49) Pinheiro, V. B.; Holliger, P. The XNA world: progress towards replication and evolution of synthetic genetic polymers. *Current opinion in chemical biology* **2012**, *16* (3-4), 245–252. DOI: 10.1016/j.cbpa.2012.05.198. Published Online: Jun. 14, 2012.
- (50) McKenzie, L. K.; El-Khoury, R.; Thorpe, J. D.; Damha, M. J.; Hollenstein, M. Recent progress in non-native nucleic acid modifications. *Chemical Society reviews* **2021**, *50* (8), 5126–5164. DOI: 10.1039/D0CS01430C.
- (51) Shaw, B. R.; Dobrikov, M.; Wang, X.; Wan, J.; He, K.; Lin, J.-L.; Li, P.; Rait, V.; Sergueeva, Z. A.; Sergueev, D. Reading, writing, and modulating genetic information with boranophosphate mimics of nucleotides, DNA, and RNA. *Annals of the New York Academy of Sciences* **2003**, *1002*, 12–29. DOI: 10.1196/annals.1281.004.
- (52) Eckstein, F. Nucleoside phosphorothioates. *Annual review of biochemistry* **1985**, *54*, 367–402. DOI: 10.1146/annurev.bi.54.070185.002055.
- (53) Kimoto, M.; Hirao, I. Genetic alphabet expansion technology by creating unnatural base pairs. *Chemical Society reviews* **2020**, *49* (21), 7602–7626. DOI: 10.1039/d0cs00457j. Published Online: Oct. 5, 2020.
- (54) Ochoa, S.; Milam, V. T. Modified Nucleic Acids: Expanding the Capabilities of Functional Oligonucleotides. *Molecules (Basel, Switzerland)* **2020**, *25* (20). DOI: 10.3390/molecules25204659. Published Online: Oct. 13, 2020.
- (55) Bege, M.; Borbás, A. The Medicinal Chemistry of Artificial Nucleic Acids and Therapeutic Oligonucleotides. *Pharmaceuticals (Basel, Switzerland)* **2022**, *15* (8). DOI: 10.3390/ph15080909. Published Online: Jul. 22, 2022.
- (56) Esquiaqui, J. M.; Sherman, E. M.; Ye, J.-D.; Fanucci, G. E. Conformational Flexibility and Dynamics of the Internal Loop and Helical Regions of the Kink-Turn Motif in the Glycine

- Riboswitch by Site-Directed Spin-Labeling. *Biochemistry* **2016**, *55* (31), 4295–4305. DOI: 10.1021/acs.biochem.6b00287. Published Online: Jul. 29, 2016.
- (57) Palluk, S.; Arlow, D. H.; Rond, T. de; Barthel, S.; Kang, J. S.; Bector, R.; Baghdassarian, H. M.; Truong, A. N.; Kim, P. W.; Singh, A. K.; Hillson, N. J.; Keasling, J. D. De novo DNA synthesis using polymerase-nucleotide conjugates. *Nature biotechnology* **2018**, *36* (7), 645–650. DOI: 10.1038/nbt.4173. Published Online: Jun. 18, 2018.
- (58) Taylor, A. I.; Houlihan, G.; Holliger, P. Beyond DNA and RNA: The Expanding Toolbox of Synthetic Genetics. *Cold Spring Harbor perspectives in biology* **2019**, *11* (6). DOI: 10.1101/cshperspect.a032490. Published Online: Jun. 3, 2019.
- (59) Duffy, K.; Arangundy-Franklin, S.; Holliger, P. Modified nucleic acids: replication, evolution, and next-generation therapeutics. *BMC biology* **2020**, *18* (1), 112. DOI: 10.1186/s12915-020-00803-6. Published Online: Sep. 2, 2020.
- (60) Mayer, G. The chemical biology of aptamers. *Angew. Chem. Int. Ed.* **2009**, *48* (15), 2672–2689. DOI: 10.1002/anie.200804643.
- (61) Wan, W. B.; Seth, P. P. The Medicinal Chemistry of Therapeutic Oligonucleotides. *Journal of medicinal chemistry* **2016**, *59* (21), 9645–9667. DOI: 10.1021/acs.jmedchem.6b00551. Published Online: Jul. 29, 2016.
- (62) Sood, A.; Shaw, B. R.; Spielvogel, B. F. Boron-containing nucleic acids. 2. Synthesis of oligodeoxynucleoside boranophosphates. *J. Am. Chem. Soc.* **1990**, *112* (24), 9000–9001. DOI: 10.1021/ja00180a066.
- (63) Sergueev, D. S.; Shaw, B. R. H<sup>-</sup>-Phosphonate Approach for Solid-Phase Synthesis of Oligodeoxyribonucleoside Boranophosphates and Their Characterization. *J. Am. Chem. Soc.* **1998**, *120* (37), 9417–9427. DOI: 10.1021/ja9814927.
- (64) Eckstein, F. Phosphorothioates, essential components of therapeutic oligonucleotides. *Nucleic acid therapeutics* **2014**, *24* (6), 374–387. DOI: 10.1089/nat.2014.0506.
- (65) Takahashi, Y.; Sato, K.; Wada, T. Solid-Phase Synthesis of Boranophosphate/Phosphorothioate/Phosphate Chimeric Oligonucleotides and Their Potential as Antisense Oligonucleotides. *The Journal of organic chemistry* **2022**, *87* (6), 3895–3909. DOI: 10.1021/acs.joc.1c01812. Published Online: Dec. 15, 2021.
- (66) Summers, J. S.; Shaw, B. R. Boranophosphates as mimics of natural phosphodiester in DNA. *Current medicinal chemistry* **2001**, *8* (10), 1147–1155. DOI: 10.2174/0929867013372409.
- (67) Kibler-Herzog, L.; Zon, G.; Uznanski, B.; Whittier, G.; Wilson, W. D. Duplex stabilities of phosphorothioate, methylphosphonate, and RNA analogs of two DNA 14-mers. *Nucleic acids research* **1991**, *19* (11), 2979–2986. DOI: 10.1093/nar/19.11.2979.
- (68) Somasunderam, A.; Thivyanathan, V.; Tanaka, T.; Li, X.; Neerathilingam, M.; Lokesh, G. L. R.; Mann, A.; Peng, Y.; Ferrari, M.; Klostergaard, J.; Gorenstein, D. G. Combinatorial selection of DNA thioaptamers targeted to the HA binding domain of human CD44. *Biochemistry* **2010**, *49* (42), 9106–9112. DOI: 10.1021/bi1009503.
- (69) Porter, K. W.; Briley, J. D.; Shaw, B. R. Direct PCR sequencing with boronated nucleotides. *Nucleic acids research* **1997**, *25* (8), 1611–1617. DOI: 10.1093/nar/25.8.1611.
- (70) Sergueev, D.; Hasan, A.; Ramaswamy, M.; Shaw, B. R. Boranophosphate Oligonucleotides: New Synthetic Approaches. *Nucleosides and Nucleotides* **1997**, *16* (7-9), 1533–1538. DOI: 10.1080/07328319708006223.
- (71) Ghadessy, F. J.; Ramsay, N.; Boudsocq, F.; Loakes, D.; Brown, A.; Iwai, S.; Vaisman, A.; Woodgate, R.; Holliger, P. Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. *Nature biotechnology* **2004**, *22* (6), 755–759. DOI: 10.1038/nbt974. Published Online: May. 23, 2004.
- (72) Gryaznov, S.; Chen, J.-K. Oligodeoxyribonucleotide N3'.fwdarw.P5' Phosphoramidates: synthesis and Hybridization Properties. *J. Am. Chem. Soc.* **1994**, *116* (7), 3143–3144. DOI: 10.1021/ja00086a062.

- (73) Gryaznov, S. M.; Lloyd, D. H.; Chen, J. K.; Schultz, R. G.; DeDionisio, L. A.; Ratmeyer, L.; Wilson, W. D. Oligonucleotide N3'--P5' phosphoramidates. *Proceedings of the National Academy of Sciences of the United States of America* **1995**, *92* (13), 5798–5802. DOI: 10.1073/pnas.92.13.5798.
- (74) Chen, J. K.; Schultz, R. G.; Lloyd, D. H.; Gryaznov, S. M. Synthesis of oligodeoxyribonucleotide N3'--P5' phosphoramidates. *Nucleic acids research* **1995**, *23* (14), 2661–2668. DOI: 10.1093/nar/23.14.2661.
- (75) Lelyveld, V. S.; Zhang, W.; Szostak, J. W. Synthesis of phosphoramidate-linked DNA by a modified DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **2020**, *117* (13), 7276–7283. DOI: 10.1073/pnas.1922400117. Published Online: Mar. 18, 2020.
- (76) Lelyveld, V. S.; O'Flaherty, D. K.; Zhou, L.; Izgu, E. C.; Szostak, J. W. DNA polymerase activity on synthetic N3'→P5' phosphoramidate DNA templates. *Nucleic acids research* **2019**, *47* (17), 8941–8949. DOI: 10.1093/nar/gkz707.
- (77) Agarwal, K. L.; Riftina, F. Synthesis and enzymatic properties of deoxyribooligonucleotides containing methyl and phenylphosphonate linkages. *Nucleic acids research* **1979**, *6* (9), 3009–3024. DOI: 10.1093/nar/6.9.3009.
- (78) Arangundy-Franklin, S.; Taylor, A. I.; Porebski, B. T.; Genna, V.; Peak-Chew, S.; Vaisman, A.; Woodgate, R.; Orozco, M.; Holliger, P. Encoded synthesis and evolution of alkyl-phosphonate nucleic acids: a synthetic genetic polymer with an uncharged backbone chemistry. *Nature chemistry* **2019**, *11* (6), 533–542. DOI: 10.1038/s41557-019-0255-4.
- (79) Murakami, A.; Blake, K. R.; Miller, P. S. Characterization of sequence-specific oligodeoxyribonucleoside methylphosphonates and their interaction with rabbit globin mRNA. *Biochemistry* **1985**, *24* (15), 4041–4046. DOI: 10.1021/bi00336a036.
- (80) Micklefield, J. *Backbone modification of nucleic acids: synthesis, structure and therapeutic applications*, Vol. 8; Bentham Science Publishers, 2001. DOI: 10.2174/0929867013372391.
- (81) Dineva, M. A.; Chakurov, S.; Bratovanova, E. K.; Devedjiev, I.; Petkov, D. D. Complete template-directed enzymatic synthesis of a potential antisense DNA containing 42 methylphosphonodiester bonds. *Bioorganic & medicinal chemistry* **1993**, *1* (6), 411–414. DOI: 10.1016/s0968-0896(00)82151-3.
- (82) Selvam, C.; Thomas, S.; Abbott, J.; Kennedy, S. D.; Rozners, E. Amides as excellent mimics of phosphate linkages in RNA. *Angewandte Chemie (International ed. in English)* **2011**, *50* (9), 2068–2070. DOI: 10.1002/anie.201007012. Published Online: Jan. 21, 2011.
- (83) Kotikam, V.; Rozners, E. Amide-Modified RNA: Using Protein Backbone to Modulate Function of Short Interfering RNAs. *Accounts of chemical research* **2020**, *53* (9), 1782–1790. DOI: 10.1021/acs.accounts.0c00249. Published Online: Jul. 13, 2020.
- (84) Baraniak, D.; Boryski, J. Triazole-Modified Nucleic Acids for the Application in Bioorganic and Medicinal Chemistry. *Biomedicines* **2021**, *9* (6). DOI: 10.3390/biomedicines9060628. Published Online: May. 31, 2021.
- (85) Varizhuk, A. M.; Kaluzhny, D. N.; Novikov, R. A.; Chizhov, A. O.; Smirnov, I. P.; Chuvilin, A. N.; Tatarinova, O. N.; Fisunov, G. Y.; Pozmogova, G. E.; Florentiev, V. L. Synthesis of triazole-linked oligonucleotides with high affinity to DNA complements and an analysis of their compatibility with biosystems. *The Journal of organic chemistry* **2013**, *78* (12), 5964–5969. DOI: 10.1021/jo400651k. Published Online: Jun. 11, 2013.
- (86) Skakuj, K.; Bujold, K. E.; Mirkin, C. A. Mercury-Free Automated Synthesis of Guanidinium Backbone Oligonucleotides. *J. Am. Chem. Soc.* **2019**, *141* (51), 20171–20176. DOI: 10.1021/jacs.9b09937. Published Online: Dec. 16, 2019.
- (87) Epple, S.; Thorpe, C.; Baker, Y. R.; El-Sagheer, A. H.; Brown, T. Consecutive 5'- and 3'-amide linkages stabilise antisense oligonucleotides and elicit an efficient RNase H response. *Chemical communications (Cambridge, England)* **2020**, *56* (41), 5496–5499. DOI: 10.1039/d0cc00444h.

- (88) Meng, M.; Schmidtgal, B.; Ducho, C. Enhanced Stability of DNA Oligonucleotides with Partially Zwitterionic Backbone Structures in Biological Media. *Molecules (Basel, Switzerland)* **2018**, *23* (11). DOI: 10.3390/molecules23112941. Published Online: Nov. 10, 2018.
- (89) El-Khoury, R.; Damha, M. J. 2'-Fluoro-arabinonucleic Acid (FANA): A Versatile Tool for Probing Biomolecular Interactions. *Accounts of chemical research* **2021**, *54* (9), 2287–2297. DOI: 10.1021/acs.accounts.1c00125. Published Online: Apr. 16, 2021.
- (90) Freund, N.; Taylor, A. I.; Arangundy-Franklin, S.; Subramanian, N.; Peak-Chew, S.-Y.; Whitaker, A. M.; Freudenthal, B. D.; Abramov, M.; Herdewijn, P.; Holliger, P. A two-residue nascent-strand steric gate controls synthesis of 2'-O-methyl- and 2'-O-(2-methoxyethyl)-RNA. *Nature chemistry* **2023**, *15* (1), 91–100. DOI: 10.1038/s41557-022-01050-8. Published Online: Oct. 13, 2022.
- (91) Herdewijn, P.; Marlière, P. Toward safe genetically modified organisms through the chemical diversification of nucleic acids. *Chemistry & biodiversity* **2009**, *6* (6), 791–808. DOI: 10.1002/cbdv.200900083.
- (92) Chaput, J. C.; Herdewijn, P. What Is XNA? *Angewandte Chemie (International ed. in English)* **2019**, *58* (34), 11570–11572. DOI: 10.1002/anie.201905999. Published Online: Jul. 19, 2019.
- (93) Soler-Bistué, A.; Zorreguieta, A.; Tolmasky, M. E. Bridged Nucleic Acids Reloaded. *Molecules (Basel, Switzerland)* **2019**, *24* (12), 2297. DOI: 10.3390/molecules24122297. Published Online: Jun. 21, 2019.
- (94) Obika, S.; Nanbu, D.; Hari, Y.; Morio, K.; In, Y.; Ishida, T.; Imanishi, T. Synthesis of 2'-O,4'-C-methyleneuridine and -cytidine. Novel bicyclic nucleosides having a fixed C3, -endo sugar puckering. *Tetrahedron letters* **1997**, *38* (50), 8735–8738. DOI: 10.1016/S0040-4039(97)10322-7.
- (95) Koshkin, A. A.; Rajwanshi, V. K.; Wengel, J. Novel convenient syntheses of LNA [2.2.1]bicyclo nucleosides. *Tetrahedron letters* **1998**, *39* (24), 4381–4384. DOI: 10.1016/S0040-4039(98)00706-0.
- (96) Koshkin, A. A.; Singh, S. K.; Nielsen, P.; Rajwanshi, V. K.; Kumar, R.; Meldgaard, M.; Olsen, C. E.; Wengel, J. LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. *Tetrahedron* **1998**, *54* (14), 3607–3630. DOI: 10.1016/S0040-4020(98)00094-5.
- (97) Hughesman, C. B.; Turner, R. F. B.; Haynes, C. A. Role of the heat capacity change in understanding and modeling melting thermodynamics of complementary duplexes containing standard and nucleobase-modified LNA. *Biochemistry* **2011**, *50* (23), 5354–5368. DOI: 10.1021/bi200223s. Published Online: May. 18, 2011.
- (98) Wahlestedt, C.; Salmi, P.; Good, L.; Kela, J.; Johnsson, T.; Hökfelt, T.; Broberger, C.; Porreca, F.; Lai, J.; Ren, K.; Ossipov, M.; Koshkin, A.; Jakobsen, N.; Skouv, J.; Oerum, H.; Jacobsen, M. H.; Wengel, J. Potent and nontoxic antisense oligonucleotides containing locked nucleic acids. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97* (10), 5633–5638. DOI: 10.1073/pnas.97.10.5633.
- (99) Kurreck, J.; Wyszko, E.; Gillen, C.; Erdmann, V. A. Design of antisense oligonucleotides stabilized by locked nucleic acids. *Nucleic acids research* **2002**, *30* (9), 1911–1918. DOI: 10.1093/nar/30.9.1911.
- (100) Veedu, R. N.; Vester, B.; Wengel, J. Enzymatic incorporation of LNA nucleotides into DNA strands. *ChemBiochem : a European journal of chemical biology* **2007**, *8* (5), 490–492. DOI: 10.1002/cbic.200600501.
- (101) Veedu, R. N.; Vester, B.; Wengel, J. In vitro incorporation of LNA nucleotides. *Nucleosides, nucleotides & nucleic acids* **2007**, *26* (8-9), 1207–1210. DOI: 10.1080/15257770701527844.
- (102) Pinheiro, V. B.; Taylor, A. I.; Cozens, C.; Abramov, M.; Renders, M.; Zhang, S.; Chaput, J. C.; Wengel, J.; Peak-Chew, S.-Y.; McLaughlin, S. H.; Herdewijn, P.; Holliger, P. Synthetic

genetic polymers capable of heredity and evolution. *Science (New York, N.Y.)* **2012**, 336 (6079), 341–344. DOI: 10.1126/science.1217622.

(103) Hoshino, H.; Kasahara, Y.; Kuwahara, M.; Obika, S. DNA Polymerase Variants with High Processivity and Accuracy for Encoding and Decoding Locked Nucleic Acid Sequences. *J. Am. Chem. Soc.* **2020**, 142 (51), 21530–21537. DOI: 10.1021/jacs.0c10902. Published Online: Dec. 11, 2020.

(104) Morihira, K.; Kasahara, Y.; Obika, S. Biological applications of xeno nucleic acids. *Molecular bioSystems* **2017**, 13 (2), 235–245. DOI: 10.1039/c6mb00538a.

(105) Rajwanshi, V. K.; Håkansson, A. E.; Sørensen, M. D.; Pitsch, S.; Singh, S. K.; Kumar, R.; Nielsen, P.; Wengel, J. The Eight Stereoisomers of LNA (Locked Nucleic Acid): A Remarkable Family of Strong RNA Binding Molecules. *Angew. Chem. Int. Ed.* **2000**, 39 (9), 1656–1659. DOI: 10.1002/(SICI)1521-3773(20000502)39:9<1656:AID-ANIE1656>3.0.CO;2-Q.

(106) Zhang, L.; Peritz, A.; Meggers, E. A simple glycol nucleic acid. *J. Am. Chem. Soc.* **2005**, 127 (12), 4174–4175. DOI: 10.1021/ja042564z.

(107) Schlegel, M. K.; Peritz, A. E.; Kittigowittana, K.; Zhang, L.; Meggers, E. Duplex formation of the simplified nucleic acid GNA. *Chembiochem : a European journal of chemical biology* **2007**, 8 (8), 927–932. DOI: 10.1002/cbic.200600435.

(108) Schlegel, M. K.; Foster, D. J.; Kel'in, A. V.; Zlatev, I.; Bisbe, A.; Jayaraman, M.; Lackey, J. G.; Rajeev, K. G.; Charissé, K.; Harp, J.; Pallan, P. S.; Maier, M. A.; Egli, M.; Manoharan, M. Chirality Dependent Potency Enhancement and Structural Impact of Glycol Nucleic Acid Modification on siRNA. *J. Am. Chem. Soc.* **2017**, 139 (25), 8537–8546. DOI: 10.1021/jacs.7b02694. Published Online: Jun. 19, 2017.

(109) Horhota, A. T.; Szostak, J. W.; McLaughlin, L. W. Glycerol nucleoside triphosphates: synthesis and polymerase substrate activities. *Organic letters* **2006**, 8 (23), 5345–5347. DOI: 10.1021/ol062232u.

(110) Chen, J. J.; Tsai, C.-H.; Cai, X.; Horhota, A. T.; McLaughlin, L. W.; Szostak, J. W. Enzymatic primer-extension with glycerol-nucleoside triphosphates on DNA templates. *PLoS one* **2009**, 4 (3), e4949. DOI: 10.1371/journal.pone.0004949. Published Online: Mar. 23, 2009.

(111) Tsai, C.-H.; Chen, J.; Szostak, J. W. Enzymatic synthesis of DNA on glycerol nucleic acid templates without stable duplex formation between product and template. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, 104 (37), 14598–14603. DOI: 10.1073/pnas.0704211104. Published Online: Sep. 4, 2007.

(112) Chemama, M.; Switzer, C. Flexible Nucleic Acids (FNAs) as Informational Molecules: Enzymatic Polymerization of fNTPs on DNA Templates and Nonenzymatic Oligomerization of RNA on FNA Templates. In *Non-Natural Nucleic Acids: Methods and Protocols*; Shank, N., Ed.; Springer eBooks Springer Protocols, Vol. 1973; Humana Press, 2019; pp 213–236. DOI: 10.1007/978-1-4939-9216-4\_14.

(113) Heuberger, B. D.; Switzer, C. A Pre-RNA Candidate Revisited: Both Enantiomers of Flexible Nucleoside Triphosphates are DNA Polymerase Substrates. *J. Am. Chem. Soc.* **2008**, 130 (2), 412–413. DOI: 10.1021/ja0770680.

(114) Schneider, K. C.; Benner, S. A. Oligonucleotides containing flexible nucleoside analogs. *J. Am. Chem. Soc.* **1990**, 112 (1), 453–455. DOI: 10.1021/ja00157a073.

(115) Merle, L.; Spach, G.; Merle, Y.; Sági, J.; Szemző, A. Some biochemical properties of an acyclic oligonucleotide analogue. A plausible ancestor of the DNA? *Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life* **1993**, 23 (2), 91–103. DOI: 10.1007/BF01581874.

(116) Chaput, J. C.; Switzer, C. Nonenzymatic oligomerization on templates containing phosphodiester-linked acyclic glycerol nucleic acid analogues. *Journal of molecular evolution* **2000**, 51 (5), 464–470. DOI: 10.1007/s002390010109.

- (117) Merle, Y.; Bonneil, E.; Merle, L.; Sági, J.; Szemző, A. Acyclic oligonucleotide analogues. *International Journal of Biological Macromolecules* **1995**, *17* (5), 239–246. DOI: 10.1016/0141-8130(95)98150-w.
- (118) Aerschot Van, A.; Verheggen, I.; Hendrix, C.; Herdewijn, P. 1,5-Anhydrohexitol Nucleic Acids, a New Promising Antisense Construct. *Angew. Chem. Int. Ed. Engl.* **1995**, *34* (12), 1338–1339. DOI: 10.1002/anie.199513381.
- (119) Groaz, E.; Herdewijn, P. Hexitol Nucleic Acid (HNA): From Chemical Design to Functional Genetic Polymer. In *Handbook of Chemical Biology of Nucleic Acids*; Sugimoto, N., Ed.; Springer Nature Singapore, 2022; pp 1–34. DOI: 10.1007/978-981-16-1313-5\_15-1.
- (120) Hendrix, C.; Rosemeyer, H.; Bouvere, B. de; van Aerschot, A.; Seela, F.; Herdewijn, P. 1',5'-Anhydrohexitol Oligonucleotides: Hybridisation and Strand Displacement with Oligoribonucleotides, Interaction with RNase H and HIV Reverse Transcriptase. *Chemistry A European J* **1997**, *3* (9), 1513–1520. DOI: 10.1002/chem.19970030920.
- (121) Kempeneers, V.; Vastmans, K.; Rozenski, J.; Herdewijn, P. Recognition of threosyl nucleotides by DNA and RNA polymerases. *Nucleic acids research* **2003**, *31* (21), 6221–6226. DOI: 10.1093/nar/gkg833.
- (122) Vastmans, K.; Froeyen, M.; Kerremans, L.; Pochet, S.; Herdewijn, P. Reverse transcriptase incorporation of 1,5-anhydrohexitol nucleotides. *Nucleic acids research* **2001**, *29* (15), 3154–3163. DOI: 10.1093/nar/29.15.3154.
- (123) Wang, J.; Verbeure, B.; Luyten, I.; Lescrinier, E.; Froeyen, M.; Hendrix, C.; Rosemeyer, H.; Seela, F.; van Aerschot, A.; Herdewijn, P. Cyclohexene Nucleic Acids (CeNA): Serum Stable Oligonucleotides that Activate RNase H and Increase Duplex Stability with Complementary RNA. *Journal of the American Chemical Society* **2000**, *122* (36), 8595–8602. DOI: 10.1021/ja000018.
- (124) Gu, P.; Schepers, G.; Rozenski, J.; van Aerschot, A.; Herdewijn, P. Base pairing properties of D- and L-cyclohexene nucleic acids (CeNA). *Oligonucleotides* **2003**, *13* (6), 479–489. DOI: 10.1089/154545703322860799.
- (125) Kempeneers, V.; Renders, M.; Froeyen, M.; Herdewijn, P. Investigation of the DNA-dependent cyclohexenyl nucleic acid polymerization and the cyclohexenyl nucleic acid-dependent DNA polymerization. *Nucleic acids research* **2005**, *33* (12), 3828–3836. DOI: 10.1093/nar/gki695. Published Online: Jul. 12, 2005.
- (126) Schöning, K.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A. Chemical etiology of nucleic acid structure: the alpha-threofuranosyl-(3'--2') oligonucleotide system. *Science (New York, N.Y.)* **2000**, *290* (5495), 1347–1351. DOI: 10.1126/science.290.5495.1347.
- (127) Schöning, K.-U.; Scholz, P.; Wu, X.; Guntha, S.; Delgado, G.; Krishnamurthy, R.; Eschenmoser, A. The  $\alpha$ -L-Threofuranosyl-(3'→2')-oligonucleotide System ('TNA'): Synthesis and Pairing Properties. *Helvetica Chimica Acta* **2002**, *85* (12), 4111–4153. DOI: 10.1002/hlca.200290000.
- (128) Orgel, L. Origin of life. A simpler nucleic acid. *Science (New York, N.Y.)* **2000**, *290* (5495), 1306–1307. DOI: 10.1126/science.290.5495.1306.
- (129) Ebert, M.-O.; Mang, C.; Krishnamurthy, R.; Eschenmoser, A.; Jaun, B. The structure of a TNA-TNA complex in solution: NMR study of the octamer duplex derived from alpha-(L)-threofuranosyl-(3'-2')-CGAATTTCG. *Journal of the American Chemical Society* **2008**, *130* (45), 15105–15115. DOI: 10.1021/ja8041959. Published Online: Oct. 18, 2008.
- (130) Pallan, P. S.; Wilds, C. J.; Wawrzak, Z.; Krishnamurthy, R.; Eschenmoser, A.; Egli, M. Why Does TNA Cross-Pair More Strongly with RNA Than with DNA? An Answer From X-ray Analysis. *Angewandte Chemie* **2003**, *115* (47), 6073–6075. DOI: 10.1002/ange.200352553.
- (131) Anosova, I.; Kowal, E. A.; Sisco, N. J.; Sau, S.; Liao, J.-Y.; Bala, S.; Rozners, E.; Egli, M.; Chaput, J. C.; van Horn, W. D. Structural Insights into Conformation Differences between DNA/TNA and RNA/TNA Chimeric Duplexes. *ChemBiochem: a European journal of chemical*

*biology* **2016**, *17* (18), 1705–1708. DOI: 10.1002/cbic.201600349. Published Online: Jul. 29, 2016.

(132) Pitsch, S.; Eschenmoser, A.; Gedin, B.; Hui, S.; Arrhenius, G. Mineral induced formation of sugar phosphates. *Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life* **1995**, *25* (4), 297–334. DOI: 10.1007/BF01581773.

(133) Chaput, J. C.; Ichida, J. K.; Szostak, J. W. DNA polymerase-mediated DNA synthesis on a TNA template. *J. Am. Chem. Soc.* **2003**, *125* (4), 856–857. DOI: 10.1021/ja028589k.

(134) Chaput, J. C.; Szostak, J. W. TNA synthesis by DNA polymerases. *Journal of the American Chemical Society* **2003**, *125* (31), 9274–9275. DOI: 10.1021/ja035917n.

(135) Yu, H.; Zhang, S.; Chaput, J. C. Darwinian evolution of an alternative genetic system provides support for TNA as an RNA progenitor. *Nature chemistry* **2012**, *4* (3), 183–187. DOI: 10.1038/nchem.1241. Published Online: Jan. 10, 2012.

(136) Liao, J.-Y.; Anosova, I.; Bala, S.; van Horn, W. D.; Chaput, J. C. A parallel stranded G-quadruplex composed of threose nucleic acid (TNA). *Biopolymers* **2017**, *107* (3). DOI: 10.1002/bip.22999.

(137) Culbertson, M. C.; Temburnikar, K. W.; Sau, S. P.; Liao, J.-Y.; Bala, S.; Chaput, J. C. Evaluating TNA stability under simulated physiological conditions. *Biorganic & medicinal chemistry letters* **2016**, *26* (10), 2418–2421. DOI: 10.1016/j.bmcl.2016.03.118. Published Online: Apr. 1, 2016.

(138) Yu, H.; Zhang, S.; Dunn, M. R.; Chaput, J. C. An efficient and faithful in vitro replication system for threose nucleic acid. *Journal of the American Chemical Society* **2013**, *135* (9), 3583–3591. DOI: 10.1021/ja3118703. Published Online: Feb. 25, 2013.

(139) Rangel, A. E.; Chen, Z.; Ayele, T. M.; Heemstra, J. M. In vitro selection of an XNA aptamer capable of small-molecule recognition. *Nucleic acids research* **2018**, *46* (16), 8057–8068. DOI: 10.1093/nar/gky667.

(140) Dunn, M. R.; Otto, C.; Fenton, K. E.; Chaput, J. C. Improving Polymerase Activity with Unnatural Substrates by Sampling Mutations in Homologous Protein Architectures. *ACS chemical biology* **2016**, *11* (5), 1210–1219. DOI: 10.1021/acscchembio.5b00949. Published Online: Feb. 17, 2016.

(141) Ichida, J. K.; Horhota, A.; Zou, K.; McLaughlin, L. W.; Szostak, J. W. High fidelity TNA synthesis by Terminator polymerase. *Nucleic acids research* **2005**, *33* (16), 5219–5225. DOI: 10.1093/nar/gki840. Published Online: Sep. 12, 2005.

(142) Dunn, M. R.; Chaput, J. C. Reverse Transcription of Threose Nucleic Acid by a Naturally Occurring DNA Polymerase. *ChemBiochem : a European journal of chemical biology* **2016**, *17* (19), 1804–1808. DOI: 10.1002/cbic.201600338. Published Online: Aug. 4, 2016.

(143) Horhota, A.; Zou, K.; Ichida, J. K.; Yu, B.; McLaughlin, L. W.; Szostak, J. W.; Chaput, J. C. Kinetic analysis of an efficient DNA-dependent TNA polymerase. *Journal of the American Chemical Society* **2005**, *127* (20), 7427–7434. DOI: 10.1021/ja0428255.

(144) Ichida, J. K.; Zou, K.; Horhota, A.; Yu, B.; McLaughlin, L. W.; Szostak, J. W. An in vitro selection system for TNA. *J. Am. Chem. Soc.* **2005**, *127* (9), 2802–2803. DOI: 10.1021/ja045364w.

(145) Zou, K.; Horhota, A.; Yu, B.; Szostak, J. W.; McLaughlin, L. W. Synthesis of alpha-L-threofuranosyl nucleoside triphosphates (tNTPs). *Organic letters* **2005**, *7* (8), 1485–1487. DOI: 10.1021/ol050081+.

(146) Zhang, S.; Chaput, J. C. Synthesis of threose nucleic acid (TNA) phosphoramidite monomers and oligonucleotide polymers. *Current protocols in nucleic acid chemistry* **2012**, *Chapter 4*, Unit4.51. DOI: 10.1002/0471142700.nc0451s50.

(147) Dunn, M. R.; Larsen, A. C.; Zahurancik, W. J.; Fahmi, N. E.; Meyers, M.; Suo, Z.; Chaput, J. C. DNA polymerase-mediated synthesis of unbiased threose nucleic acid (TNA) polymers requires 7-deazaguanine to suppress G:G mispairing during TNA transcription. *Journal of the*

*American Chemical Society* **2015**, *137* (12), 4014–4017. DOI: 10.1021/ja511481n. Published Online: Mar. 20, 2015.

(148) Mei, H.; Liao, J.-Y.; Jimenez, R. M.; Wang, Y.; Bala, S.; McCloskey, C.; Switzer, C.; Chaput, J. C. Synthesis and Evolution of a Threose Nucleic Acid Aptamer Bearing 7-Deaza-7-Substituted Guanosine Residues. *Journal of the American Chemical Society* **2018**, *140* (17), 5706–5713. DOI: 10.1021/jacs.7b13031. Published Online: Apr. 23, 2018.

(149) Hoogsteen, K. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Cryst* **1963**, *16* (9), 907–916. DOI: 10.1107/S0365110X63002437.

(150) Skelly, J. V.; Edwards, K. J.; Jenkins, T. C.; Neidle, S. Crystal structure of an oligonucleotide duplex containing G.G base pairs: influence of mispairing on DNA backbone conformation. *Proceedings of the National Academy of Sciences of the United States of America* **1993**, *90* (3), 804–808. DOI: 10.1073/pnas.90.3.804.

(151) Chim, N.; Shi, C.; Sau, S. P.; Nikoomezar, A.; Chaput, J. C. Structural basis for TNA synthesis by an engineered TNA polymerase. *Nature communications* **2017**, *8* (1), 1810. DOI: 10.1038/s41467-017-02014-0. Published Online: Nov. 27, 2017.

(152) Nikoomezar, A.; Vallejo, D.; Yik, E. J.; Chaput, J. C. Programmed Allelic Mutagenesis of a DNA Polymerase with Single Amino Acid Resolution. *ACS synthetic biology* **2020**, *9* (7), 1873–1881. DOI: 10.1021/acssynbio.0c00236. Published Online: Jun. 30, 2020.

(153) Hajjar, M.; Chim, N.; Liu, C.; Herdewijn, P.; Chaput, J. C. Crystallographic analysis of engineered polymerases synthesizing phosphonomethylthreosyl nucleic acid. *Nucleic acids research* **2022**, *50* (17), 9663–9674. DOI: 10.1093/nar/gkac792.

(154) Medina, E.; Yik, E. J.; Herdewijn, P.; Chaput, J. C. Functional Comparison of Laboratory-Evolved XNA Polymerases for Synthetic Biology. *ACS synthetic biology* **2021**, *10* (6), 1429–1437. DOI: 10.1021/acssynbio.1c00048. Published Online: May. 24, 2021.

(155) Pfeiffer, F.; Rosenthal, M.; Siegl, J.; Ewers, J.; Mayer, G. Customised nucleic acid libraries for enhanced aptamer selection and performance. *Current opinion in biotechnology* **2017**, *48*, 111–118. DOI: 10.1016/j.copbio.2017.03.026. Published Online: Apr. 22, 2017.

(156) Switzer, C.; Moroney, S. E.; Benner, S. A. Enzymatic incorporation of a new base pair into DNA and RNA. *J. Am. Chem. Soc.* **1989**, *111* (21), 8322–8323. DOI: 10.1021/ja00203a067.

(157) Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature* **1990**, *343* (6253), 33–37. DOI: 10.1038/343033a0.

(158) Hoshika, S.; Leal, N. A.; Kim, M.-J.; Kim, M.-S.; Karalkar, N. B.; Kim, H.-J.; Bates, A. M.; Watkins, N. E.; SantaLucia, H. A.; Meyer, A. J.; DasGupta, S.; Piccirilli, J. A.; Ellington, A. D.; SantaLucia, J.; Georgiadis, M. M.; Benner, S. A. Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science (New York, N.Y.)* **2019**, *363* (6429), 884–887. DOI: 10.1126/science.aat0971.

(159) Morales, J. C.; Kool, E. T. Efficient replication between non-hydrogen-bonded nucleoside shape analogs. *Nature structural biology* **1998**, *5* (11), 950–954. DOI: 10.1038/2925.

(160) Seo, Y. J.; Hwang, G. T.; Ordoukhanian, P.; Romesberg, F. E. Optimization of an unnatural base pair toward natural-like replication. *J. Am. Chem. Soc.* **2009**, *131* (9), 3246–3252. DOI: 10.1021/ja807853m.

(161) Li, L.; Degardin, M.; Lavergne, T.; Malyshev, D. A.; Dhimi, K.; Ordoukhanian, P.; Romesberg, F. E. Natural-like replication of an unnatural base pair for the expansion of the genetic alphabet and biotechnology applications. *Journal of the American Chemical Society* **2014**, *136* (3), 826–829. DOI: 10.1021/ja408814g. Published Online: Oct. 23, 2013.

(162) Kimoto, M.; Yamashige, R.; Yokoyama, S.; Hirao, I. PCR amplification and transcription for site-specific labeling of large RNA molecules by a two-unnatural-base-pair system. *Journal*



*of nucleic acids* **2012**, 2012, 230943. DOI: 10.1155/2012/230943. Published Online: Jun. 26, 2012.

(163) Hottin, A.; Marx, A. Structural Insights into the Processing of Nucleobase-Modified Nucleotides by DNA Polymerases. *Accounts of chemical research* **2016**, 49 (3), 418–427. DOI: 10.1021/acs.accounts.5b00544. Published Online: Mar. 5, 2016.

(164) Kuwahara, M.; Suto, Y.; Minezaki, S.; Kitagata, R.; Nagashima, J.; Sawai, H. Substrate property and incorporation accuracy of various dATP analogs during enzymatic polymerization using thermostable DNA polymerases. *Nucleic acids symposium series (2004)* **2006** (50), 31–32. DOI: 10.1093/nass/nrl016.

(165) Hasegawa, T.; Shoji, A.; Kuwahara, M.; Ozaki, H.; Sawai, H. Synthesis and property of DNA labeled with fluorescent acridone. *Nucleic acids symposium series (2004)* **2006** (50), 145–146. DOI: 10.1093/nass/nrl072.

(166) Jäger, S.; Rasched, G.; Kornreich-Leshem, H.; Engeser, M.; Thum, O.; Famulok, M. A versatile toolbox for variable DNA functionalization at high density. *J. Am. Chem. Soc.* **2005**, 127 (43), 15071–15082. DOI: 10.1021/ja051725b.

(167) Jäger, S.; Famulok, M. Generation and enzymatic amplification of high-density functionalized DNA double strands. *Angew. Chem. Int. Ed.* **2004**, 43 (25), 3337–3340. DOI: 10.1002/anie.200453926.

(168) Brudno, Y.; Liu, D. R. Recent progress toward the templated synthesis and directed evolution of sequence-defined synthetic polymers. *Chemistry & biology* **2009**, 16 (3), 265–276. DOI: 10.1016/j.chembiol.2009.02.004.

(169) Kashani-Sabet, M. Ribozyme therapeutics. *The journal of investigative dermatology. Symposium proceedings* **2002**, 7 (1), 76–78. DOI: 10.1046/j.1523-1747.2002.19642.x.

(170) Fantoni, N. Z.; El-Sagheer, A. H.; Brown, T. A Hitchhiker's Guide to Click-Chemistry with Nucleic Acids. *Chemical reviews* **2021**, 121 (12), 7122–7154. DOI: 10.1021/acs.chemrev.0c00928. Published Online: Jan. 14, 2021.

(171) Samanta, B.; Seikowski, J.; Höbartner, C. Fluorogenic Labeling of 5-Formylpyrimidine Nucleotides in DNA and RNA. *Angew. Chem. Int. Ed.* **2016**, 55 (5), 1912–1916. DOI: 10.1002/anie.201508893. Published Online: Dec. 17, 2015.

(172) Walunj, M. B.; Tanpure, A. A.; Srivatsan, S. G. Post-transcriptional labeling by using Suzuki-Miyaura cross-coupling generates functional RNA probes. *Nucleic acids research* **2018**, 46 (11), e65. DOI: 10.1093/nar/gky185.

(173) Kore, A. R.; Senthilvelan, A.; Shanmugasundaram, M. Highly chemoselective palladium-catalyzed Sonogashira coupling of 5-iodouridine-5'-triphosphates with propargylamine: a new efficient method for the synthesis of 5-aminopropargyl-uridine-5'-triphosphates. *Tetrahedron letters* **2012**, 53 (24), 3070–3072. DOI: 10.1016/j.tetlet.2012.04.023.

(174) Busskamp, H.; Batroff, E.; Niederwieser, A.; Abdel-Rahman, O. S.; Winter, R. F.; Wittmann, V.; Marx, A. Efficient labelling of enzymatically synthesized vinyl-modified DNA by an inverse-electron-demand Diels-Alder reaction. *Chem. Commun.* **2014**, 50 (74), 10827–10829. DOI: 10.1039/c4cc04332d.

(175) Rao, H.; Tanpure, A. A.; Sawant, A. A.; Srivatsan, S. G. Enzymatic incorporation of an azide-modified UTP analog into oligoribonucleotides for post-transcriptional chemical functionalization. *Nature protocols* **2012**, 7 (6), 1097–1112. DOI: 10.1038/nprot.2012.046. Published Online: May. 10, 2012.

(176) Schoch, J.; Jäschke, A. Synthesis and enzymatic incorporation of norbornene-modified nucleoside triphosphates for Diels–Alder bioconjugation. *RSC Adv.* **2013**, 3 (13), 4181. DOI: 10.1039/c3ra40373d.

(177) Ren, X.; Gerowska, M.; El-Sagheer, A. H.; Brown, T. Enzymatic incorporation and fluorescent labelling of cyclooctyne-modified deoxyuridine triphosphates in DNA. *Bioorganic & medicinal chemistry* **2014**, 22 (16), 4384–4390. DOI: 10.1016/j.bmc.2014.05.050. Published Online: Jun. 5, 2014.

- (178) Alexander Rich. *Horizons in Biochemistry: On the Problems of Evolution and Biochemical Information Transfer*; Academic Press, 1962.
- (179) Switzer, C. Y.; Moroney, S. E.; Benner, S. A. Enzymatic recognition of the base pair between isocytidine and isoguanosine. *Biochemistry* **1993**, *32* (39), 10489–10496. DOI: 10.1021/bi00090a027.
- (180) Moran, S.; Ren, R. X.-F.; Rumney, S.; Kool, E. T. Difluorotoluene, a Nonpolar Isostere for Thymine, Codes Specifically and Efficiently for Adenine in DNA Replication. *J. Am. Chem. Soc.* **1997**, *119* (8), 2056–2057. DOI: 10.1021/ja963718g.
- (181) McMinn, D. L.; Ogawa, A. K.; Wu, Y.; Liu, J.; Schultz, P. G.; Romesberg, F. E. Efforts toward Expansion of the Genetic Alphabet: DNA Polymerase Recognition of a Highly Stable, Self-Pairing Hydrophobic Base. *J. Am. Chem. Soc.* **1999**, *121* (49), 11585–11586. DOI: 10.1021/ja9925150.
- (182) Henry, A. A.; Romesberg, F. E. Beyond A, C, G and T: augmenting nature's alphabet. *Current opinion in chemical biology* **2003**, *7* (6), 727–733. DOI: 10.1016/j.cbpa.2003.10.011.
- (183) Ishikawa, M.; Hirao, I.; Yokoyama, S. Synthesis of 3-(2-deoxy- $\beta$ -d-ribofuranosyl)pyridin-2-one and 2-amino-6-(N,N-dimethylamino)-9-(2-deoxy- $\beta$ -d-ribofuranosyl)purine derivatives for an unnatural base pair. *Tetrahedron letters* **2000**, *41* (20), 3931–3934. DOI: 10.1016/S0040-4039(00)00520-7.
- (184) Ohtsuki, T.; Kimoto, M.; Ishikawa, M.; Mitsui, T.; Hirao, I.; Yokoyama, S. Unnatural base pairs for specific transcription. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98* (9), 4922–4925. DOI: 10.1073/pnas.091532698.
- (185) Fujiwara, T.; Kimoto, M.; Sugiyama, H.; Hirao, I.; Yokoyama, S. Synthesis of 6-(2-thienyl)purine nucleoside derivatives that form unnatural base pairs with pyridin-2-one nucleosides. *Bioorganic & medicinal chemistry letters* **2001**, *11* (16), 2221–2223. DOI: 10.1016/S0960-894X(01)00415-2.
- (186) Martinot, T. A.; Benner, S. A. Artificial genetic systems: exploiting the "aromaticity" formalism to improve the tautomeric ratio for isoguanosine derivatives. *The Journal of organic chemistry* **2004**, *69* (11), 3972–3975. DOI: 10.1021/jo0497959.
- (187) Hutter, D.; Benner, S. A. Expanding the genetic alphabet: non-epimerizing nucleoside with the pyDDA hydrogen-bonding pattern. *The Journal of organic chemistry* **2003**, *68* (25), 9839–9842. DOI: 10.1021/jo034900k.
- (188) Biondi, E.; Benner, S. A. Artificially Expanded Genetic Information Systems for New Aptamer Technologies. *Biomedicines* **2018**, *6* (2). DOI: 10.3390/biomedicines6020053. Published Online: May. 9, 2018.
- (189) Yang, Z.; Chen, F.; Chamberlin, S. G.; Benner, S. A. Expanded genetic alphabets in the polymerase chain reaction. *Angew. Chem. Int. Ed.* **2010**, *49* (1), 177–180. DOI: 10.1002/anie.200905173.
- (190) Yang, Z.; Chen, F.; Alvarado, J. B.; Benner, S. A. Amplification, mutation, and sequencing of a six-letter synthetic genetic system. *J. Am. Chem. Soc.* **2011**, *133* (38), 15105–15112. DOI: 10.1021/ja204910n. Published Online: Sep. 6, 2011.
- (191) Leal, N. A.; Kim, H.-J.; Hoshika, S.; Kim, M.-J.; Carrigan, M. A.; Benner, S. A. Transcription, reverse transcription, and analysis of RNA containing artificial genetic components. *ACS synthetic biology* **2015**, *4* (4), 407–413. DOI: 10.1021/sb500268n. Published Online: Aug. 19, 2014.
- (192) Hirao, I.; Kimoto, M.; Mitsui, T.; Fujiwara, T.; Kawai, R.; Sato, A.; Harada, Y.; Yokoyama, S. An unnatural hydrophobic base pair system: site-specific incorporation of nucleotide analogs into DNA and RNA. *Nature methods* **2006**, *3* (9), 729–735. DOI: 10.1038/nmeth915.
- (193) Kimoto, M.; Sato, A.; Kawai, R.; Yokoyama, S.; Hirao, I. Site-specific incorporation of functional components into RNA by transcription using unnatural base pair systems. *Nucleic acids symposium series (2004)* **2009** (53), 73–74. DOI: 10.1093/nass/nrp037.
- (194) Someya, T.; Ando, A.; Kimoto, M.; Hirao, I. Site-specific labeling of RNA by combining genetic alphabet expansion transcription and copper-free click chemistry. *Nucleic acids*

*research* **2015**, *43* (14), 6665–6676. DOI: 10.1093/nar/gkv638. Published Online: Jun. 29, 2015.

(195) Mitsui, T.; Kimoto, M.; Sato, A.; Yokoyama, S.; Hirao, I. An unnatural hydrophobic base, 4-propynylpyrrole-2-carbaldehyde, as an efficient pairing partner of 9-methylimidazo(4,5)-bpyridine. *Bioorganic & medicinal chemistry letters* **2003**, *13* (24), 4515–4518. DOI: 10.1016/j.bmcl.2003.09.059.

(196) Hirao, I.; Mitsui, T.; Kimoto, M.; Yokoyama, S. An efficient unnatural base pair for PCR amplification. *J. Am. Chem. Soc.* **2007**, *129* (50), 15549–15555. DOI: 10.1021/ja073830m. Published Online: Nov. 21, 2007.

(197) Kimoto, M.; Kawai, R.; Mitsui, T.; Yokoyama, S.; Hirao, I. An unnatural base pair system for efficient PCR amplification and functionalization of DNA molecules. *Nucleic acids research* **2009**, *37* (2), e14. DOI: 10.1093/nar/gkn956. Published Online: Dec. 10, 2008.

(198) Yamashige, R.; Kimoto, M.; Takezawa, Y.; Sato, A.; Mitsui, T.; Yokoyama, S.; Hirao, I. Highly specific unnatural base pair systems as a third base pair for PCR amplification. *Nucleic acids research* **2012**, *40* (6), 2793–2806. DOI: 10.1093/nar/gkr1068. Published Online: Nov. 24, 2011.

(199) Henry, A. A.; Yu, C.; Romesberg, F. E. Determinants of unnatural nucleobase stability and polymerase recognition. *J. Am. Chem. Soc.* **2003**, *125* (32), 9638–9646. DOI: 10.1021/ja035398o.

(200) Matsuda, S.; Henry, A. A.; Romesberg, F. E. Optimization of unnatural base pair packing for polymerase recognition. *J. Am. Chem. Soc.* **2006**, *128* (19), 6369–6375. DOI: 10.1021/ja057575m.

(201) Wu, Y.; Ogawa, A. K.; Berger, M.; McMinn, D. L.; Schultz, P. G.; Romesberg, F. E. Efforts toward Expansion of the Genetic Alphabet: Optimization of Interbase Hydrophobic Interactions. *J. Am. Chem. Soc.* **2000**, *122* (32), 7621–7632. DOI: 10.1021/ja0009931.

(202) Tae, E. L.; Wu, Y.; Xia, G.; Schultz, P. G.; Romesberg, F. E. Efforts toward expansion of the genetic alphabet: replication of DNA with three base pairs. *J. Am. Chem. Soc.* **2001**, *123* (30), 7439–7440. DOI: 10.1021/ja010731e.

(203) Berger, M.; Luzzi, S. D.; Henry, A. A.; Romesberg, F. E. Stability and selectivity of unnatural DNA with five-membered-ring nucleobase analogues. *J. Am. Chem. Soc.* **2002**, *124* (7), 1222–1226. DOI: 10.1021/ja012090t.

(204) Matsuda, S.; Romesberg, F. E. Optimization of interstrand hydrophobic packing interactions within unnatural DNA base pairs. *J. Am. Chem. Soc.* **2004**, *126* (44), 14419–14427. DOI: 10.1021/ja047291m.

(205) Leconte, A. M.; Matsuda, S.; Hwang, G. T.; Romesberg, F. E. Efforts towards expansion of the genetic alphabet: pyridone and methyl pyridone nucleobases. *Angew. Chem. Int. Ed.* **2006**, *45* (26), 4326–4329. DOI: 10.1002/anie.200601272.

(206) Matsuda, S.; Fillo, J. D.; Henry, A. A.; Rai, P.; Wilkens, S. J.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Schultz, P. G.; Spraggon, G.; Romesberg, F. E. Efforts toward expansion of the genetic alphabet: structure and replication of unnatural base pairs. *Journal of the American Chemical Society* **2007**, *129* (34), 10466–10473. DOI: 10.1021/ja072276d. Published Online: Aug. 8, 2007.

(207) Leconte, A. M.; Hwang, G. T.; Matsuda, S.; Capek, P.; Hari, Y.; Romesberg, F. E. Discovery, characterization, and optimization of an unnatural base pair for expansion of the genetic alphabet. *Journal of the American Chemical Society* **2008**, *130* (7), 2336–2343. DOI: 10.1021/ja078223d. Published Online: Jan. 25, 2008.

(208) Hwang, G. T.; Romesberg, F. E. Unnatural substrate repertoire of A, B, and X family DNA polymerases. *J. Am. Chem. Soc.* **2008**, *130* (44), 14872–14882. DOI: 10.1021/ja803833h. Published Online: Oct. 11, 2008.

(209) Malyshev, D. A.; Seo, Y. J.; Ordoukhanian, P.; Romesberg, F. E. PCR with an expanded genetic alphabet. *J. Am. Chem. Soc.* **2009**, *131* (41), 14620–14621. DOI: 10.1021/ja906186f.

- (210) Betz, K.; Malyshev, D. A.; Lavergne, T.; Welte, W.; Diederichs, K.; Dwyer, T. J.; Ordoukhanian, P.; Romesberg, F. E.; Marx, A. KlenTaq polymerase replicates unnatural base pairs by inducing a Watson-Crick geometry. *Nature chemical biology* **2012**, *8* (7), 612–614. DOI: 10.1038/nchembio.966. Published Online: Jun. 3, 2012.
- (211) Betz, K.; Malyshev, D. A.; Lavergne, T.; Welte, W.; Diederichs, K.; Romesberg, F. E.; Marx, A. Structural insights into DNA replication without hydrogen bonds. *J. Am. Chem. Soc.* **2013**, *135* (49), 18637–18643. DOI: 10.1021/ja409609j. Published Online: Nov. 27, 2013.
- (212) Dien, V. T.; Holcomb, M.; Feldman, A. W.; Fischer, E. C.; Dwyer, T. J.; Romesberg, F. E. Progress Toward a Semi-Synthetic Organism with an Unrestricted Expanded Genetic Alphabet. *Journal of the American Chemical Society* **2018**, *140* (47), 16115–16123. DOI: 10.1021/jacs.8b08416. Published Online: Nov. 12, 2018.
- (213) Domnick, C.; Eggert, F.; Kath-Schorr, S. Site-specific enzymatic introduction of a norbornene modified unnatural base into RNA and application in post-transcriptional labeling. *Chemical communications (Cambridge, England)* **2015**, *51* (39), 8253–8256. DOI: 10.1039/c5cc01765c.
- (214) Eggert, F.; Kath-Schorr, S. A cyclopropene-modified nucleotide for site-specific RNA labeling using genetic alphabet expansion transcription. *Chemical communications (Cambridge, England)* **2016**, *52* (45), 7284–7287. DOI: 10.1039/c6cc02321e. Published Online: May. 16, 2016.
- (215) Bornewasser, L.; Domnick, C.; Kath-Schorr, S. Stronger together for in-cell translation: natural and unnatural base modified mRNA. *Chemical science* **2022**, *13* (17), 4753–4761. DOI: 10.1039/D2SC00670G. Published Online: Mar. 3, 2022.
- (216) Dörrenhaus, R.; Wagner, P. K.; Kath-Schorr, S. Two are not enough: synthetic strategies and applications of unnatural base pairs. *Biological chemistry* **2023**, *404* (10), 883–896. DOI: 10.1515/hsz-2023-0169. Published Online: Jun. 26, 2023.
- (217) Wei, D.; Li, X.; Wang, Y.; Yu, H. Functional XNA and Biomedical Application. In *Handbook of Chemical Biology of Nucleic Acids*; Sugimoto, N., Ed.; Springer Nature Singapore, 2023; pp 2173–2201. DOI: 10.1007/978-981-19-9776-1\_70.
- (218) Tu, T.; Huan, S.; Ke, G.; Zhang, X. Functional Xeno Nucleic Acids for Biomedical Application. *Chemical research in Chinese universities* **2022**, 1–7. DOI: 10.1007/s40242-022-2186-7. Published Online: Jul. 5, 2022.
- (219) James E. Summerton. Morpholino, siRNA, and S-DNA Compared: Impact of Structure and Mechanism of Action on Off-Target Effects and Sequence Specificity. *Current Topics in Medicinal Chemistry* **2007**, *10* (7), 651–660.
- (220) Moulton, J. D. Using morpholinos to control gene expression. *Current protocols in nucleic acid chemistry* **2007**, *Chapter 4* (1), Unit 4.30. DOI: 10.1002/0471142700.nc0430s27.
- (221) Hudziak, R. M.; Barofsky, E.; Barofsky, D. F.; Weller, D. L.; Huang, S. B.; Weller, D. D. Resistance of morpholino phosphorodiamidate oligomers to enzymatic degradation. *Antisense & nucleic acid drug development* **1996**, *6* (4), 267–272. DOI: 10.1089/oli.1.1996.6.267.
- (222) Das, A.; Ghosh, A.; Sinha, S. C5-pyrimidine-functionalized morpholino oligonucleotides exhibit differential binding affinity, target specificity and lipophilicity. *Organic & biomolecular chemistry* **2023**, *21* (6), 1242–1253. DOI: 10.1039/d2ob01759h. Published Online: Feb. 8, 2023.
- (223) Bhadra, J.; Pattanayak, S.; Sinha, S. Synthesis of Morpholino Monomers, Chlorophosphoramidate Monomers, and Solid-Phase Synthesis of Short Morpholino Oligomers. *Current protocols in nucleic acid chemistry* **2015**, *62*, 4.65.1-4.65.26. DOI: 10.1002/0471142700.nc0465s62. Published Online: Sep. 1, 2015.
- (224) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science (New York, N.Y.)* **1991**, *254* (5037), 1497–1500. DOI: 10.1126/science.1962210.
- (225) Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Freier, S. M.; Driver, D. A.; Berg, R. H.; Kim, S. K.; Norden, B.; Nielsen, P. E. PNA hybridizes to complementary

oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature* **1993**, *365* (6446), 566–568. DOI: 10.1038/365566a0.

(226) Demidov, V. V.; Potaman, V. N.; Frank-Kamenetskii, M. D.; Egholm, M.; Buchard, O.; Sönnichsen, S. H.; Nielsen, P. E. Stability of peptide nucleic acids in human serum and cellular extracts. *Biochemical pharmacology* **1994**, *48* (6), 1310–1313. DOI: 10.1016/0006-2952(94)90171-6.

(227) Periyalagan, A.; Hong, I. S. A novel synthetic method of peptide nucleic acid (PNA) oligomers using Boc/Cbz-protected PNA trimer blocks in the solution phase. *Bulletin Korean Chem Soc* **2022**, *43* (5), 714–723. DOI: 10.1002/bkcs.12513.

(228) Rosenbaum, D. M.; Liu, D. R. Efficient and sequence-specific DNA-templated polymerization of peptide nucleic acid aldehydes. *Journal of the American Chemical Society* **2003**, *125* (46), 13924–13925. DOI: 10.1021/ja038058b.

(229) Luo, M.; Groaz, E.; Froeyen, M.; Pezo, V.; Jaziri, F.; Leonczak, P.; Schepers, G.; Rozenski, J.; Marlière, P.; Herdewijn, P. Invading *Escherichia coli* Genetics with a Xenobiotic Nucleic Acid Carrying an Acyclic Phosphonate Backbone (ZNA). *J. Am. Chem. Soc.* **2019**, *141* (27), 10844–10851. DOI: 10.1021/jacs.9b04714. Published Online: Jun. 28, 2019.

(230) Liu, C.; Cozens, C.; Jaziri, F.; Rozenski, J.; Maréchal, A.; Dumbre, S.; Pezo, V.; Marlière, P.; Pinheiro, V. B.; Groaz, E.; Herdewijn, P. Phosphonomethyl Oligonucleotides as Backbone-Modified Artificial Genetic Polymers. *Journal of the American Chemical Society* **2018**, *140* (21), 6690–6699. DOI: 10.1021/jacs.8b03447. Published Online: May. 17, 2018.

(231) Rejman, D.; Snásel, J.; Liboska, R.; Tocík, Z.; Paces, O.; Králíková, S.; Rinnová, M.; Kois, P.; Rosenberg, I. Oligonucleotides with isopolar phosphonate internucleotide linkage: a new perspective for antisense compounds? *Nucleosides, nucleotides & nucleic acids* **2001**, *20* (4-7), 819–823. DOI: 10.1081/NCN-100002437.

(232) McCloskey, C. M.; Li, Q.; Yik, E. J.; Chim, N.; Ngor, A. K.; Medina, E.; Grubisic, I.; Co Ting Keh, L.; Poplin, R.; Chaput, J. C. Evolution of Functionally Enhanced  $\alpha$ -L-Threofuranosyl Nucleic Acid Aptamers. *ACS synthetic biology* **2021**, *10* (11), 3190–3199. DOI: 10.1021/acssynbio.1c00481. Published Online: Nov. 5, 2021.

(233) Majumdar, B.; Sarma, D.; Yu, Y.; Lozoya-Colinas, A.; Chaput, J. C. Increasing the functional density of threose nucleic acid. *RSC Chem. Biol.* **2023**. DOI: 10.1039/D3CB00159H.

(234) Li, Q.; Maola, V. A.; Chim, N.; Hussain, J.; Lozoya-Colinas, A.; Chaput, J. C. Synthesis and Polymerase Recognition of Threose Nucleic Acid Triphosphates Equipped with Diverse Chemical Functionalities. *Journal of the American Chemical Society* **2021**, *143* (42), 17761–17768. DOI: 10.1021/jacs.1c08649. Published Online: Oct. 12, 2021.

(235) Ramaraj, T.; Angel, T.; Dratz, E. A.; Jesaitis, A. J.; Mumey, B. Antigen-antibody interface properties: composition, residue interactions, and features of 53 non-redundant structures. *Biochimica et biophysica acta* **2012**, *1824* (3), 520–532. DOI: 10.1016/j.bbapap.2011.12.007. Published Online: Jan. 10, 2012.

(236) Hollenstein, M.; Hipolito, C. J.; Lam, C. H.; Perrin, D. M. A self-cleaving DNA enzyme modified with amines, guanidines and imidazoles operates independently of divalent metal cations (M<sup>2+</sup>). *Nucleic acids research* **2009**, *37* (5), 1638–1649. DOI: 10.1093/nar/gkn1070. Published Online: Jan. 19, 2009.

(237) Vaught, J. D.; Bock, C.; Carter, J.; Fitzwater, T.; Otis, M.; Schneider, D.; Rolando, J.; Waugh, S.; Wilcox, S. K.; Eaton, B. E. Expanding the chemistry of DNA for in vitro selection. *Journal of the American Chemical Society* **2010**, *132* (12), 4141–4151. DOI: 10.1021/ja908035g.

(238) Rohloff, J. C.; Gelinas, A. D.; Jarvis, T. C.; Ochsner, U. A.; Schneider, D. J.; Gold, L.; Janjic, N. Nucleic Acid Ligands With Protein-like Side Chains: Modified Aptamers and Their Use as Diagnostic and Therapeutic Agents. *Molecular therapy. Nucleic acids* **2014**, *3* (10), e201. DOI: 10.1038/mtna.2014.49. Published Online: Oct. 7, 2014.

- (239) Klöcker, N.; Weissenboeck, F. P.; Rentmeister, A. Covalent labeling of nucleic acids. *Chemical Society reviews* **2020**, *49* (23), 8749–8773. DOI: 10.1039/d0cs00600a. Published Online: Oct. 21, 2020.
- (240) Carell, T.; Vrabel, M. Bioorthogonal Chemistry—Introduction and Overview corrected. *Topics in current chemistry (Cham)* **2016**, *374* (1), 9. DOI: 10.1007/s41061-016-0010-x. Published Online: Feb. 1, 2016.
- (241) Plücker, O.; Siegl, J.; Bryant, L. L.; Mayer, G. Dynamic changes in DNA populations revealed by split-combine selection. *Chemical science* **2020**, *11* (35), 9577–9583. DOI: 10.1039/d0sc01952f. Published Online: Aug. 14, 2020.
- (242) Perrone, D.; Marchesi, E.; Preti, L.; Navacchia, M. L. Modified Nucleosides, Nucleotides and Nucleic Acids via Click Azide-Alkyne Cycloaddition for Pharmacological Applications. *Molecules (Basel, Switzerland)* **2021**, *26* (11). DOI: 10.3390/molecules26113100. Published Online: May. 22, 2021.
- (243) Patterson, D. M.; Nazarova, L. A.; Prescher, J. A. Finding the right (bioorthogonal) chemistry. *ACS chemical biology* **2014**, *9* (3), 592–605. DOI: 10.1021/cb400828a. Published Online: Jan. 30, 2014.
- (244) Rigolot, V.; Biot, C.; Lion, C. To View Your Biomolecule, Click inside the Cell. *Angew. Chem. Int. Ed.* **2021**, *60* (43), 23084–23105. DOI: 10.1002/anie.202101502. Published Online: Jul. 9, 2021.
- (245) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed.* **2001**, *40* (11), 2004–2021. DOI: 10.1002/1521-3773(20010601)40:11<2004:AID-ANIE2004>3.0.CO;2-5.
- (246) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective “Ligation” of Azides and Terminal Alkynes. *Angew. Chem. Int. Ed.* **2002**, *41* (14), 2596–2599. DOI: 10.1002/1521-3773(20020715)41:14<2596:AID-ANIE2596>3.0.CO;2-4.
- (247) Tornøe, C. W.; Christensen, C.; Meldal, M. Peptidotriazoles on solid phase: 1,2,3-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *The Journal of organic chemistry* **2002**, *67* (9), 3057–3064. DOI: 10.1021/jo011148j.
- (248) Presolski, S. I.; Hong, V. P.; Finn, M. G. Copper-Catalyzed Azide-Alkyne Click Chemistry for Bioconjugation. *Current protocols in chemical biology* **2011**, *3* (4), 153–162. DOI: 10.1002/9780470559277.ch110148. Published Online: Dec. 1, 2011.
- (249) Pitié, M.; Pratviel, G. Activation of DNA carbon-hydrogen bonds by metal complexes. *Chemical reviews* **2010**, *110* (2), 1018–1059. DOI: 10.1021/cr900247m.
- (250) Paredes, E.; Das, S. R. Click chemistry for rapid labeling and ligation of RNA. *Chembiochem : a European journal of chemical biology* **2011**, *12* (1), 125–131. DOI: 10.1002/cbic.201000466.
- (251) Paredes, E.; Das, S. R. Optimization of acetonitrile co-solvent and copper stoichiometry for pseudo-ligandless click chemistry with nucleic acids. *Bioorganic & medicinal chemistry letters* **2012**, *22* (16), 5313–5316. DOI: 10.1016/j.bmcl.2012.06.027. Published Online: Jun. 29, 2012.
- (252) Eltepu, L.; Jayaraman, M.; Rajeev, K. G.; Manoharan, M. An immobilized and reusable Cu(I) catalyst for metal ion-free conjugation of ligands to fully deprotected oligonucleotides through click reaction. *Chem. Commun.* **2013**, *49* (2), 184–186. DOI: 10.1039/c2cc36811k. Published Online: Nov. 21, 2012.
- (253) Pfeiffer, F.; Tolle, F.; Rosenthal, M.; Brändle, G. M.; Ewers, J.; Mayer, G. Identification and characterization of nucleobase-modified aptamers by click-SELEX. *Nature protocols* **2018**, *13* (5), 1153–1180. DOI: 10.1038/nprot.2018.023. Published Online: Apr. 26, 2018.
- (254) Siegl, J.; Nikolin, C.; Phung, N. L.; Thoms, S.; Blume, C.; Mayer, G. Split-Combine Click-SELEX Reveals Ligands Recognizing the Transplant Rejection Biomarker CXCL9. *ACS chemical biology* **2022**, *17* (1), 129–137. DOI: 10.1021/acscchembio.1c00789. Published Online: Jan. 12, 2022.

- (255) Salic, A.; Mitchison, T. J. A chemical method for fast and sensitive detection of DNA synthesis in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105* (7), 2415–2420. DOI: 10.1073/pnas.0712168105. Published Online: Feb. 12, 2008.
- (256) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. A strain-promoted 3 + 2 azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **2004**, *126* (46), 15046–15047. DOI: 10.1021/ja044996f.
- (257) Kim, E.; Koo, H. Biomedical applications of copper-free click chemistry: in vitro, in vivo, and ex vivo. *Chemical science* **2019**, *10* (34), 7835–7851. DOI: 10.1039/c9sc03368h. Published Online: Aug. 16, 2019.
- (258) Knall, A.-C.; Slugovc, C. Inverse electron demand Diels-Alder (IEDDA)-initiated conjugation: a (high) potential click chemistry scheme. *Chemical Society reviews* **2013**, *42* (12), 5131–5142. DOI: 10.1039/C3CS60049A.
- (259) Kozma, E.; Demeter, O.; Kele, P. Bio-orthogonal Fluorescent Labelling of Biopolymers through Inverse-Electron-Demand Diels-Alder Reactions. *Chembiochem : a European journal of chemical biology* **2017**, *18* (6), 486–501. DOI: 10.1002/cbic.201600607. Published Online: Feb. 1, 2017.
- (260) Wiessler, M.; Waldeck, W.; Kliem, C.; Pipkorn, R.; Braun, K. The Diels-Alder-reaction with inverse-electron-demand, a very efficient versatile click-reaction concept for proper ligation of variable molecular partners. *International journal of medical sciences* **2009**, *7* (1), 19–28. DOI: 10.7150/ijms.7.19. Published Online: Dec. 5, 2009.
- (261) Oliveira, B. L.; Guo, Z.; Bernardes, G. J. L. Inverse electron demand Diels-Alder reactions in chemical biology. *Chemical Society reviews* **2017**, *46* (16), 4895–4950. DOI: 10.1039/C7CS00184C.
- (262) Zhang, F.-G.; Chen, Z.; Tang, X.; Ma, J.-A. Triazines: Syntheses and Inverse Electron-demand Diels-Alder Reactions. *Chemical reviews* **2021**, *121* (23), 14555–14593. DOI: 10.1021/acs.chemrev.1c00611. Published Online: Sep. 29, 2021.
- (263) Ploschik, D.; Rönicke, F.; Beike, H.; Strasser, R.; Wagenknecht, H.-A. DNA Primer Extension with Cyclopropenylated 7-Deaza-2'-deoxyadenosine and Efficient Bioorthogonal Labeling in Vitro and in Living Cells. *Chembiochem : a European journal of chemical biology* **2018**, *19* (18), 1949–1953. DOI: 10.1002/cbic.201800354. Published Online: Aug. 15, 2018.
- (264) Domnick, C.; Hagelueken, G.; Eggert, F.; Schiemann, O.; Kath-Schorr, S. Posttranscriptional spin labeling of RNA by tetrazine-based cycloaddition. *Organic & biomolecular chemistry* **2019**, *17* (7), 1805–1808. DOI: 10.1039/c8ob02597e.
- (265) Eggert, F.; Kulikov, K.; Domnick, C.; Leifels, P.; Kath-Schorr, S. Illuminated by foreign letters - Strategies for site-specific cyclopropene modification of large functional RNAs via in vitro transcription. *Methods (San Diego, Calif.)* **2017**, *120*, 17–27. DOI: 10.1016/j.ymeth.2017.04.021. Published Online: Apr. 26, 2017.
- (266) Pyka, A. M.; Domnick, C.; Braun, F.; Kath-Schorr, S. Diels-Alder cycloadditions on synthetic RNA in mammalian cells. *Bioconjugate chemistry* **2014**, *25* (8), 1438–1443. DOI: 10.1021/bc500302y. Published Online: Aug. 7, 2014.
- (267) Ganz, D.; Harijan, D.; Wagenknecht, H.-A. Labelling of DNA and RNA in the cellular environment by means of bioorthogonal cycloaddition chemistry. *RSC Chem. Biol.* **2020**, *1* (3), 86–97. DOI: 10.1039/d0cb00047g. Published Online: Jun. 2, 2020.
- (268) Egli, M.; Manoharan, M. Chemistry, structure and function of approved oligonucleotide therapeutics. *Nucleic acids research* **2023**, *51* (6), 2529–2573. DOI: 10.1093/nar/gkad067.
- (269) Zhou, S.; Chen, W.; Cole, J.; Zhu, G. Delivery of nucleic acid therapeutics for cancer immunotherapy. *Medicine in drug discovery* **2020**, *6*. DOI: 10.1016/j.medidd.2020.100023. Published Online: Mar. 24, 2020.
- (270) Sharma, V. K.; Rungta, P.; Prasad, A. K. Nucleic acid therapeutics: basic concepts and recent developments. *RSC Adv.* **2014**, *4* (32), 16618. DOI: 10.1039/c3ra47841f.

- (271) Diafa, S.; Hollenstein, M. Generation of Aptamers with an Expanded Chemical Repertoire. *Molecules (Basel, Switzerland)* **2015**, *20* (9), 16643–16671. DOI: 10.3390/molecules200916643. Published Online: Sep. 14, 2015.
- (272) Paunovska, K.; Loughrey, D.; Dahlman, J. E. Drug delivery systems for RNA therapeutics. *Nature reviews. Genetics* **2022**, *23* (5), 265–280. DOI: 10.1038/s41576-021-00439-4. Published Online: Jan. 4, 2022.
- (273) Zhu, Y.; Zhu, L.; Wang, X.; Jin, H. RNA-based therapeutics: an overview and prospectus. *Cell death & disease* **2022**, *13* (7), 644. DOI: 10.1038/s41419-022-05075-2. Published Online: Jul. 23, 2022.
- (274) Wang, F.; Liu, L. S.; Lau, C. H.; Han Chang, T. J.; Tam, D. Y.; Leung, H. M.; Tin, C.; Lo, P. K. Synthetic  $\alpha$ -l-Threose Nucleic Acids Targeting Bcl-2 Show Gene Silencing and in Vivo Antitumor Activity for Cancer Therapy. *ACS applied materials & interfaces* **2019**, *11* (42), 38510–38518. DOI: 10.1021/acsami.9b14324. Published Online: Oct. 8, 2019.
- (275) Liu, L. S.; Leung, H. M.; Tam, D. Y.; Lo, T. W.; Wong, S. W.; Lo, P. K.  $\alpha$ -l-Threose Nucleic Acids as Biocompatible Antisense Oligonucleotides for Suppressing Gene Expression in Living Cells. *ACS applied materials & interfaces* **2018**, *10* (11), 9736–9743. DOI: 10.1021/acsami.8b01180. Published Online: Mar. 8, 2018.
- (276) Matsuda, S.; Bala, S.; Liao, J.-Y.; Datta, D.; Mikami, A.; Woods, L.; Harp, J. M.; Gilbert, J. A.; Bisbe, A.; Manoharan, R. M.; Kim, M.; Theile, C. S.; Guenther, D. C.; Jiang, Y.; Agarwal, S.; Maganti, R.; Schlegel, M. K.; Zlatev, I.; Charisse, K.; Rajeev, K. G.; Castoreno, A.; Maier, M.; Janas, M. M.; Egli, M.; Chaput, J. C.; Manoharan, M. Shorter Is Better: The  $\alpha$ -(l)-Threofuranosyl Nucleic Acid Modification Improves Stability, Potency, Safety, and Ago2 Binding and Mitigates Off-Target Effects of Small Interfering RNAs. *J. Am. Chem. Soc.* **2023**, *145* (36), 19691–19706. DOI: 10.1021/jacs.3c04744. Published Online: Aug. 28, 2023.
- (277) Mook, O. R.; Baas, F.; Wissel, M. B. de; Fluiter, K. Evaluation of locked nucleic acid-modified small interfering RNA in vitro and in vivo. *Molecular cancer therapeutics* **2007**, *6* (3), 833–843. DOI: 10.1158/1535-7163.MCT-06-0195.
- (278) Whitehead, K. A.; Dahlman, J. E.; Langer, R. S.; Anderson, D. G. Silencing or stimulation? siRNA delivery and the immune system. *Annual review of chemical and biomolecular engineering* **2011**, *2*, 77–96. DOI: 10.1146/annurev-chembioeng-061010-114133.
- (279) Gote, V.; Bolla, P. K.; Kommineni, N.; Butreddy, A.; Nukala, P. K.; Palakurthi, S. S.; Khan, W. A Comprehensive Review of mRNA Vaccines. *International Journal of Molecular Sciences* **2023**, *24* (3). DOI: 10.3390/ijms24032700. Published Online: Jan. 31, 2023.
- (280) Ho, L. L. Y.; Schiess, G. H. A.; Miranda, P.; Weber, G.; Astakhova, K. Pseudouridine and N1-methylpseudouridine as potent nucleotide analogues for RNA therapy and vaccine development. *RSC Chem. Biol.* **2024**, *5* (5), 418–425. DOI: 10.1039/D4CB00022F. Published Online: Mar. 19, 2024.
- (281) Panigaj, M.; Johnson, M. B.; Ke, W.; McMillan, J.; Goncharova, E. A.; Chandler, M.; Afonin, K. A. Aptamers as Modular Components of Therapeutic Nucleic Acid Nanotechnology. *ACS nano* **2019**, *13* (11), 12301–12321. DOI: 10.1021/acs.nano.9b06522. Published Online: Nov. 5, 2019.
- (282) Zhang, Y.; Lai, B. S.; Juhas, M. Recent Advances in Aptamer Discovery and Applications. *Molecules (Basel, Switzerland)* **2019**, *24* (5). DOI: 10.3390/molecules24050941. Published Online: Mar. 7, 2019.
- (283) Liu, Q.; Zhang, W.; Chen, S.; Zhuang, Z.; Zhang, Y.; Jiang, L.; Lin, J. S. SELEX tool: a novel and convenient gel-based diffusion method for monitoring of aptamer-target binding. *Journal of biological engineering* **2020**, *14*, 1. DOI: 10.1186/s13036-019-0223-y. Published Online: Jan. 13, 2020.
- (284) Catuogno, S.; Esposito, C. L. Aptamer Cell-Based Selection: Overview and Advances. *Biomedicines* **2017**, *5* (3). DOI: 10.3390/biomedicines5030049. Published Online: Aug. 14, 2017.



- (285) Ellington, A. D.; Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **1990**, *346* (6287), 818–822. DOI: 10.1038/346818a0.
- (286) Tuerk, C.; Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (New York, N.Y.)* **1990**, *249* (4968), 505–510. DOI: 10.1126/science.2200121.
- (287) Wang, T.; Chen, C.; Larcher, L. M.; Barrero, R. A.; Veedu, R. N. Three decades of nucleic acid aptamer technologies: Lessons learned, progress and opportunities on aptamer development. *Biotechnology advances* **2019**, *37* (1), 28–50. DOI: 10.1016/j.biotechadv.2018.11.001. Published Online: Nov. 5, 2018.
- (288) Odeh, F.; Nsairat, H.; Alshaer, W.; Ismail, M. A.; Esawi, E.; Qaqish, B.; Bawab, A. A.; Ismail, S. I. Aptamers Chemistry: Chemical Modifications and Conjugation Strategies. *Molecules (Basel, Switzerland)* **2019**, *25* (1). DOI: 10.3390/molecules25010003. Published Online: Dec. 18, 2019.
- (289) Elskens, J. P.; Elskens, J. M.; Madder, A. Chemical Modification of Aptamers for Increased Binding Affinity in Diagnostic Applications: Current Status and Future Prospects. *IJMS* **2020**, *21* (12), 4522. DOI: 10.3390/ijms21124522.
- (290) Volk, D. E.; Lokesh, G. L. R. Development of Phosphorothioate DNA and DNA Thioaptamers. *Biomedicines* **2017**, *5* (3). DOI: 10.3390/biomedicines5030041. Published Online: Jul. 13, 2017.
- (291) Mann, A. P.; Somasunderam, A.; Nieves-Alicea, R.; Li, X.; Hu, A.; Sood, A. K.; Ferrari, M.; Gorenstein, D. G.; Tanaka, T. Identification of thioaptamer ligand against E-selectin: potential application for inflamed vasculature targeting. *PLoS one* **2010**, *5* (9). DOI: 10.1371/journal.pone.0013050. Published Online: Sep. 30, 2010.
- (292) Somasunderam, A.; Ferguson, M. R.; Rojo, D. R.; Thiviyanathan, V.; Li, X.; O'Brien, W. A.; Gorenstein, D. G. Combinatorial selection, inhibition, and antiviral activity of DNA thioaptamers targeting the RNase H domain of HIV-1 reverse transcriptase. *Biochemistry* **2005**, *44* (30), 10388–10395. DOI: 10.1021/bi0507074.
- (293) Sung, H. J.; Choi, S.; Lee, J. W.; Ok, C. Y.; Bae, Y.-S.; Kim, Y.-H.; Lee, W.; Heo, K.; Kim, I.-H. Inhibition of human neutrophil activity by an RNA aptamer bound to interleukin-8. *Biomaterials* **2014**, *35* (1), 578–589. DOI: 10.1016/j.biomaterials.2013.09.107. Published Online: Oct. 13, 2013.
- (294) Alves Ferreira-Bravo, I.; Cozens, C.; Holliger, P.; DeStefano, J. J. Selection of 2'-deoxy-2'-fluoroarabinonucleotide (FANA) aptamers that bind HIV-1 reverse transcriptase with picomolar affinity. *Nucleic acids research* **2015**, *43* (20), 9587–9599. DOI: 10.1093/nar/gkv1057. Published Online: Oct. 17, 2015.
- (295) Burmeister, P. E.; Lewis, S. D.; Silva, R. F.; Preiss, J. R.; Horwitz, L. R.; Pendergrast, P. S.; McCauley, T. G.; Kurz, J. C.; Epstein, D. M.; Wilson, C.; Keefe, A. D. Direct in vitro selection of a 2'-O-methyl aptamer to VEGF. *Chemistry & biology* **2005**, *12* (1), 25–33. DOI: 10.1016/j.chembiol.2004.10.017.
- (296) Chen, Z.; Luo, H.; Gubu, A.; Yu, S.; Zhang, H.; Dai, H.; Zhang, Y.; Zhang, B.; Ma, Y.; Lu, A.; Zhang, G. Chemically modified aptamers for improving binding affinity to the target proteins via enhanced non-covalent bonding. *Frontiers in cell and developmental biology* **2023**, *11*, 1091809. DOI: 10.3389/fcell.2023.1091809. Published Online: Feb. 23, 2023.
- (297) Ng, E. W. M.; Shima, D. T.; Calias, P.; Cunningham, E. T.; Guyer, D. R.; Adamis, A. P. Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *Nature reviews. Drug discovery* **2006**, *5* (2), 123–132. DOI: 10.1038/nrd1955.
- (298) Mullard, A. FDA approves second RNA aptamer. *Nature reviews. Drug discovery* **2023**, *22* (10), 774. DOI: 10.1038/d41573-023-00148-z.
- (299) Ruckman, J.; Green, L. S.; Beeson, J.; Waugh, S.; Gillette, W. L.; Henninger, D. D.; Claesson-Welsh, L.; Janjić, N. 2'-Fluoropyrimidine RNA-based aptamers to the 165-amino acid form of vascular endothelial growth factor (VEGF165). Inhibition of receptor binding and VEGF-induced vascular permeability through interactions requiring the exon 7-encoded domain.

*Journal of Biological Chemistry* **1998**, *273* (32), 20556–20567. DOI: 10.1074/jbc.273.32.20556.

(300) Oliveira, R.; Pinho, E.; Sousa, A. L.; DeStefano, J. J.; Azevedo, N. F.; Almeida, C. Improving aptamer performance with nucleic acid mimics: de novo and post-SELEX approaches. *Trends in biotechnology* **2022**, *40* (5), 549–563. DOI: 10.1016/j.tibtech.2021.09.011. Published Online: Oct. 28, 2021.

(301) Kasahara, Y.; Irisawa, Y.; Ozaki, H.; Obika, S.; Kuwahara, M. 2',4'-BNA/LNA aptamers: CE-SELEX using a DNA-based library of full-length 2'-O,4'-C-methylene-bridged/linked bicyclic ribonucleotides. *Bioorganic & medicinal chemistry letters* **2013**, *23* (5), 1288–1292. DOI: 10.1016/j.bmcl.2012.12.093. Published Online: Jan. 16, 2013.

(302) Li, X.; Li, Z.; Yu, H. Selection of threose nucleic acid aptamers to block PD-1/PD-L1 interaction for cancer immunotherapy. *Chem. Commun.* **2020**, *56* (93), 14653–14656. DOI: 10.1039/D0CC06032A. Published Online: Nov. 6, 2020.

(303) Dunn, M. R.; McCloskey, C. M.; Buckley, P.; Rhea, K.; Chaput, J. C. Generating Biologically Stable TNA Aptamers that Function with High Affinity and Thermal Stability. *Journal of the American Chemical Society* **2020**, *142* (17), 7721–7724. DOI: 10.1021/jacs.0c00641. Published Online: Apr. 20, 2020.

(304) Dunn, M. R.; Chaput, J. C. An In Vitro Selection Protocol for Threose Nucleic Acid (TNA) Using DNA Display. *Current protocols in nucleic acid chemistry* **2014**, *57*, 9.8.1-19. DOI: 10.1002/0471142700.nc0908s57. Published Online: Jun. 24, 2014.

(305) Tolle, F.; Brändle, G. M.; Matzner, D.; Mayer, G. A Versatile Approach Towards Nucleobase-Modified Aptamers. *Angew. Chem. Int. Ed.* **2015**, *54* (37), 10971–10974. DOI: 10.1002/anie.201503652. Published Online: Jul. 23, 2015.

(306) Hirao, I.; Kimoto, M.; Lee, K. H. DNA aptamer generation by ExSELEX using genetic alphabet expansion with a mini-hairpin DNA stabilization method. *Biochimie* **2018**, *145*, 15–21. DOI: 10.1016/j.biochi.2017.09.007. Published Online: Sep. 13, 2017.

(307) Kimoto, M.; Yamashige, R.; Matsunaga, K.; Yokoyama, S.; Hirao, I. Generation of high-affinity DNA aptamers using an expanded genetic alphabet. *Nature biotechnology* **2013**, *31* (5), 453–457. DOI: 10.1038/nbt.2556. Published Online: Apr. 7, 2013.

(308) Matsunaga, K.; Kimoto, M.; Hirao, I. High-Affinity DNA Aptamer Generation Targeting von Willebrand Factor A1-Domain by Genetic Alphabet Expansion for Systematic Evolution of Ligands by Exponential Enrichment Using Two Types of Libraries Composed of Five Different Bases. *Journal of the American Chemical Society* **2017**, *139* (1), 324–334. DOI: 10.1021/jacs.6b10767. Published Online: Dec. 28, 2016.

(309) Kimoto, M.; Matsunaga, K.; Hirao, I. DNA Aptamer Generation by Genetic Alphabet Expansion SELEX (ExSELEX) Using an Unnatural Base Pair System. *Methods in molecular biology (Clifton, N.J.)* **2016**, *1380*, 47–60. DOI: 10.1007/978-1-4939-3197-2\_4.

(310) Wang, H.; Zhu, W.; Wang, C.; Li, X.; Wang, L.; Huo, B.; Mei, H.; Zhu, A.; Zhang, G.; Li, L. Locating, tracing and sequencing multiple expanded genetic letters in complex DNA context via a bridge-base approach. *Nucleic acids research* **2023**, *51* (9), e52. DOI: 10.1093/nar/gkad218.

(311) Esawi, E.; Nsairat, H.; Mahmoud, I. S.; Lafi, Z.; Al-Kadash, A.; Al-Ragheb, B. A.; Ismail, S. I.; Alhaer, W. Clinical use and future perspective of aptamers. In *Aptamers Engineered Nanocarriers for Cancer Therapy*, Elsevier, 2023; pp 481–520. DOI: 10.1016/B978-0-323-85881-6.00013-0.

(312) Strauss, S.; Nickels, P. C.; Strauss, M. T.; Jimenez Sabinina, V.; Ellenberg, J.; Carter, J. D.; Gupta, S.; Janjic, N.; Jungmann, R. Modified aptamers enable quantitative sub-10-nm cellular DNA-PAINT imaging. *Nature methods* **2018**, *15* (9), 685–688. DOI: 10.1038/s41592-018-0105-0. Published Online: Aug. 20, 2018.

(313) Kim, D.-H.; Seo, J.-M.; Shin, K.-J.; Yang, S.-G. Design and clinical developments of aptamer-drug conjugates for targeted cancer therapy. *Biomaterials research* **2021**, *25* (1), 42. DOI: 10.1186/s40824-021-00244-4. Published Online: Nov. 25, 2021.

- (314) Kumar Kulabhusan, P.; Hussain, B.; Yüce, M. Current Perspectives on Aptamers as Diagnostic Tools and Therapeutic Agents. *Pharmaceutics* **2020**, *12* (7). DOI: 10.3390/pharmaceutics12070646. Published Online: Jul. 9, 2020.
- (315) Belmont, P.; Constant, J. F.; Demeunynck, M. Nucleic acid conformation diversity: from structure to function and regulation. *Chem. Soc. Rev.* **2001**, *30* (1), 70–81. DOI: 10.1039/a904630e.
- (316) Nishimura, Y.; Torigoe, C.; Tsuboi, M. Salt induced B $\rightarrow$ A transition of poly(dG).poly(dC) and the stabilization of A form by its methylation. *Nucleic acids research* **1986**, *14* (6), 2737–2748. DOI: 10.1093/nar/14.6.2737.
- (317) Ivanov, V. I.; Minchenkova, L. E.; Minyat, E. E.; Frank-Kamenetskii, M. D.; Schyolkina, A. K. The B to A transition of DNA in solution. *Journal of molecular biology* **1974**, *87* (4), 817–833. DOI: 10.1016/0022-2836(74)90086-2.
- (318) Feig, M.; Pettitt, B. M. A molecular simulation picture of DNA hydration around A- and B-DNA. *Biopolymers* **1998**, *48* (4), 199. DOI: 10.1002/(SICI)1097-0282(1998)48:4<199:AID-BIP2>3.0.CO;2-5.
- (319) Egli, M. DNA-cation interactions: quo vadis? *Chemistry & biology* **2002**, *9* (3), 277–286. DOI: 10.1016/s1074-5521(02)00116-3.
- (320) Potaman VN, S. R. *DNA: Alternative Conformations and Biology*; Madame Curie Bioscience Database; Landes Bioscience, 2000-2013.
- (321) Delgado, J. L.; Vance, N. R.; Kerns, R. J. *Crystal structure of DNA dodecamer D(CGCGAATTCGCG)*, 2018. DOI: 10.2210/pdb6cq3/pdb.
- (322) Tanaka, Y.; Fujii, S.; Hiroaki, H.; Sakata, T.; Tanaka, T.; Uesugi, S.; Tomita, K.-I.; Kyogoku, Y. *A'-Form RNA Double Helix in the Single CRystal Structure of r(UGAGCUUCGGCUC)*, 1998. DOI: 10.2210/pdb413d/pdb.
- (323) Luo, Z.; Dauter, M.; Dauter, Z. *Z-DNA dodecamer d(CGCGCGCGCGCG)2 at 0.75 Å resolution solved by P-SAD*, 2014. DOI: 10.2210/pdb4ocb/pdb.
- (324) Wang, A. H.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; van Boom, J. H.; van der Marel, G.; Rich, A. Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **1979**, *282* (5740), 680–686. DOI: 10.1038/282680a0.
- (325) Krall, J. B.; Nichols, P. J.; Henen, M. A.; Vicens, Q.; Vögeli, B. Structure and Formation of Z-DNA and Z-RNA. *Molecules (Basel, Switzerland)* **2023**, *28* (2). DOI: 10.3390/molecules28020843. Published Online: Jan. 14, 2023.
- (326) Hall, K.; Cruz, P.; Tinoco, I.; Jovin, T. M.; van de Sande, J. H. 'Z-RNA'--a left-handed RNA double helix. *Nature* **1984**, *311* (5986), 584–586. DOI: 10.1038/311584a0.
- (327) Peck, L. J.; Nordheim, A.; Rich, A.; Wang, J. C. Flipping of cloned d(pCpG)n.d(pCpG)n DNA sequences from right- to left-handed helical structure by salt, Co(III), or negative supercoiling. *Proceedings of the National Academy of Sciences of the United States of America* **1982**, *79* (15), 4560–4564.
- (328) Rich, A.; Nordheim, A.; Wang, A. H. The chemistry and biology of left-handed Z-DNA. *Annual review of biochemistry* **1984**, *53*, 791–846. DOI: 10.1146/annurev.bi.53.070184.004043.
- (329) Rich, A.; Zhang, S. Timeline: Z-DNA: the long road to biological function. *Nature reviews. Genetics* **2003**, *4* (7), 566–572. DOI: 10.1038/nrg1115.
- (330) Fujii, S.; Wang, A. H.; van der Marel, G.; van Boom, J. H.; Rich, A. Molecular structure of (m<sup>5</sup> dC-dG)<sub>3</sub>: the role of the methyl group on 5-methyl cytosine in stabilizing Z-DNA. *Nucleic acids research* **1982**, *10* (23), 7879–7892. DOI: 10.1093/nar/10.23.7879.
- (331) Ross, W. S.; Hardin, C. C.; Tinoco, I.; Rao, S. N.; Pearlman, D. A.; Kollman, P. A. Effects of nucleotide bromination on the stabilities of Z-RNA and Z-DNA: a molecular mechanics/thermodynamic perturbation study. *Biopolymers* **1989**, *28* (11), 1939–1957. DOI: 10.1002/bip.360281111.
- (332) Barciszewski, J.; Jurczak, J.; Porowski, S.; Specht, T.; Erdmann, V. A. The role of water structure in conformational changes of nucleic acids in ambient and high-pressure conditions.

*European journal of biochemistry* **1999**, *260* (2), 293–307. DOI: 10.1046/j.1432-1327.1999.00184.x.

(333) Balasubramaniam, T.; Ishizuka, T.; Xiao, C.-D.; Bao, H.-L.; Xu, Y. 2'-O-Methyl-8-methylguanosine as a Z-Form RNA Stabilizer for Structural and Functional Study of Z-RNA. *Molecules (Basel, Switzerland)* **2018**, *23* (10). DOI: 10.3390/molecules23102572. Published Online: Oct. 9, 2018.

(334) Xu, Y.; Ikeda, R.; Sugiyama, H. 8-Methylguanosine: a powerful Z-DNA stabilizer. *J. Am. Chem. Soc.* **2003**, *125* (44), 13519–13524. DOI: 10.1021/ja036233i.

(335) Hardin, C. C.; Zarling, D. A.; Puglisi, J. D.; Trulson, M. O.; Davis, P. W.; Tinoco, I. Stabilization of Z-RNA by chemical bromination and its recognition by anti-Z-DNA antibodies. *Biochemistry* **1987**, *26* (16), 5191–5199. DOI: 10.1021/bi00390a044.

(336) Nordheim, A.; Tesser, P.; Azorin, F.; Kwon, Y. H.; Möller, A.; Rich, A. Isolation of *Drosophila* proteins that bind selectively to left-handed Z-DNA. *Proceedings of the National Academy of Sciences of the United States of America* **1982**, *79* (24), 7729–7733. DOI: 10.1073/pnas.79.24.7729.

(337) Herbert, A.; Alfken, J.; Kim, Y. G.; Mian, I. S.; Nishikura, K.; Rich, A. A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94* (16), 8421–8426. DOI: 10.1073/pnas.94.16.8421.

(338) Herbert, A.; Lowenhaupt, K.; Spitzner, J.; Rich, A. Chicken double-stranded RNA adenosine deaminase has apparent specificity for Z-DNA. *Proceedings of the National Academy of Sciences of the United States of America* **1995**, *92* (16), 7550–7554. DOI: 10.1073/pnas.92.16.7550.

(339) Tang, Q. Z-nucleic acids: Uncovering the functions from past to present. *European journal of immunology* **2022**, *52* (11), 1700–1711. DOI: 10.1002/eji.202249968. Published Online: Oct. 10, 2022.

(340) Schwartz, T.; Behlke, J.; Lowenhaupt, K.; Heinemann, U.; Rich, A. Structure of the DLM-1-Z-DNA complex reveals a conserved family of Z-DNA-binding proteins. *Nature structural biology* **2001**, *8* (9), 761–765. DOI: 10.1038/nsb0901-761.

(341) Fu, Y.; Comella, N.; Tognazzi, K.; Brown, L. F.; Dvorak, H. F.; Kocher, O. Cloning of DLM-1, a novel gene that is up-regulated in activated macrophages, using RNA differential display. *Gene* **1999**, *240* (1), 157–163. DOI: 10.1016/s0378-1119(99)00419-9.

(342) Zhong, Y.; Zhong, X.; Qiao, L.; Wu, H.; Liu, C.; Zhang, T. Z $\alpha$  domain proteins mediate the immune response. *Front. Immunol.* **2023**, *14*. DOI: 10.3389/fimmu.2023.1241694.

(343) Deigendesch, N.; Koch-Nolte, F.; Rothenburg, S. ZBP1 subcellular localization and association with stress granules is controlled by its Z-DNA binding domains. *Nucleic acids research* **2006**, *34* (18), 5007–5020. DOI: 10.1093/nar/gkl575. Published Online: Sep. 20, 2006.

(344) Hao, Y.; Yang, B.; Yang, J.; Shi, X.; Yang, X.; Zhang, D.; Zhao, D.; Yan, W.; Chen, L.; Zheng, H.; Zhang, K.; Liu, X. ZBP1: A Powerful Innate Immune Sensor and Double-Edged Sword in Host Immunity. *International Journal of Molecular Sciences* **2022**, *23* (18). DOI: 10.3390/ijms231810224. Published Online: Sep. 6, 2022.

(345) Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; Žídek, A.; Green, T.; Tunyasuvunakool, K.; Petersen, S.; Jumper, J.; Clancy, E.; Green, R.; Vora, A.; Lutfi, M.; Figurnov, M.; Cowie, A.; Hobbs, N.; Kohli, P.; Kleywegt, G.; Birney, E.; Hassabis, D.; Velankar, S. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic acids research* **2022**, *50* (D1), D439–D444. DOI: 10.1093/nar/gkab1061.

(346) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.;

- Berghammer, T.; Bodenstern, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, 596 (7873), 583–589. DOI: 10.1038/s41586-021-03819-2. Published Online: Jul. 15, 2021.
- (347) Schwartz, T.; Behlke, J.; Lowenhaupt, K.; Heinemann, U.; Rich, A. *Crystal Structure of the DNA-Binding Domain Zalpha of DLM-1 Bound to Z-DNA*, 2001. DOI: 10.2210/pdb1j75/pdb.
- (348) Rebsamen, M.; Heinz, L. X.; Meylan, E.; Michallet, M.-C.; Schroder, K.; Hofmann, K.; Vazquez, J.; Benedict, C. A.; Tschopp, J. DAI/ZBP1 recruits RIP1 and RIP3 through RIP homotypic interaction motifs to activate NF-kappaB. *EMBO reports* **2009**, 10 (8), 916–922. DOI: 10.1038/embor.2009.109. Published Online: Jul. 10, 2009.
- (349) Takaoka, A.; Wang, Z.; Choi, M. K.; Yanai, H.; Negishi, H.; Ban, T.; Lu, Y.; Miyagishi, M.; Kodama, T.; Honda, K.; Ohba, Y.; Taniguchi, T. DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* **2007**, 448 (7152), 501–505. DOI: 10.1038/nature06013. Published Online: Jul. 8, 2007.
- (350) Zhang, T.; Yin, C.; Boyd, D. F.; Quarato, G.; Ingram, J. P.; Shubina, M.; Ragan, K. B.; Ishizuka, T.; Crawford, J. C.; Tummers, B.; Rodriguez, D. A.; Xue, J.; Peri, S.; Kaiser, W. J.; López, C. B.; Xu, Y.; Upton, J. W.; Thomas, P. G.; Green, D. R.; Balachandran, S. Influenza Virus Z-RNAs Induce ZBP1-Mediated Necroptosis. *Cell* **2020**, 180 (6), 1115-1129.e13. DOI: 10.1016/j.cell.2020.02.050.
- (351) Nogusa, S.; Thapa, R. J.; Dillon, C. P.; Liedmann, S.; Oguin, T. H.; Ingram, J. P.; Rodriguez, D. A.; Kosoff, R.; Sharma, S.; Sturm, O.; Verbist, K.; Gough, P. J.; Bertin, J.; Hartmann, B. M.; Sealfon, S. C.; Kaiser, W. J.; Mocarski, E. S.; López, C. B.; Thomas, P. G.; Oberst, A.; Green, D. R.; Balachandran, S. RIPK3 Activates Parallel Pathways of MLKL-Driven Necroptosis and FADD-Mediated Apoptosis to Protect against Influenza A Virus. *Cell host & microbe* **2016**, 20 (1), 13–24. DOI: 10.1016/j.chom.2016.05.011. Published Online: Jun. 16, 2016.
- (352) Lin, J.; Kumari, S.; Kim, C.; Van, T.-M.; Wachsmuth, L.; Polykratis, A.; Pasparakis, M. RIPK1 counteracts ZBP1-mediated necroptosis to inhibit inflammation. *Nature* **2016**, 540 (7631), 124–128. DOI: 10.1038/nature20558. Published Online: Nov. 7, 2016.
- (353) Koehler, H.; Cotsmire, S.; Zhang, T.; Balachandran, S.; Upton, J. W.; Langland, J.; Kalman, D.; Jacobs, B. L.; Mocarski, E. S. Vaccinia virus E3 prevents sensing of Z-RNA to block ZBP1-dependent necroptosis. *Cell host & microbe* **2021**, 29 (8), 1266-1276.e5. DOI: 10.1016/j.chom.2021.05.009. Published Online: Jun. 29, 2021.
- (354) Guo, H.; Gilley, R. P.; Fisher, A.; Lane, R.; Landsteiner, V. J.; Ragan, K. B.; Dovey, C. M.; Carette, J. E.; Upton, J. W.; Mocarski, E. S.; Kaiser, W. J. Species-independent contribution of ZBP1/DAI/DLM-1-triggered necroptosis in host defense against HSV1. *Cell death & disease* **2018**, 9 (8), 816. DOI: 10.1038/s41419-018-0868-3. Published Online: Jul. 26, 2018.
- (355) Maelfait, J.; Liverpool, L.; Bridgeman, A.; Ragan, K. B.; Upton, J. W.; Rehwinkel, J. Sensing of viral and endogenous RNA by ZBP1/DAI induces necroptosis. *The EMBO journal* **2017**, 36 (17), 2529–2543. DOI: 10.15252/embj.201796476. Published Online: Jul. 17, 2017.
- (356) Banoth, B.; Tuladhar, S.; Karki, R.; Sharma, B. R.; Briard, B.; Kesavardhana, S.; Burton, A.; Kanneganti, T.-D. ZBP1 promotes fungi-induced inflammasome activation and pyroptosis, apoptosis, and necroptosis (PANoptosis). *The Journal of biological chemistry* **2020**, 295 (52), 18276–18283. DOI: 10.1074/jbc.RA120.015924. Published Online: Oct. 27, 2020.
- (357) Sridharan, H.; Ragan, K. B.; Guo, H.; Gilley, R. P.; Landsteiner, V. J.; Kaiser, W. J.; Upton, J. W. Murine cytomegalovirus IE3-dependent transcription is required for DAI/ZBP1-mediated necroptosis. *EMBO reports* **2017**, 18 (8), 1429–1441. DOI: 10.15252/embr.201743947. Published Online: Jun. 12, 2017.
- (358) Karki, R.; Lee, S.; Mall, R.; Pandian, N.; Wang, Y.; Sharma, B. R.; Malireddi, R. S.; Yang, D.; Trifkovic, S.; Steele, J. A.; Connelly, J. P.; Vishwanath, G.; Sasikala, M.; Reddy, D. N.; Vogel, P.; Pruetz-Miller, S. M.; Webby, R.; Jonsson, C. B.; Kanneganti, T.-D. ZBP1-dependent inflammatory cell death, PANoptosis, and cytokine storm disrupt IFN therapeutic efficacy

during coronavirus infection. *Science immunology* **2022**, *7* (74), eabo6294. DOI: 10.1126/sciimmunol.abo6294. Published Online: Aug. 26, 2022.

(359) Jiao, H.; Wachsmuth, L.; Kumari, S.; Schwarzer, R.; Lin, J.; Eren, R. O.; Fisher, A.; Lane, R.; Young, G. R.; Kassiotis, G.; Kaiser, W. J.; Pasparakis, M. Z-nucleic-acid sensing triggers ZBP1-dependent necroptosis and inflammation. *Nature* **2020**, *580* (7803), 391–395. DOI: 10.1038/s41586-020-2129-8. Published Online: Mar. 25, 2020.

(360) Karki, R.; Sundaram, B.; Sharma, B. R.; Lee, S.; Malireddi, R. K. S.; Nguyen, L. N.; Christgen, S.; Zheng, M.; Wang, Y.; Samir, P.; Neale, G.; Vogel, P.; Kanneganti, T.-D. ADAR1 restricts ZBP1-mediated immune response and PANoptosis to promote tumorigenesis. *Cell reports* **2021**, *37* (3), 109858. DOI: 10.1016/j.celrep.2021.109858.

(361) Baik, J. Y.; Liu, Z.; Jiao, D.; Kwon, H.-J.; Yan, J.; Kadigamuwa, C.; Choe, M.; Lake, R.; Kruhlak, M.; Tandon, M.; Cai, Z.; Choksi, S.; Liu, Z.-G. ZBP1 not RIPK1 mediates tumor necroptosis in breast cancer. *Nature communications* **2021**, *12* (1), 2666. DOI: 10.1038/s41467-021-23004-3. Published Online: May. 11, 2021.

(362) Hannah Depmeier. *Towards the Synthesis of L- $\alpha$ -Threofuranosyl Nucleoside Triphosphates and Oligonucleotides with an Expanded Genetic Alphabet*, 2019.

(363) Merkel, M.; Arndt, S.; Ploschik, D.; Cserép, G. B.; Wenge, U.; Kele, P.; Wagenknecht, H.-A. Scope and Limitations of Typical Copper-Free Bioorthogonal Reactions with DNA: Reactive 2'-Deoxyuridine Triphosphates for Postsynthetic Labeling. *The Journal of organic chemistry* **2016**, *81* (17), 7527–7538. DOI: 10.1021/acs.joc.6b01205. Published Online: Aug. 24, 2016.

(364) Yang, J.; Liang, Y.; Šečkutė, J.; Houk, K. N.; Devaraj, N. K. Synthesis and reactivity comparisons of 1-methyl-3-substituted cyclopropene mini-tags for tetrazine bioorthogonal reactions. *Chemistry A European J* **2014**, *20* (12), 3365–3375. DOI: 10.1002/chem.201304225. Published Online: Feb. 24, 2014.

(365) Ravasco, J. M. J. M.; Monteiro, C. M.; Trindade, A. F. Cyclopropenes: a new tool for the study of biological systems. *Org. Chem. Front.* **2017**, *4* (6), 1167–1198. DOI: 10.1039/C7QO00054E.

(366) Niimi, K.; Mori, H.; Miyazaki, E.; Osaka, I.; Kakizoe, H.; Takimiya, K.; Adachi, C. 2,2'-Binaphtho[2,3-b]furanyl: a versatile organic semiconductor with a furan-furan junction. *Chemical communications (Cambridge, England)* **2012**, *48* (47), 5892–5894. DOI: 10.1039/C2CC31960H. Published Online: May. 8, 2012.

(367) Sau, S. P.; Fahmi, N. E.; Liao, J.-Y.; Bala, S.; Chaput, J. C. A Scalable Synthesis of  $\alpha$ -L-Threose Nucleic Acid Monomers. *The Journal of organic chemistry* **2016**, *81* (6), 2302–2307. DOI: 10.1021/acs.joc.5b02768. Published Online: Feb. 26, 2016.

(368) Dumbre, S. G.; Jang, M.-Y.; Herdewijn, P. Synthesis of  $\alpha$ -L-threose nucleoside phosphonates via regioselective sugar protection. *The Journal of organic chemistry* **2013**, *78* (14), 7137–7144. DOI: 10.1021/jo400907g. Published Online: Jul. 3, 2013.

(369) Lavergne, T.; Lamichhane, R.; Malyshev, D. A.; Li, Z.; Li, L.; Sperling, E.; Williamson, J. R.; Millar, D. P.; Romesberg, F. E. FRET Characterization of Complex Conformational Changes in a Large 16S Ribosomal RNA Fragment Site-Specifically Labeled Using Unnatural Base Pairs. *ACS chemical biology* **2016**, *11* (5), 1347–1353. DOI: 10.1021/acscchembio.5b00952. Published Online: Mar. 4, 2016.

(370) Srivatsan, S. G.; Tor, Y. Synthesis and enzymatic incorporation of a fluorescent pyrimidine ribonucleotide. *Nature protocols* **2007**, *2* (6), 1547–1555. DOI: 10.1038/nprot.2007.222.

(371) Yoshikawa, M.; Kato, T.; Takenishi, T. A novel method for phosphorylation of nucleosides to 5'-nucleotides. *Tetrahedron letters* **1967**, *50*, 5065–5068. DOI: 10.1016/S0040-4039(01)89915-9.

(372) Ludwig, J. A new route to nucleoside 5'-triphosphates. *Acta biochimica et biophysica; Academiae Scientiarum Hungaricae* **1981**, *16* (3-4), 131–133.

(373) Xia, Y.; Xu, Q.; Herron, A. N.; Wang, W.; Shao, L.; Li, G.; an Liu; Cheng, J. Research progress in natural N-glycosides and synthetic methodologies for the formation of N-glycosidic bonds. *Sci Tradit Chin Med* **2024**, 2 (2), 82–94. DOI: 10.1097/st9.0000000000000035.

(374) Zou, R.; Robins, M. J. High-yield regioselective synthesis of 9-glycosyl guanine nucleosides and analogues via coupling with 2- N -acetyl-6- O -diphenylcarbamoylguanine. *Can. J. Chem.* **1987**, 65 (6), 1436–1437. DOI: 10.1139/v87-243.

(375) Beyer, H. M.; Gonschorek, P.; Samodelov, S. L.; Meier, M.; Weber, W.; Zurbriggen, M. D. AQUA Cloning: A Versatile and Simple Enzyme-Free Cloning Approach. *PloS one* **2015**, 10 (9), e0137652. DOI: 10.1371/journal.pone.0137652. Published Online: Sep. 11, 2015.

(376) Nikoomanzar, A.; Dunn, M. R.; Chaput, J. C. Engineered Polymerases with Altered Substrate Specificity: Expression and Purification. *Current protocols in nucleic acid chemistry* **2017**, 69, 4.75.1-4.75.20. DOI: 10.1002/cpnc.33. Published Online: Jun. 19, 2017.

(377) Tabor, S.; Richardson, C. C. Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and Escherichia coli DNA polymerase I. *Proceedings of the National Academy of Sciences of the United States of America* **1989**, 86 (11), 4076–4080.

(378) Berendsen, H.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **1995**, 91 (1-3), 43–56. DOI: 10.1016/0010-4655(95)00042-E.

(379) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; Mackerell, A. D. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone  $\phi$ ,  $\psi$  and side-chain  $\chi(1)$  and  $\chi(2)$  dihedral angles. *Journal of chemical theory and computation* **2012**, 8 (9), 3257–3273. DOI: 10.1021/ct300400x. Published Online: Jul. 18, 2012.

(380) Domnick, C.; Eggert, F.; Wuebben, C.; Bornewasser, L.; Hagelueken, G.; Schiemann, O.; Kath-Schorr, S. EPR Distance Measurements on Long Non-coding RNAs Empowered by Genetic Alphabet Expansion Transcription. *Angewandte Chemie (International ed. in English)* **2020**, 59 (20), 7891–7896. DOI: 10.1002/anie.201916447. Published Online: Mar. 13, 2020.

(381) Anhäuser, L.; Klöcker, N.; Muttach, F.; Mäsing, F.; Špaček, P.; Studer, A.; Rentmeister, A. A Benzophenone-Based Photocaging Strategy for the N7 Position of Guanosine. *Angew. Chem. Int. Ed.* **2020**, 59 (8), 3161–3165. DOI: 10.1002/anie.201914573. Published Online: Dec. 30, 2019.

(382) Gobbi, S.; Cavalli, A.; Negri, M.; Schewe, K. E.; Belluti, F.; Piazza, L.; Hartmann, R. W.; Recanatini, M.; Bisi, A. Imidazolylmethylbenzophenones as highly potent aromatase inhibitors. *Journal of medicinal chemistry* **2007**, 50 (15), 3420–3422. DOI: 10.1021/jm0702938. Published Online: Jun. 22, 2007.

(383) Tayama, E.; Kimura, H. Asymmetric Sommelet-Hauser rearrangement of N-benzylic ammonium salts. *Angew. Chem. Int. Ed.* **2007**, 46 (46), 8869–8871. DOI: 10.1002/anie.200703832.

(384) Ortiz-Trankina, L. N.; Crain, J.; Williams, C.; Mack, J. Developing benign syntheses using ion pairs via solvent-free mechanochemistry. *Green Chem.* **2020**, 22 (11), 3638–3642. DOI: 10.1039/D0GC01116A.

(385) Pautus, S.; Sehr, P.; Lewis, J.; Fortuné, A.; Wolkerstorfer, A.; Szolar, O.; Guilligay, D.; Lunardi, T.; Décout, J.-L.; Cusack, S. New 7-methylguanine derivatives targeting the influenza polymerase PB2 cap-binding domain. *Journal of medicinal chemistry* **2013**, 56 (21), 8915–8930. DOI: 10.1021/jm401369y. Published Online: Nov. 6, 2013.

(386) Hamm, M. L.; Billig, K. Synthesis, oligonucleotide incorporation and base pair stability of 7-methyl-8-oxo-2'-deoxyguanosine. *Organic & biomolecular chemistry* **2006**, 4 (22), 4068–4070. DOI: 10.1039/b612597b.

(387) Sarma, D.; Majumdar, B.; Chaput, J. C. Alternative Synthesis of  $\alpha$ -l-Threofuranosyl Guanosine 3'-triphosphate. *Synlett* **2024**, 35 (06), 706–710. DOI: 10.1055/s-0041-1738452.

- (388) Rubin, M.; Rubina, M.; Gevorgyan, V. Transition metal chemistry of cyclopropenes and cyclopropanes. *Chemical reviews* **2007**, *107* (7), 3117–3179. DOI: 10.1021/cr050988l.
- (389) Hollenstein, M. Nucleoside triphosphates--building blocks for the modification of nucleic acids. *Molecules (Basel, Switzerland)* **2012**, *17* (11), 13569–13591. DOI: 10.3390/molecules171113569. Published Online: Nov. 15, 2012.
- (390) Knoblauch, B. 5-Substituted UTP derivatives as P2Y2 receptor agonists. *European Journal of Medicinal Chemistry* **1999**, *34* (10), 809–824. DOI: 10.1016/S0223-5234(99)00211-1.
- (391) Cozzone, P. J.; Jardetzky, O. Phosphorus-31 Fourier transform nuclear magnetic resonance study of mononucleotides and dinucleotides. 1. Chemical shifts. *Biochemistry* **1976**, *15* (22), 4853–4859. DOI: 10.1021/bi00667a016.
- (392) Gillerman, I.; Fischer, B. An improved one-pot synthesis of nucleoside 5'-triphosphate analogues. *Nucleosides, nucleotides & nucleic acids* **2010**, *29* (3), 245–256. DOI: 10.1080/15257771003709569.
- (393) Borsenberger, V.; Kukwikila, M.; Howorka, S. Synthesis and enzymatic incorporation of modified deoxyuridine triphosphates. *Organic & biomolecular chemistry* **2009**, *7* (18), 3826–3835. DOI: 10.1039/b906956a. Published Online: Jul. 21, 2009.
- (394) El-Tayeb, A.; Qi, A.; Müller, C. E. Synthesis and structure-activity relationships of uracil nucleotide derivatives and analogues as agonists at human P2Y2, P2Y4, and P2Y6 receptors. *Journal of medicinal chemistry* **2006**, *49* (24), 7076–7087. DOI: 10.1021/jm060848j.
- (395) Shepard, S. M.; Jessen, H. J.; Cummins, C. C. Beyond Triphosphates: Reagents and Methods for Chemical Oligophosphorylation. *J. Am. Chem. Soc.* **2022**, *144* (17), 7517–7530. DOI: 10.1021/jacs.1c07990. Published Online: Apr. 26, 2022.
- (396) Crauste, C.; Périgaud, C.; Peyrottes, S. Insights into the soluble PEG-supported synthesis of cytosine-containing nucleoside 5'-mono-, di-, and triphosphates. *The Journal of organic chemistry* **2009**, *74* (23), 9165–9172. DOI: 10.1021/jo901931z.
- (397) Ripp, A.; Singh, J.; Jessen, H. J. Rapid Synthesis of Nucleoside Triphosphates and Analogues. *Current protocols in nucleic acid chemistry* **2020**, *81* (1), e108. DOI: 10.1002/cpnc.108.
- (398) Singh, J.; Ripp, A.; Haas, T. M.; Qiu, D.; Keller, M.; Wender, P. A.; Siegel, J. S.; Baldrige, K. K.; Jessen, H. J. Synthesis of Modified Nucleoside Oligophosphates Simplified: Fast, Pure, and Protecting Group Free. *J. Am. Chem. Soc.* **2019**, *141* (38), 15013–15017. DOI: 10.1021/jacs.9b08273.
- (399) Ludwig, J.; Eckstein, F. Rapid and efficient synthesis of nucleoside 5'-O-(1-thiotriphosphates), 5'-triphosphates and 2',3'-cyclophosphorothioates using 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one. *J. Org. Chem.* **1989**, *54* (3), 631–635. DOI: 10.1021/jo00264a024.
- (400) Rachwalak, M.; Romanowska, J.; Sobkowski, M.; Stawinski, J. Nucleoside Di- and Triphosphates as a New Generation of Anti-HIV Pronucleotides. Chemical and Biological Aspects. *Applied Sciences* **2021**, *11* (5), 2248. DOI: 10.3390/app11052248.
- (401) KARCHER, S. J. Recombinant DNA Cloning. In *Molecular Biology*, Elsevier, 1995; pp 45–134. DOI: 10.1016/B978-012397720-5.50036-0.
- (402) Patterson, D. M.; Nazarova, L. A.; Xie, B.; Kamber, D. N.; Prescher, J. A. Functionalized cyclopropenes as bioorthogonal chemical reporters. *J. Am. Chem. Soc.* **2012**, *134* (45), 18638–18643. DOI: 10.1021/ja3060436. Published Online: Nov. 1, 2012.
- (403) Wu, H.; Devaraj, N. K. Advances in Tetrazine Bioorthogonal Chemistry Driven by the Synthesis of Novel Tetrazines and Dienophiles. *Accounts of chemical research* **2018**, *51* (5), 1249–1259. DOI: 10.1021/acs.accounts.8b00062. Published Online: Apr. 11, 2018.
- (404) Kajiyama, T.; Kuwahara, M.; Goto, M.; Kambara, H. Optimization of pyrosequencing reads by superior successive incorporation efficiency of improved 2'-deoxyadenosine-5'-triphosphate analogs. *Analytical biochemistry* **2011**, *416* (1), 8–17. DOI: 10.1016/j.ab.2011.04.039. Published Online: May. 4, 2011.



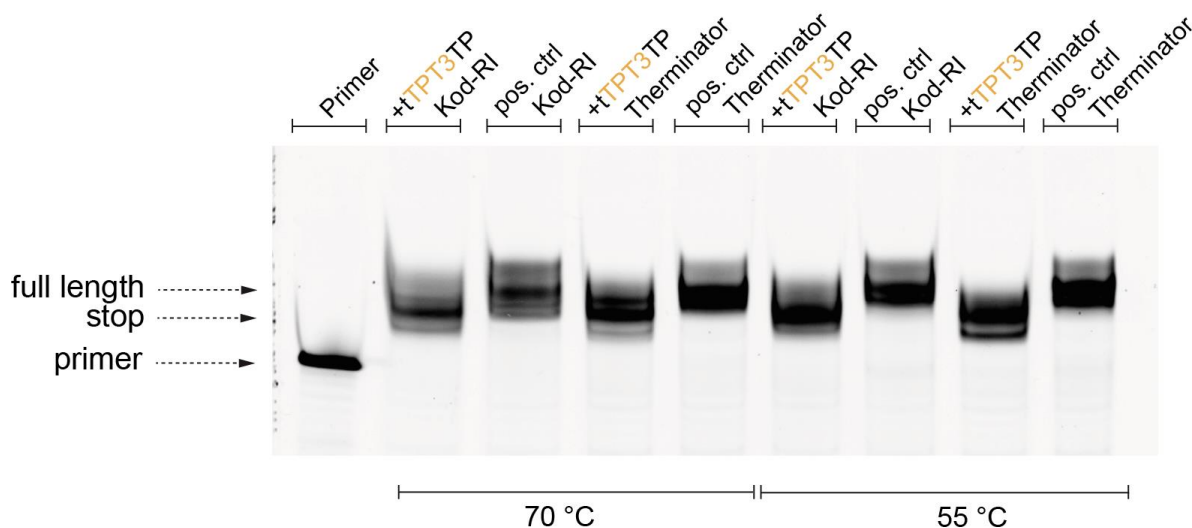
- (405) Ganz, D.; Geng, P.; Wagenknecht, H.-A. The Efficiency of Metabolic Labeling of DNA by Diels-Alder Reactions with Inverse Electron Demand: Correlation with the Size of Modified 2'-Deoxyuridines. *ACS chemical biology* **2023**, *18* (5), 1054–1059. DOI: 10.1021/acscchembio.3c00079. Published Online: Mar. 15, 2023.
- (406) Thapa, R. J.; Ingram, J. P.; Ragan, K. B.; Nogusa, S.; Boyd, D. F.; Benitez, A. A.; Sridharan, H.; Kosoff, R.; Shubina, M.; Landsteiner, V. J.; Andrade, M.; Vogel, P.; Sigal, L. J.; tenOever, B. R.; Thomas, P. G.; Upton, J. W.; Balachandran, S. DAI Senses Influenza A Virus Genomic RNA and Activates RIPK3-Dependent Cell Death. *Cell host & microbe* **2016**, *20* (5), 674–681. DOI: 10.1016/j.chom.2016.09.014. Published Online: Oct. 13, 2016.
- (407) Kuriakose, T.; Man, S. M.; Malireddi, R. K. S.; Karki, R.; Kesavardhana, S.; Place, D. E.; Neale, G.; Vogel, P.; Kanneganti, T.-D. ZBP1/DAI is an innate sensor of influenza virus triggering the NLRP3 inflammasome and programmed cell death pathways. *Science immunology* **2016**, *1* (2). DOI: 10.1126/sciimmunol.aag2045. Published Online: Aug. 12, 2016.
- (408) Ha, S. C.; Kim, D.; Hwang, H.-Y.; Rich, A.; Kim, Y.-G.; Kim, K. K. The crystal structure of the second Z-DNA binding domain of human DAI (ZBP1) in complex with Z-DNA reveals an unusual binding mode to Z-DNA. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105* (52), 20671–20676. DOI: 10.1073/pnas.0810463106. Published Online: Dec. 18, 2008.
- (409) Nichols, P. J.; Bevers, S.; Henen, M.; Kieft, J. S.; Vicens, Q.; Vögeli, B. Recognition of non-CpG repeats in Alu and ribosomal RNAs by the Z-RNA binding domain of ADAR1 induces A-Z junctions. *Nature communications* **2021**, *12* (1), 793. DOI: 10.1038/s41467-021-21039-0. Published Online: Feb. 4, 2021.
- (410) Placido, D.; Brown, B. A.; Lowenhaupt, K.; Rich, A.; Athanasiadis, A. A left-handed RNA double helix bound by the Z alpha domain of the RNA-editing enzyme ADAR1. *Structure (London, England : 1993)* **2007**, *15* (4), 395–404. DOI: 10.1016/j.str.2007.03.001.
- (411) Hellman, L. M.; Fried, M. G. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature protocols* **2007**, *2* (8), 1849–1861. DOI: 10.1038/nprot.2007.249.
- (412) Jerabek-Willemsen, M.; André, T.; Wanner, R.; Roth, H. M.; Duhr, S.; Baaske, P.; Breitsprecher, D. MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure* **2014**, *1077*, 101–113. DOI: 10.1016/j.molstruc.2014.03.009.
- (413) Klán, P.; Šolomek, T.; Bochet, C. G.; Blanc, A.; Givens, R.; Rubina, M.; Popik, V.; Kostikov, A.; Wirz, J. Photoremovable protecting groups in chemistry and biology: reaction mechanisms and efficacy. *Chemical reviews* **2013**, *113* (1), 119–191. DOI: 10.1021/cr300177k. Published Online: Dec. 21, 2012.
- (414) Fulmer, G. R.; Miller, A. J. M.; Sherden, N. H.; Gottlieb, H. E.; Nudelman, A.; Stoltz, B. M.; Bercaw, J. E.; Goldberg, K. I. NMR Chemical Shifts of Trace Impurities: Common Laboratory Solvents, Organics, and Gases in Deuterated Solvents Relevant to the Organometallic Chemist. *Organometallics* **2010**, *29* (9), 2176–2179. DOI: 10.1021/om100106e.
- (415) Yang, J.; Šečkutė, J.; Cole, C. M.; Devaraj, N. K. Live-cell imaging of cyclopropene tags with fluorogenic tetrazine cycloadditions. *Angew. Chem. Int. Ed.* **2012**, *51* (30), 7476–7479. DOI: 10.1002/anie.201202122. Published Online: Jun. 13, 2012.
- (416) Serebryany, V.; Beigelman, L. An efficient preparation of protected ribonucleosides for phosphoramidite RNA synthesis. *Tetrahedron letters* **2002**, *43* (11), 1983–1985. DOI: 10.1016/S0040-4039(02)00181-8.
- (417) Kibbe, W. A. OligoCalc: an online oligonucleotide properties calculator. *Nucleic acids research* **2007**, *35* (Web Server issue), W43-6. DOI: 10.1093/nar/gkm234. Published Online: Apr. 22, 2007.
- (418) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological

force fields. *Journal of computational chemistry* **2010**, *31* (4), 671–690. DOI: 10.1002/jcc.21367.

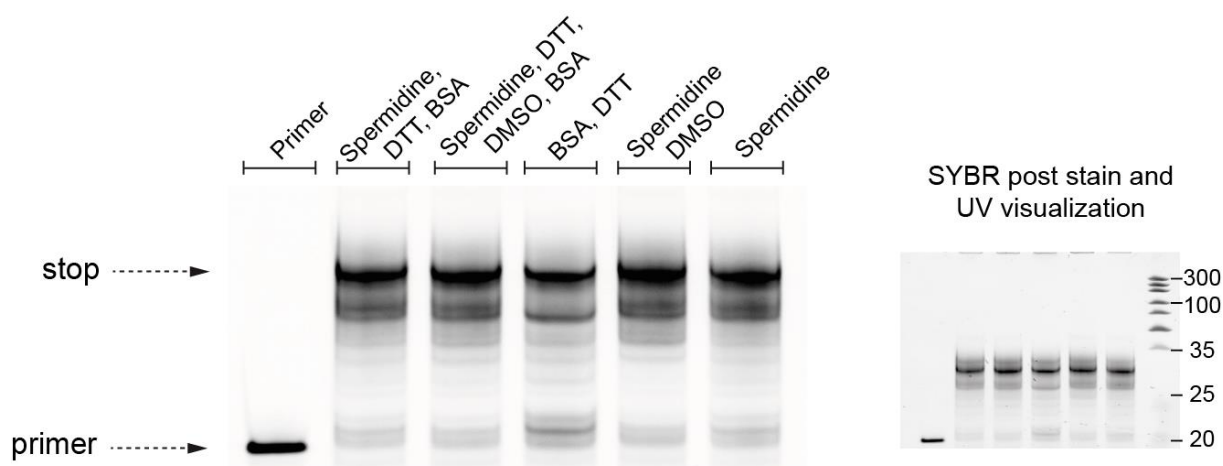
(419) Yu, W.; He, X.; Vanommeslaeghe, K.; Mackerell, A. D. Extension of the CHARMM General Force Field to sulfonyl-containing compounds and its utility in biomolecular simulations. *Journal of computational chemistry* **2012**, *33* (31), 2451–2468. DOI: 10.1002/jcc.23067. Published Online: Jul. 23, 2012.

## 7 Appendix

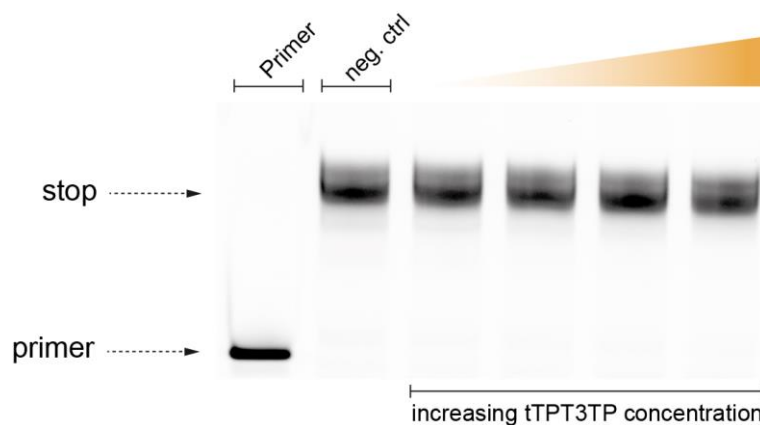
### 7.1 Supplementary figures and data



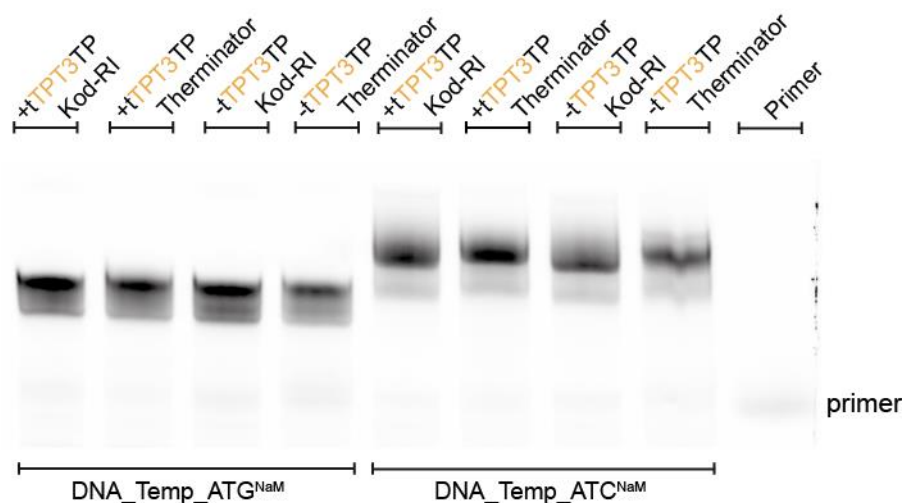
**Figure S 1: dPAGE results of Therminator and Kod-RI polymerase-mediated synthesis towards tTPT3-modified exTNA at different temperatures.** Fluorescence scan ( $\lambda_{\text{excitation}} = 477 \text{ nm}$ ,  $\lambda_{\text{emission}} = 535 \text{ nm}$ ) of the 20 % dPAGE analysis of primer extension assays using the DNA\_Temp\_shortATGC<sup>NaM</sup> DNA template, tNTPs, and tTPT3TP. The positive controls contain an unmodified template and tNTPs with exclusively canonical nucleobases. The primer only and positive controls serve as size markers.



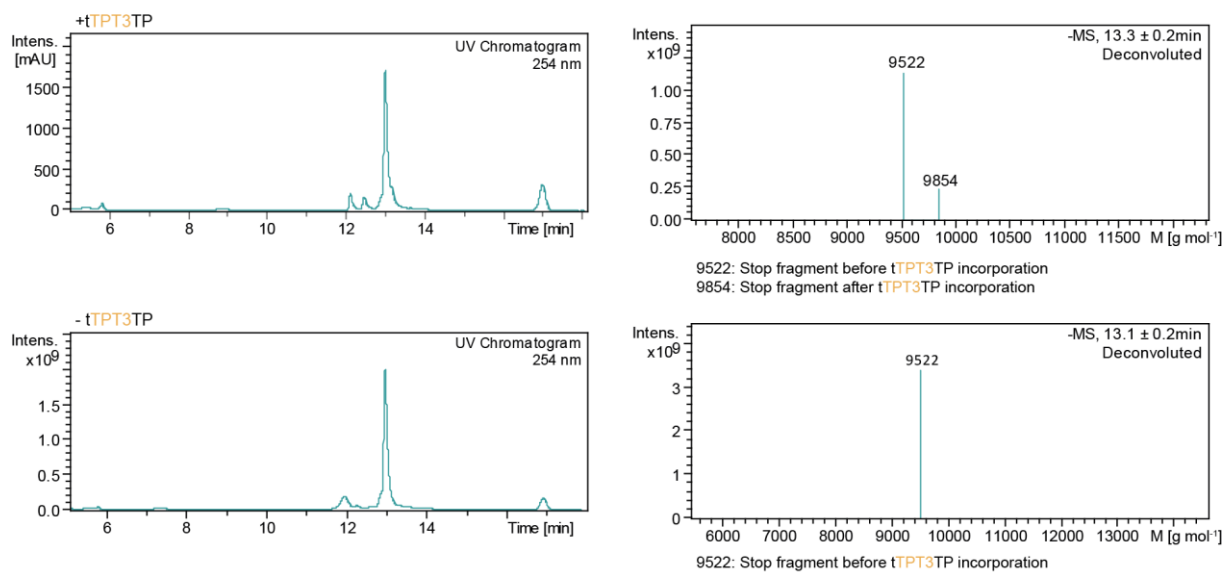
**Figure S 2: dPAGE results of Therminator polymerase-mediated synthesis towards tTPT3-modified exTNA using different additives.** Left: Fluorescence scan ( $\lambda_{\text{excitation}} = 477 \text{ nm}$ ,  $\lambda_{\text{emission}} = 535 \text{ nm}$ ) of the 20 % dPAGE analysis of primer extension assays using the DNA\_Temp\_midATGC<sup>NaM</sup> DNA template (41 nt in length, modification at position 34), tNTPs, and tTPT3TP. The additives were added to the PEx with the following final concentrations:  $100 \mu\text{g mL}^{-1}$  BSA,  $1 \text{ mM}$  DTT,  $2 \%$  ( $v/v$ ) DMSO,  $2 \text{ mM}$  spermidine. The primer serves as a size marker. The product size was determined using the GeneRuler ULR DNA ladder, followed by SYBR safe post-staining and UV visualization. The product is migrating below the 35 bp band of the ladder, indicating that only stop fragment is obtained by all PEx, irrespective of the additives used.



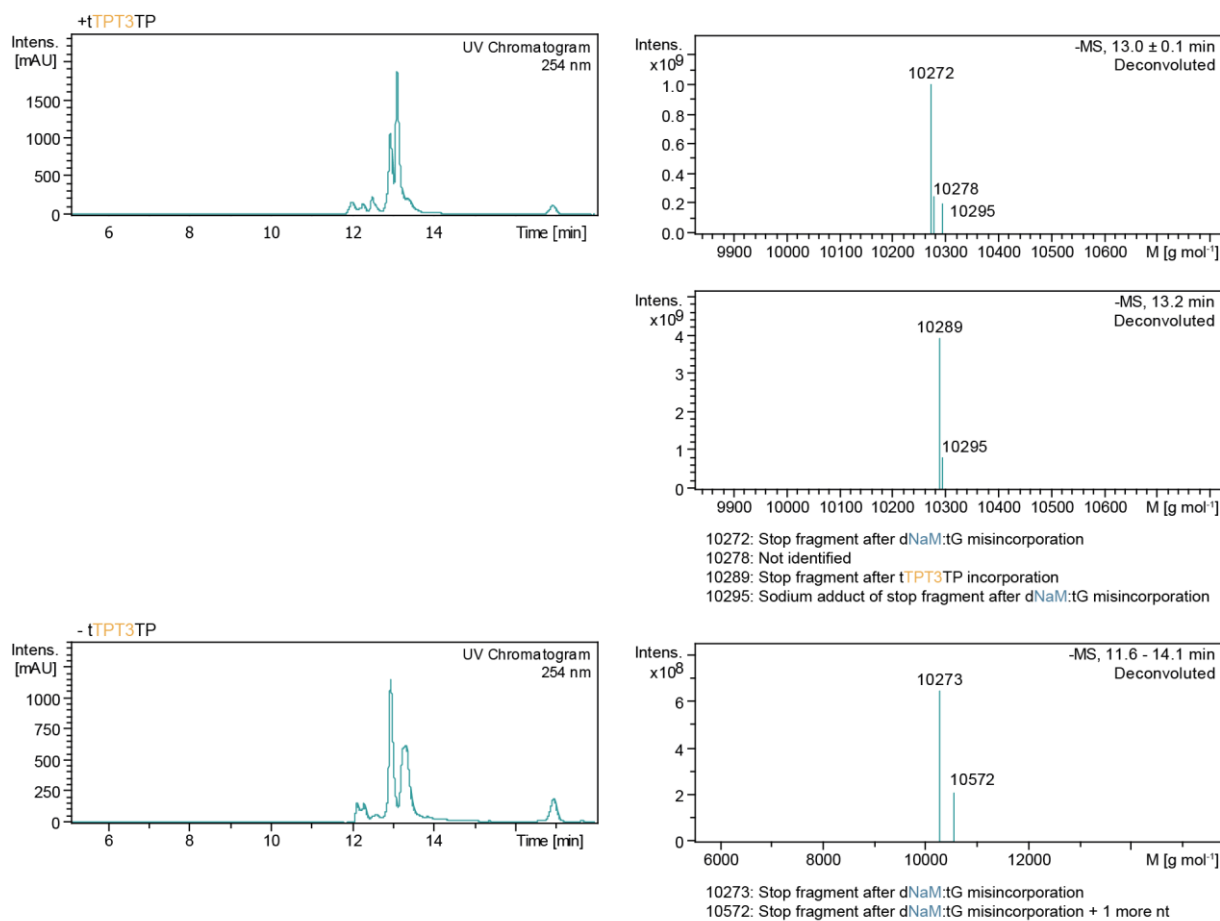
**Figure S 3: dPAGE results of Therminator polymerase-mediated synthesis towards TPT3-modified exTNA using different tTPT3TP concentrations (i.e. 0.25 mM, 0.5 mM, 0.75 mM, 1.0 mM).** Fluorescence scan ( $\lambda_{\text{excitation}} = 477 \text{ nm}$ ,  $\lambda_{\text{emission}} = 535 \text{ nm}$ ) of the 20 % dPAGE analysis of primer extension assays using the DNA\_Temp\_midATG<sup>NaM</sup> template, tNTPs and tTPT3TP. The negative control does not contain tTPT3TP resulting in a stop fragment at the site of the UB. The primer only and negative control serve as size markers.



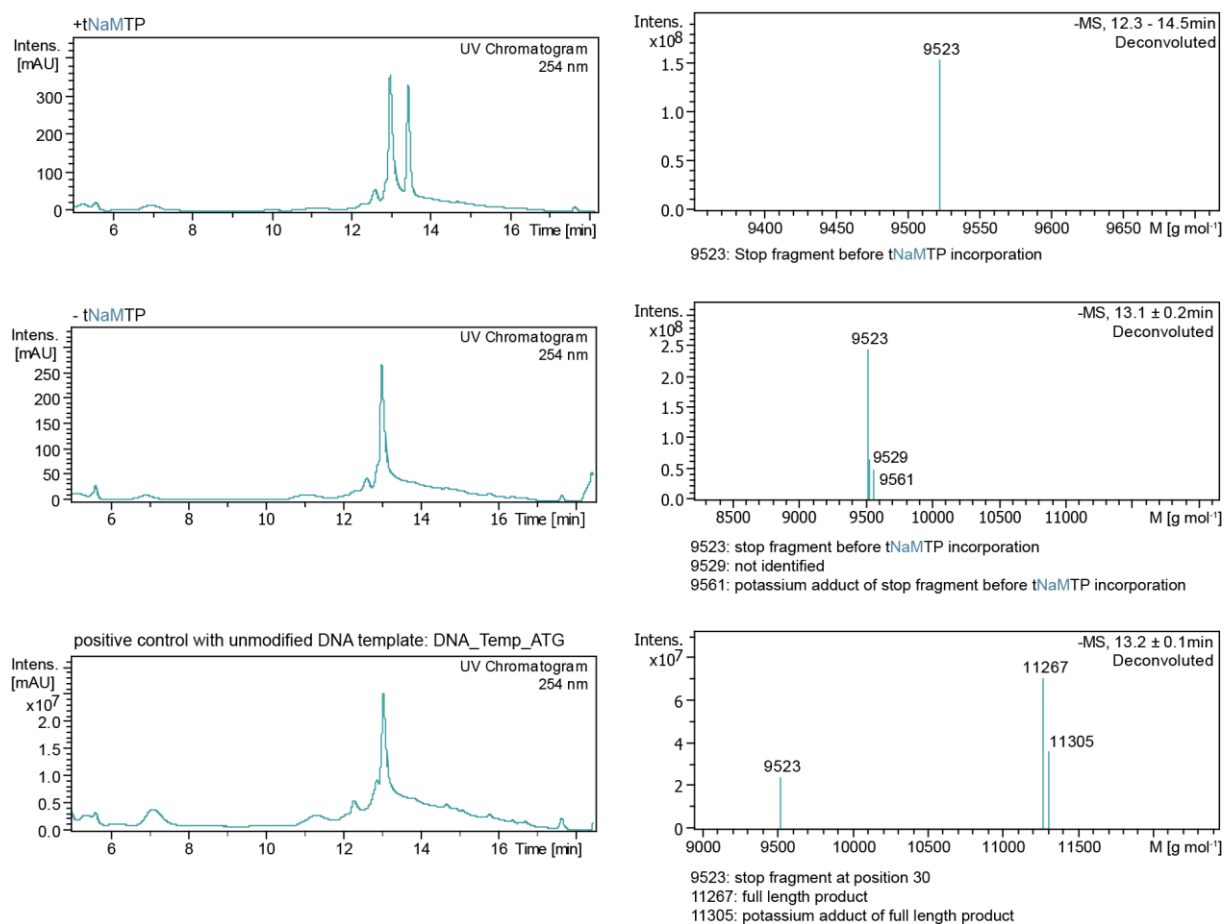
**Figure S 4: Complete gel of the 20 % dPAGE of PEx with Therminator and Kod-RI polymerases using DNA\_pTemp\_midATG<sup>NaM</sup> and DNA\_pTemp\_midATC<sup>NaM</sup>.** PEx were supplemented with the respective TNA nucleotides and tTPT3TP. The negative controls do not contain tTPT3TP, resulting in a stop fragment at the site of the UB. The primer only and negative controls serve as size markers. The dPAGE was imaged by fluorescence scanning ( $\lambda_{\text{excitation}} = 477 \text{ nm}$ ,  $\lambda_{\text{emission}} = 535 \text{ nm}$ ).



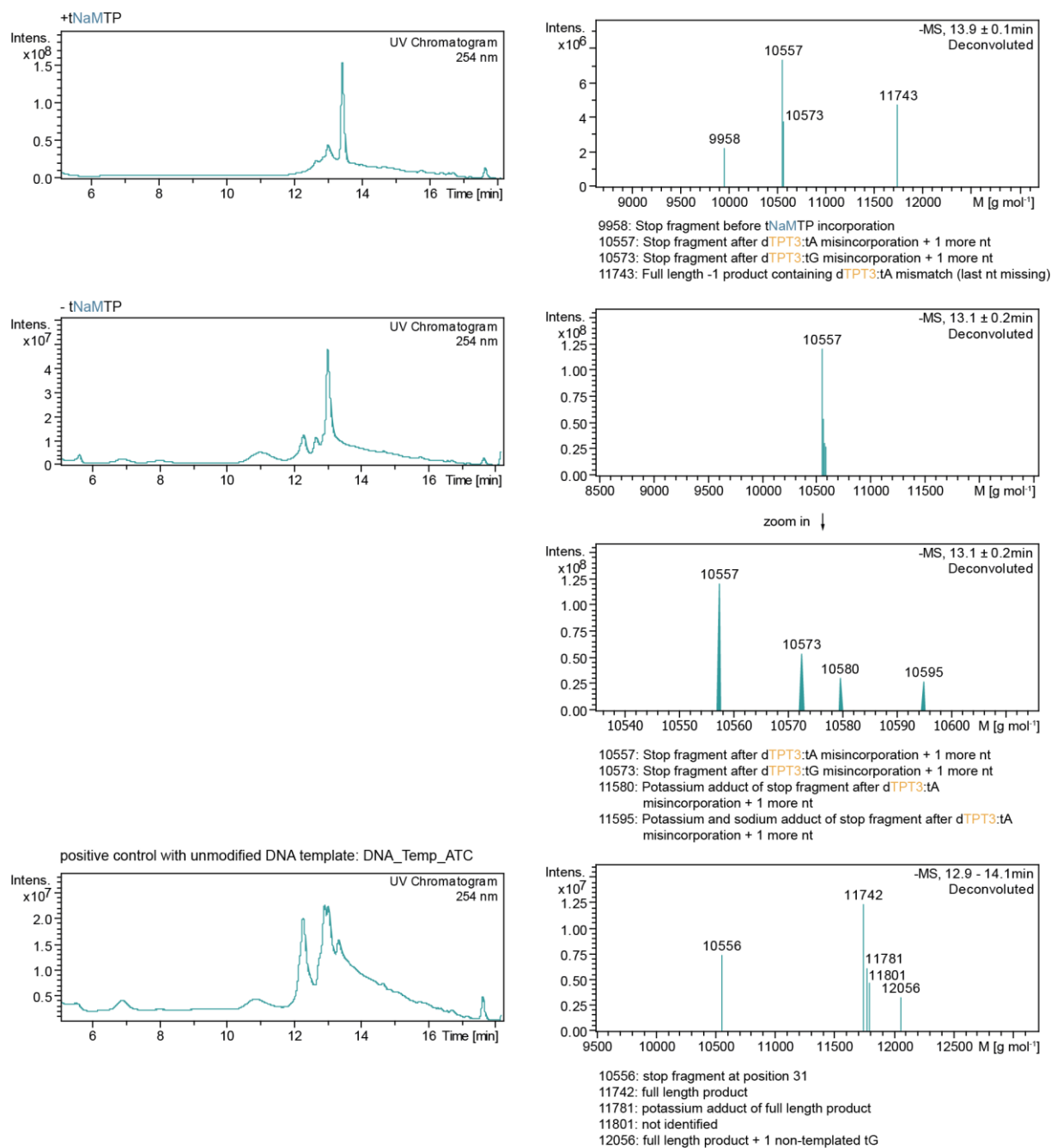
**Figure S 5: UV chromatograms and deconvoluted ESI- mass spectra of purified primer extension assays using Terminator polymerase employing DNA template DNA\_pTemp\_midATG<sup>NaM</sup> and tTPT3TP.** As canonical nucleotides, tTTP, tATP and tCTP were added to the PEx. In the negative control, tTPT3TP was omitted. Determined molecular weights of each detected peak were assigned underneath.



**Figure S 6: UV chromatograms and deconvoluted ESI mass spectra of purified primer extension assays using Therminator polymerase employing DNA template DNA\_pTemp\_midATC<sup>NaM</sup> and tTPT3TP.** As canonical nucleotides, tTTP, tATP and tGTP were added to the PEx. In the negative control, tTPT3TP was omitted. Determined molecular weights of each detected peak were assigned underneath.

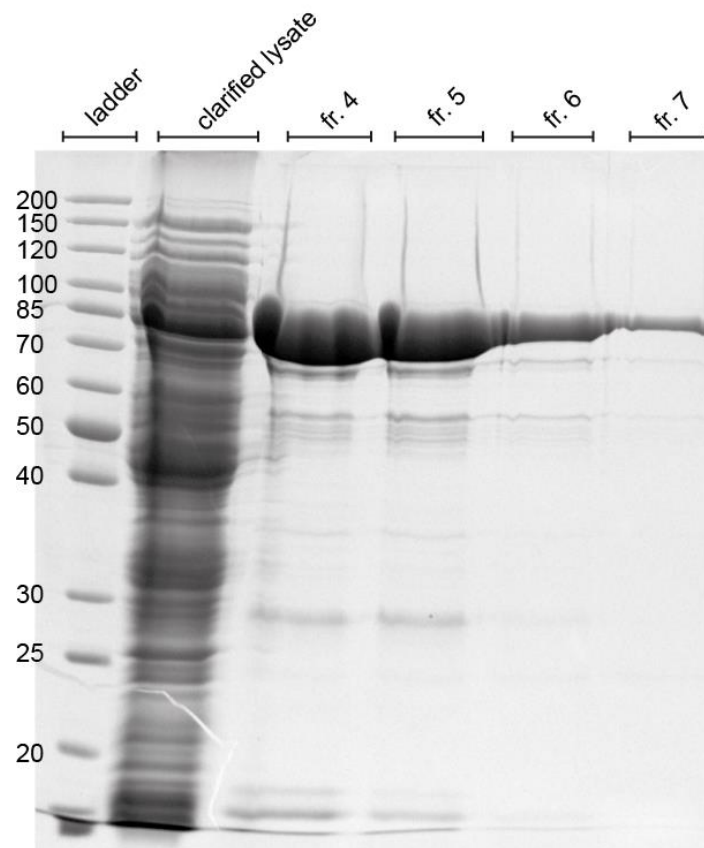


**Figure S 7: UV chromatograms and deconvoluted ESI<sup>-</sup> mass spectra of purified primer extension assays using Therminator polymerase employing DNA template DNA\_pTemp\_midATG<sup>TPT3</sup> and tNaMTP.** As canonical nucleotides, tTTP, tATP and tCTP were added to the PEx. In the negative control, tNaMTP was omitted. In the positive control, unmodified DNA\_pTemp\_midATG was used. Determined molecular weights of each detected peak were assigned underneath.

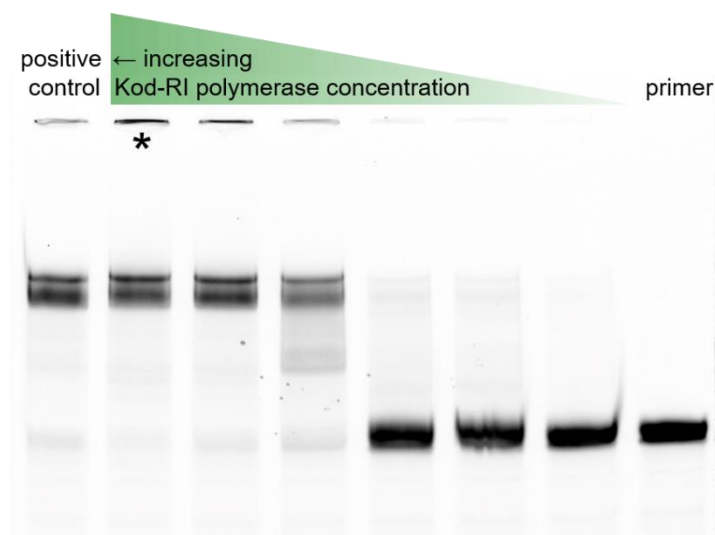


**Figure S 8: UV chromatograms and deconvoluted ESI<sup>-</sup> mass spectra of purified primer extension assays using Therminator polymerase employing DNA template DNA\_pTemp\_midATC<sup>TPT3</sup> and tNaMTP.** As canonical nucleotides, tTTP, tATP and tGTP were added to the PEx. In the negative control, tNaMTP was omitted. In the positive control, unmodified DNA\_pTemp\_midATC was used. Determined molecular weights of each detected peak were assigned underneath.

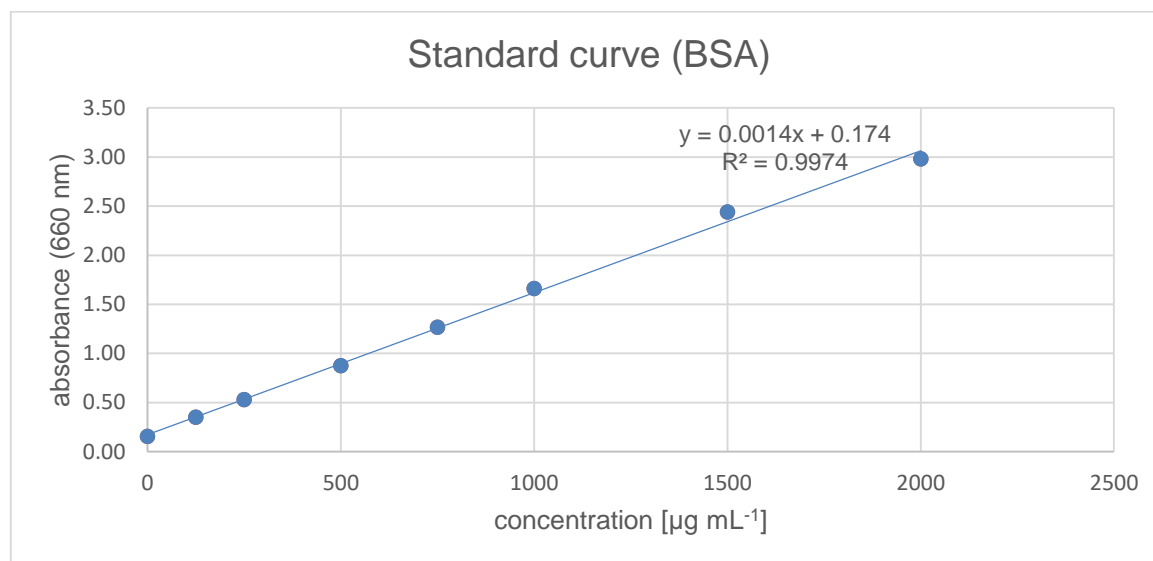




**Figure S 9: Exemplary SDS-PAGE (10 %) of the pooled fractions from one affinity chromatography purification run of His-tagged Kod-RI polymerase.** Ladder: PageRuler unstained protein ladder, 10-200 kDa. Molecular weight of the target protein: 90 kDa. The clarified lysate was loaded onto the column and represents the purity of the sample before affinity chromatography.



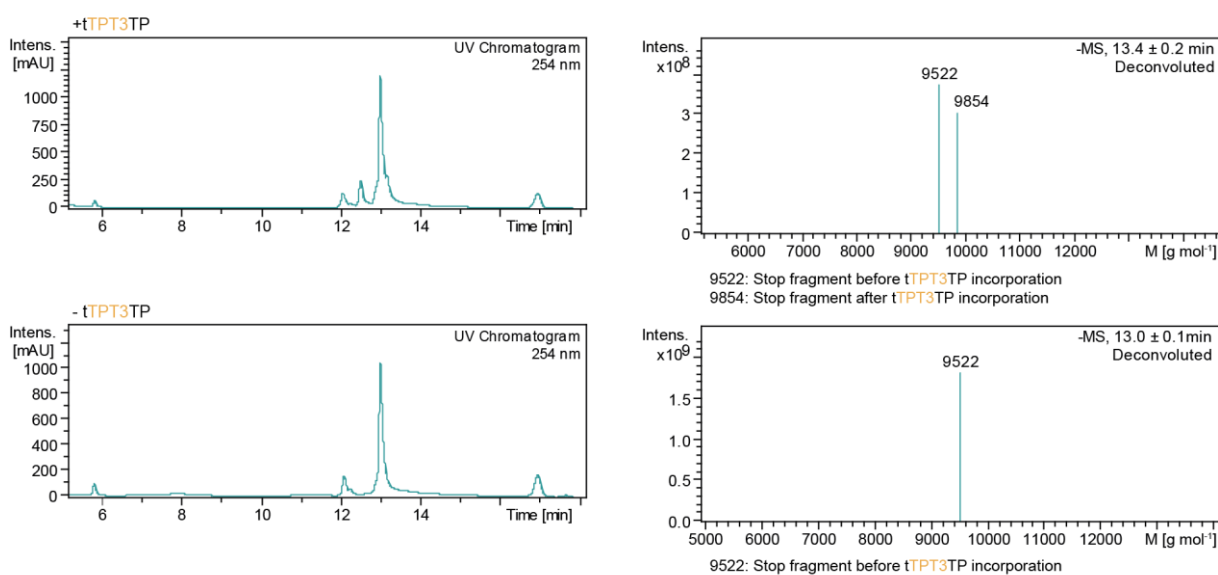
**Figure S 10: Denaturing 15 % PAGE analysis of the Kod-RI polymerase activity assay with 3 hours incubation at 55 °C.** Primer extension assays were performed with DNA\_Temp\_longATGC. The selected working concentration of Kod-RI polymerase for TNA primer extension assays is labeled with an asterisk. The positive control (left lane) was performed with commercially available Therminator polymerase. The right lane contains primer only and serves as a size marker. The image shows the fluorescence scan ( $\lambda_{\text{excitation}} = 477 \text{ nm}$  and  $\lambda_{\text{emission}} = 535 \text{ nm}$ ) of the gel.



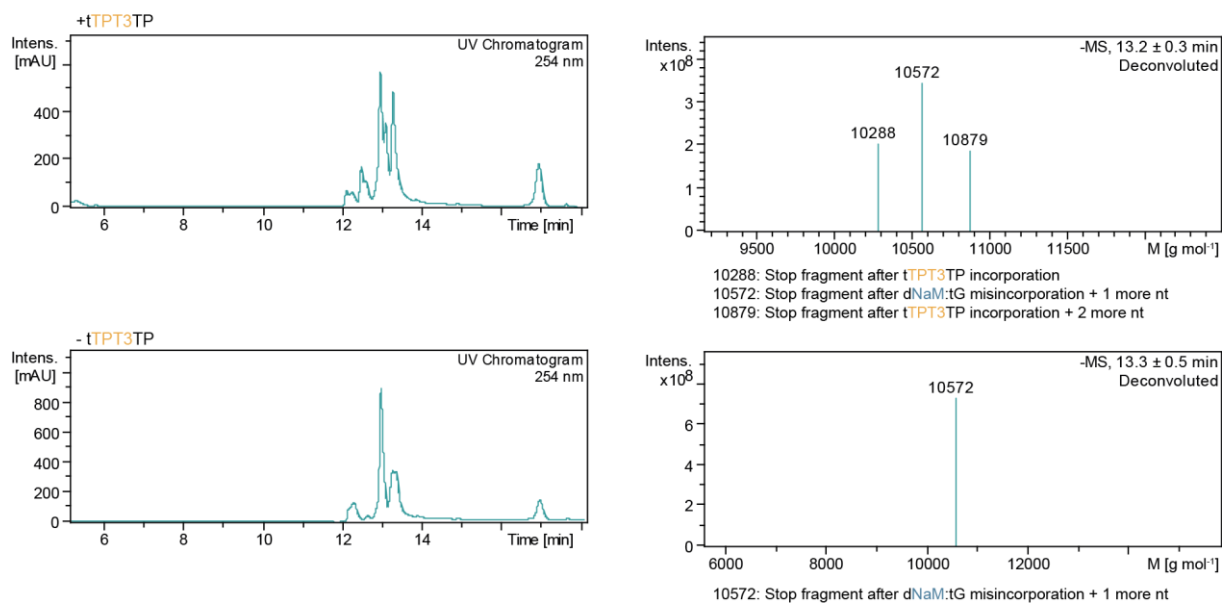
**Figure S 11: Bovine Serum albumin standard curve prepared with 8 different concentrations using the Pierce 660 nm Protein assay kit.** The blue line is the linear regression for the entire set of standard points and gives the equation  $y = 0.00014 x + 0.174$ .

**Table S 1: Protein concentration of the Kod-RI polymerase stock.** The concentration was determined using the Pierce 660 nm Protein assay kit. The absorbance of the sample was measured at 660 nm. The concentration of the protein sample was determined by interpolation using the BSA standard curve (Figure S11).

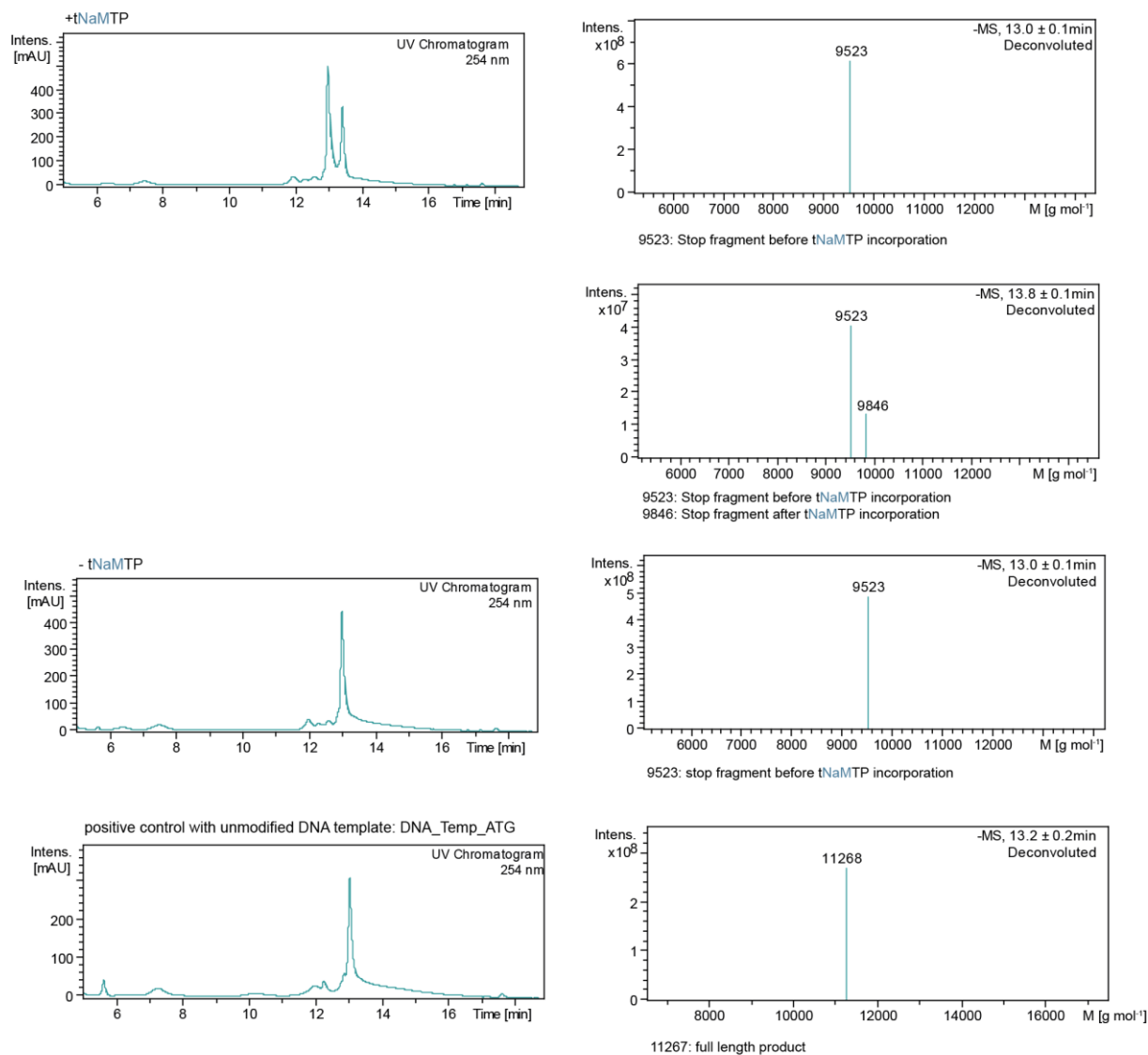
	Absorbance [660 nm]	Protein concentration [ $\mu\text{g mL}^{-1}$ ]
Kod-RI polymerase	2.3362	1544.39



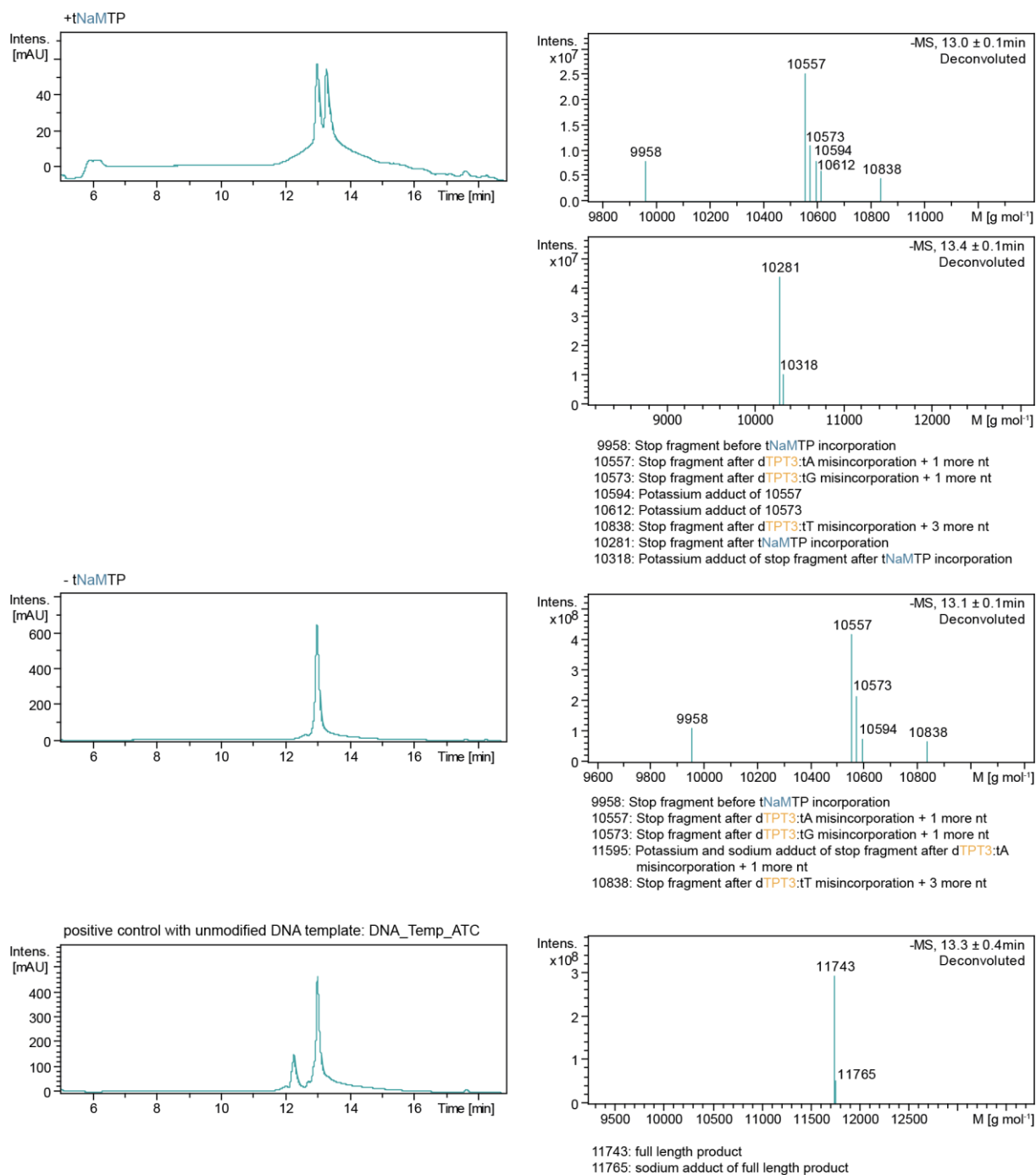
**Figure S 12: UV chromatograms and deconvoluted ESI mass spectra of purified primer extension assays using Kod-RI polymerase employing DNA template DNA\_pTemp\_midATG<sup>NaM</sup> and tTPT3TP.** As canonical nucleotides, tTTP, tATP and tCTP were added to the PEx. In the negative control, tTPT3TP was omitted. Determined molecular weights of each detected peak were assigned underneath.



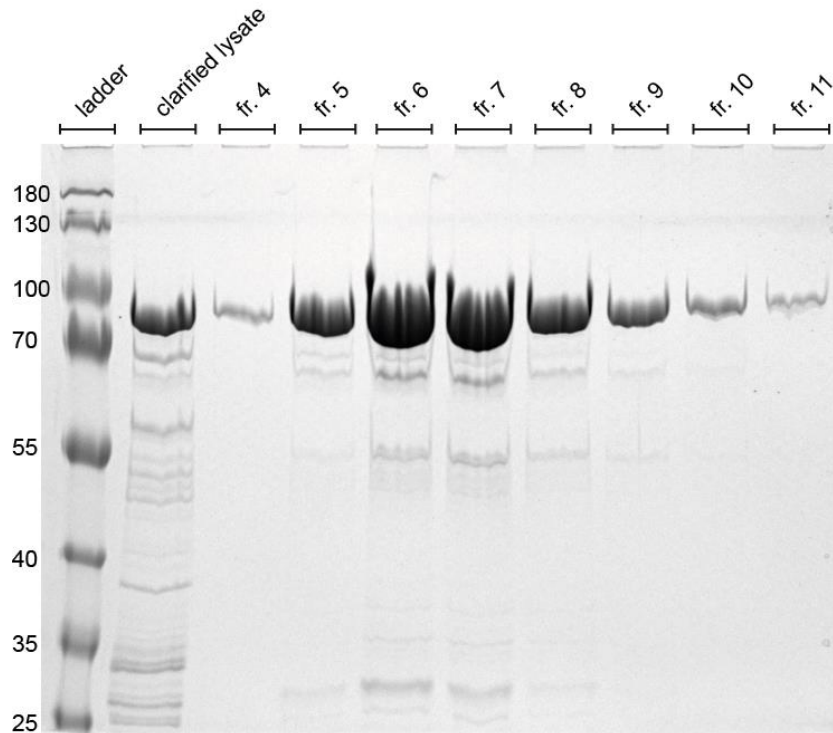
**Figure S 13: UV chromatograms and deconvoluted ESI<sup>-</sup> mass spectra of purified primer extension assays using Kod-RI polymerase employing DNA template DNA\_pTemp\_midATC<sup>NaM</sup> and tTPT3TP.** As canonical nucleotides, tTTP, tATP and tGTP were added to the PEx. In the negative control, tTPT3TP was omitted. Determined molecular weights of each detected peak were assigned underneath.



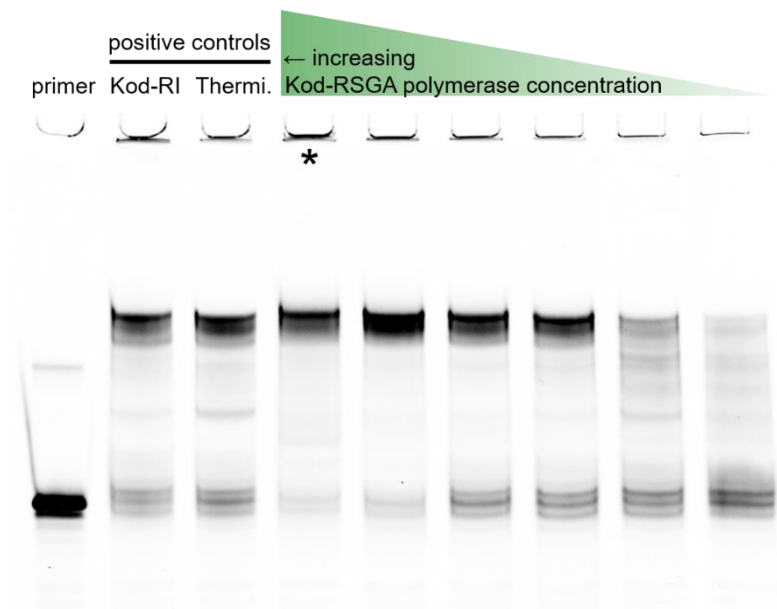
**Figure S 14: UV chromatograms and deconvoluted ESI<sup>-</sup> mass spectra of purified primer extension assays using Kod-RI polymerase employing DNA template DNA\_pTemp\_midATG<sup>TPT3</sup> and tNaMTP.** As canonical nucleotides, tTTP, tATP and tCTP were added to the PEx. In the negative control, tNaMTP was omitted. In the positive control, unmodified DNA\_pTemp\_midATG was used. Determined molecular weights of each detected peak were assigned underneath.



**Figure S 15: UV chromatograms and deconvoluted ESI mass spectra of purified primer extension assays using Kod-RI polymerase employing DNA template DNA\_pTemp\_midATC<sup>TPT3</sup> and tNaMTP.** As canonical nucleotides, tTTP, tATP and tGTP were added to the PEX. In the negative control, tNaMTP was omitted. In the positive control, unmodified DNA\_pTemp\_midATC was used. Determined molecular weights of each detected peak were assigned underneath.



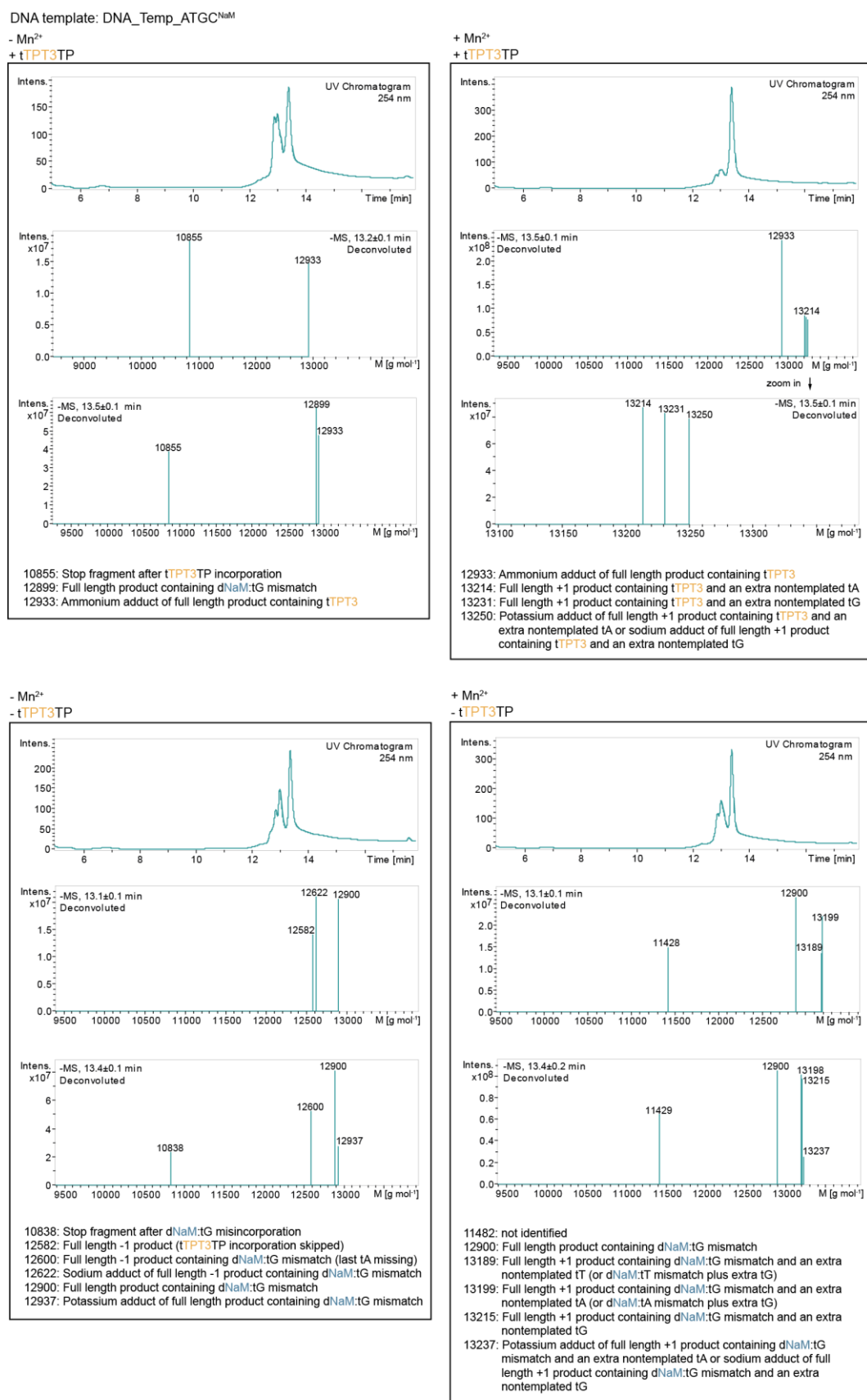
**Figure S 16: Exemplary SDS-PAGE (10 %) of the pooled fractions from one affinity chromatography purification run of His-tagged Kod-RSGA polymerase.** Ladder: PageRuler prestained protein ladder, 10-180 kDa. Molecular weight of the target protein: 90 kDa. The clarified lysate was loaded onto the column and represents the purity of the sample before affinity chromatography.



**Figure S 17: Denaturing 15 % PAGE analysis of the Kod-RSGA polymerase activity assay with 3 hours incubation at 55 °C.** Primer extension assays were performed with DNA\_Temp\_longATGC. The selected working concentration of Kod-RSGA polymerase for TNA primer extension assays is labeled with an asterisk. The positive control 1 (second lane from left) was performed with Kod-RI polymerase, the positive control 2 (third lane from left) was performed with commercially available Therminator (Thermi.) polymerase. The left lane contains primer only and serves as a size marker. The image shows the fluorescence scan ( $\lambda_{\text{excitation}} = 477 \text{ nm}$  and  $\lambda_{\text{emission}} = 535 \text{ nm}$ ) of the gel.

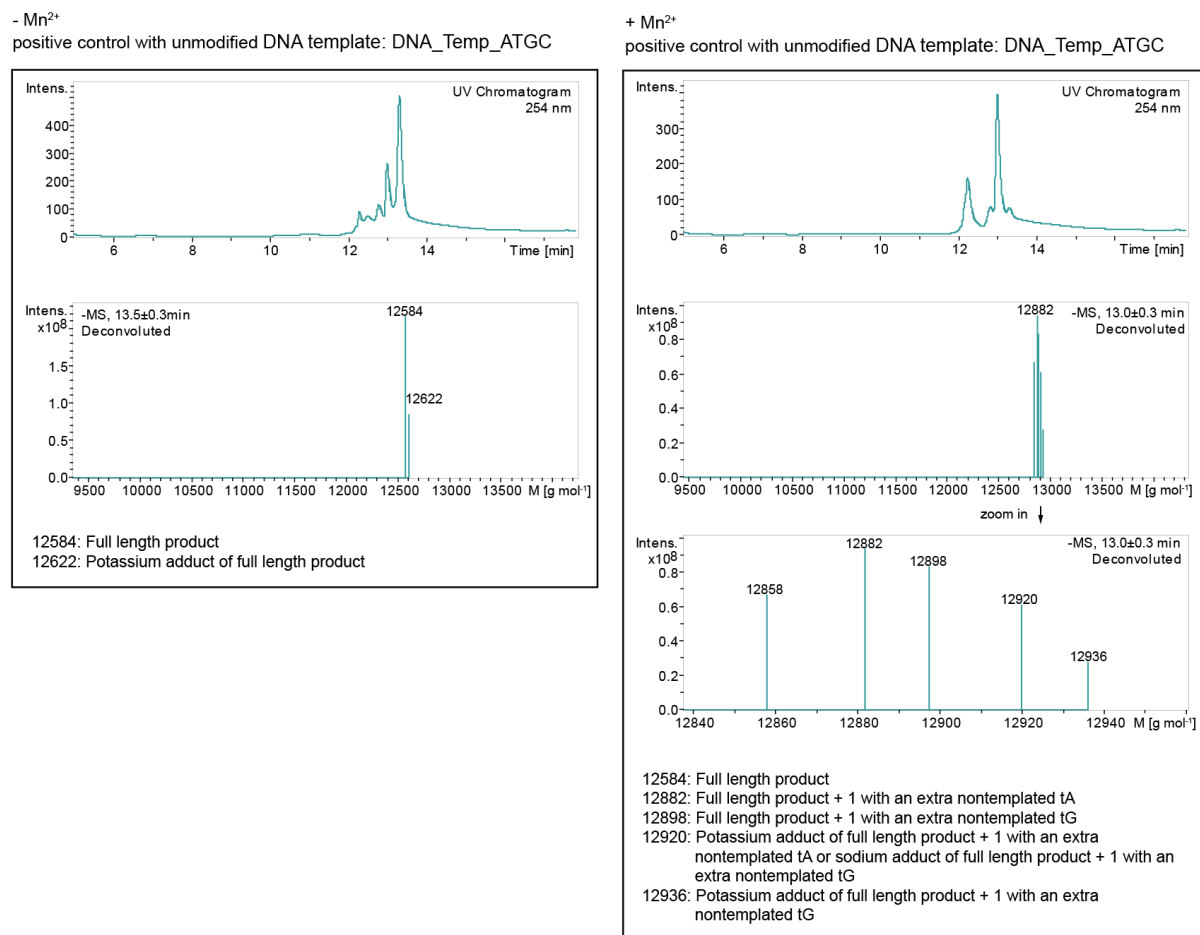
**Table S 2: Protein concentration of the Kod-RSGA polymerase stock.** The concentration was determined using the Pierce 660 nm Protein assay kit. The absorbance of the sample was measured at 660 nm. The concentration of the protein sample was determined by interpolation using the BSA standard curve (Figure S11).

	<b>Absorbance [660 nm]</b>	<b>Protein concentration [<math>\mu\text{g mL}^{-1}</math>]</b>
Kod-RSGA polymerase	3.3285	2253.19



**Figure S 18: UV chromatograms and deconvoluted ESI<sup>-</sup> mass spectra of purified primer extension assays using Kod-RSGA polymerase employing DNA template DNA\_pTemp\_mid2ATGC<sup>NaM</sup> and tTPT3TP. PEx were either performed without MnCl<sub>2</sub> or with 1 mM MnCl<sub>2</sub> supplementation. In negative the controls, tTPT3TP was omitted. Determined molecular weights of each detected peak were assigned underneath.**



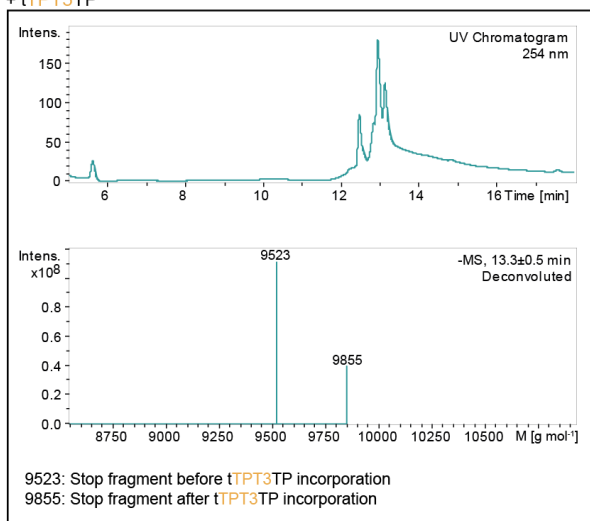


**Figure S 19: UV chromatograms and deconvoluted ESI<sup>-</sup> mass spectra of purified primer extension assays using Kod-RSGA polymerase employing DNA template DNA\_pTemp\_mid2ATGC and canonical tNTPs.** PEx were either performed without MnCl<sub>2</sub> or with 1 mM MnCl<sub>2</sub> supplementation. Determined molecular weights of each detected peak were assigned underneath.

DNA template: DNA\_Temp\_ATG<sup>NaM</sup>

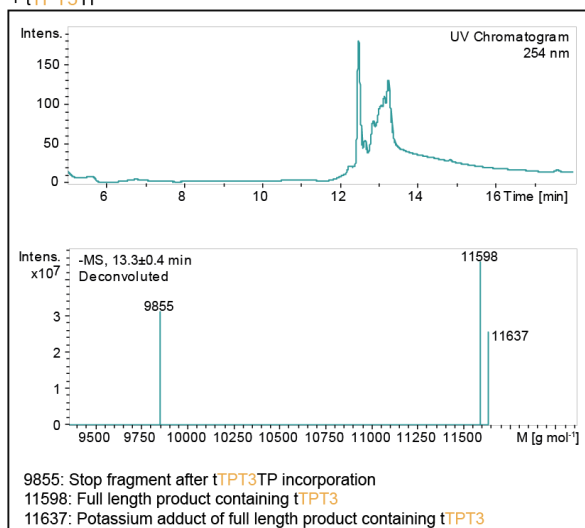
- Mn<sup>2+</sup>

+ tTPT3TP



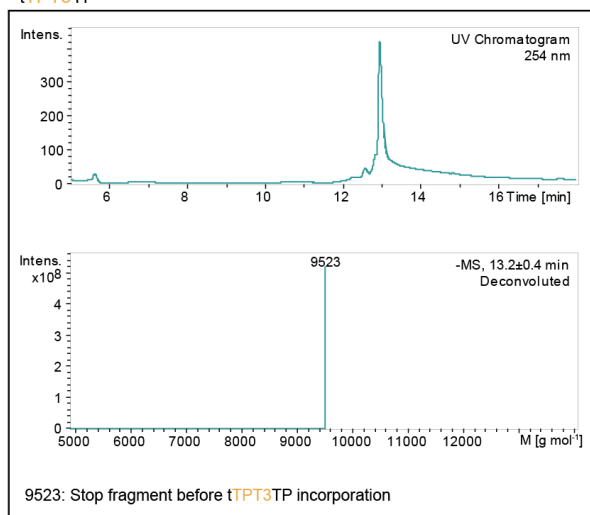
+ Mn<sup>2+</sup>

+ tTPT3TP



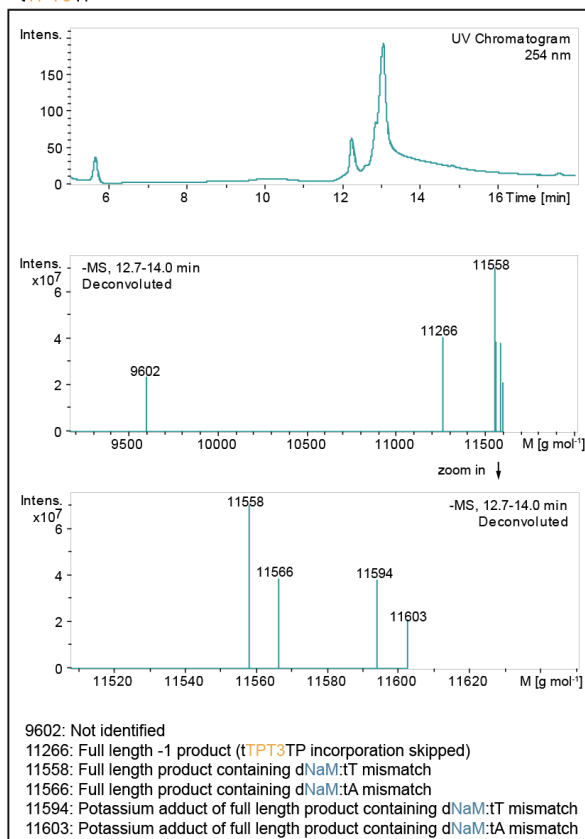
- Mn<sup>2+</sup>

- tTPT3TP

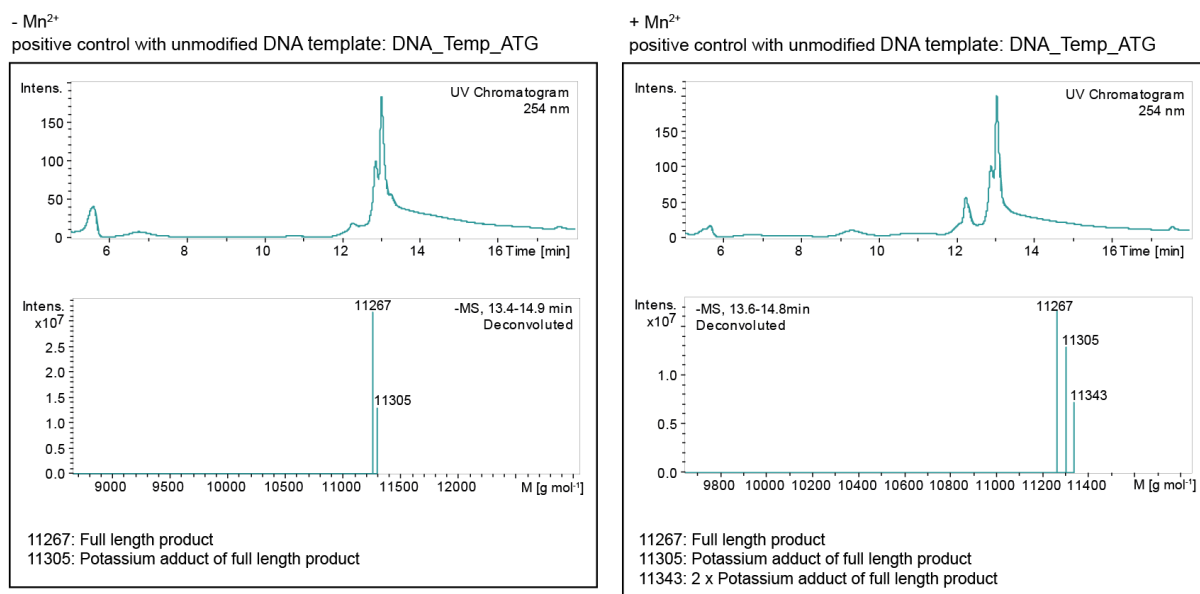


+ Mn<sup>2+</sup>

- tTPT3TP



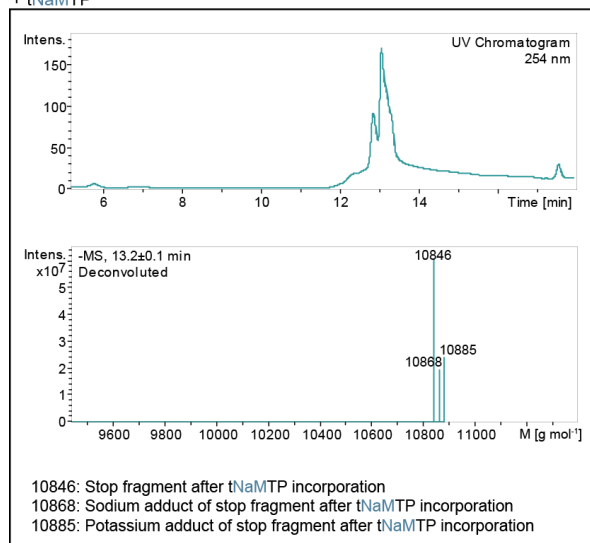
**Figure S 20: UV chromatograms and deconvoluted ESI mass spectra of purified primer extension assays using Kod-RSGA polymerase employing DNA template DNA\_pTemp\_midATG<sup>NaM</sup> and tTPT3TP. PEx were either performed without MnCl<sub>2</sub> or with 1 mM MnCl<sub>2</sub> supplementation. In the negative controls, tTPT3TP was omitted. Determined molecular weights of each detected peak were assigned underneath.**



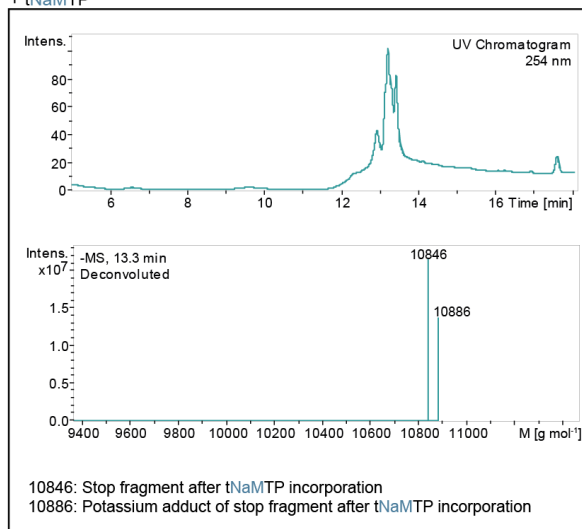
**Figure S 21: UV chromatograms and deconvoluted ESI<sup>-</sup> mass spectra of purified primer extension assays using Kod-RSGA polymerase employing DNA template DNA\_pTemp\_midATG and canonical tNTPs.** PEx were either performed without MnCl<sub>2</sub> or with 1 mM MnCl<sub>2</sub> supplementation. Determined molecular weights of each detected peak were assigned underneath.

DNA template: DNA\_Temp\_ATGC<sup>TPT3</sup>

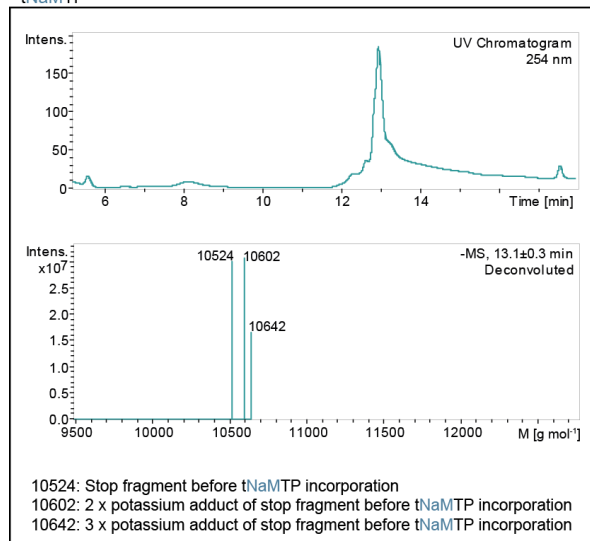
- Mn<sup>2+</sup>  
+ tNaMTP



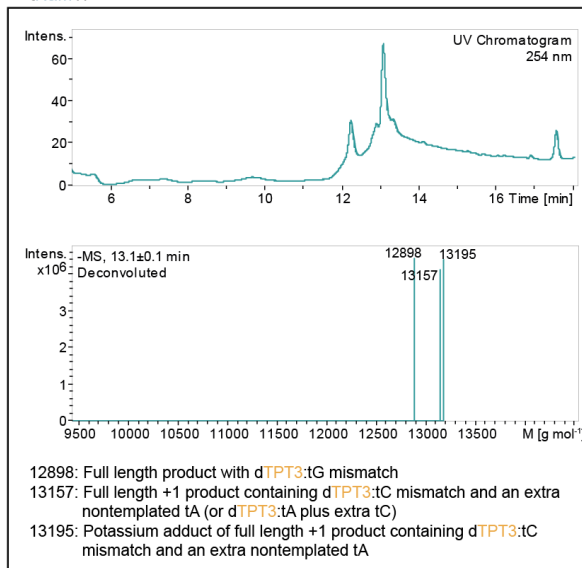
+ Mn<sup>2+</sup>  
+ tNaMTP



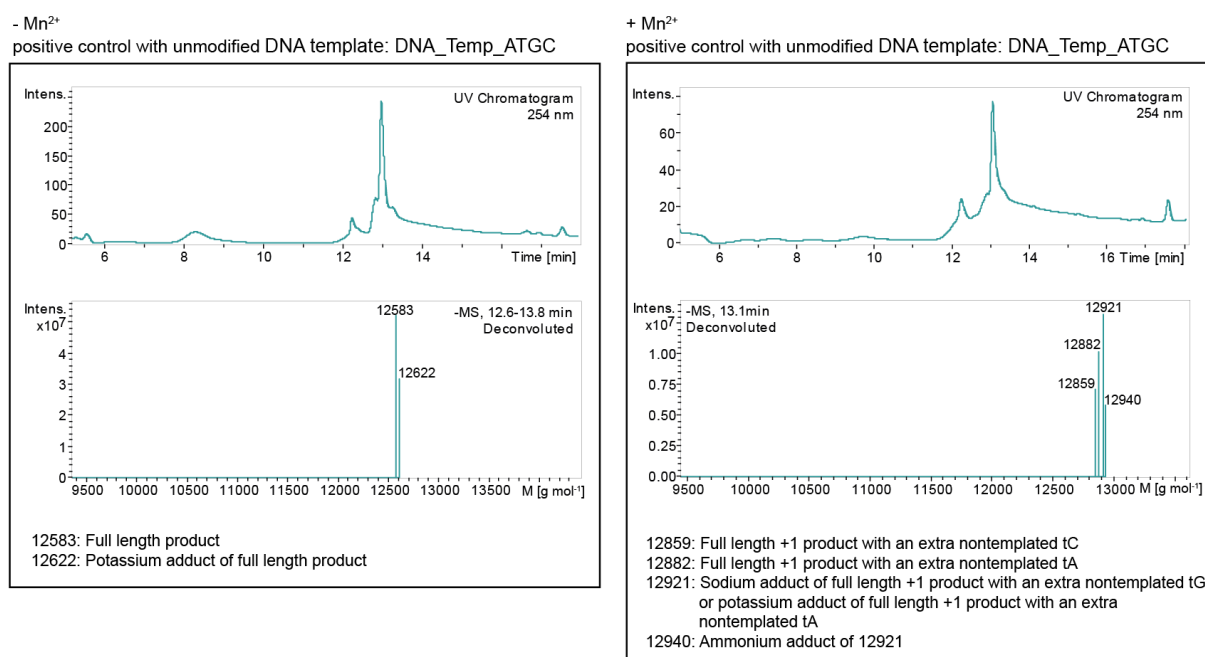
- Mn<sup>2+</sup>  
- tNaMTP



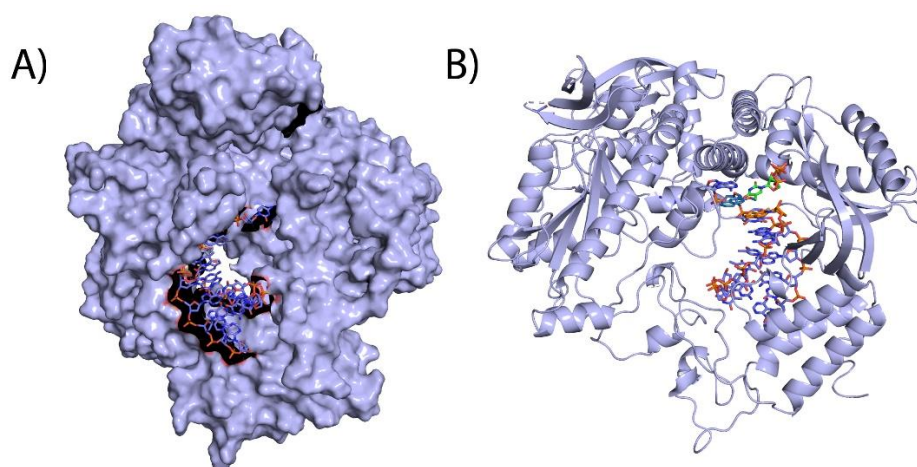
+ Mn<sup>2+</sup>  
- tNaMTP



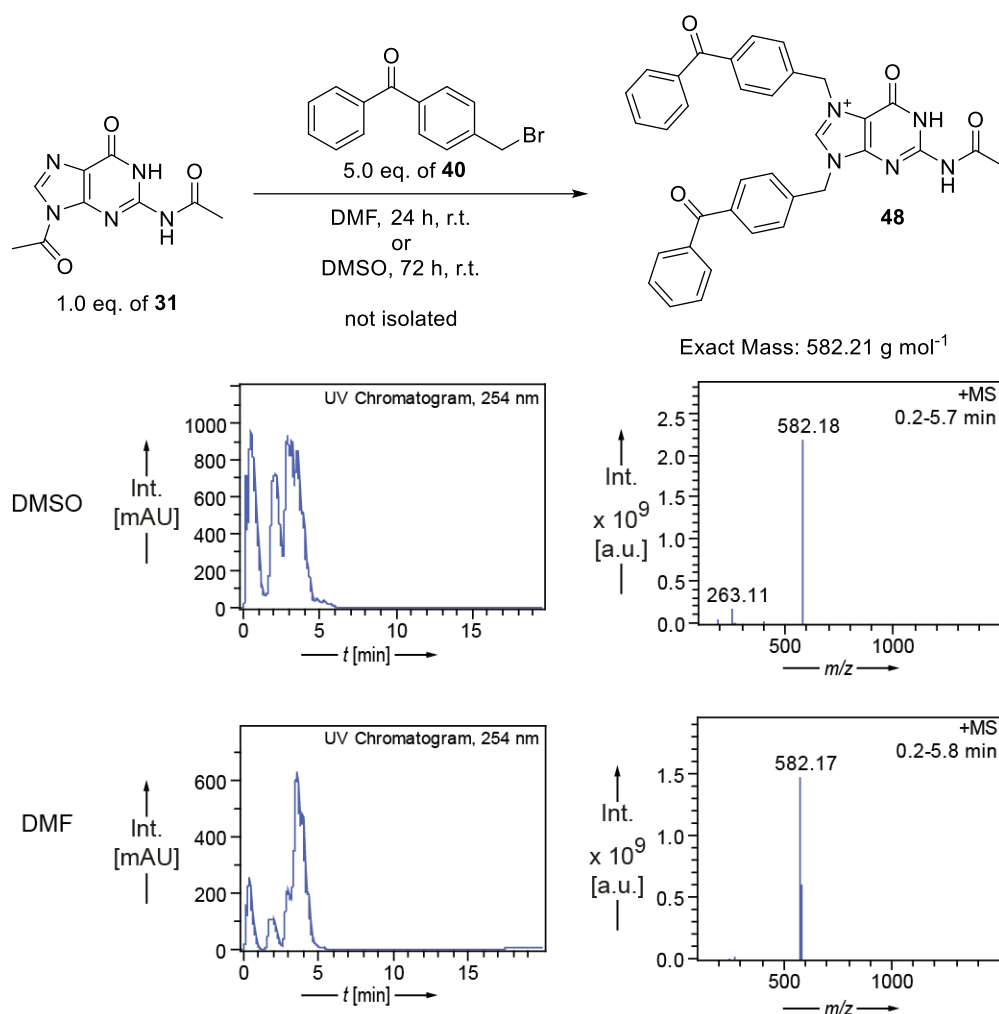
**Figure S 22: UV chromatograms and deconvoluted ESI<sup>-</sup> mass spectra of purified primer extension assays using Kod-RSGA polymerase employing DNA template DNA\_pTemp\_mid2ATGC<sup>TPT3</sup> and tNaMTP. PEx were either performed without MnCl<sub>2</sub> or with 1 mM MnCl<sub>2</sub> supplementation. In the negative controls, tNaMTP was omitted. Determined molecular weights of each detected peak were assigned underneath.**



**Figure S 23: UV chromatograms and deconvoluted ESI- mass spectra of purified primer extension assays using Kod-RSGA polymerase employing DNA template DNA\_pTemp\_mid2ATGC and canonical tNTPs.** PEx were either performed without MnCl<sub>2</sub> or with 1 mM MnCl<sub>2</sub> supplementation. Determined molecular weights of each detected peak were assigned underneath.



**Figure S 24: Representative snapshot (cluster analysis) of 200 ns MD simulation of Kod-RSGA polymerase in complex with a DNA template containing dNaM and a nascent TNA oligonucleotide containing tTPT3.** The DNA:TNA hybrid is buried within the Kod-RSGA polymerase active site in its closed ternary complex. A) Surface representation. B) Cartoon representation.



**Figure S 25:** LC-MS analysis of the *N*<sup>7</sup>-functionalization of *N*<sup>2,9</sup>-diacetylated guanine **31** using 5.0 equivalents of BP-Br (**40**) as a simplified test system. The benzophenone moiety is attached at the *N*<sup>7</sup> and *N*<sup>9</sup> nitrogens of guanine.

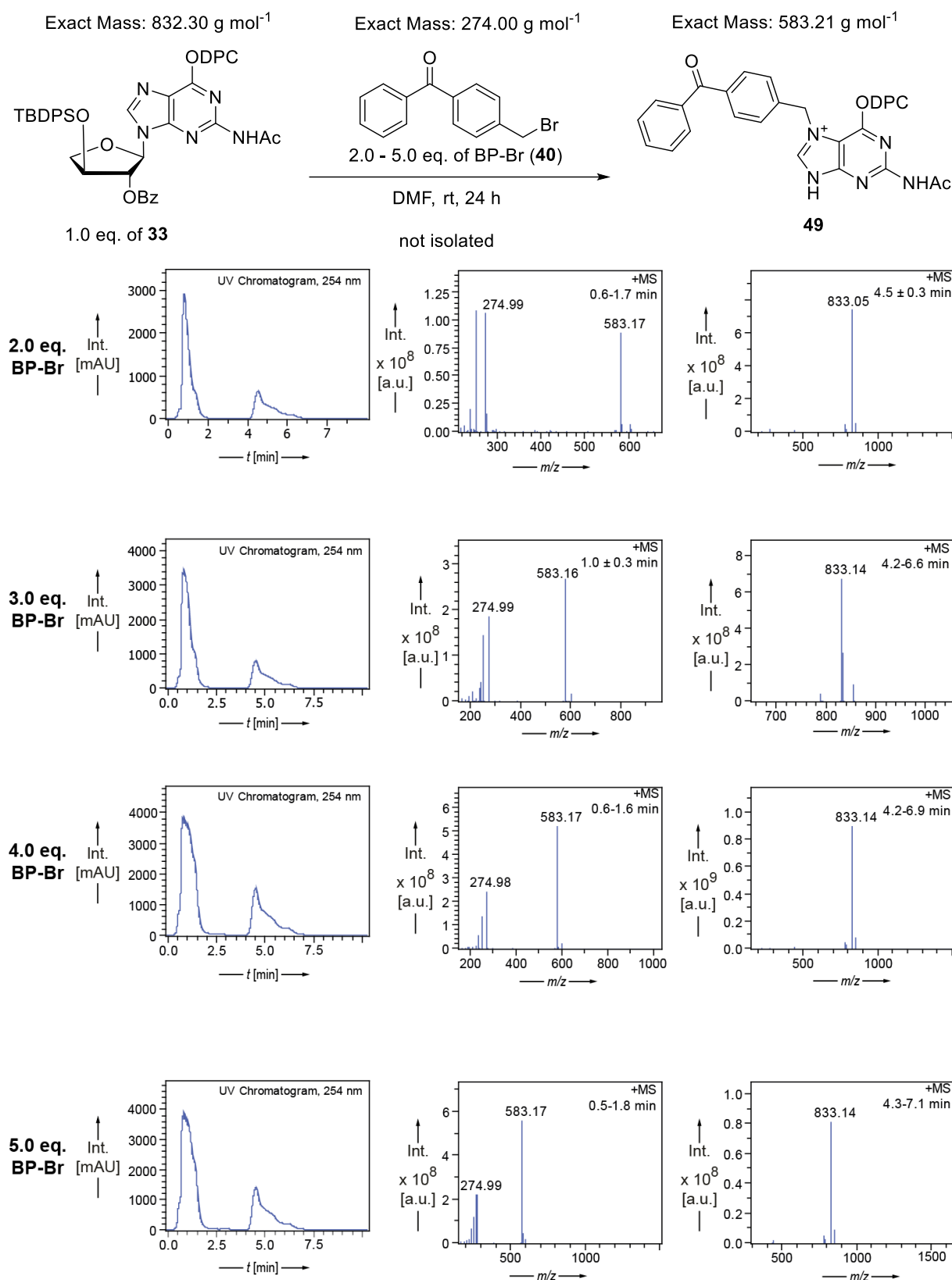
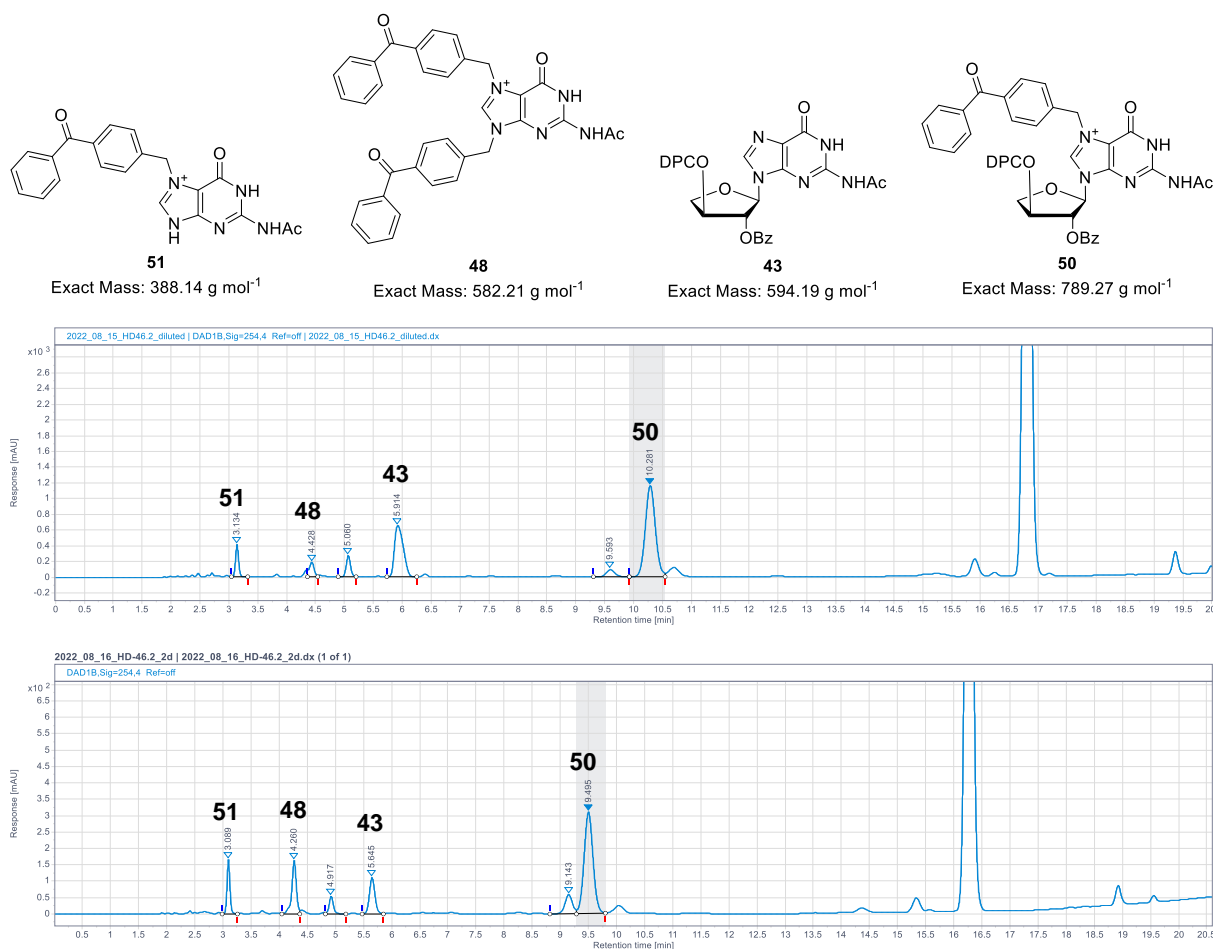


Figure S 26: LC-MS analysis of the *N*-functionalization of threoguanosine 33 with different amounts of equivalents of BP-Br(40). Attachment of the benzophenone moiety results in the cleavage of the *N*-glycosidic bond.



**Figure S 27: RP-HPLC analysis of the crude reaction mixture of the *N'*-functionalization of threoguanosine 43.** The upper UV chromatogram ( $\lambda = 254$  nm) was measured after 24 h of reaction time. The lower UV chromatogram ( $\lambda = 254$  nm) was measured after 48 h of reaction time. The peaks were assigned to the corresponding compounds by LC-MS analysis.

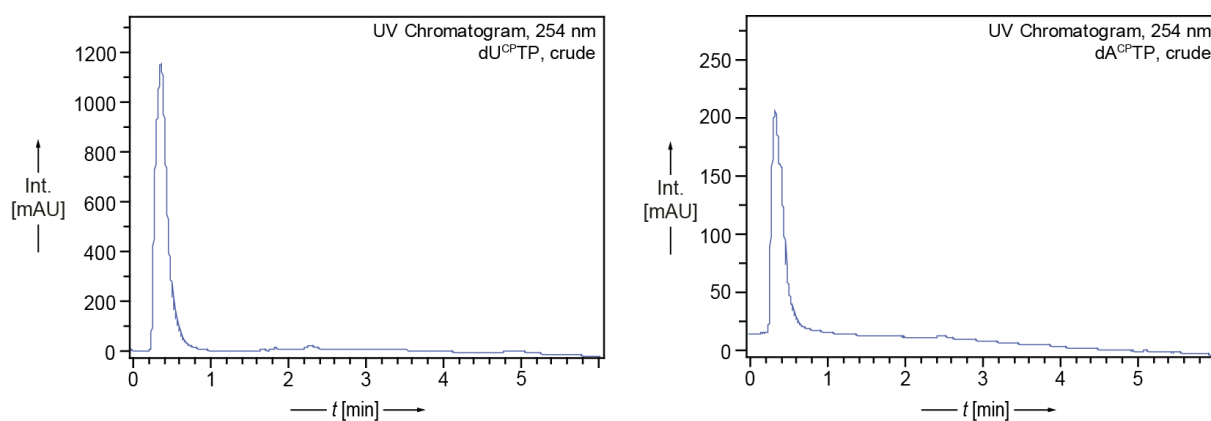
**Table S 3: Area under the curve (AUC) analysis of the crude reaction mixture of the *N'*-benzophenone-functionalization of threoguanosine 43 after 24 h reaction time.** The values shown here correspond to the upper UV chromatogram of Figure S27.

peak number	retention time [min]	AUC	AUC [%]	compound
1	3.134	1510.212	5.80	<b>51</b>
2	4.428	1069.930	4.11	<b>48</b>
3	5.060	1417.744	5.45	unknown
4	5.915	6963.404	26.75	<b>43</b>
5	9.539	967.361	3.72	unknown
6	10.28	14098.953	54.169	<b>50</b>

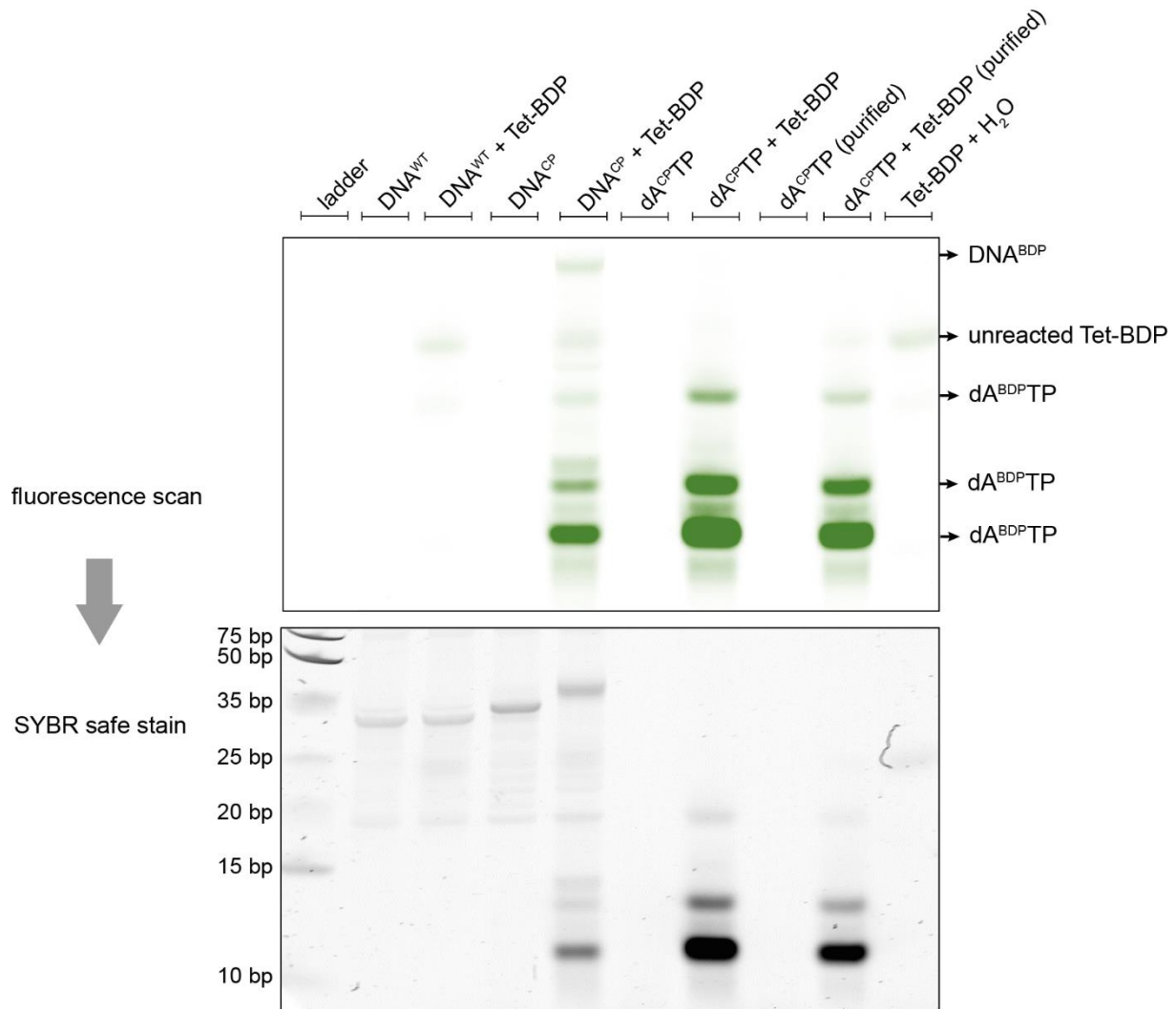


**Table S 4: Area under the curve (AUC) analysis of the crude reaction mixture of the *N*<sup>7</sup>-benzophenone-functionalization of threoguanosine 43 after 48 h reaction time.** The values shown here correspond to the bottom UV chromatogram of Figure S27.

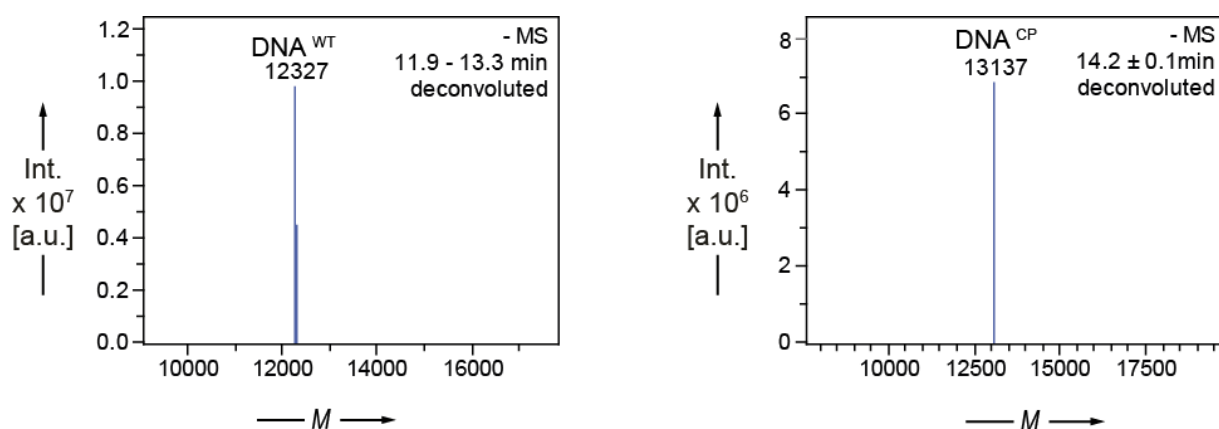
peak number	retention time [min]	AUC	AUC [%]	compound
1	3.089	5.83.024	8.95	<b>51</b>
2	4.260	918.421	14.10	<b>48</b>
3	4.917	274.920	4.22	unknown
4	5.645	745.255	11.44	<b>43</b>
5	9.143	531.642	8.16	unknown
6	9.495	3463.824	53.15	<b>50</b>



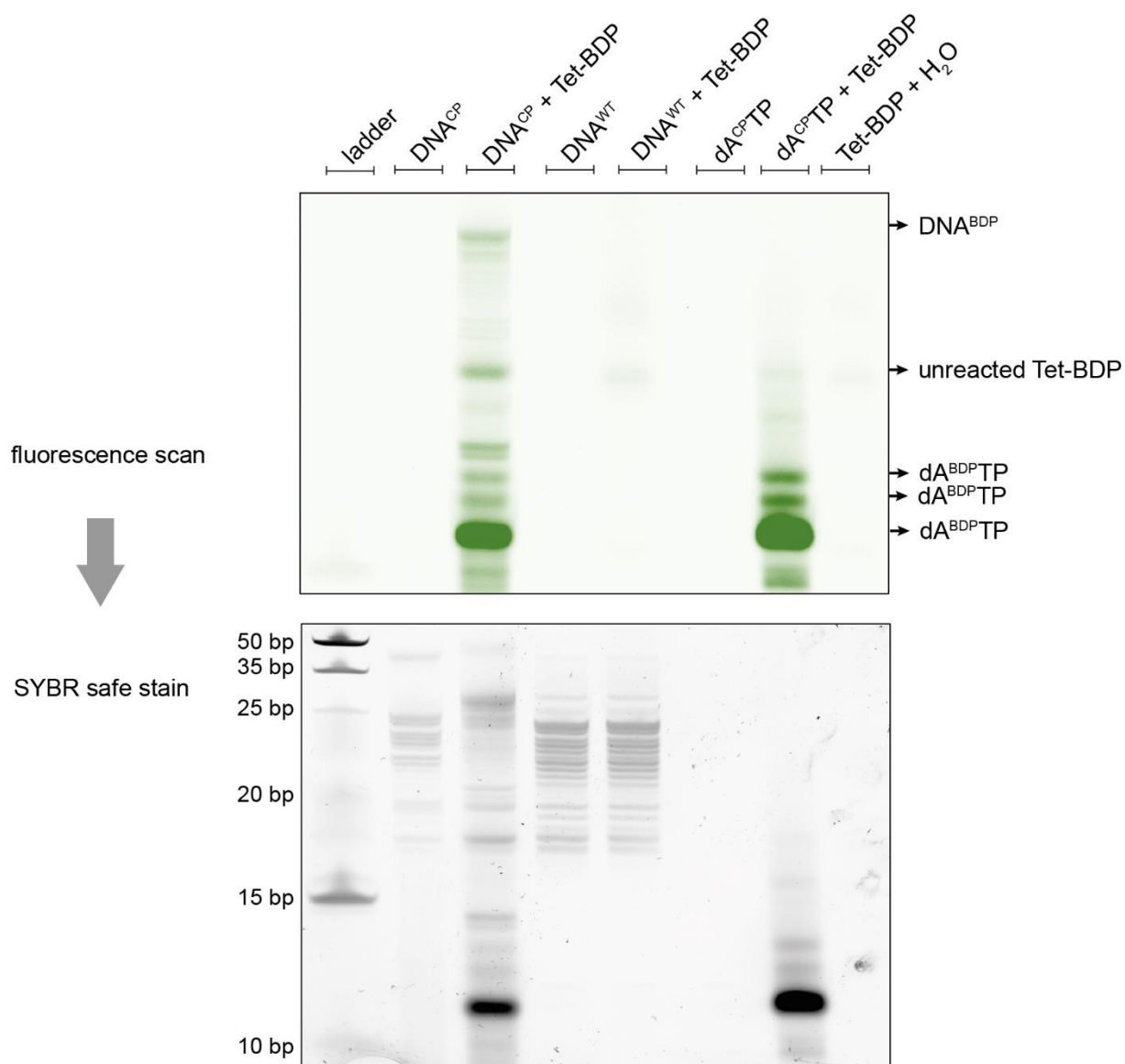
**Figure S 28: UV traces ( $\lambda = 254$  nm) of the crude products obtained after triphosphate synthesis of dU<sup>CPTP</sup> (58, left) and dA<sup>CPTP</sup> (59, right).**



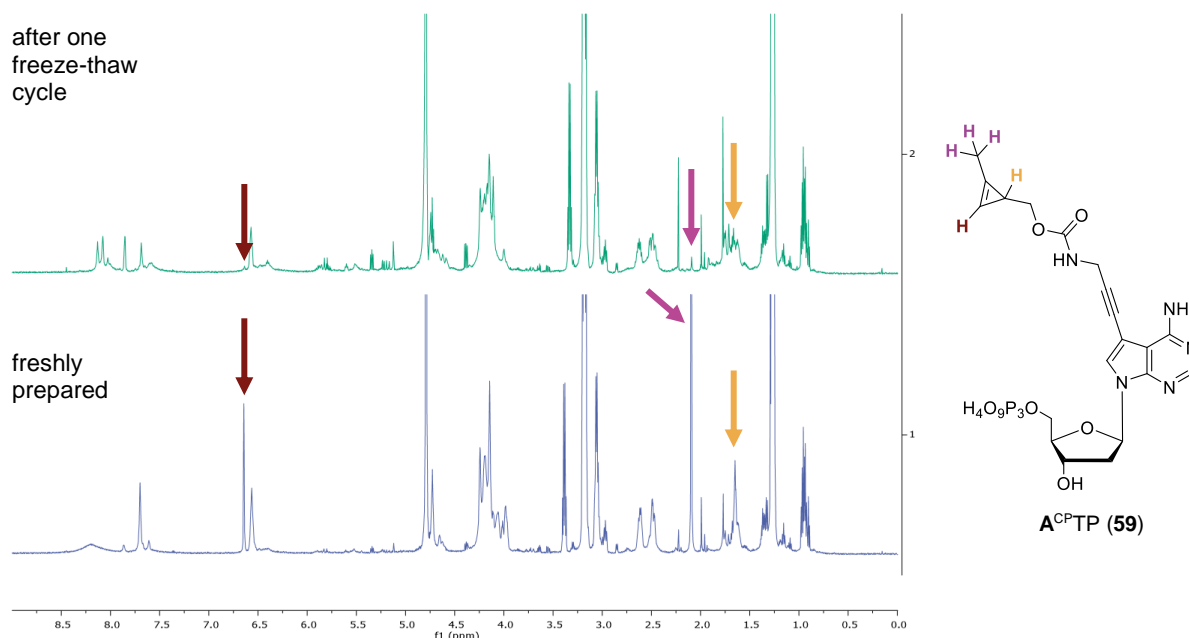
**Figure S 29: Full image of the 20 % dPAGE analysis of primer extension assays with dA<sup>CP</sup>TP using Taq polymerase and further downstream functionalization of DNA<sup>CP</sup> with Tet-BDP.** The upper image depicts the fluorescence scan ( $\lambda_{\text{excitation}} = 477 \text{ nm}$ ,  $\lambda_{\text{emission}} = 535 \text{ nm}$ ) and the lower image displays the UV visualization after SYBR safe post-staining. Ladder: GeneRuler Ultra Low Range DNA Ladder.



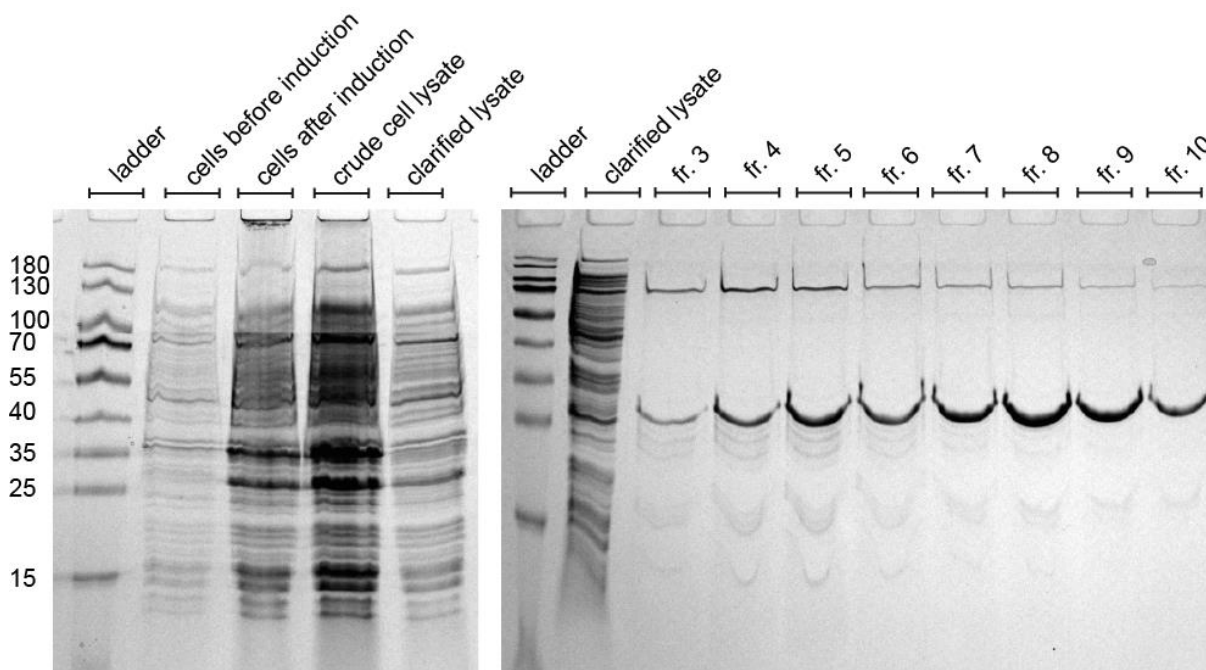
**Figure S 30: ESI- mass spectra of DNA<sup>WT</sup> and DNA<sup>CP</sup>.** M calculated for full-length DNA<sup>WT</sup>: 12326. M calculated for full-length DNA<sup>CP</sup>: 13135.



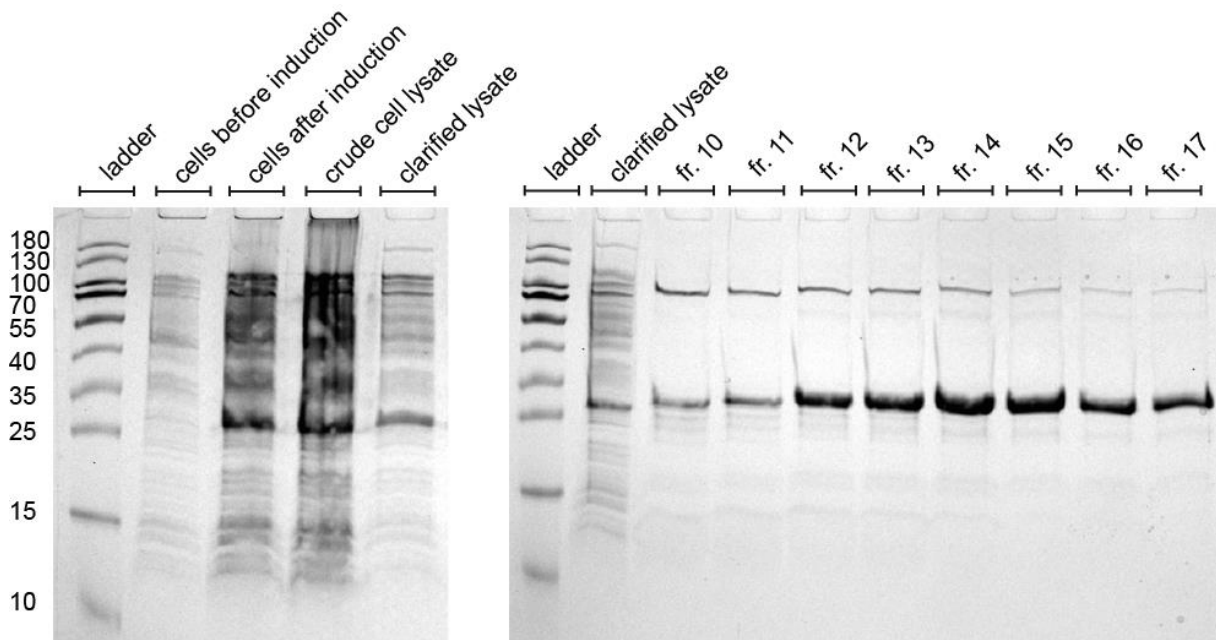
**Figure S 31: 20 % dPAGE analysis of primer extension assays with  $dA^{CPTP}$  using OneTaq polymerase and further downstream functionalization of  $DNA^{CP}$  with Tet-BDP.** The upper image depicts the fluorescence scan ( $\lambda_{excitation} = 477 \text{ nm}$ ,  $\lambda_{emission} = 535 \text{ nm}$ ) and the lower image displays the UV visualization after SYBR safe post-staining. Ladder: GeneRuler Ultra Low Range DNA Ladder.



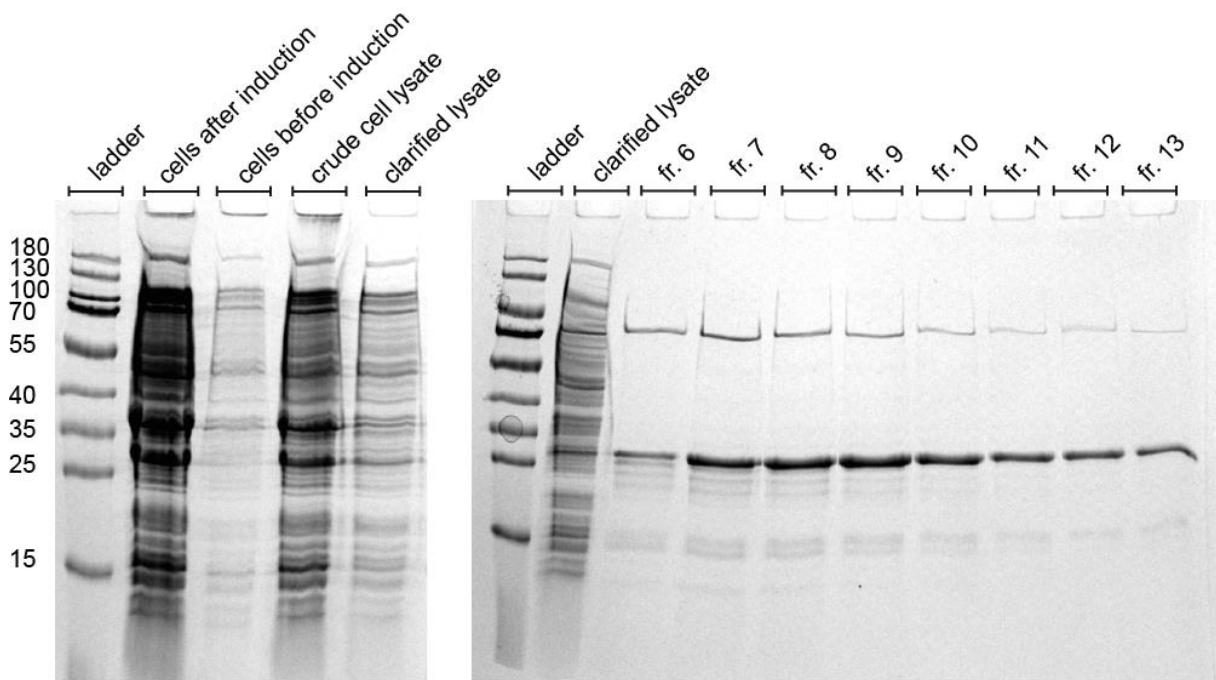
**Figure S 32: Evaluation of the stability of A<sup>C</sup>PTP (59) by NMR.** Upper spectrum: <sup>1</sup>H NMR spectrum of freshly prepared A<sup>C</sup>PTP (59). Bottom spectrum: <sup>1</sup>H NMR spectrum of A<sup>C</sup>PTP (59) after one freeze-thaw cycle. The arrows highlight the main changes in the spectrum with respect to the peaks corresponding to the protons attached to the cyclopropene moiety.



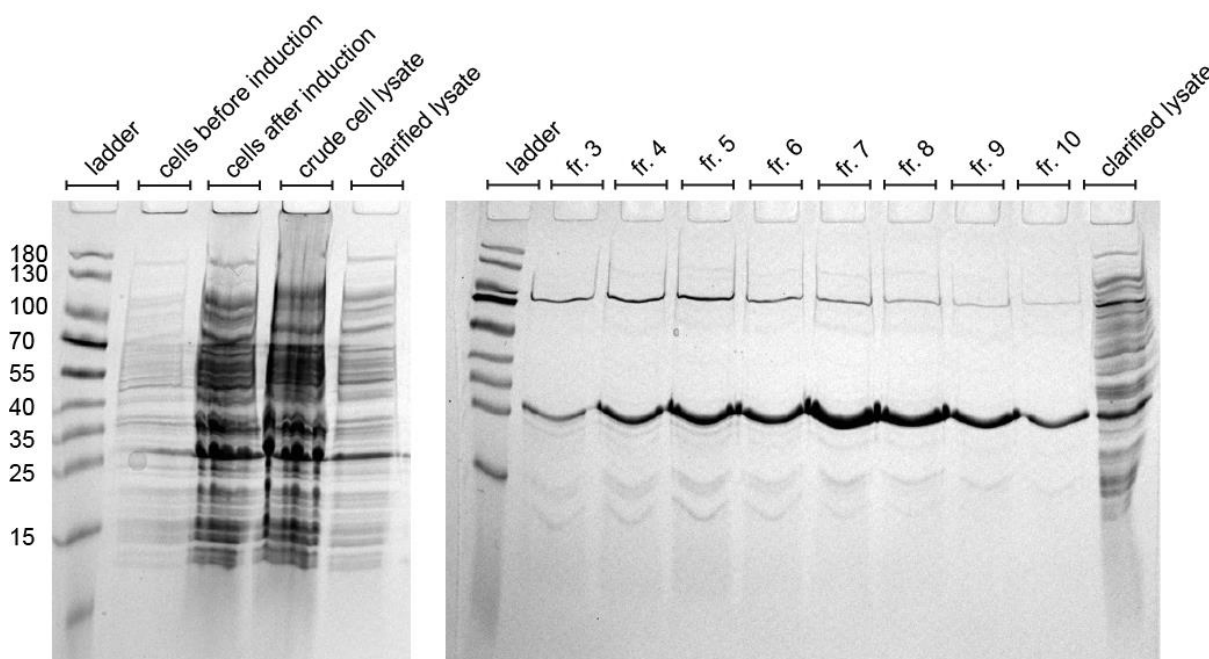
**Figure S 33: SDS-PAGE (15 %) of the expression and IMAC purification of His-tagged mZBP1<sup>WT</sup>.** Ladder: PageRuler Prestained Protein Ladder, 10-180 kDa. Molecular weight of the target protein: 24 kDa. The clarified lysate was loaded onto the column and represents the purity of the sample before affinity chromatography.



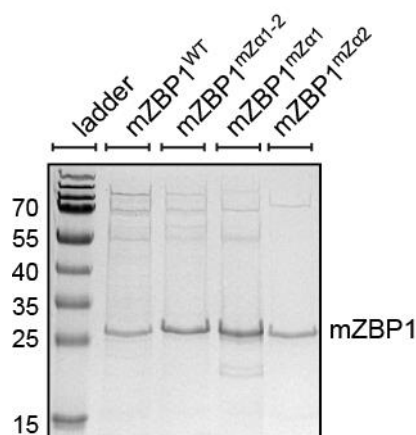
**Figure S 34: SDS-PAGE (15 %) of the expression and IMAC purification of His-tagged mZBP1<sup>mZ $\alpha$ 1</sup>.** Ladder: PageRuler Prestained Protein Ladder, 10-180 kDa. Molecular weight of the target protein: 24 kDa. The clarified lysate was loaded onto the column and represents the purity of the sample before affinity chromatography.



**Figure S 35: SDS-PAGE (15 %) of the expression and IMAC purification of His-tagged mZBP1<sup>mZ $\alpha$ 2</sup>.** Ladder: PageRuler Prestained Protein Ladder, 10-180 kDa. Molecular weight of the target protein: 24 kDa. The clarified lysate was loaded onto the column and represents the purity of the sample before affinity chromatography.



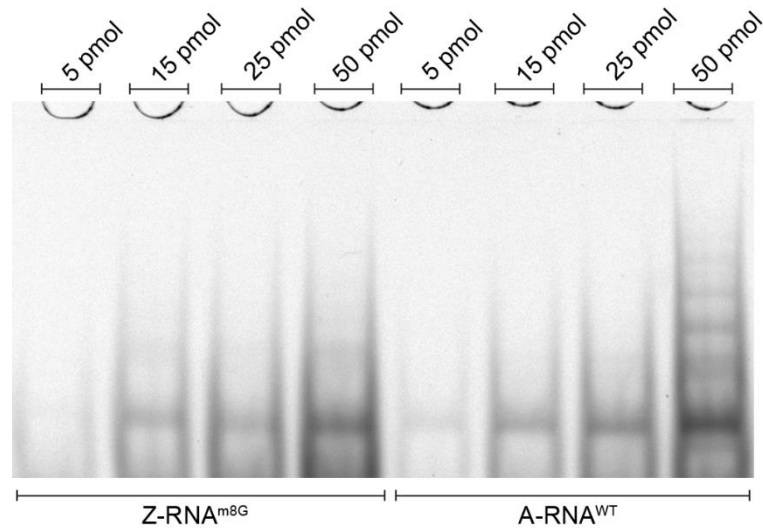
**Figure S 36: SDS-PAGE (15 %) of the expression and IMAC purification of His-tagged mZBP1<sup>mZα1-2</sup>.** Ladder: PageRuler Prestained Protein Ladder, 10-180 kDa. Molecular weight of the target protein: 24 kDa. The clarified lysate was loaded onto the column and represents the purity of the sample before affinity chromatography.



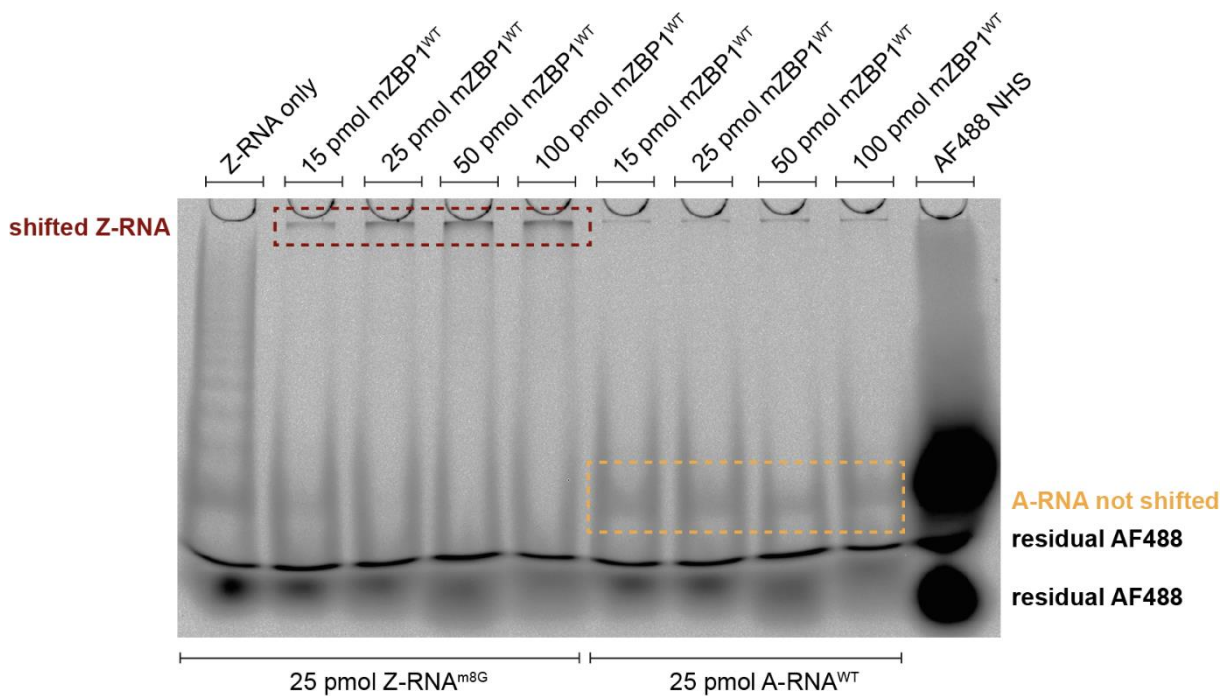
**Figure S 37: Complete 15 % SDS-PAGE of the different stocks of expressed and purified mZBP1 variants.** Ladder: PageRuler Prestained Protein Ladder, 10-180 kDa. Molecular weight of the target protein: 24 kDa.

**Table S 5: Protein concentrations of the distinct mZBP1 variant stocks.** The concentrations were determined using the Pierce 660 nm Protein assay kit. The absorbance of the samples was measured at 660 nm. The concentration of the protein sample was determined by interpolation using the BSA standard curve (Figure S11).

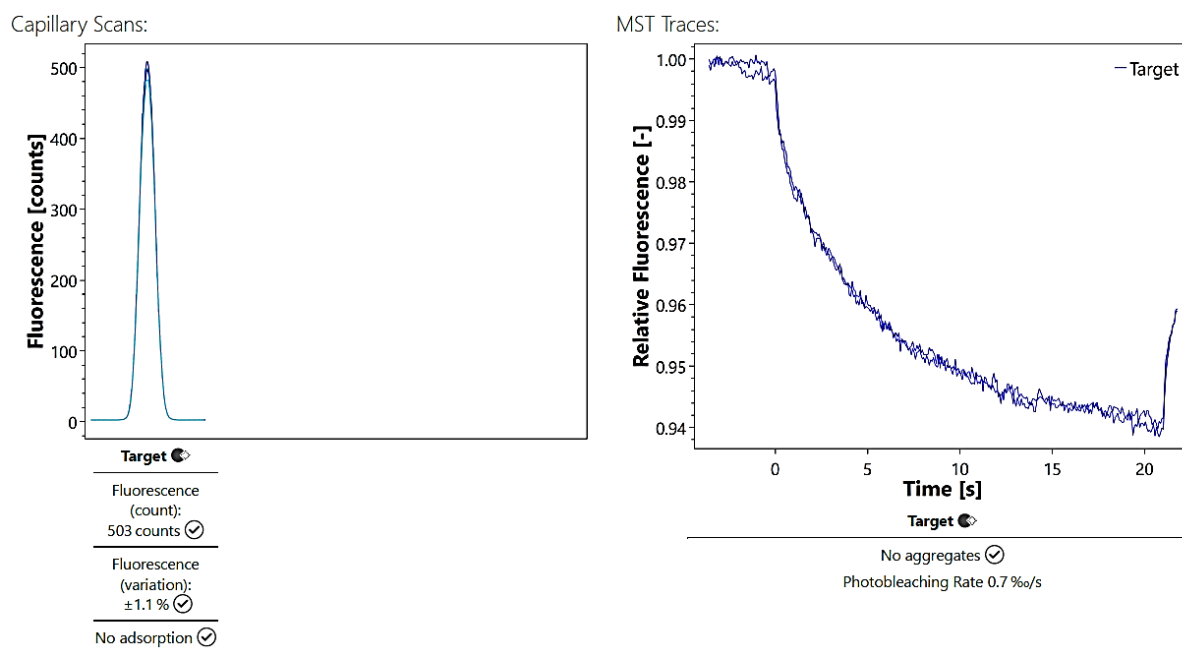
mZBP1 variant	Absorbance [660 nm]	Protein concentration [ $\mu\text{g mL}^{-1}$ ]
mZBP1 <sup>WT</sup>	0.7406	404.71
mZBP1 <sup>mZα1</sup>	1.2086	738.96
mZBP1 <sup>mZα2</sup>	0.5671	280.79
mZBP1 <sup>mZα1-2</sup>	0.7628	420.56



**Figure S 38: 8 % nPAGE of titrating amounts of AF88-labeled Z-RNA<sup>m8G</sup> and A-RNA<sup>WT</sup>.** Different amounts of Z-RNA<sup>m8G</sup> and A-RNA<sup>WT</sup> (i.e. 5 pmol, 15 pmol, 25 pmol, and 50 pmol) were run on nPAGE to determine the minimum detectable amount. Visualization was performed by fluorescence scanning ( $\lambda_{\text{excitation}} = 477 \text{ nm}$ ,  $\lambda_{\text{emission}} = 535 \text{ nm}$ ).



**Figure S 39: Complete 8 % nPAGE of comparative interaction studies of mZBP1<sup>WT</sup> with Z-RNA<sup>m8G</sup> and A-RNA<sup>WT</sup>.** A fixed amount of 25 pmol of fluorescently-labeled Z-RNA<sup>m8G</sup> or A-RNA<sup>WT</sup> was incubated with titrating amounts of mZBP1<sup>WT</sup> (i.e. 15 pmol, 25 pmol, 50 pmol, and 100 pmol). Visualization was performed by fluorescence scanning ( $\lambda_{\text{excitation}} = 477 \text{ nm}$ ,  $\lambda_{\text{emission}} = 535 \text{ nm}$ ).



**Figure S 40: Microscale thermophoresis pretests.** Scans with 20 nm AF488-labeled Z-RNA<sup>m8G</sup> in interaction buffer showed no unspecific adsorption to the capillaries and no aggregation. In addition, a suitable fluorescence intensity within the optimal range between 200 and 1500 counts was confirmed.



## 7.2 NMR spectra

This paragraph contains all NMR spectra of known and novel compounds described in chapter 5.2.7. Solvent residual signals (sol.) and assignable impurities are marked with asterisks.

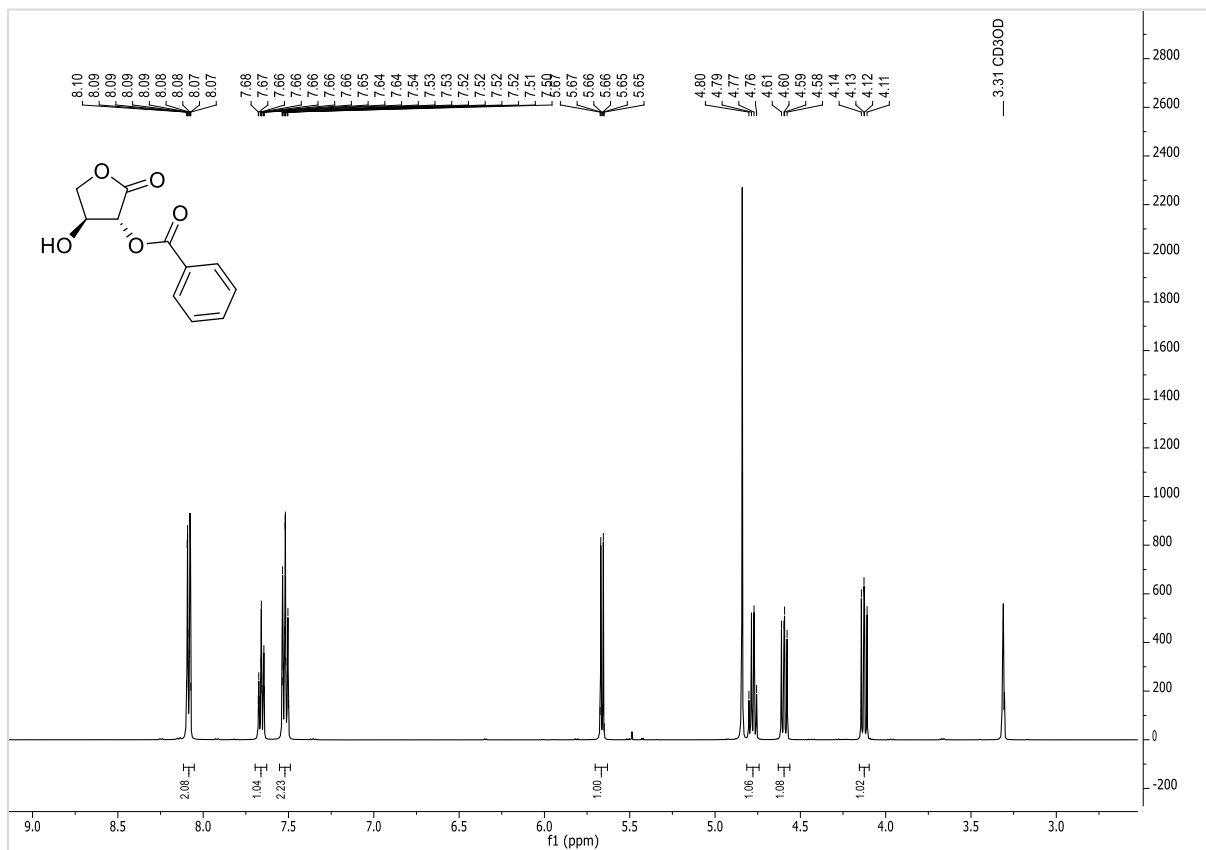


Figure S 41: <sup>1</sup>H-NMR spectrum of compound 14 (CD<sub>3</sub>OD, 400 MHz, r.t.).

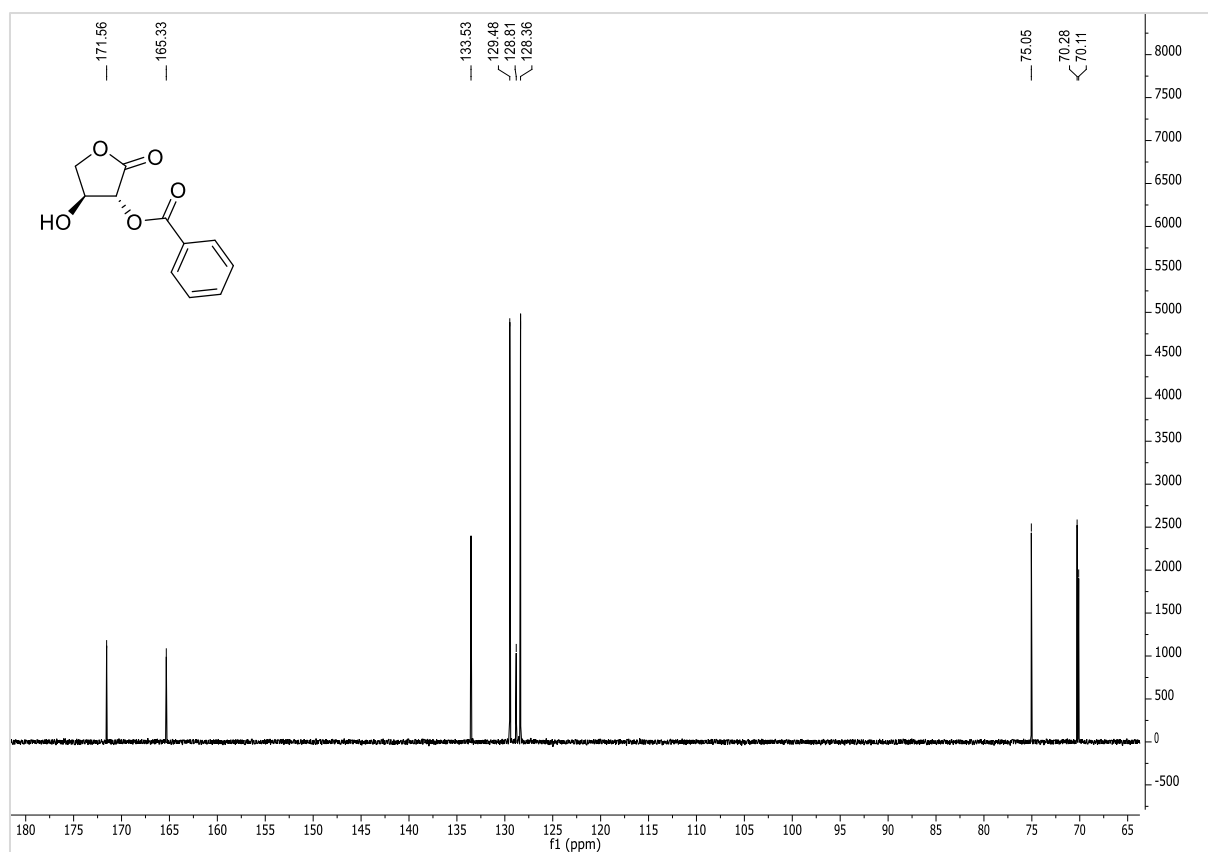


Figure S 42: <sup>13</sup>C-NMR spectrum of compound 14 (CD<sub>3</sub>OD, 126 MHz, r.t.).

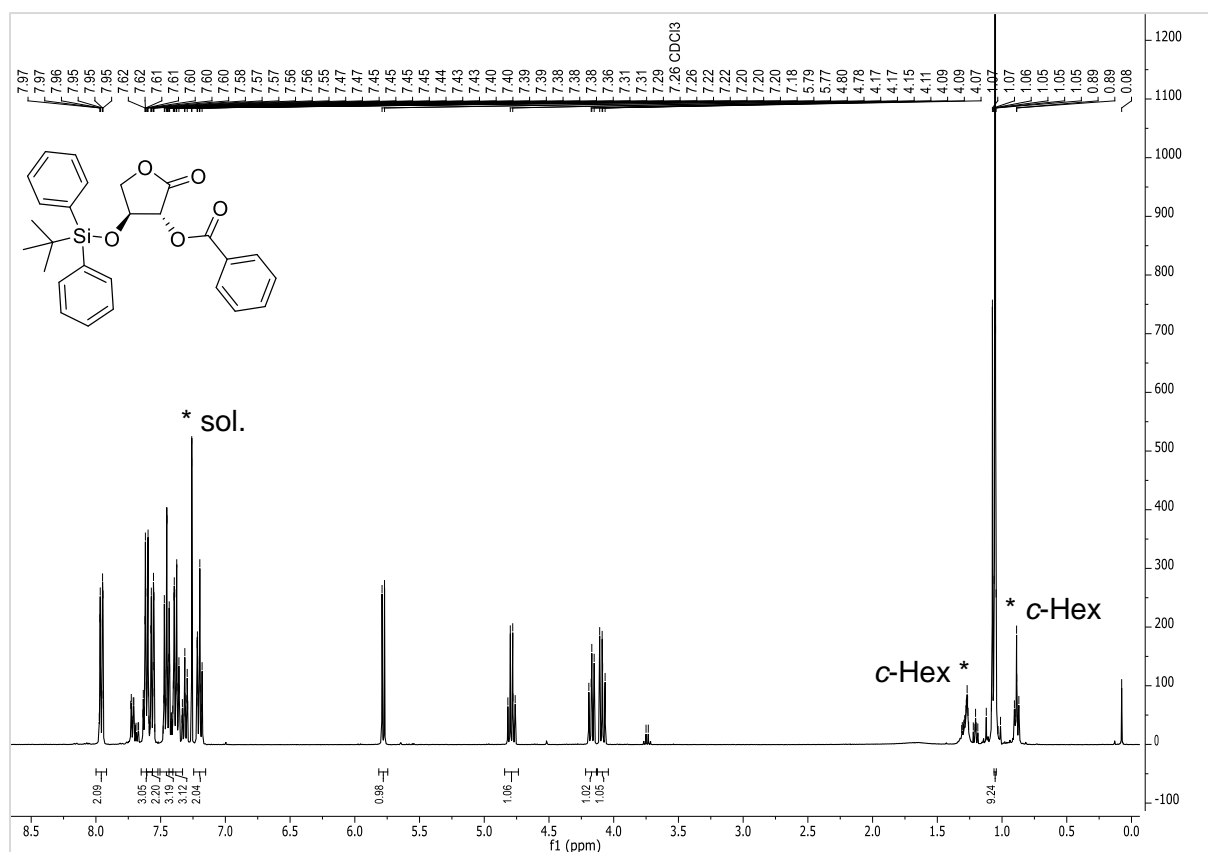
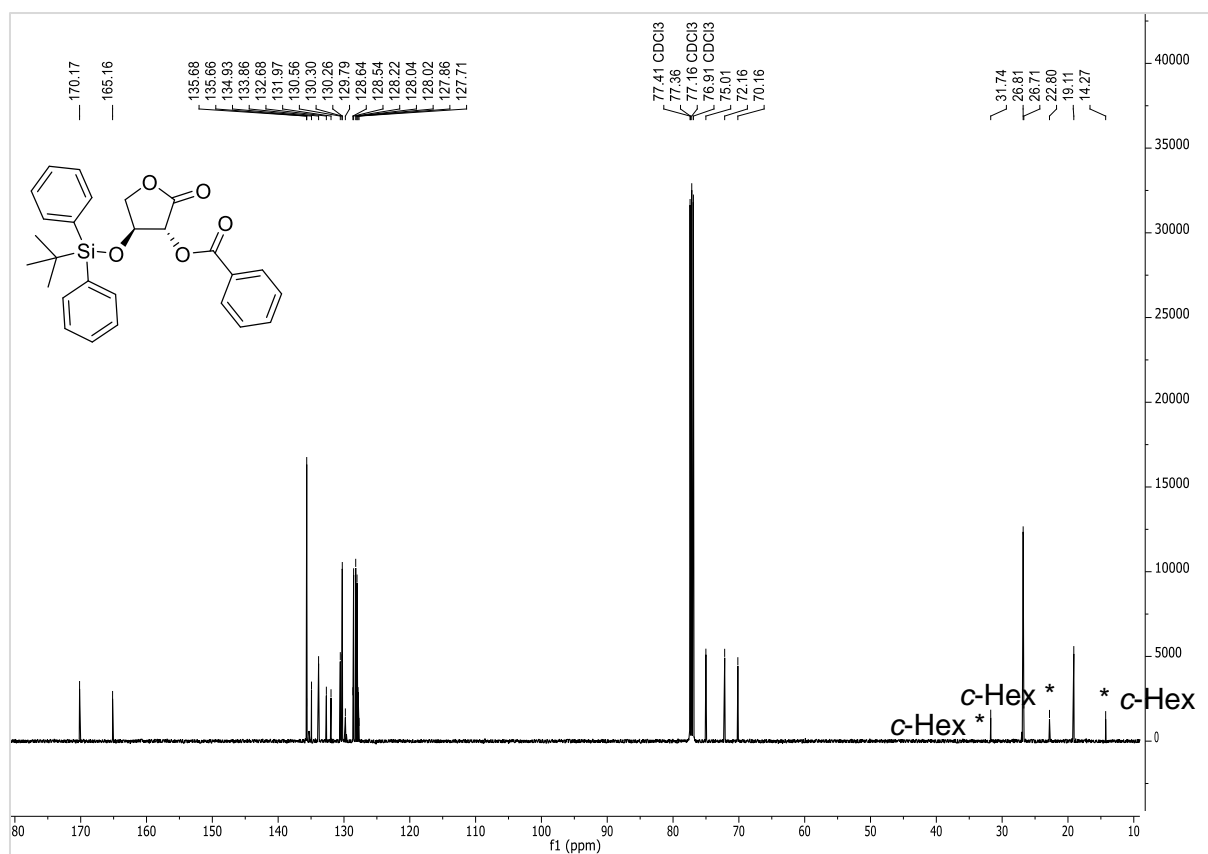
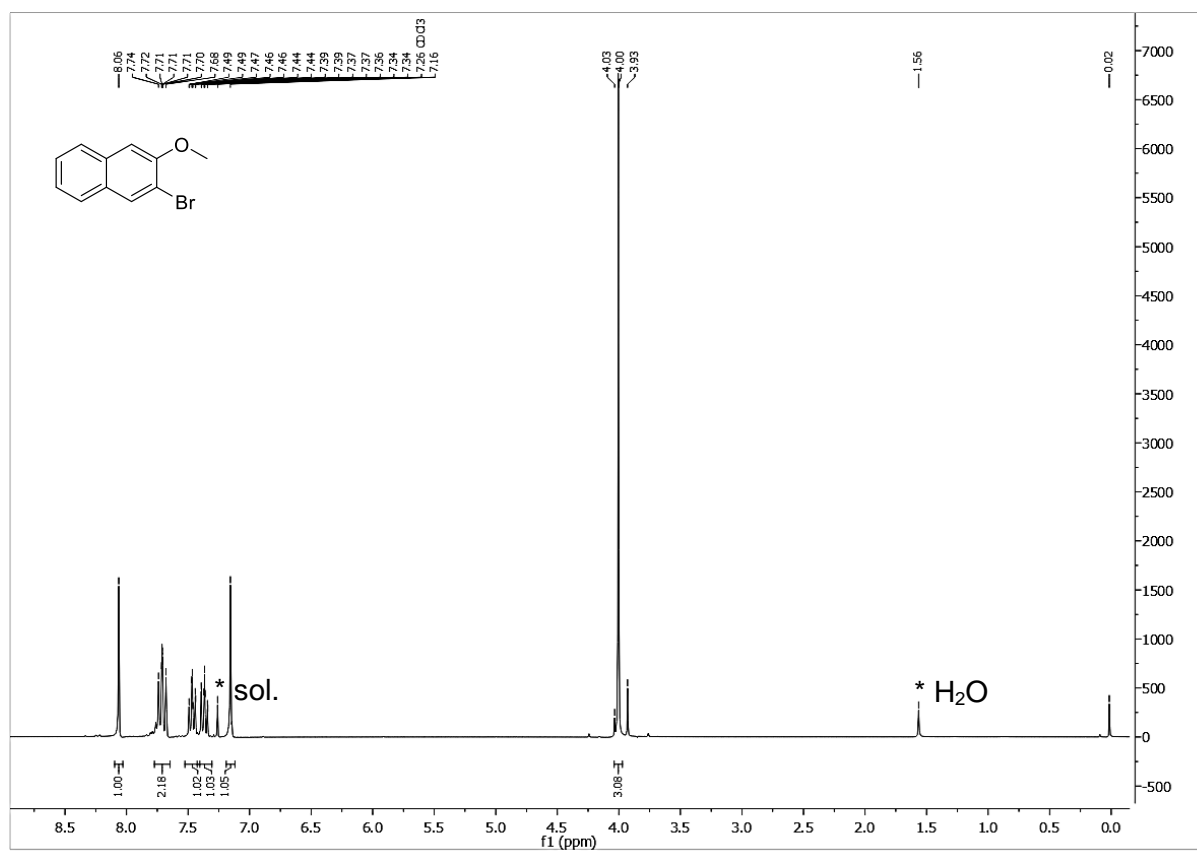


Figure S 43: <sup>1</sup>H-NMR spectrum of compound 12 (CDCl<sub>3</sub>, 400 MHz, r.t.).



**Figure S 44:**  $^{13}\text{C}$ -NMR spectrum of compound 12 ( $\text{CDCl}_3$ , 126 MHz, r.t.).



**Figure S 45:**  $^1\text{H}$ -NMR spectrum of compound 10 ( $\text{CDCl}_3$ , 300 MHz, r.t.).

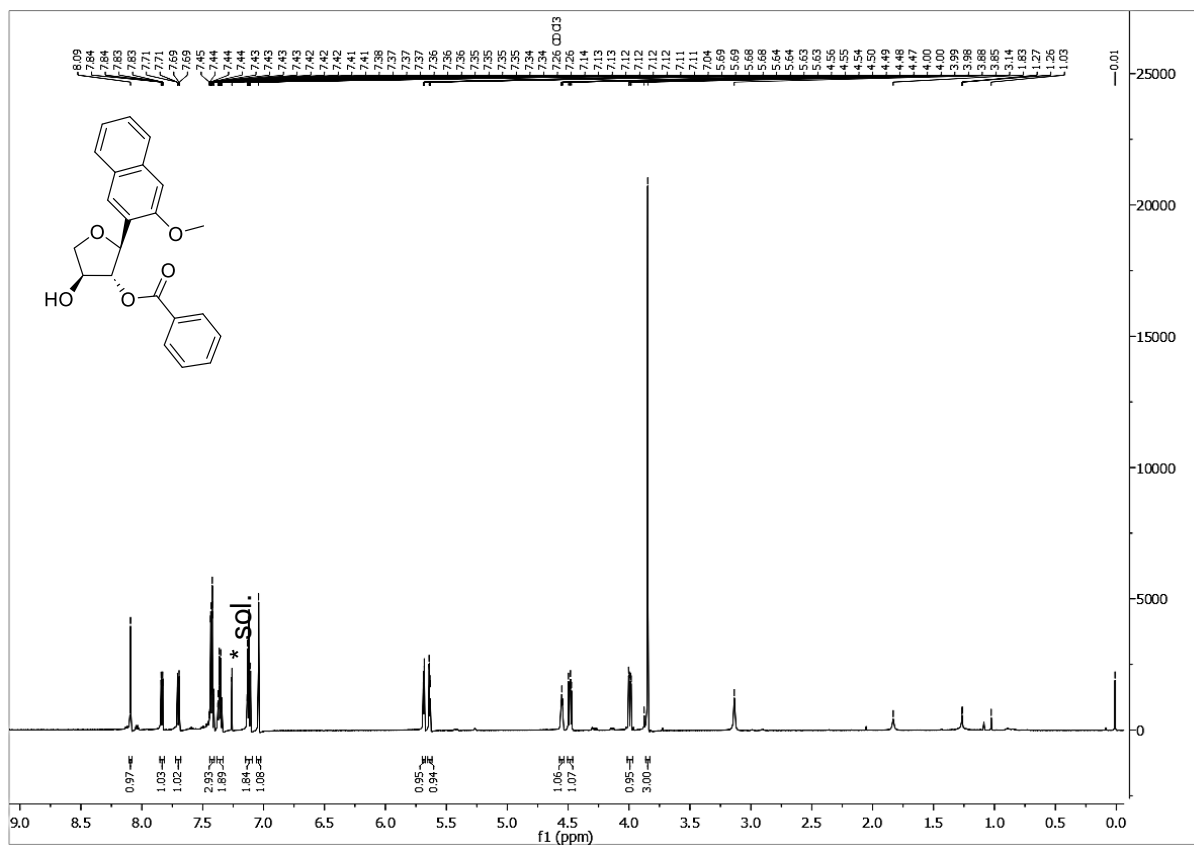


Figure S 46: <sup>1</sup>H-NMR spectrum of compound 16 (CDCl<sub>3</sub>, 600 MHz, r.t.).

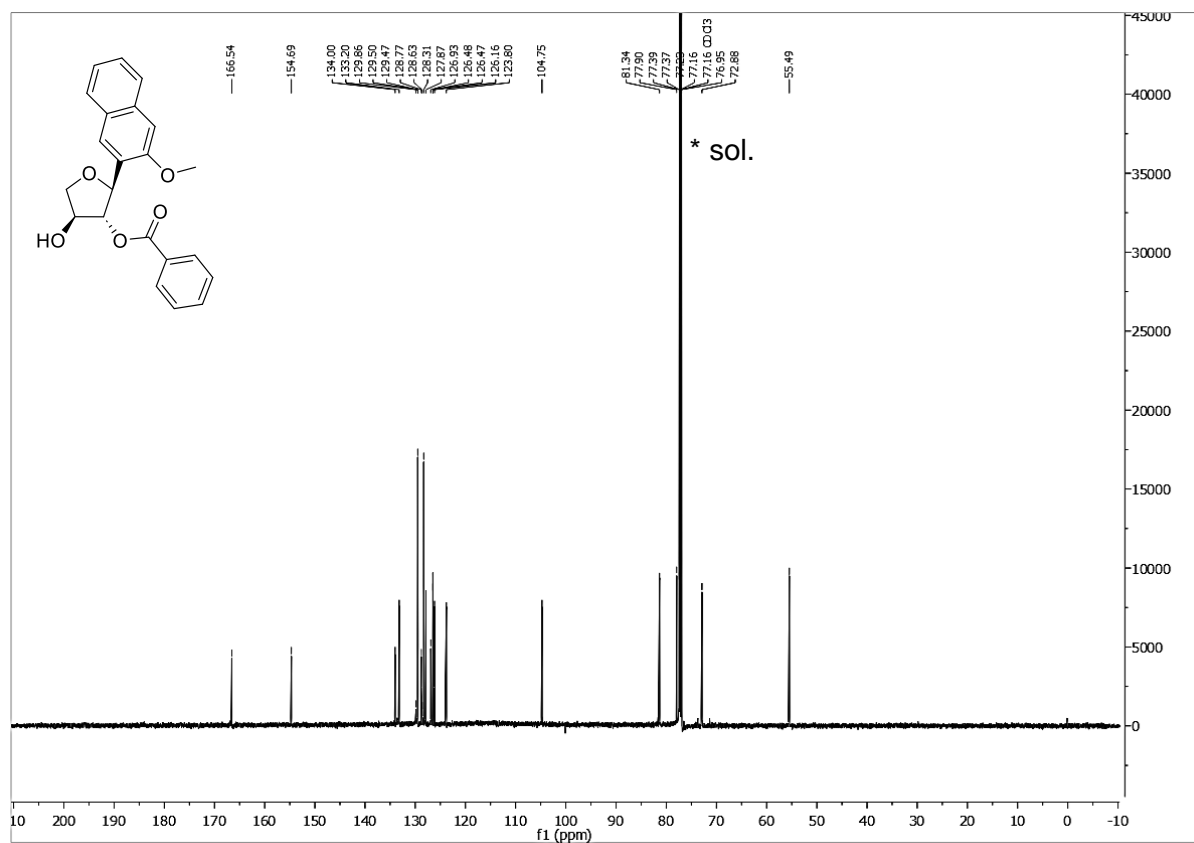


Figure S 47: <sup>13</sup>C-NMR spectrum of compound 16 (CDCl<sub>3</sub>, 151 MHz, r.t.).

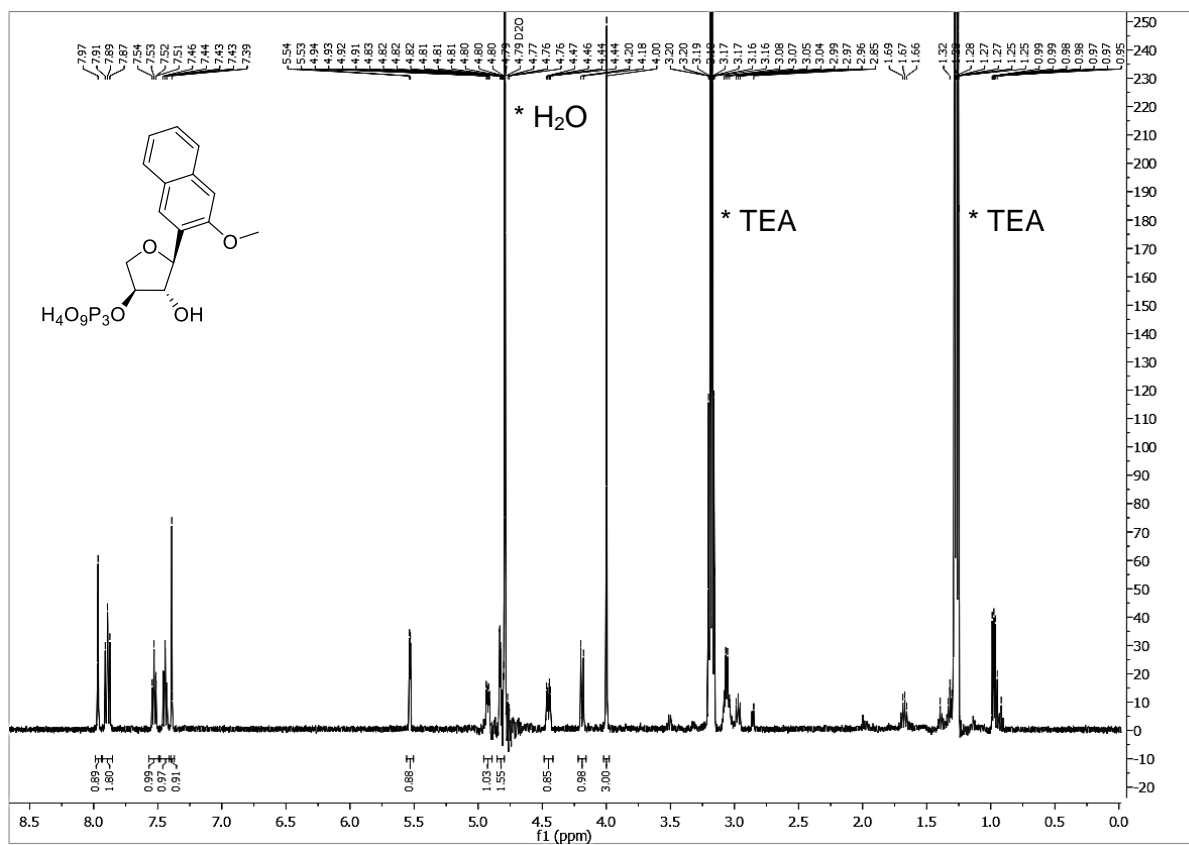


Figure S 48:  $^1\text{H-NMR}$  spectrum of compound 6 ( $\text{D}_2\text{O}$ , 500 MHz, r.t.).

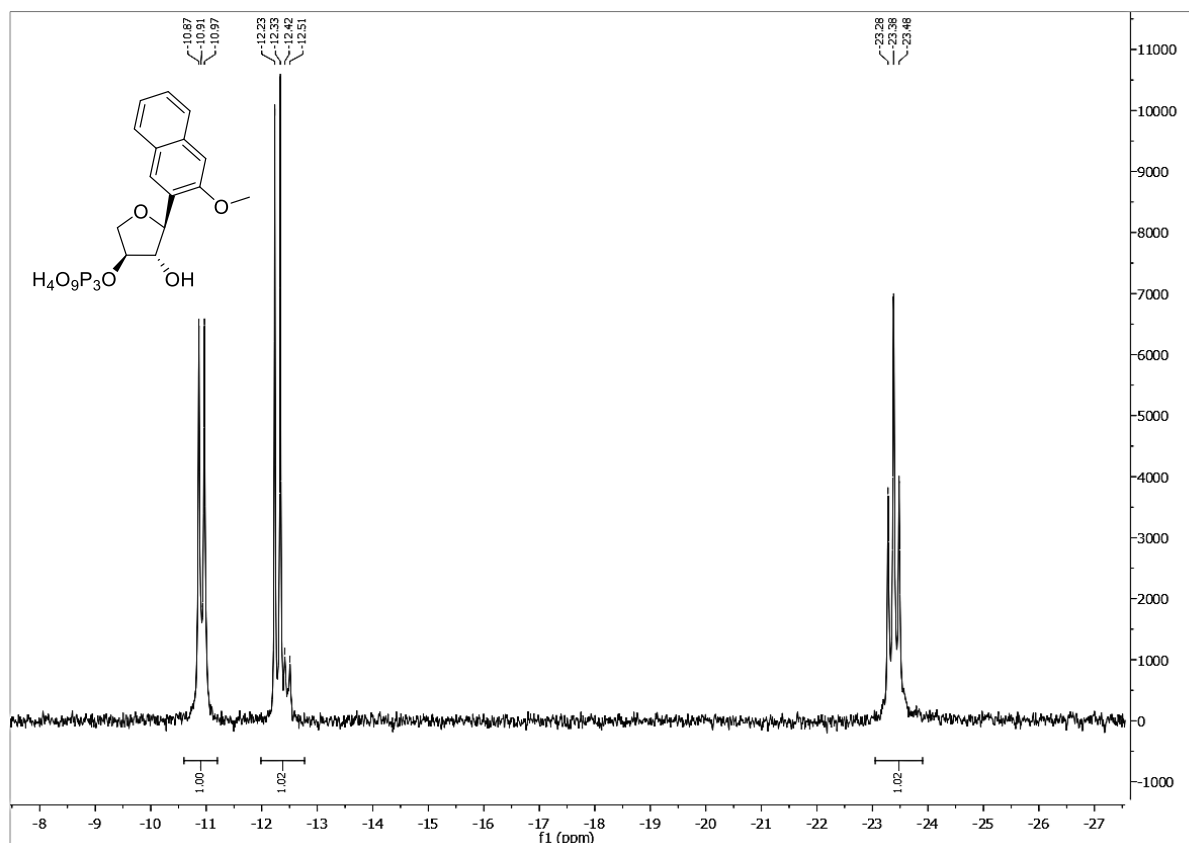
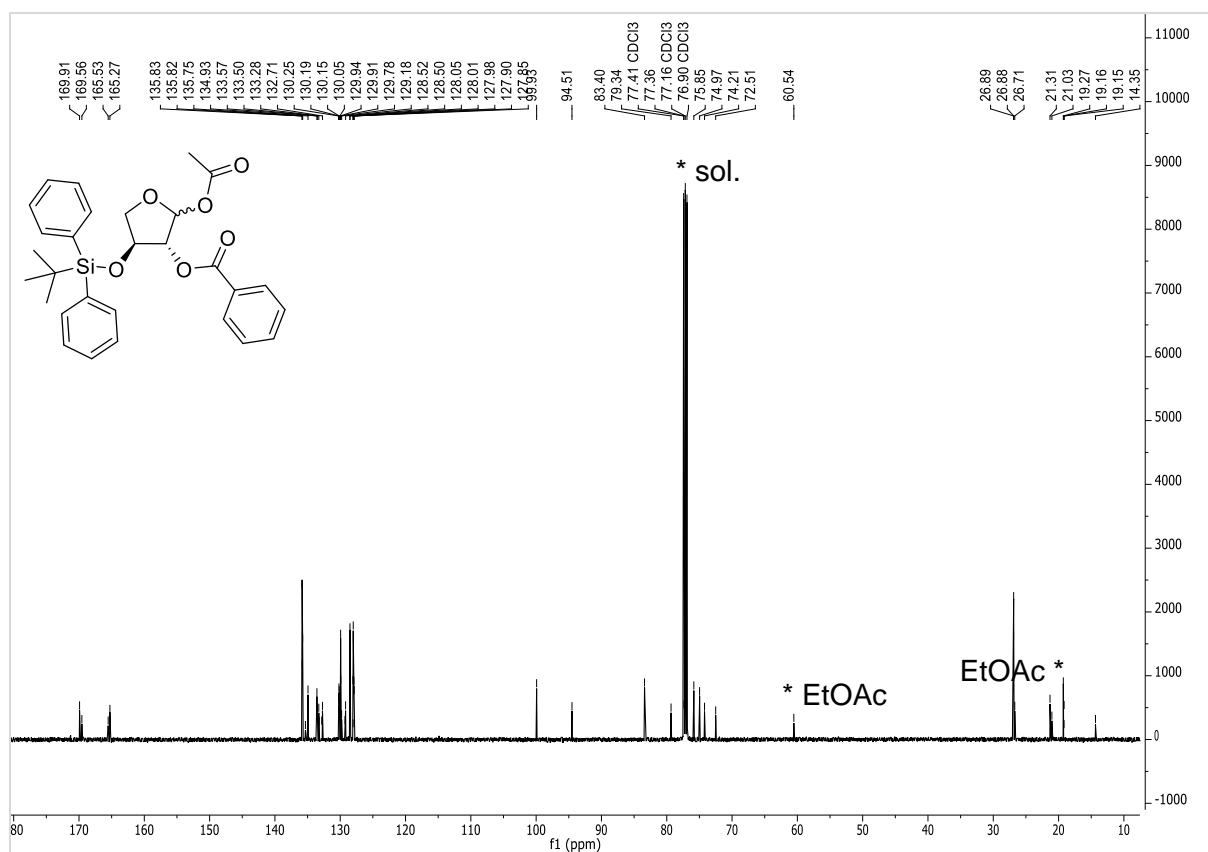
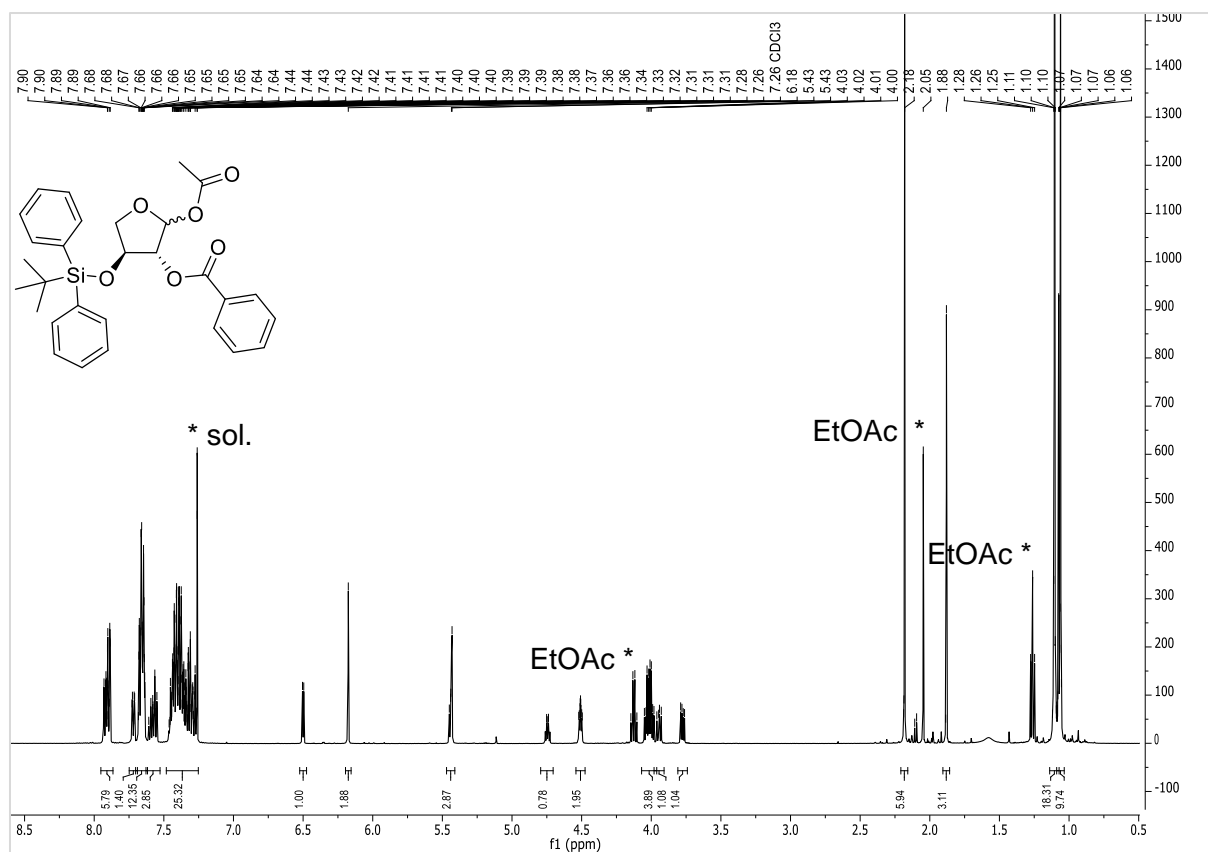


Figure S 49:  $^{31}\text{P-NMR}$  spectrum of compound 6 ( $\text{D}_2\text{O}$ , 202 MHz, r.t.).



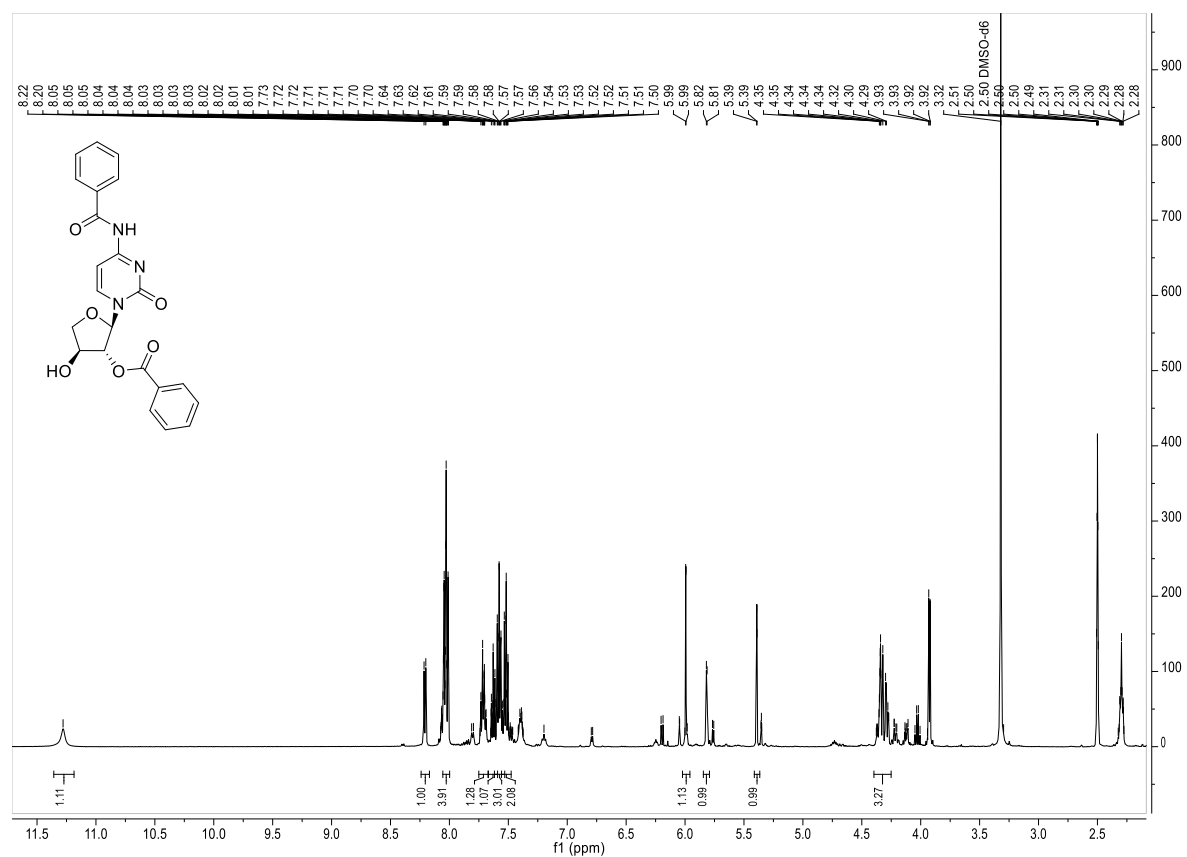


Figure S 52: <sup>1</sup>H-NMR spectrum of compound 24 (DMSO-*d*<sub>6</sub>, 500 MHz, r.t.).

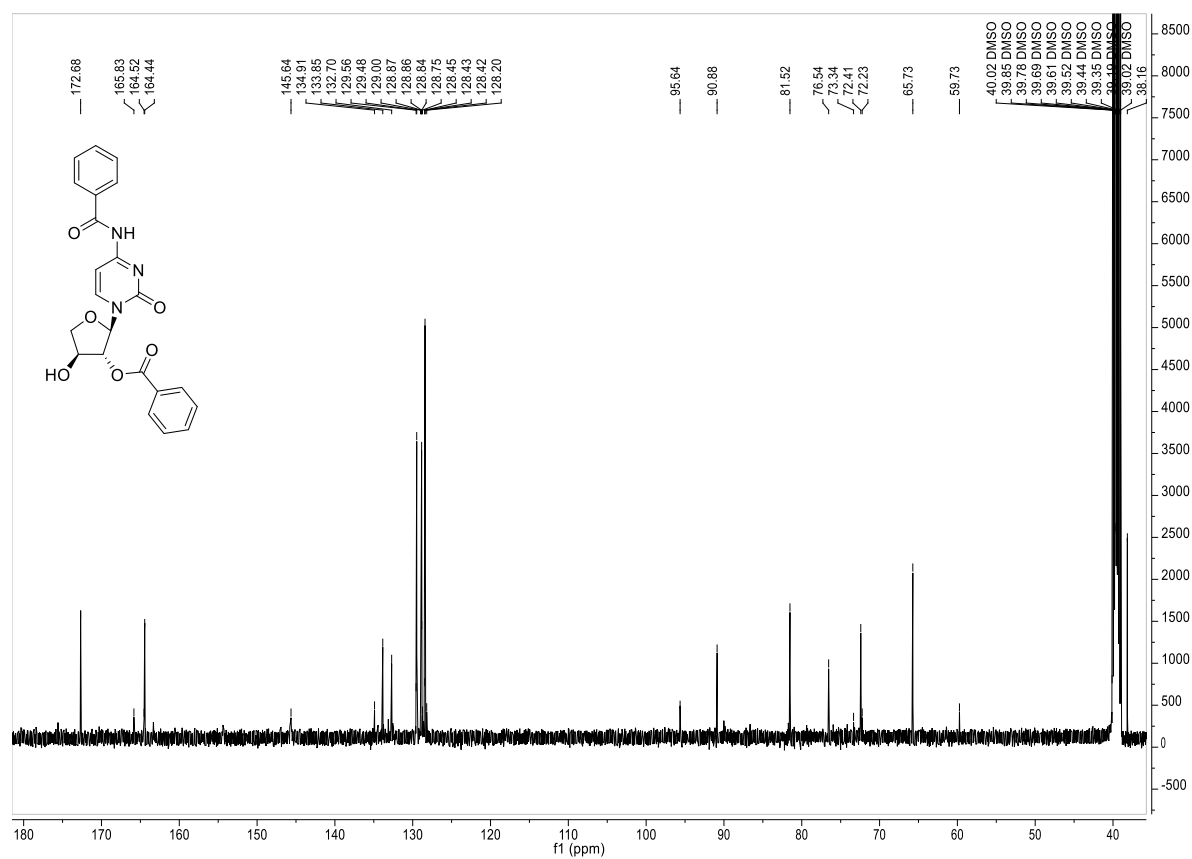


Figure S 53: <sup>13</sup>C-NMR spectrum of compound 24 (DMSO-*d*<sub>6</sub>, 126 MHz, r.t.).

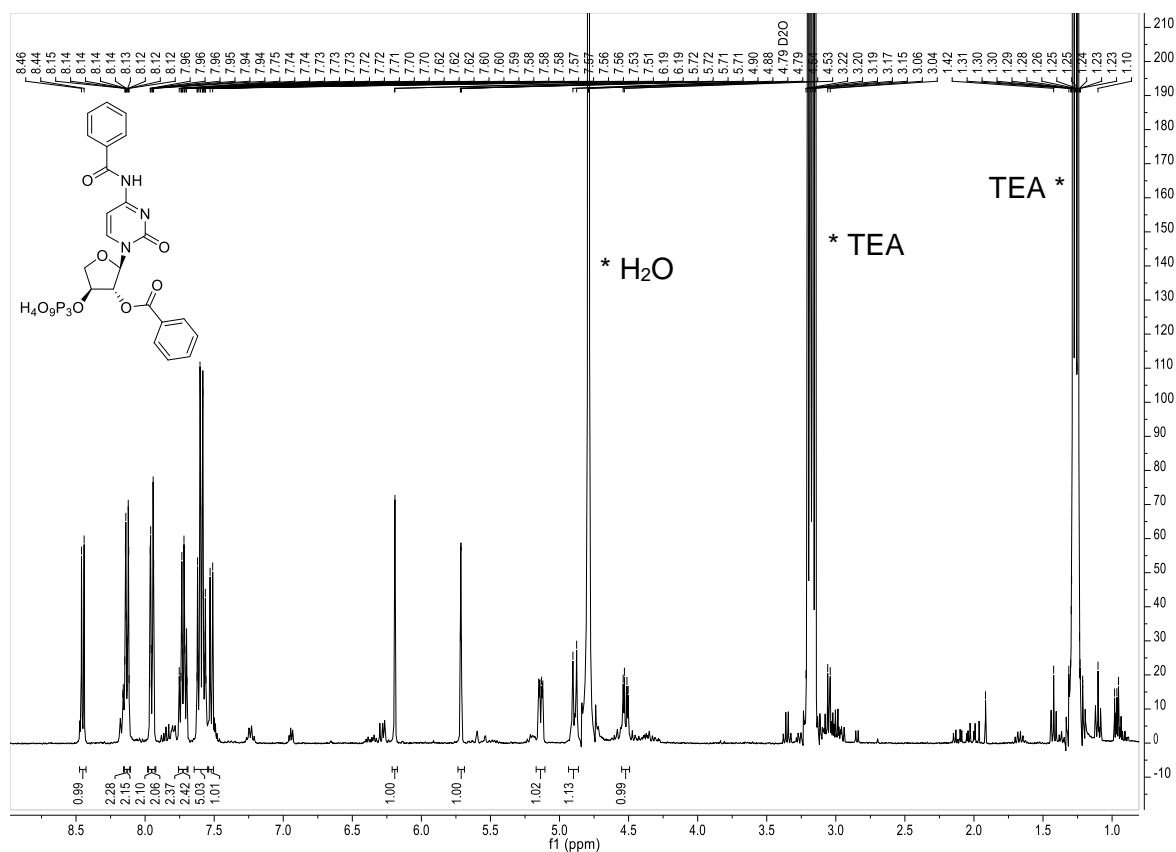


Figure S 54:  $^1\text{H-NMR}$  spectrum of compound 25 ( $\text{D}_2\text{O}$ , 400 MHz, r.t.).

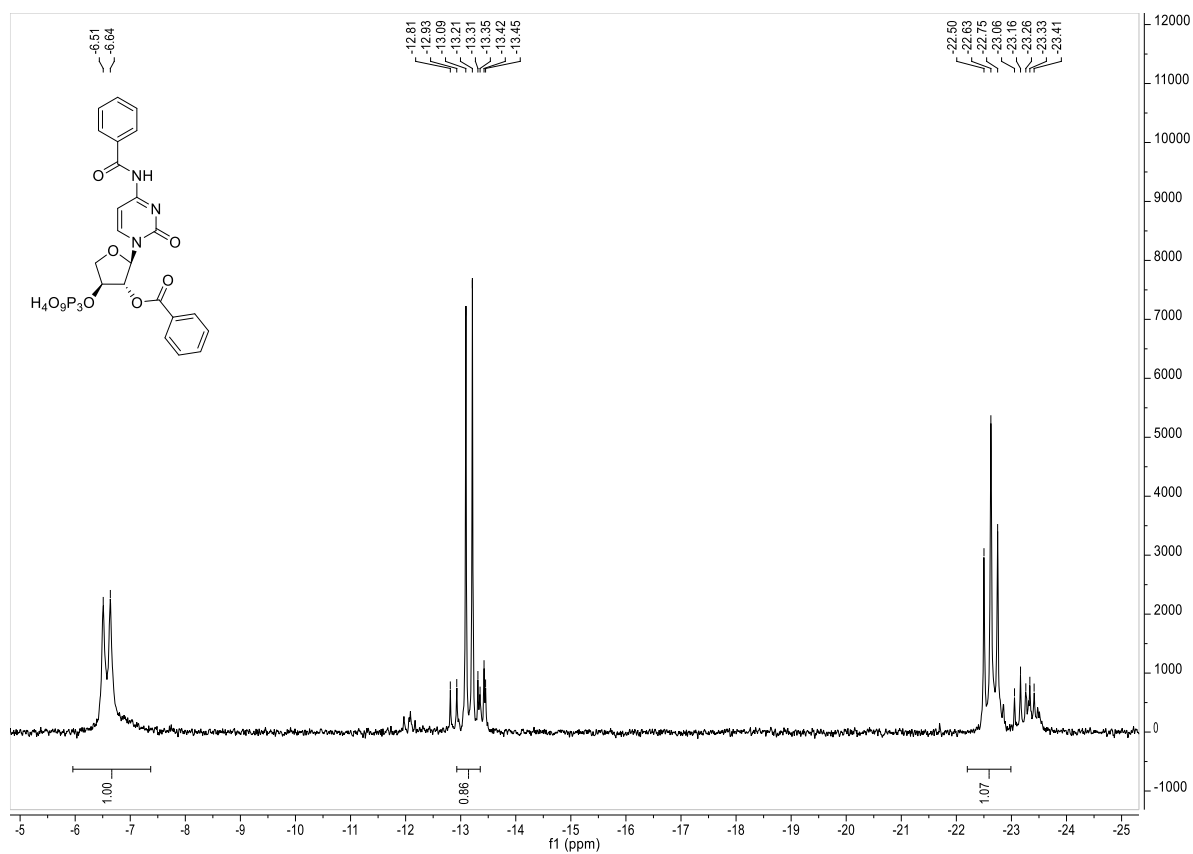


Figure S 55:  $^{31}\text{P-NMR}$  spectrum of compound 25 ( $\text{D}_2\text{O}$ , 162 MHz, r.t.).



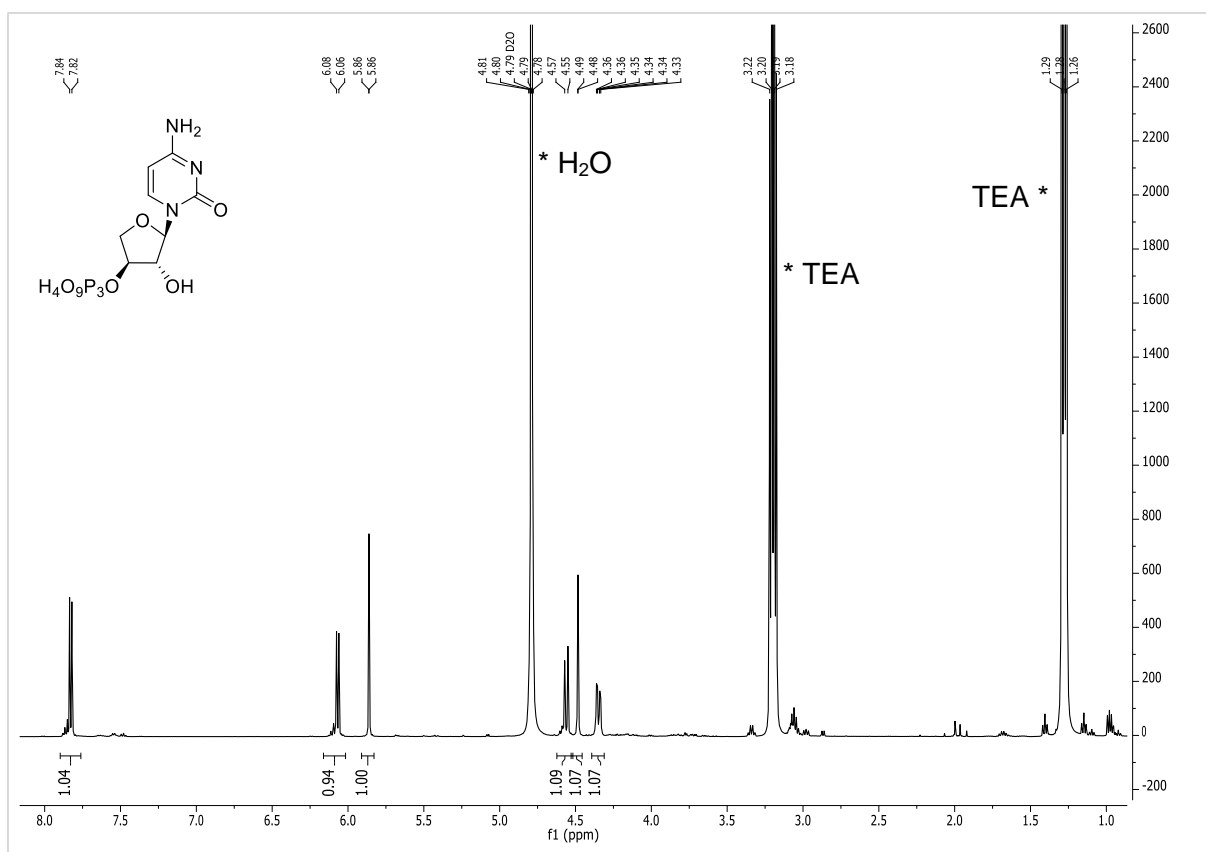


Figure S 56: <sup>1</sup>H-NMR Spectrum of compound 7 (D<sub>2</sub>O, 500 MHz, r.t.).

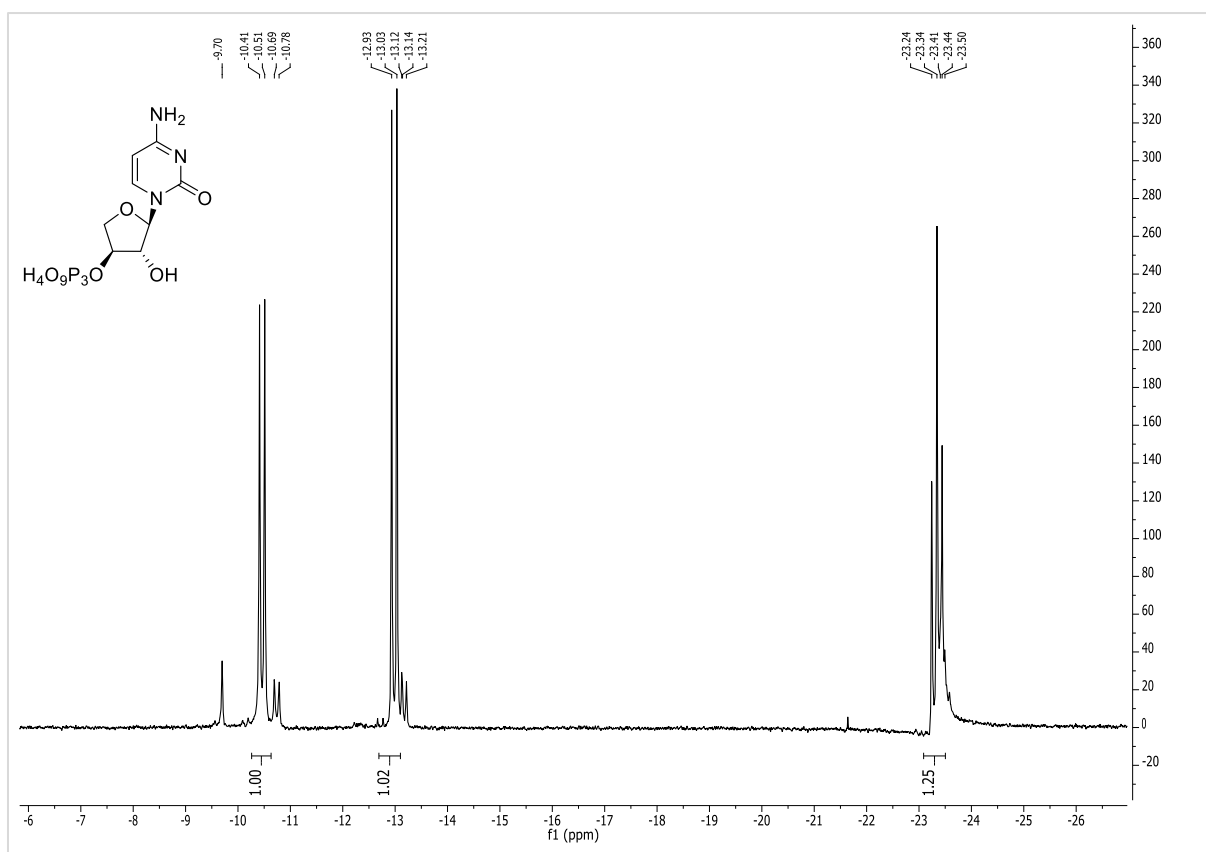
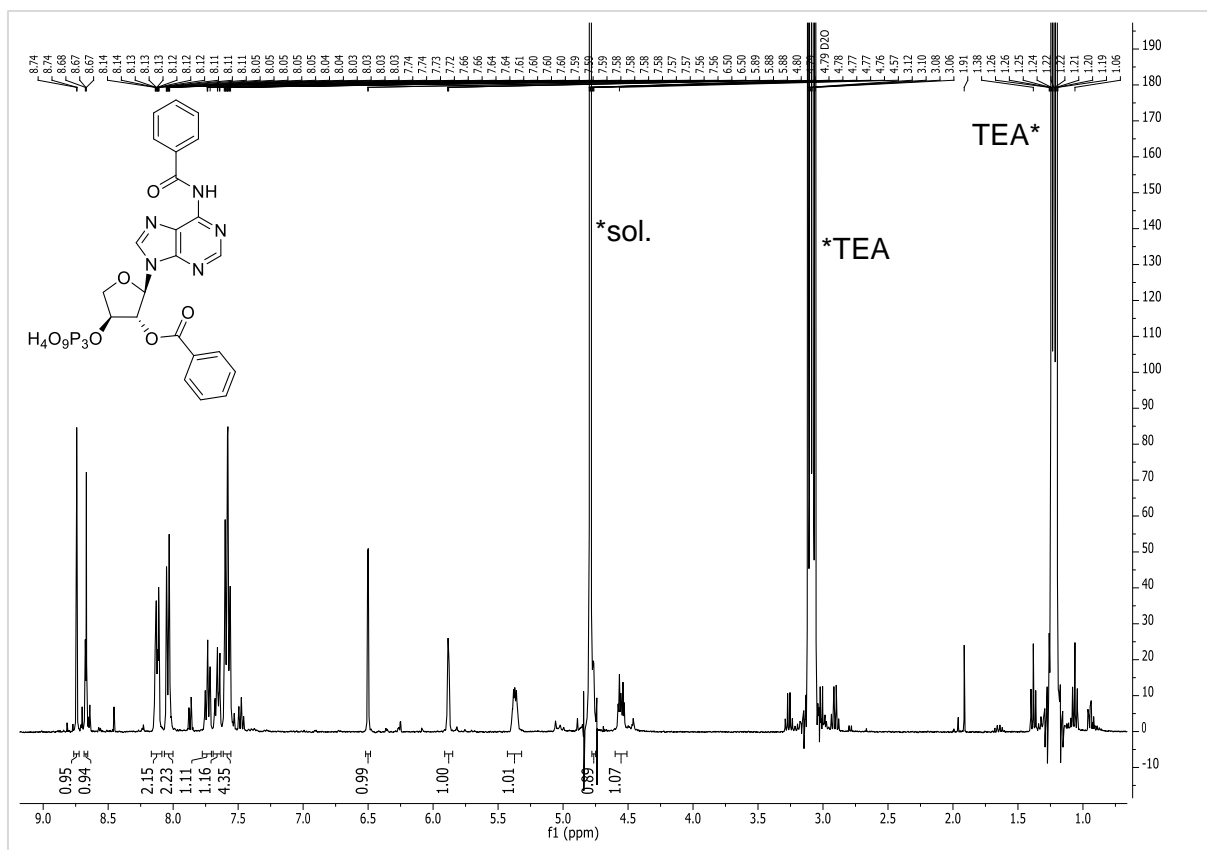
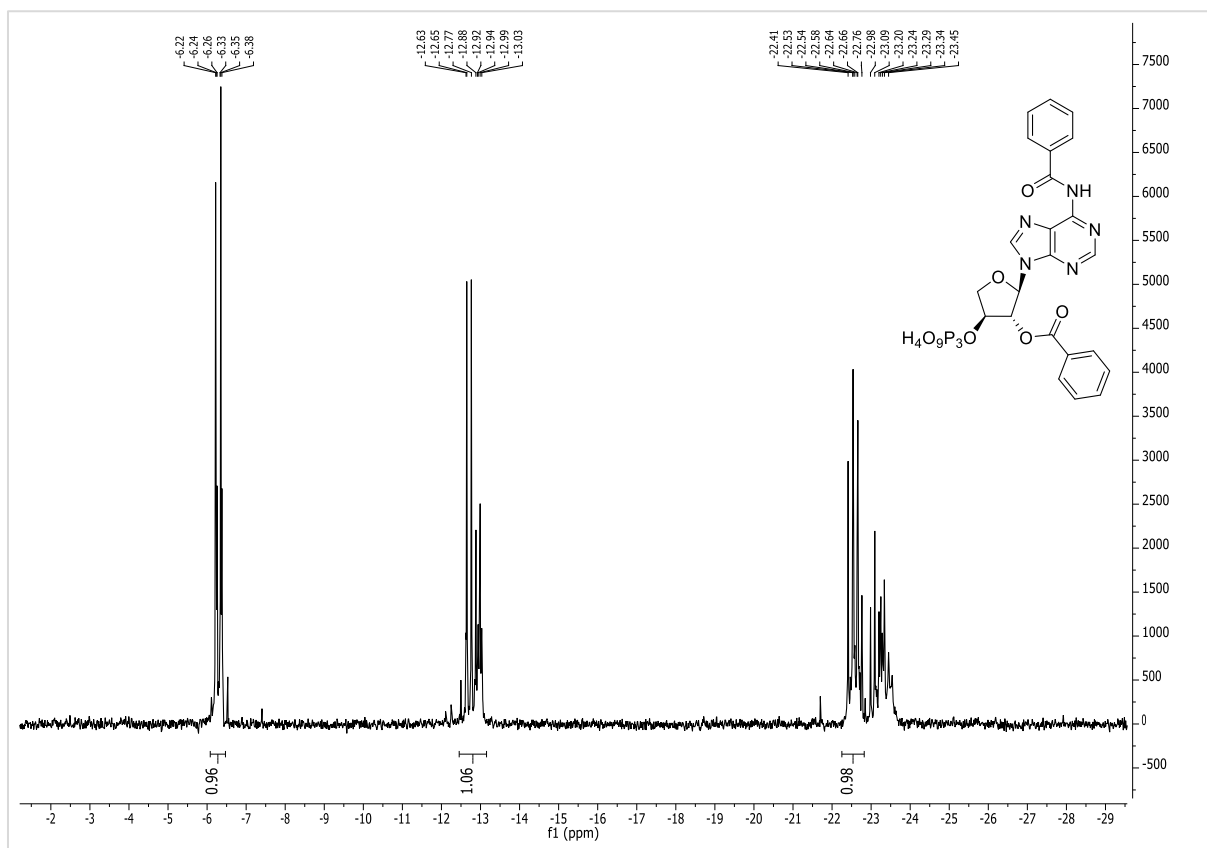


Figure S 57: <sup>31</sup>P-NMR Spectrum of compound 7 (D<sub>2</sub>O, 202 MHz, r.t.).



Figure S 60:  $^1\text{H-NMR}$  spectrum of compound 29 ( $\text{D}_2\text{O}$ , 400 Hz, r.t.).Figure S 61:  $^{31}\text{P-NMR}$  spectrum of compound 29 ( $\text{D}_2\text{O}$ , 126 Hz, r.t.).

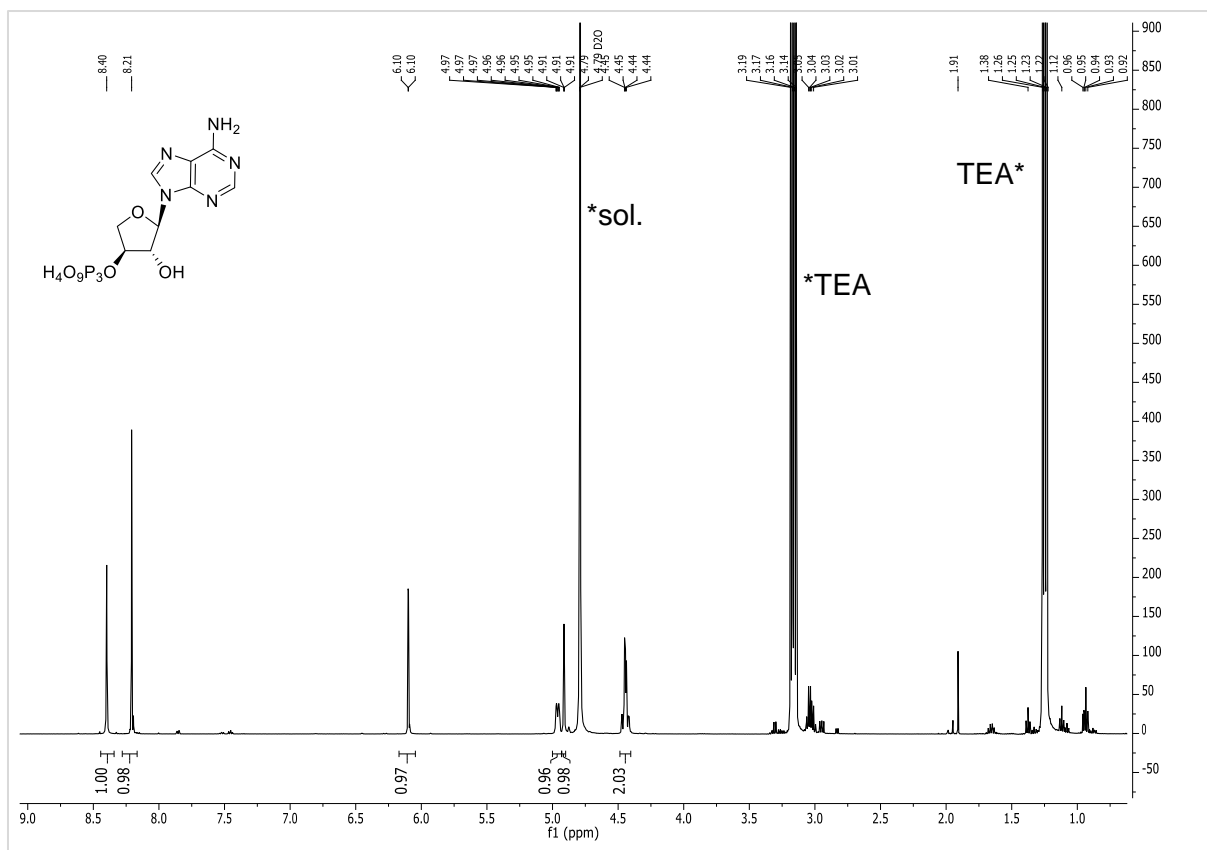


Figure S 62: <sup>1</sup>H-NMR spectrum of compound 8 (D<sub>2</sub>O, 500 MHz, r.t.).

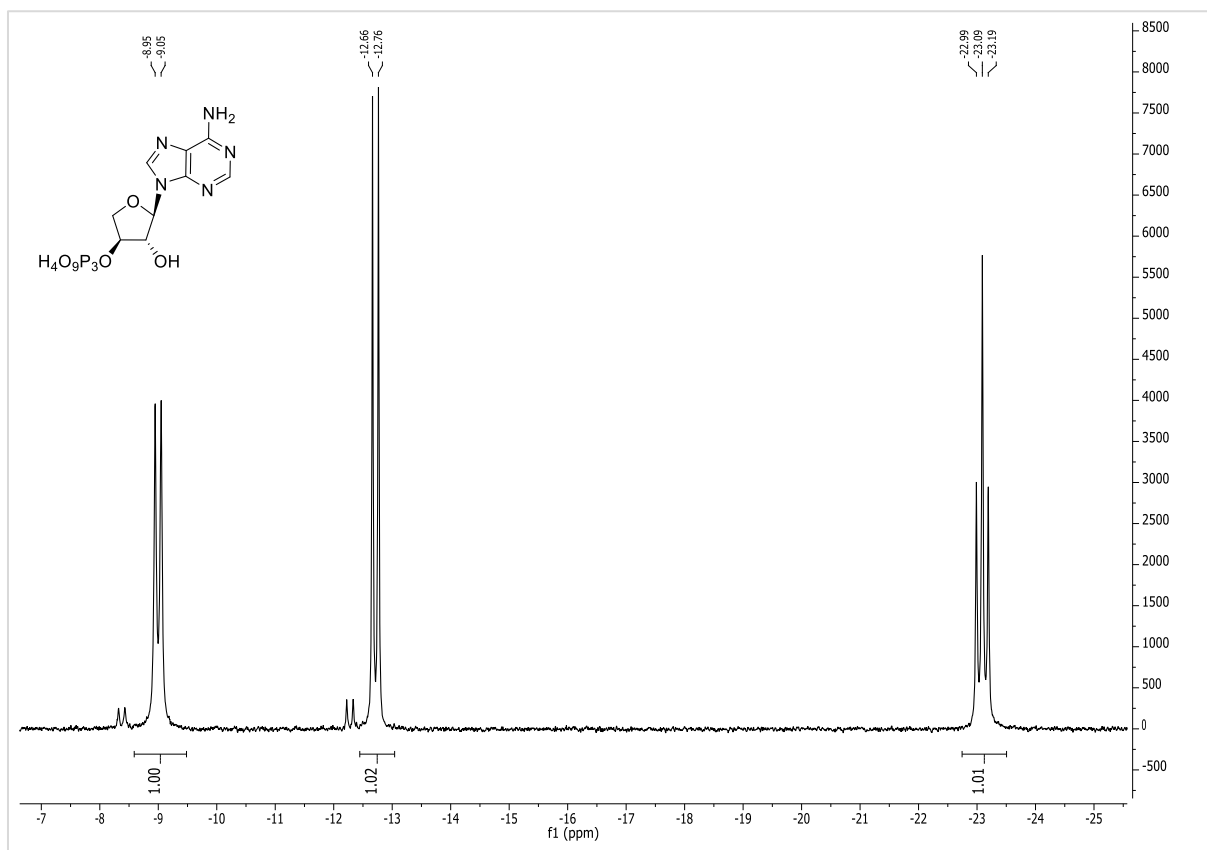


Figure S 63: <sup>31</sup>P-NMR spectrum of compound 8 (D<sub>2</sub>O, 202 MHz, r.t.).

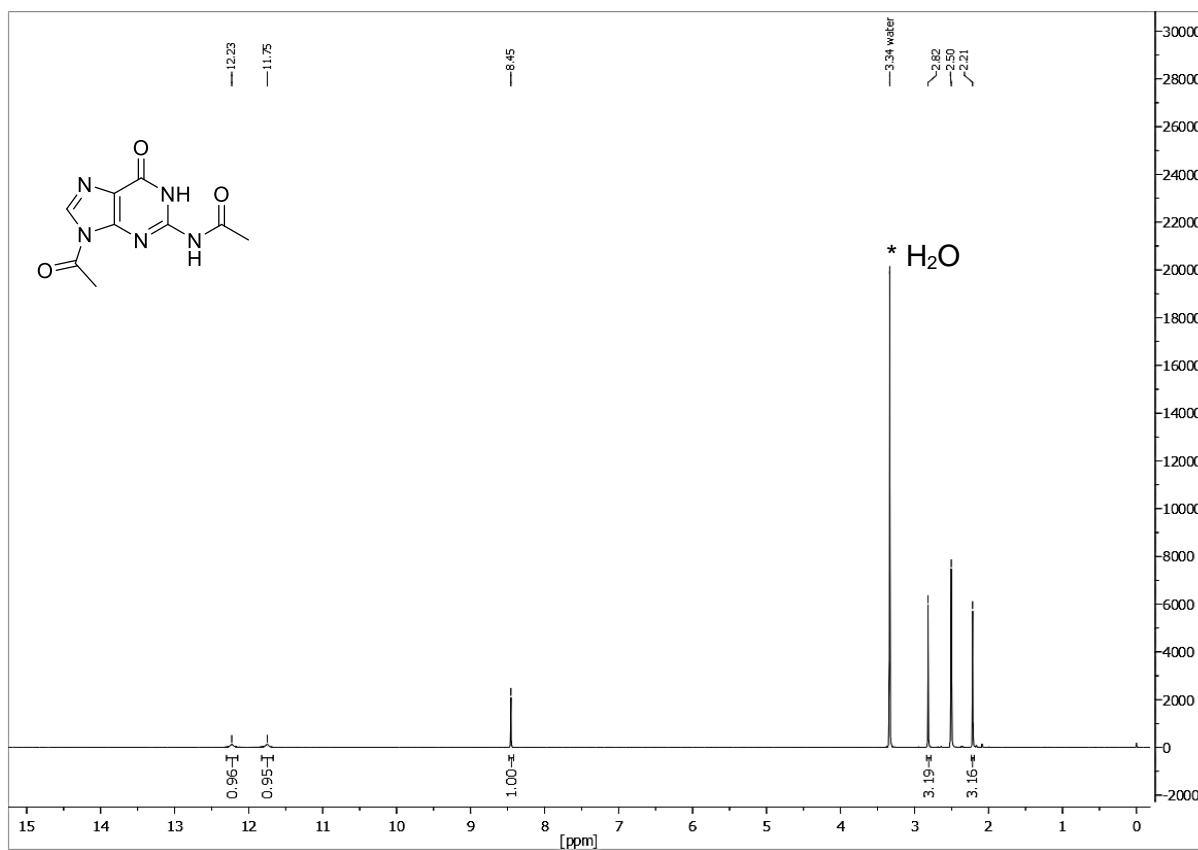


Figure S 64:  $^1\text{H-NMR}$  spectrum of compound 31 (DMSO- $d_6$ , 499 MHz, r.t.).

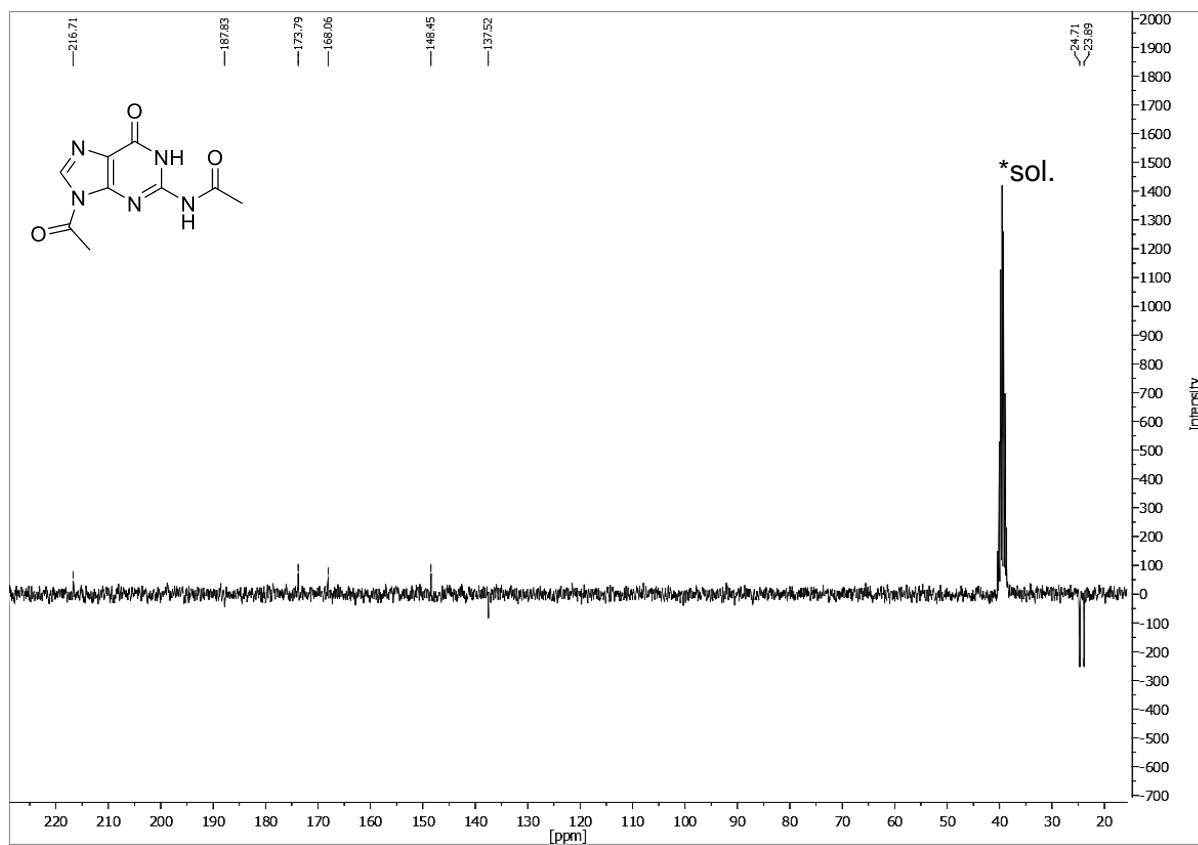


Figure S 65:  $^{13}\text{C-NMR}$  spectrum of compound 31 (DMSO- $d_6$ , 75 MHz, r.t.).

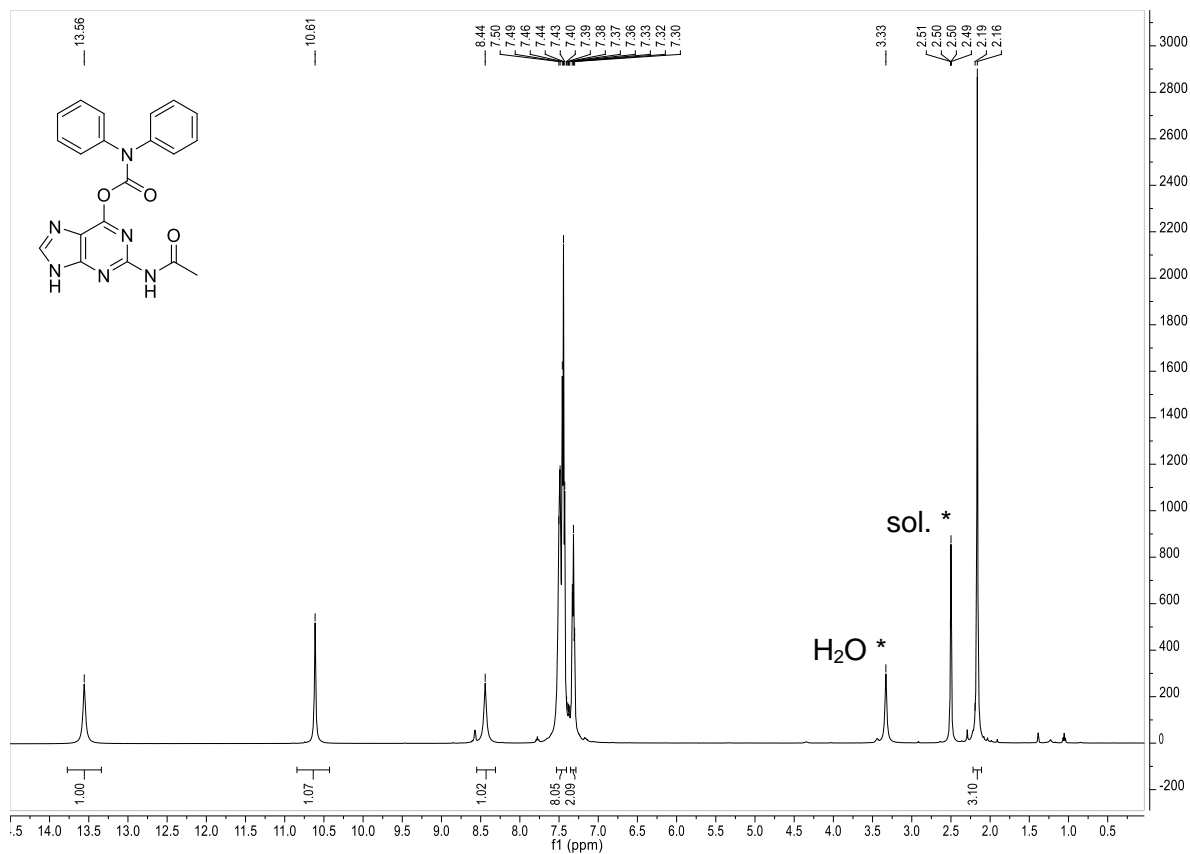


Figure S 66: <sup>1</sup>H-NMR spectrum of compound 32 (DMSO-*d*<sub>6</sub>, 500 MHz, r.t.).

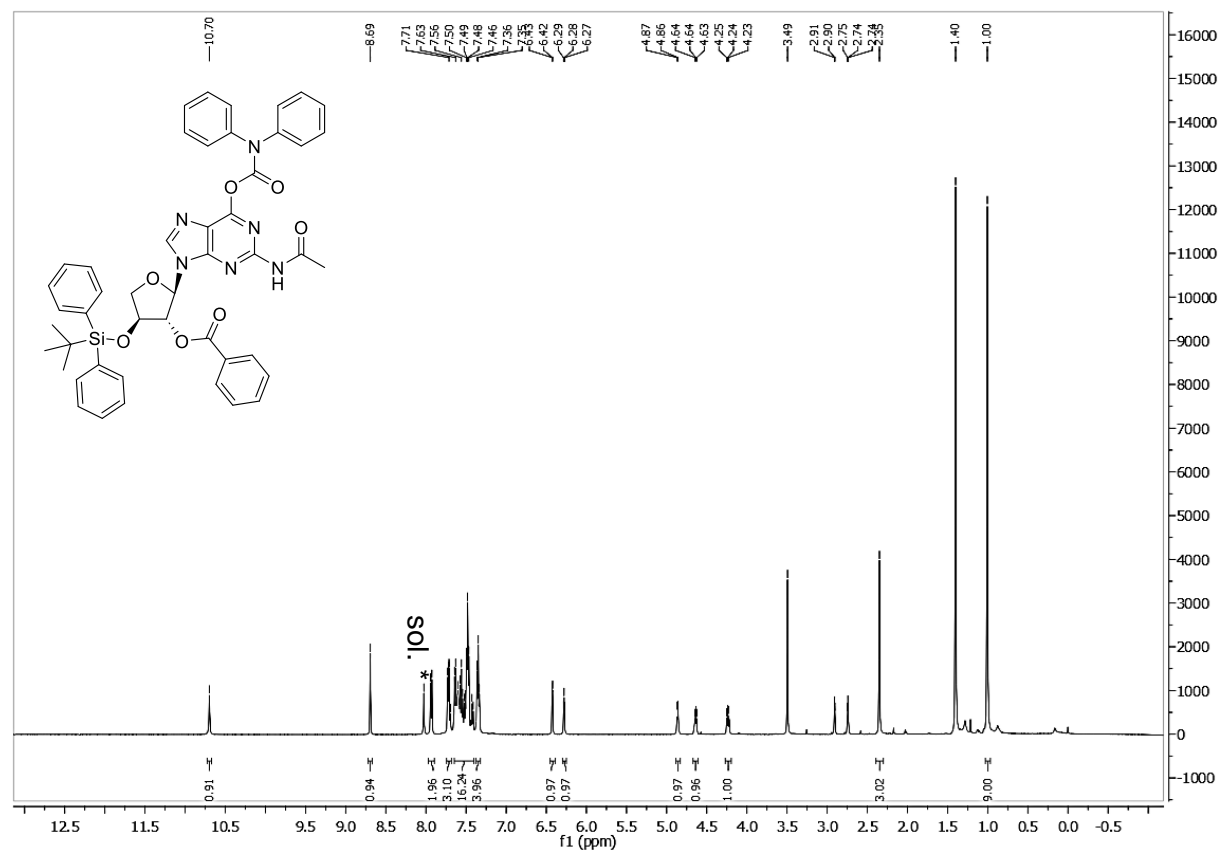


Figure S 67: <sup>1</sup>H-NMR spectrum of compound 33 (DMF-*d*<sub>2</sub>, 500 MHz, r.t.).

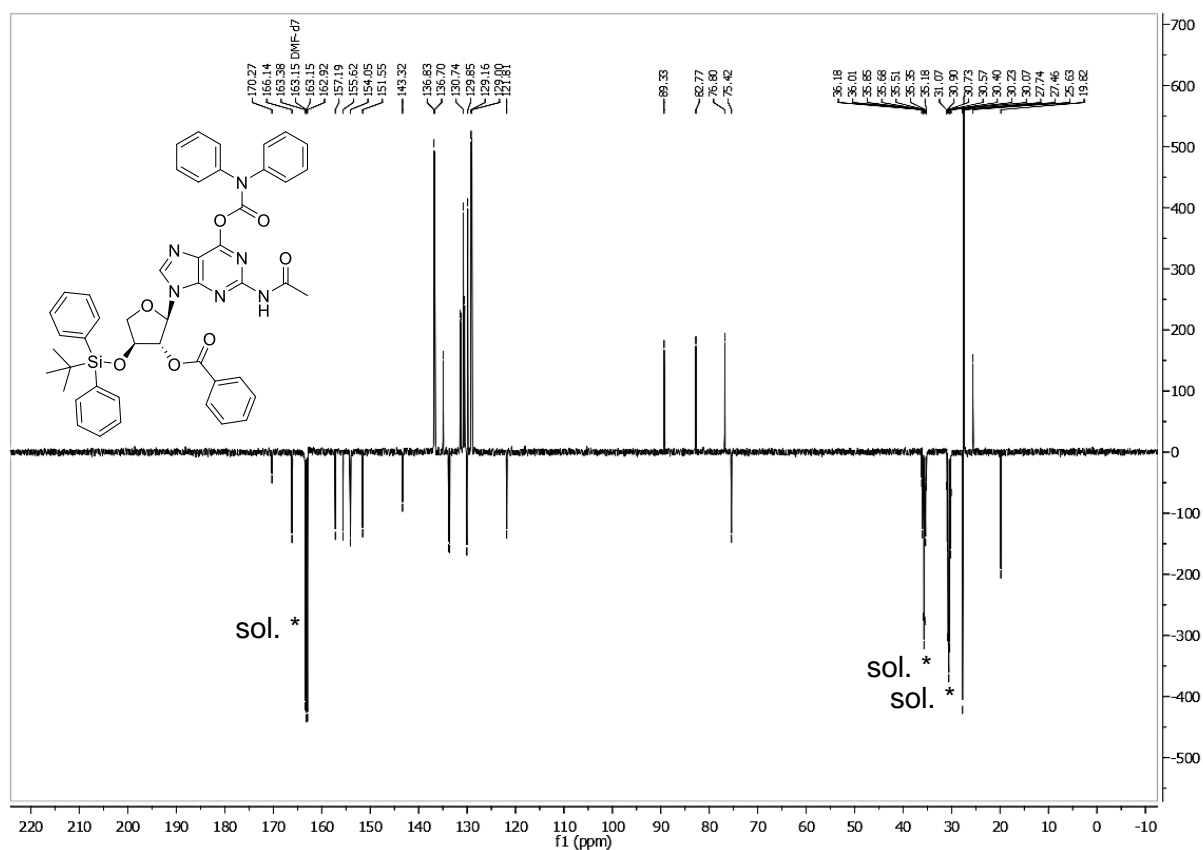


Figure S 68:  $^{13}\text{C}$ -NMR spectrum of compound 33 (DMF- $d_7$ , 126 MHz, r.t.).

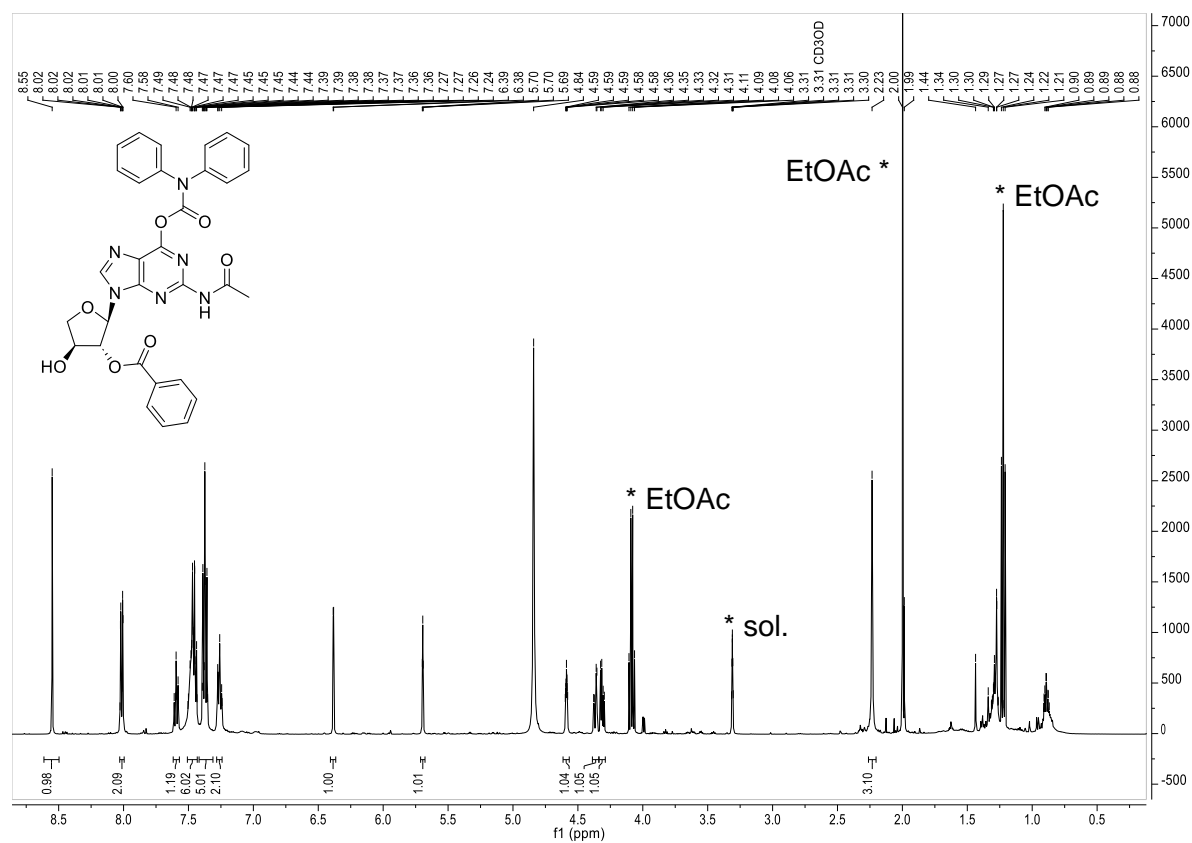


Figure S 69:  $^1\text{H}$ -NMR spectrum of compound 34 (CD $_3$ OD, 500 MHz, r.t.).

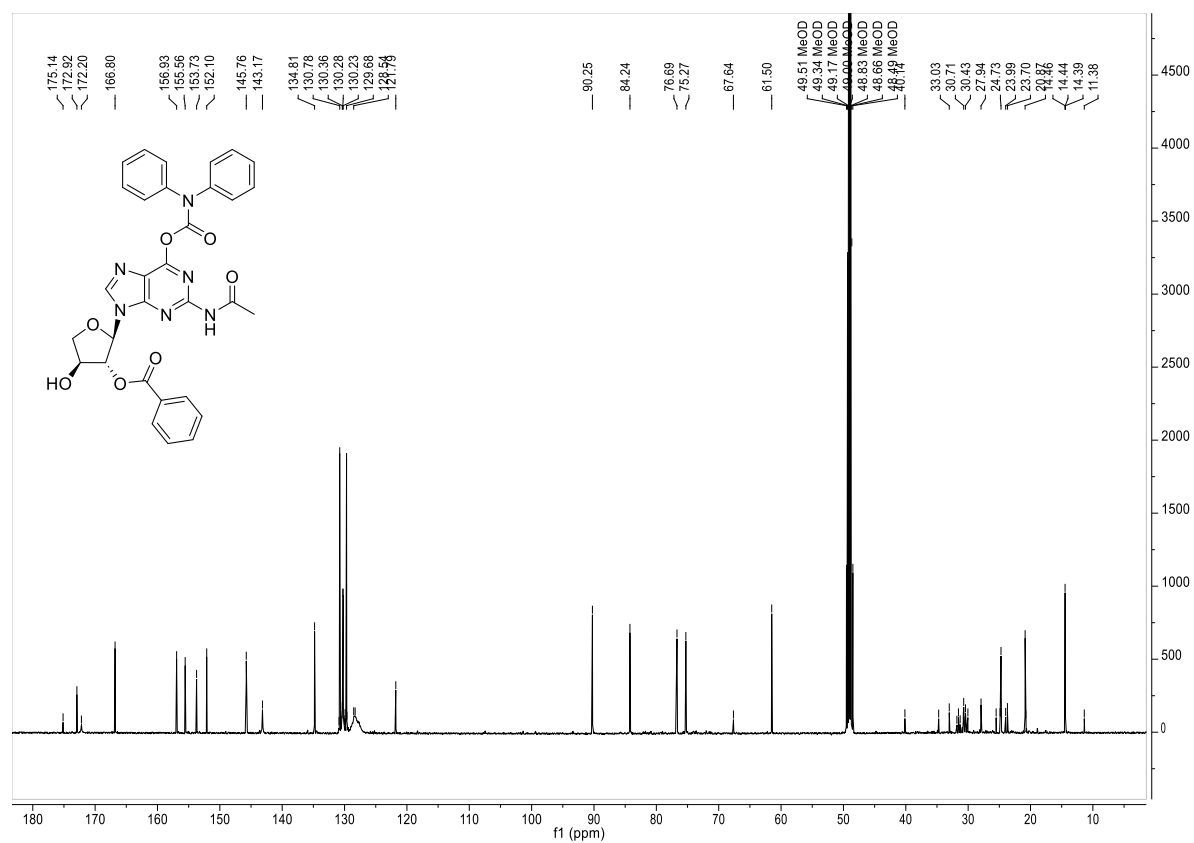


Figure S 70: <sup>13</sup>C-NMR spectrum of compound 34 (CD<sub>3</sub>OD, 126 MHz, r.t.).

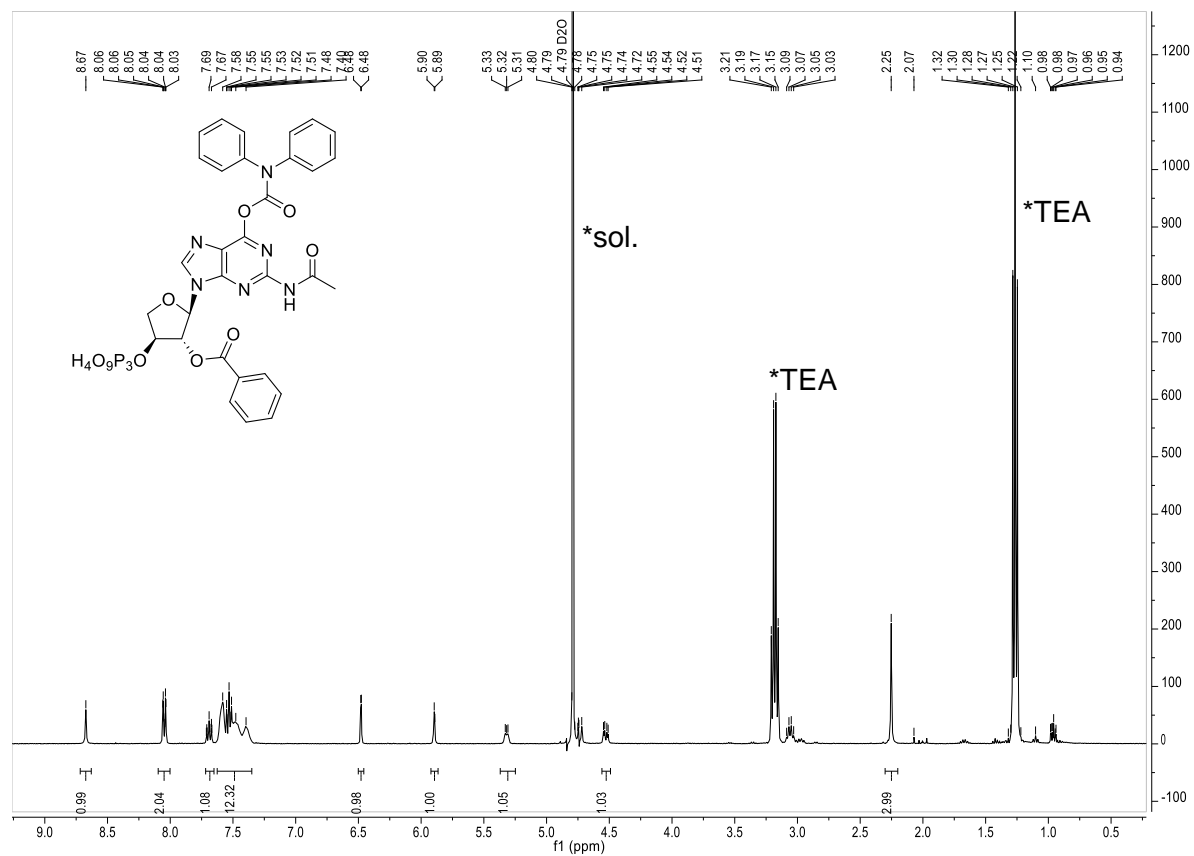
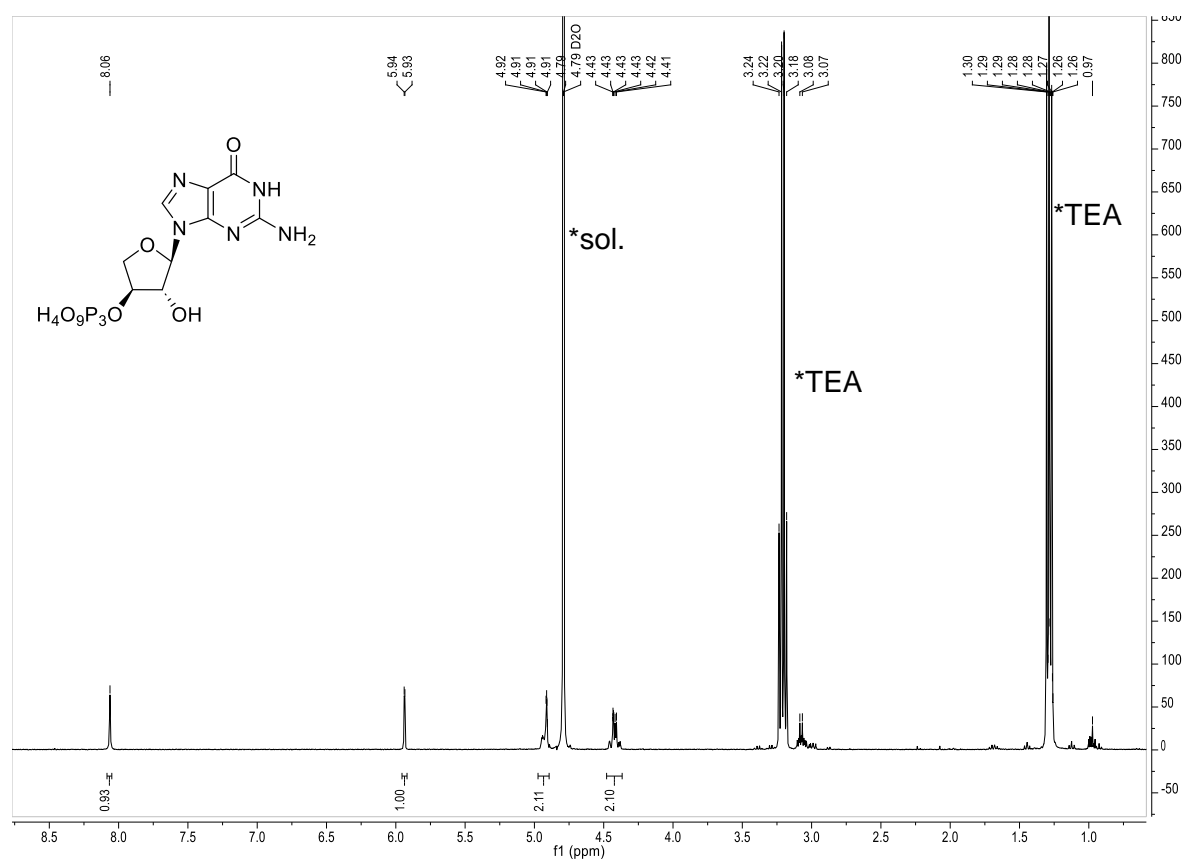
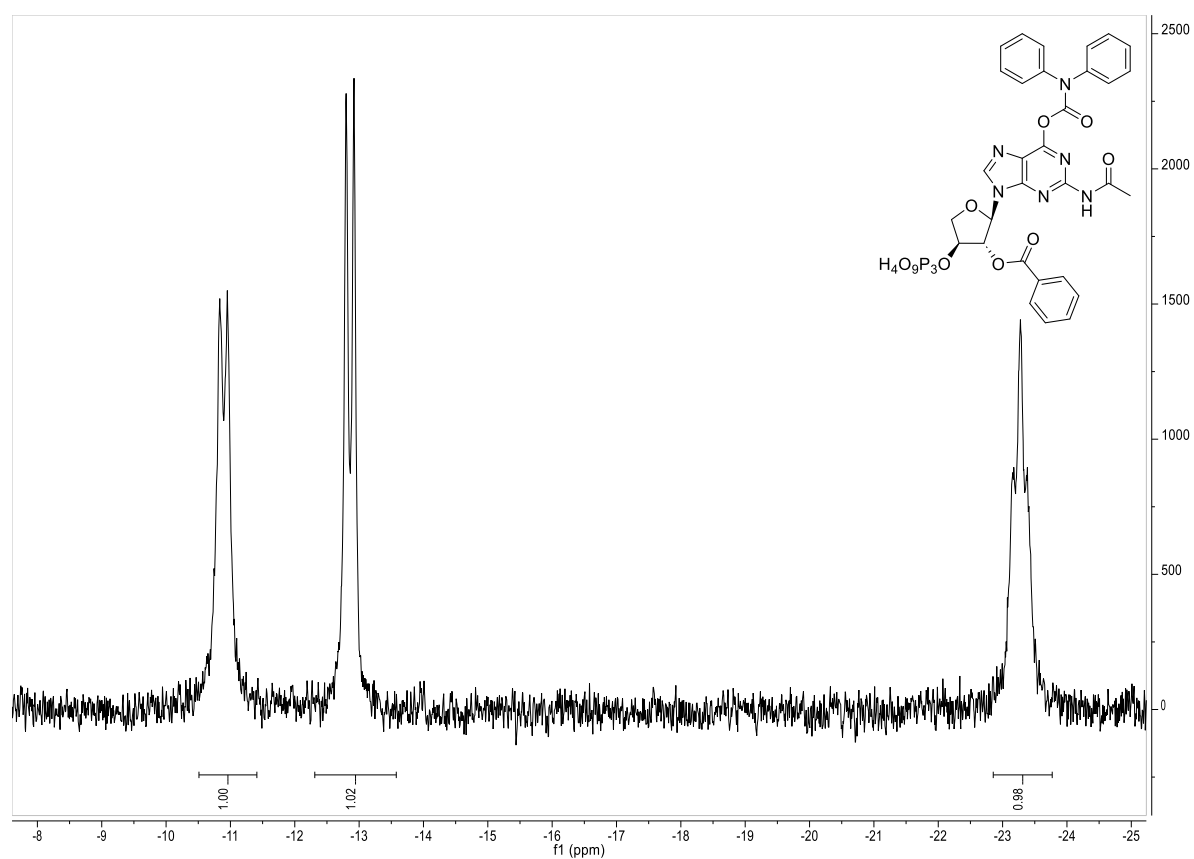


Figure S 71: <sup>1</sup>H-NMR spectrum of compound 35 (D<sub>2</sub>O, 400 MHz, r.t.).





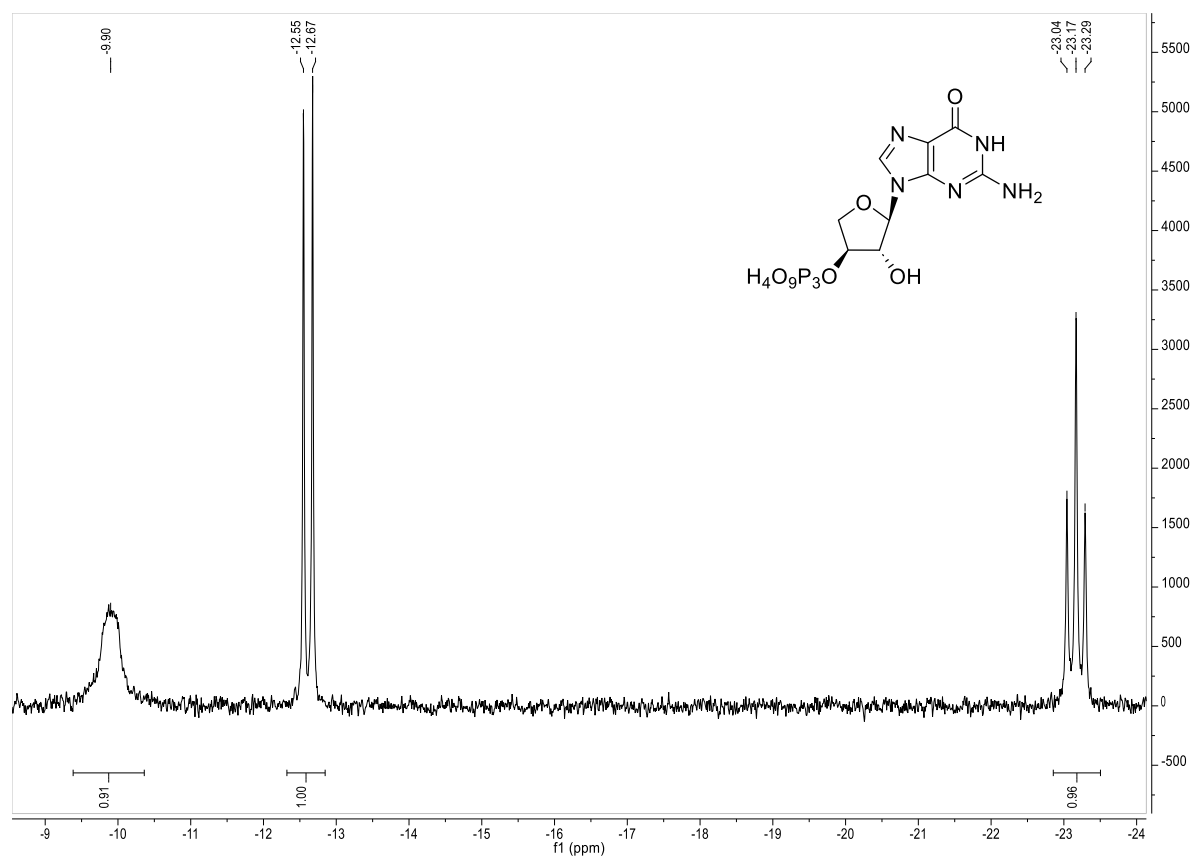


Figure S 74:  $^{31}\text{P}$ -NMR spectrum of compound 9 ( $\text{D}_2\text{O}$ , 162 MHz, r.t.).

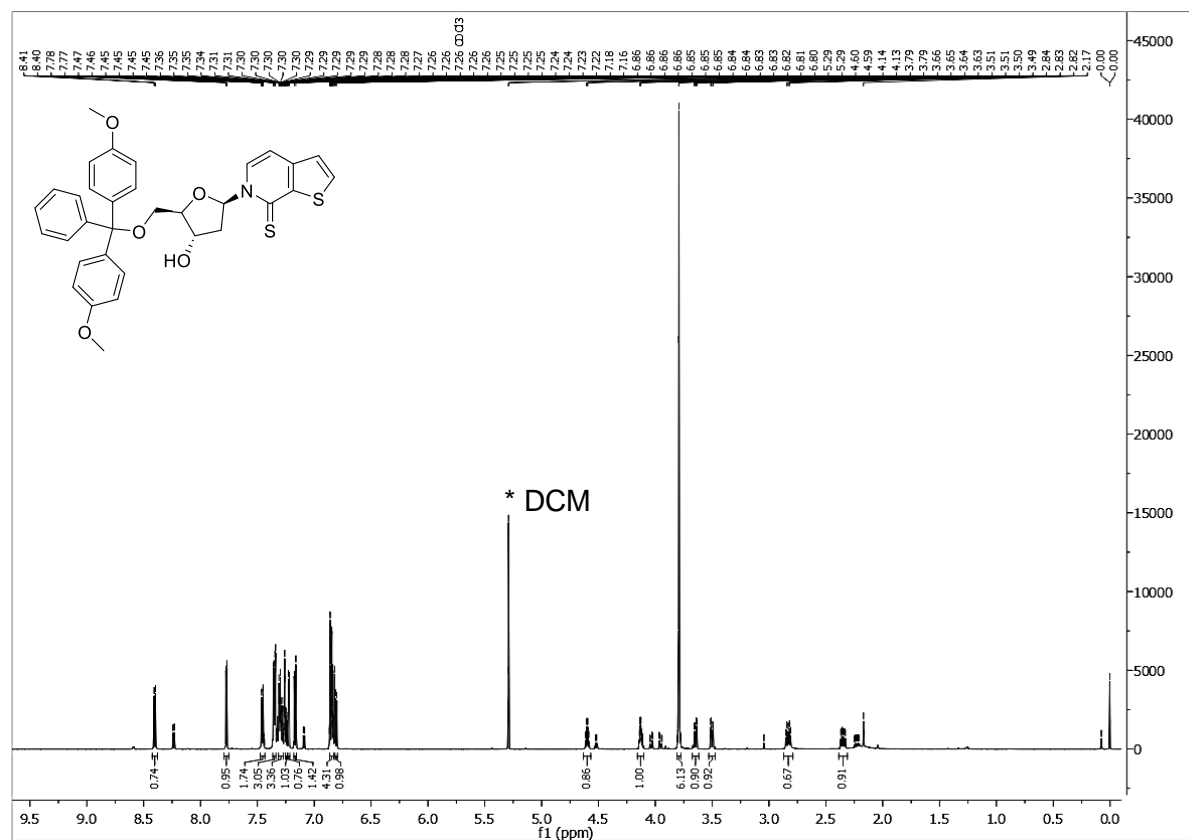


Figure S 75:  $^1\text{H}$ -NMR spectrum of compound 38 ( $\text{CDCl}_3$ , 600 MHz, r.t.).

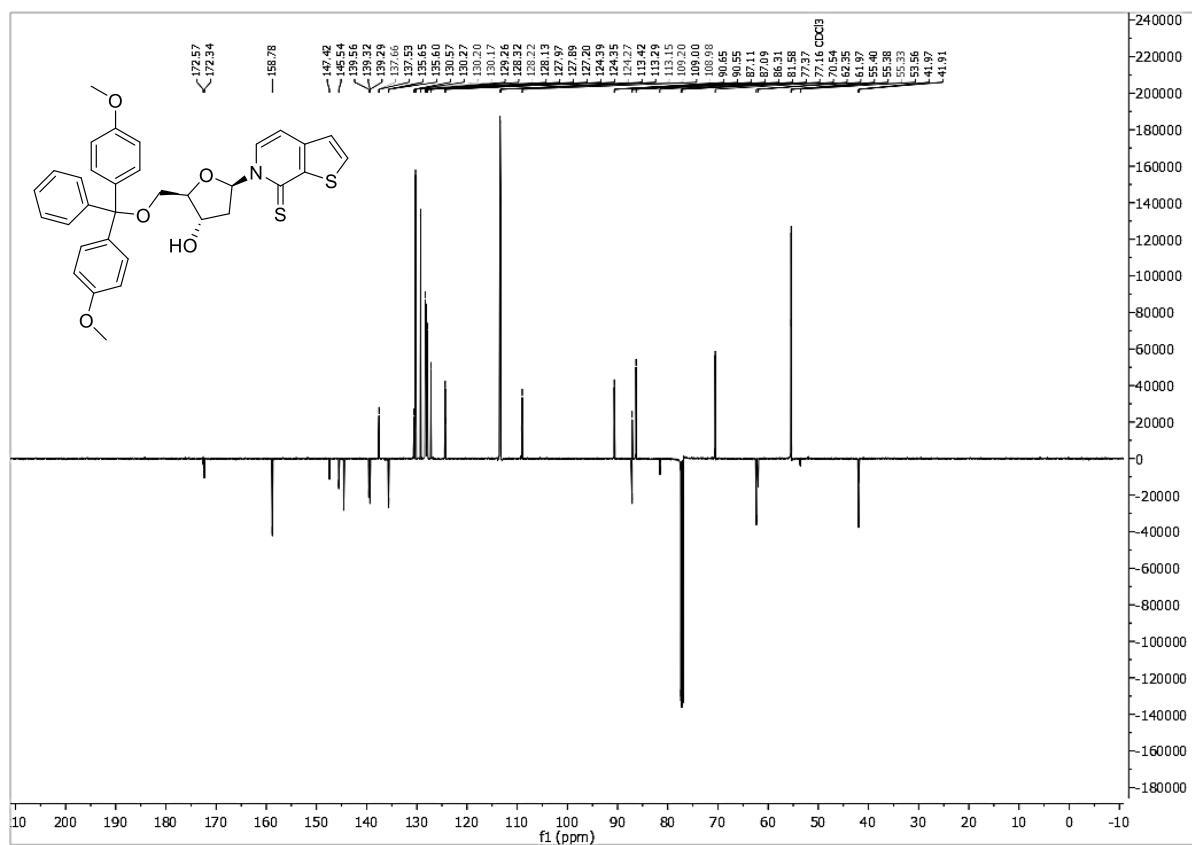


Figure S 76:  $^{13}\text{C}$ -NMR spectrum of compound 38 ( $\text{CDCl}_3$ , 151 MHz, r.t.).

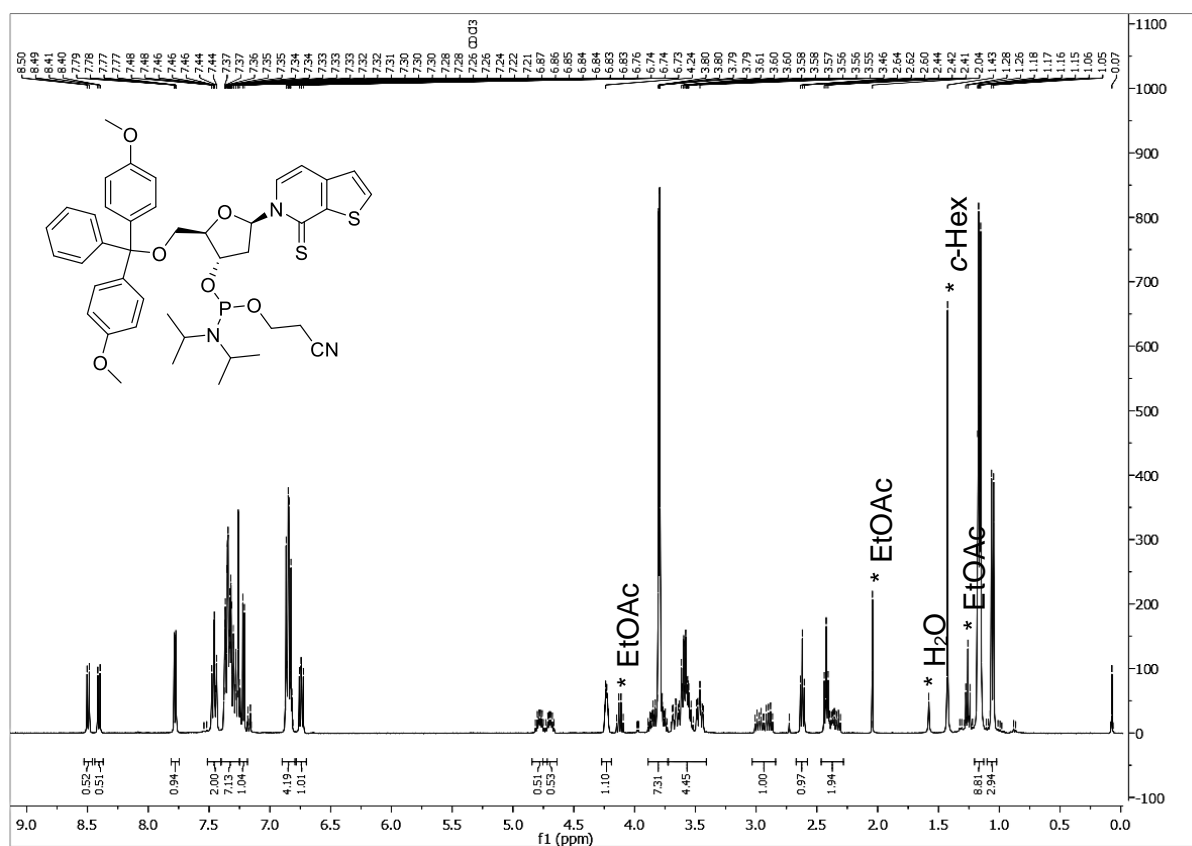


Figure S 77:  $^1\text{H}$ -NMR spectrum of compound 36 ( $\text{CDCl}_3$ , 400 MHz, r.t.).

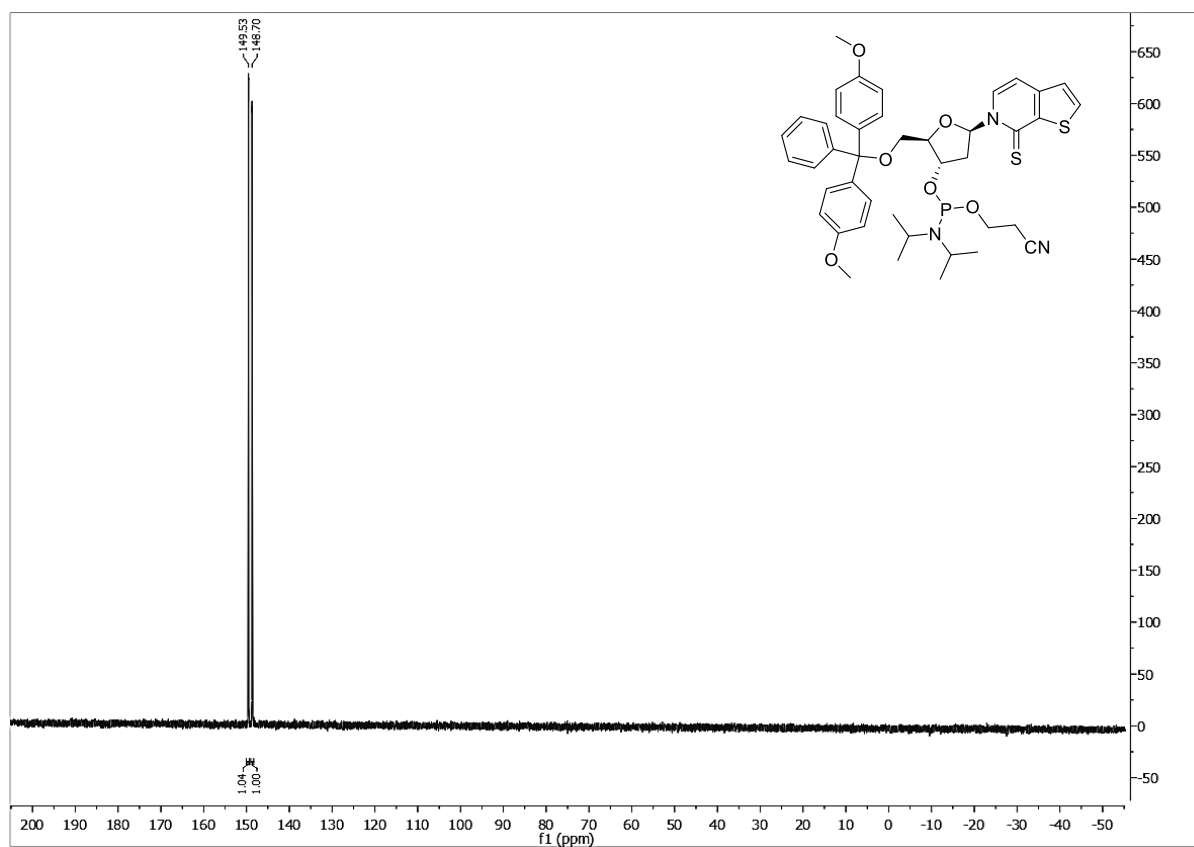


Figure S 78:  $^{31}\text{P}$ -NMR spectrum of compound 36 ( $\text{CDCl}_3$ , 162 MHz, r.t.).

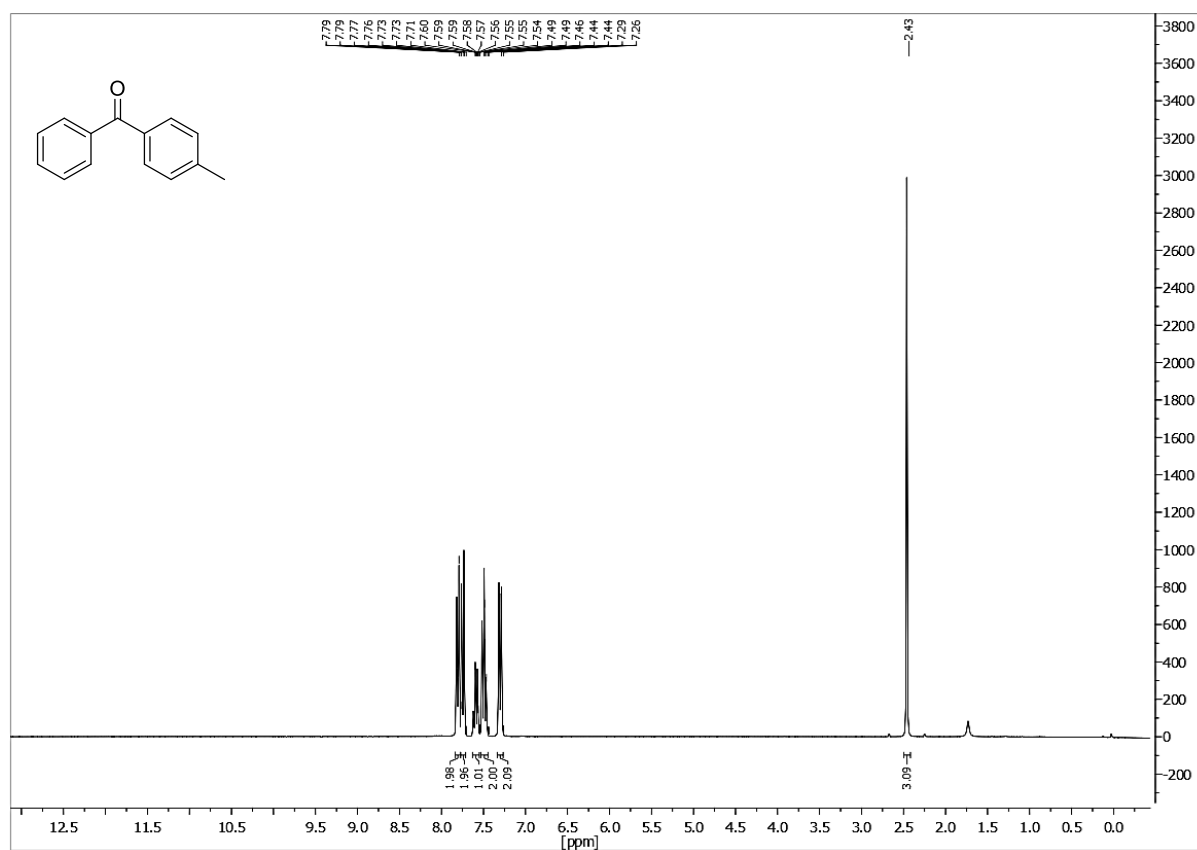


Figure S 79:  $^1\text{H}$ -NMR spectrum of compound 44 ( $\text{CDCl}_3$ , 300 MHz, r.t.).

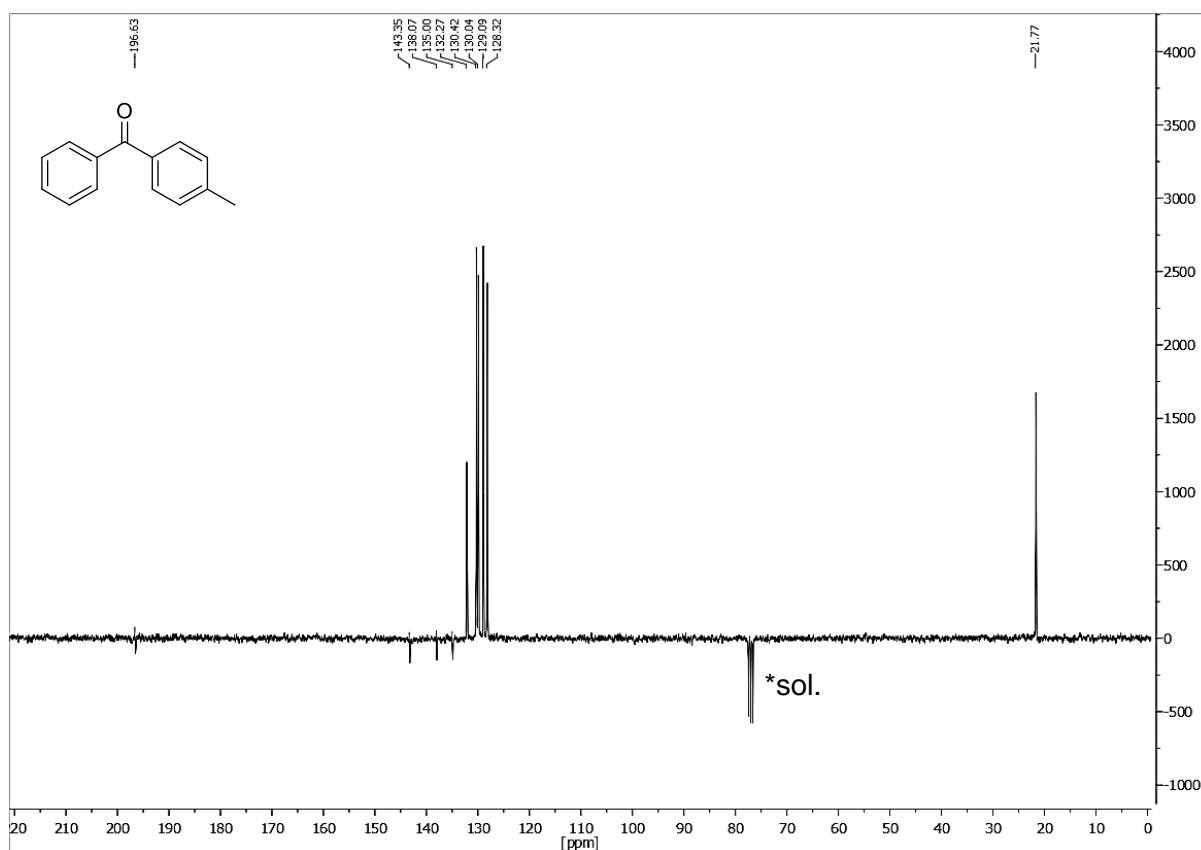


Figure S 80: <sup>13</sup>C-NMR spectrum of compound 44 (CDCl<sub>3</sub>, 75 MHz, r.t.).

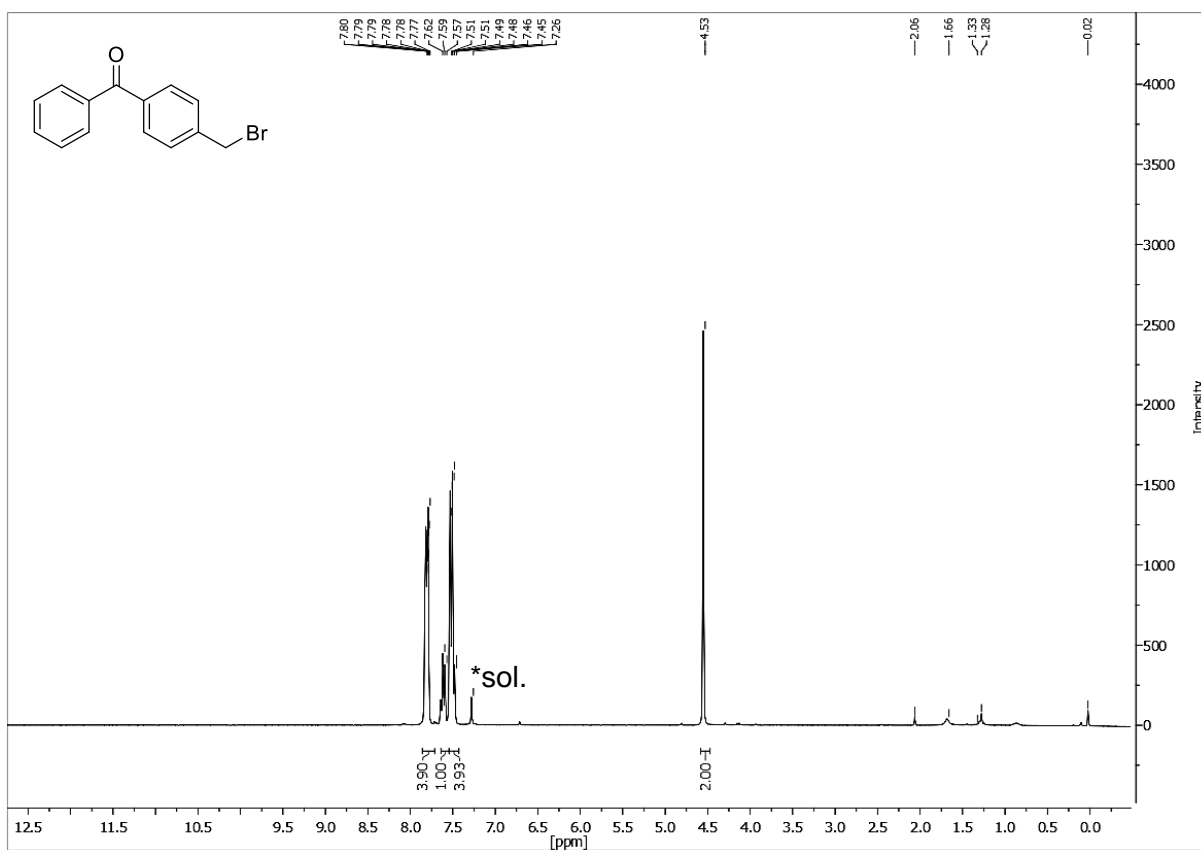


Figure S 81: <sup>1</sup>H-NMR spectrum of compound 40 (CDCl<sub>3</sub>, 300 MHz, r.t.).

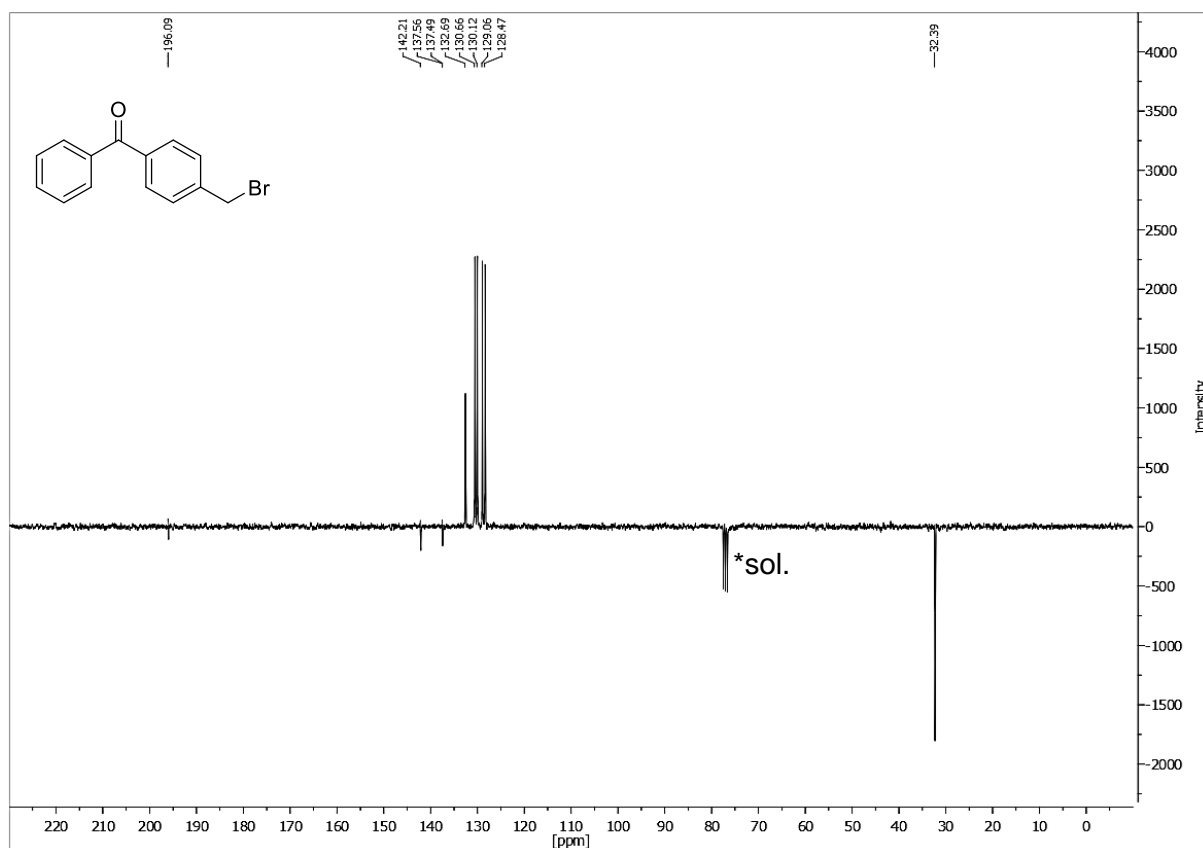


Figure S 82:  $^{13}\text{C-NMR}$  spectrum of compound 40 ( $\text{CDCl}_3$ , 75 MHz, r.t.).

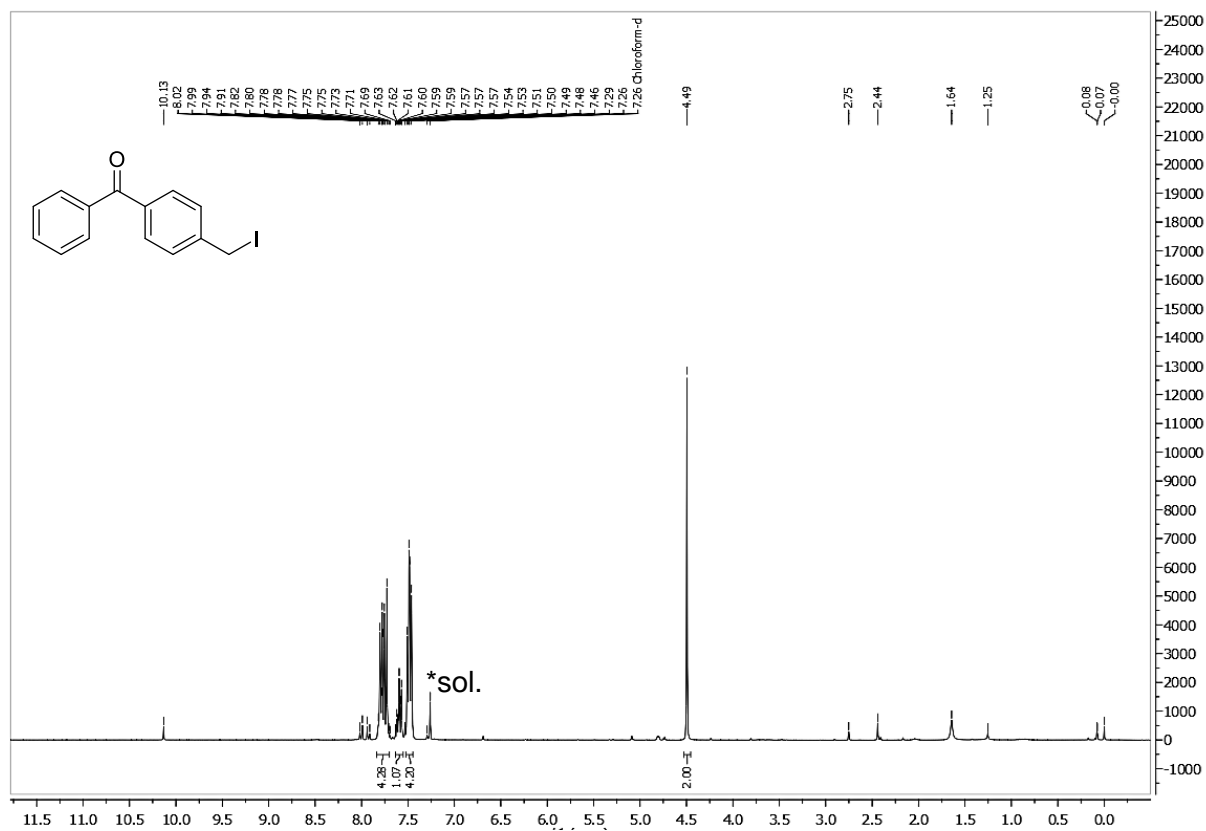


Figure S 83:  $^1\text{H-NMR}$  spectrum of compound 47 ( $\text{CDCl}_3$ , 300 MHz, r.t.).

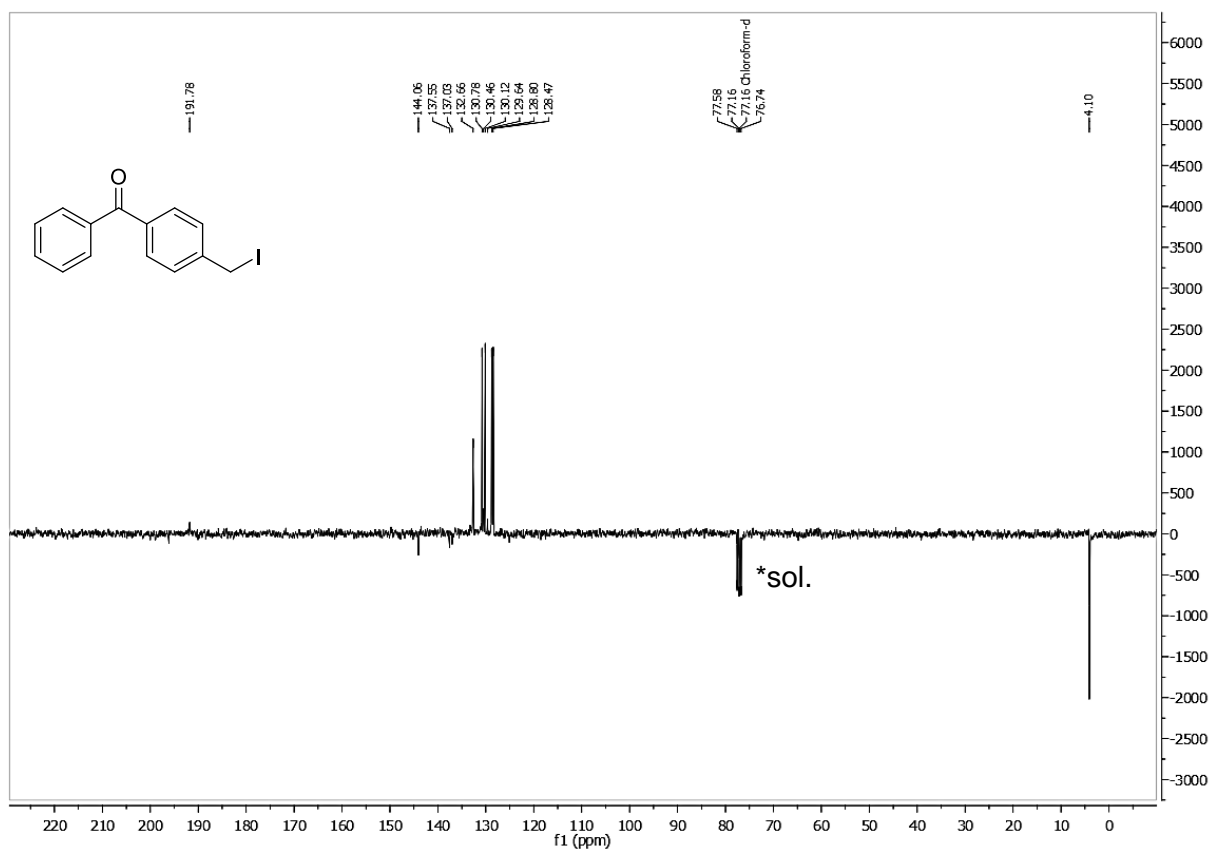


Figure S 84: <sup>13</sup>C-NMR spectrum of compound 47 (CDCl<sub>3</sub>, 75 MHz, r.t.).

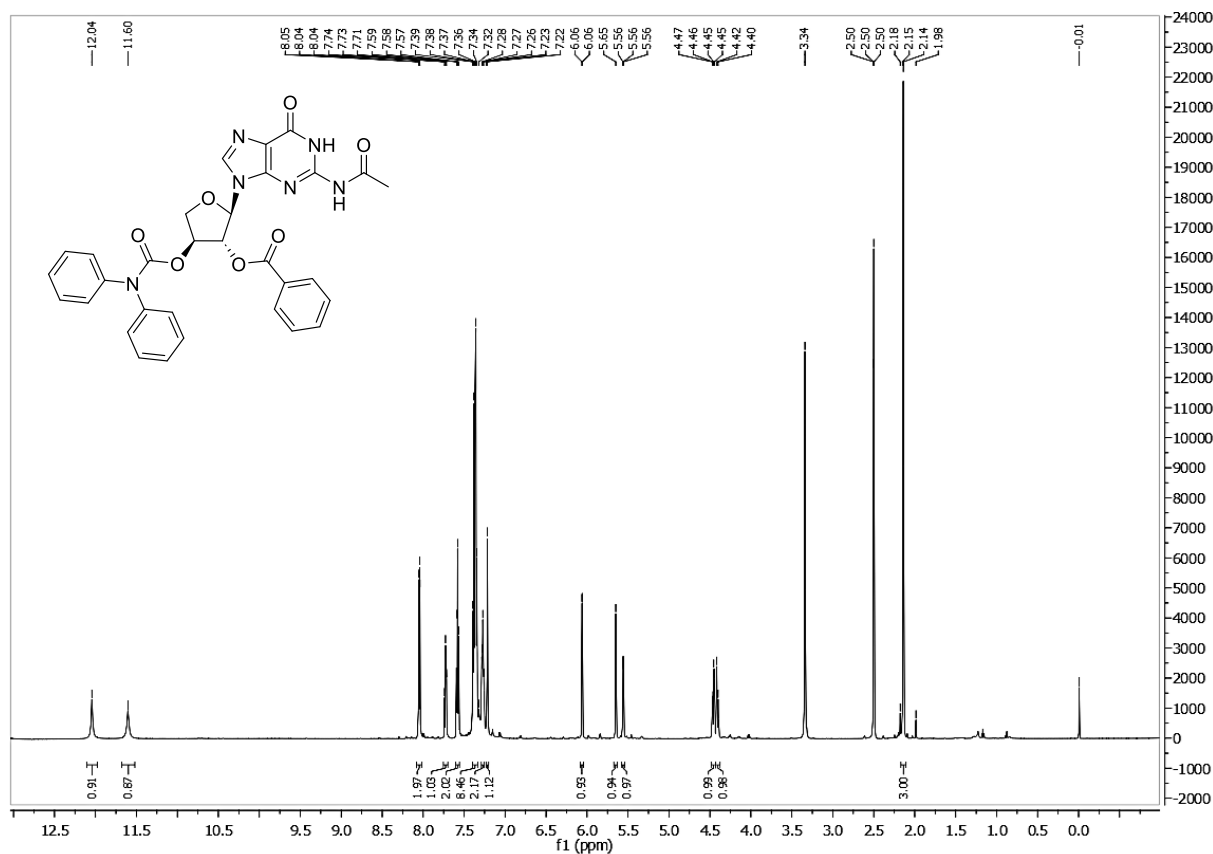


Figure S 85: <sup>1</sup>H-NMR spectrum of compound 43 (DMSO-*d*<sub>6</sub>, 600 MHz, r.t.).

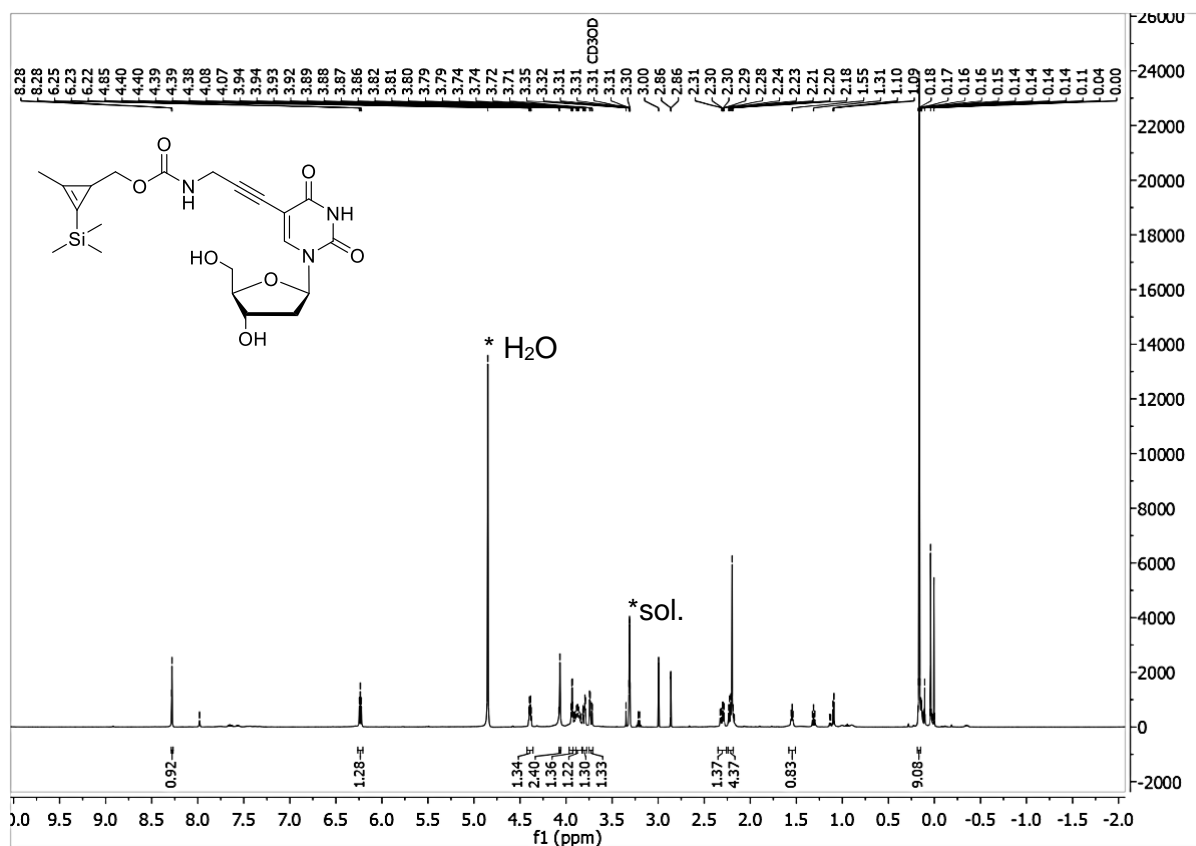


Figure S 86:  $^1\text{H-NMR}$  spectrum of compound 65 ( $\text{CD}_3\text{OD}$ , 500 MHz, r.t.).

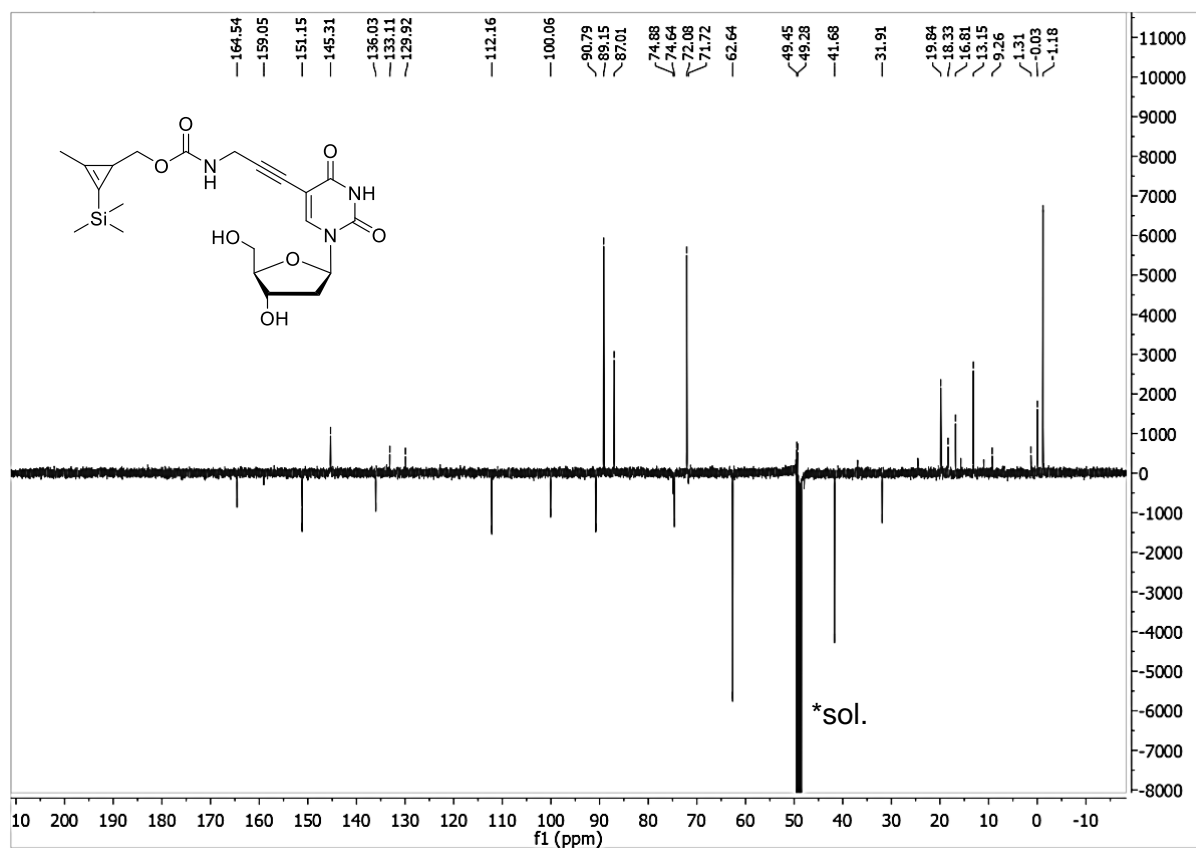


Figure S 87:  $^{13}\text{C-NMR}$  spectrum of compound 65 ( $\text{CD}_3\text{OD}$ , 126 MHz, r.t.).



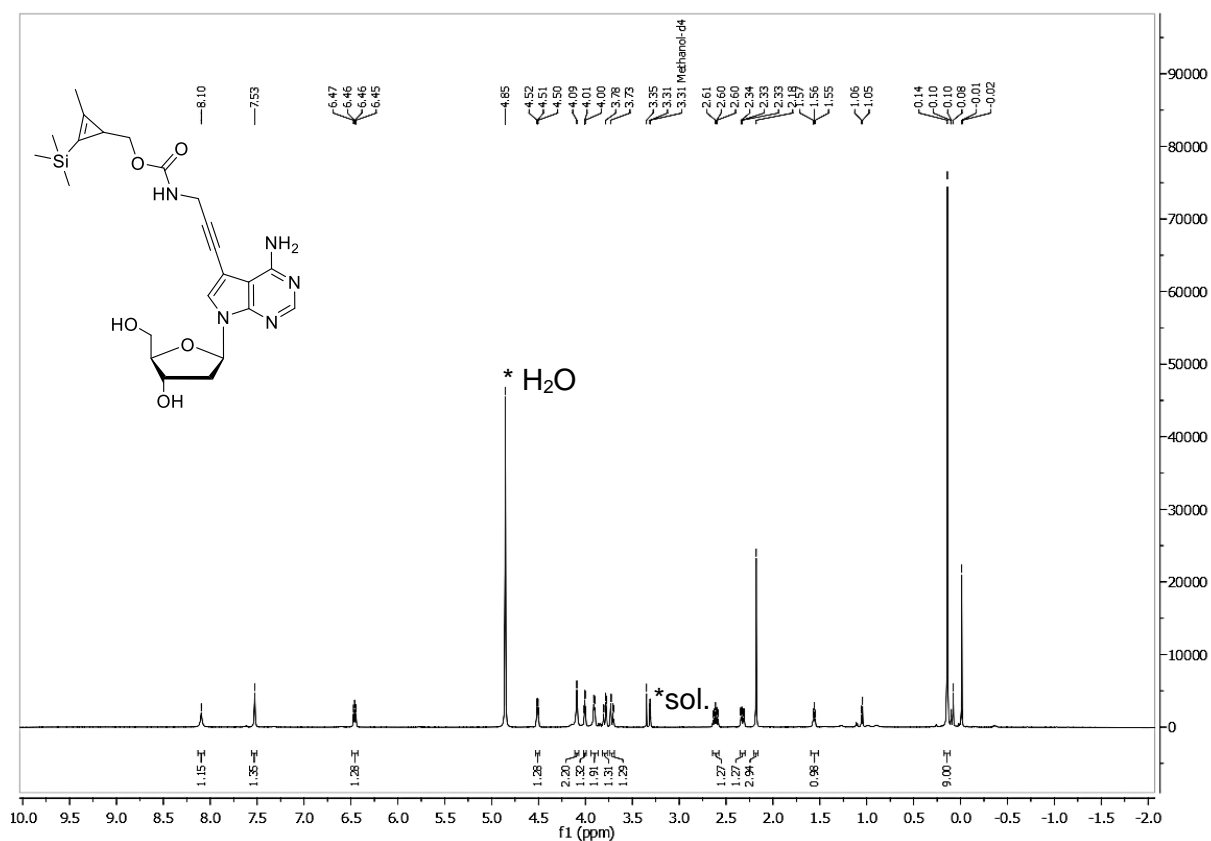


Figure S 88:  $^1\text{H-NMR}$  spectrum of compound 66 ( $\text{CD}_3\text{OD}$ , 500 MHz, r.t.).

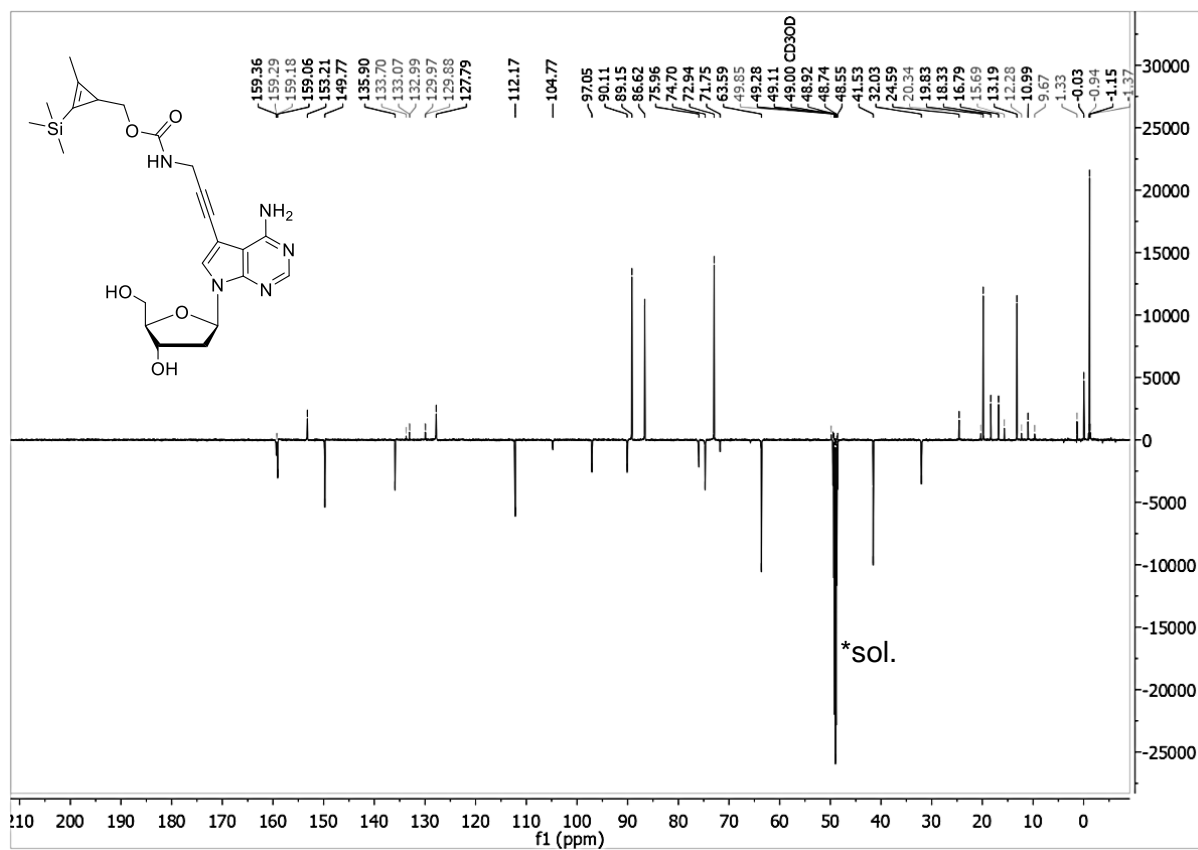
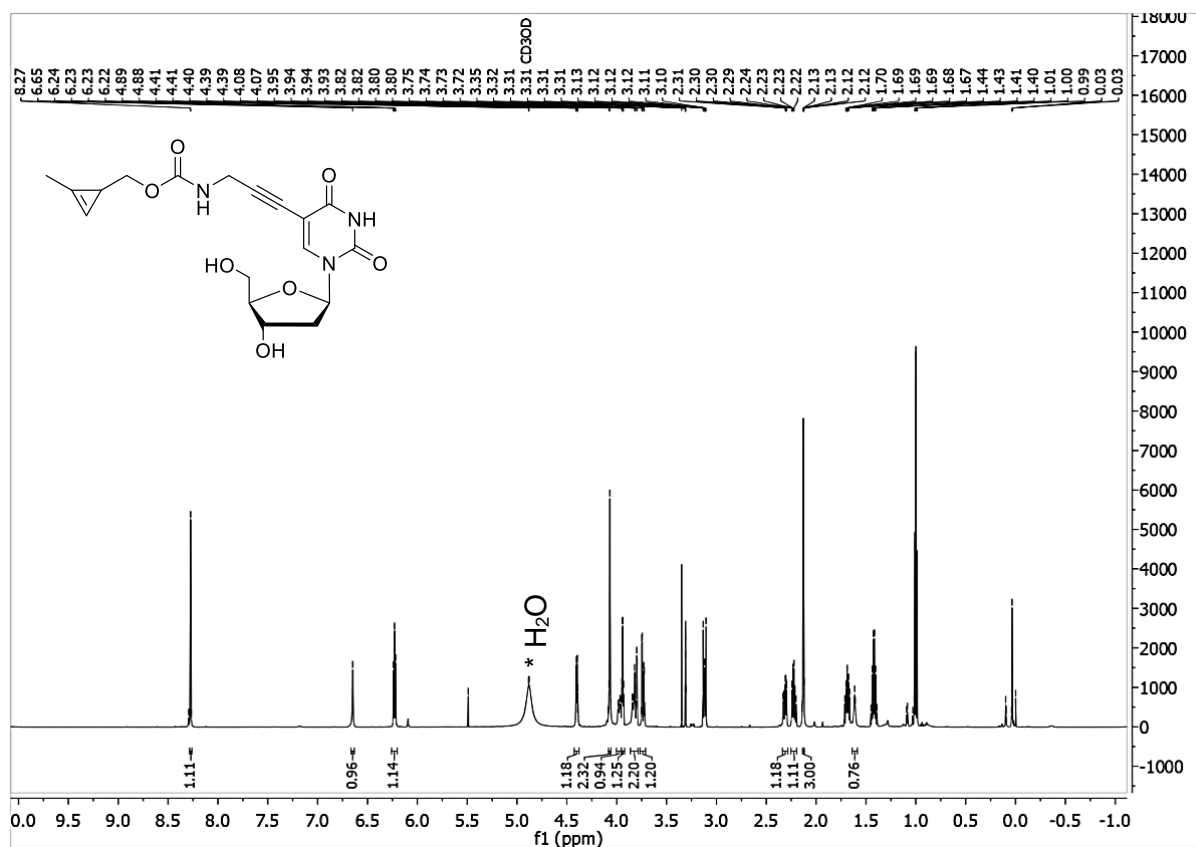
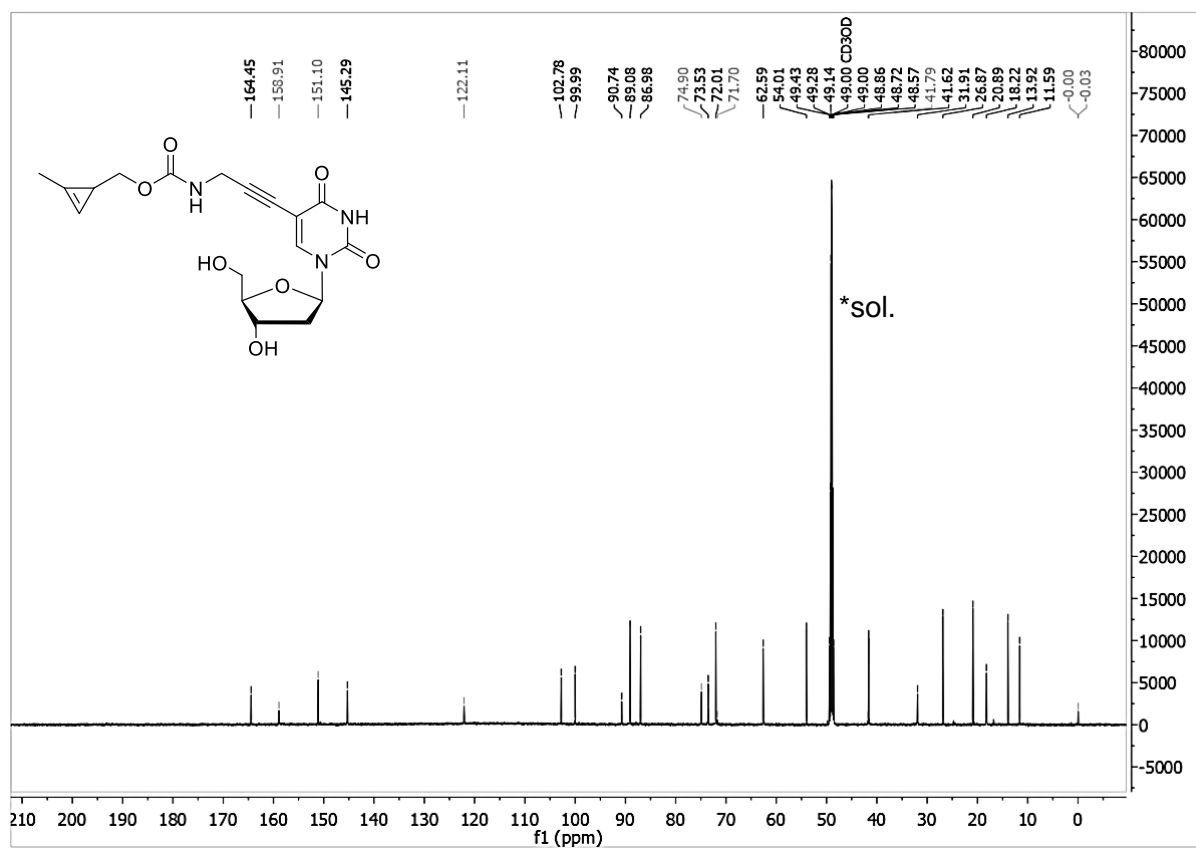
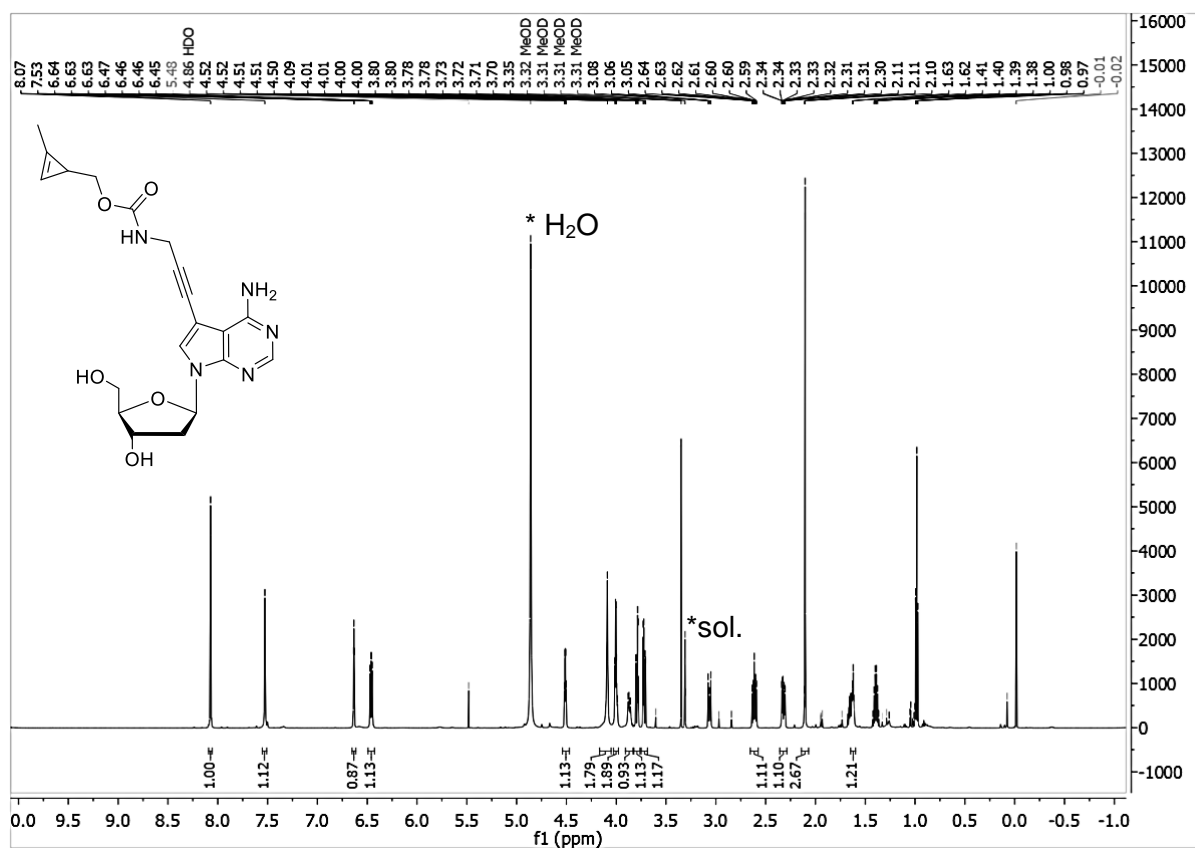
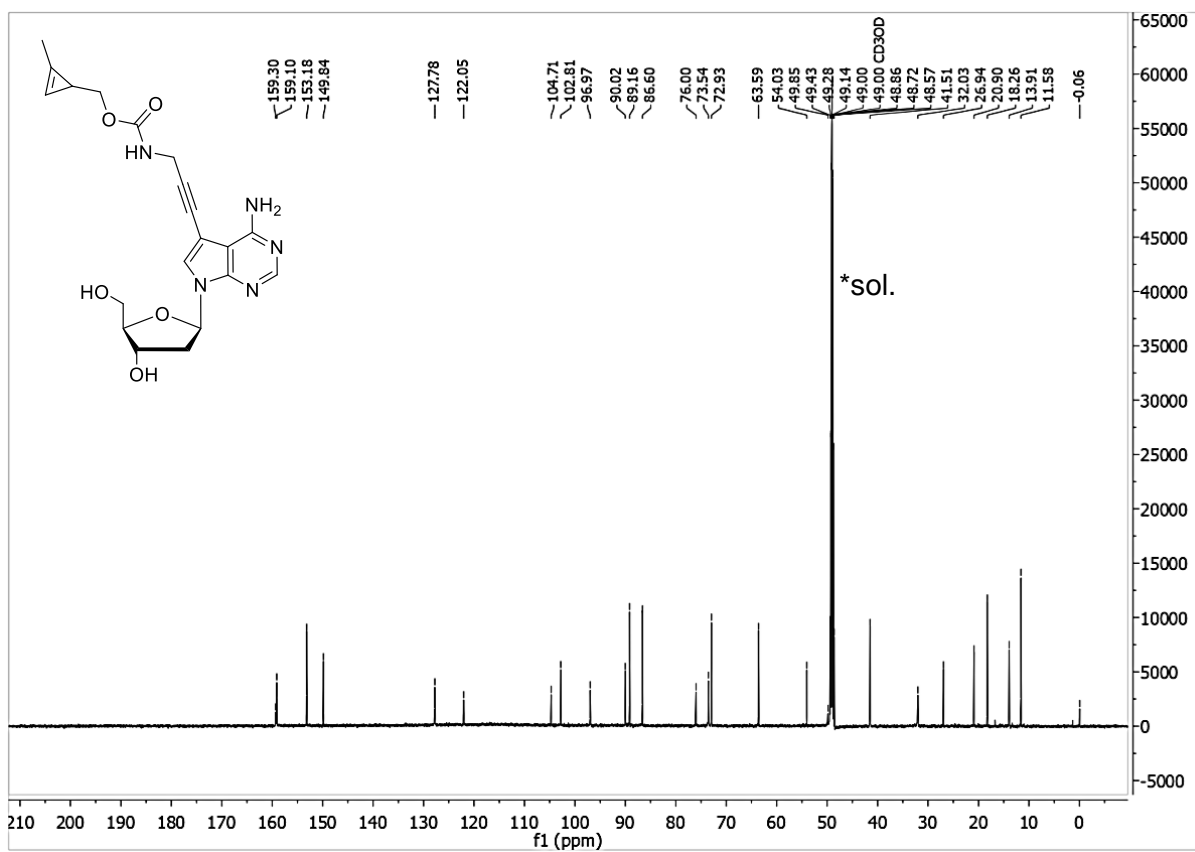


Figure S 89:  $^{13}\text{C-NMR}$  spectrum of compound 66 ( $\text{CD}_3\text{OD}$ , 126 MHz, r.t.).

Figure S 90: <sup>1</sup>H-NMR spectrum of compound 60 (CD<sub>3</sub>OD, 600 MHz, r.t.).Figure S 91: <sup>13</sup>C-NMR spectrum of compound 60 (CD<sub>3</sub>OD, 151 MHz, r.t.).

Figure S 92: <sup>1</sup>H-NMR spectrum of compound 61 (CD<sub>3</sub>OD, 600 MHz, r.t.).Figure S 93: <sup>13</sup>C-NMR spectrum of compound 61 (CD<sub>3</sub>OD, 151 MHz, r.t.).

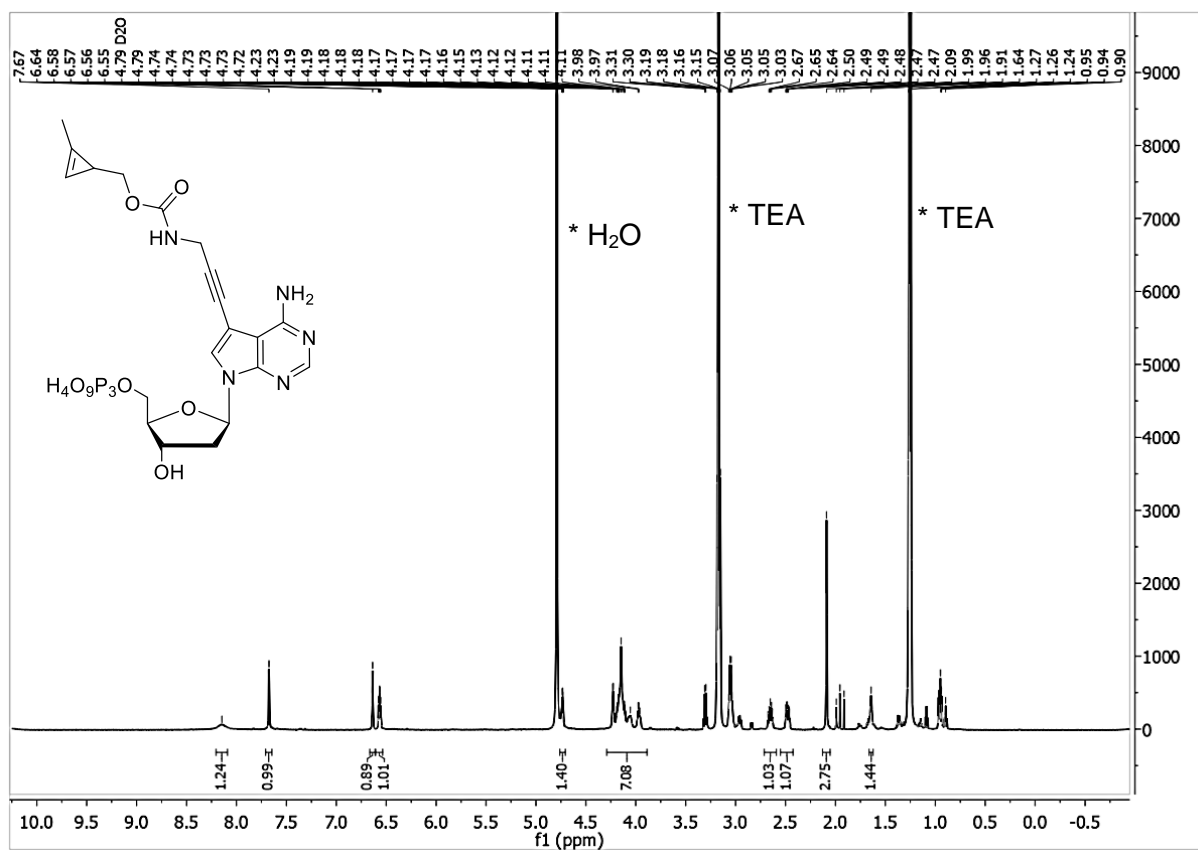


Figure S 94: <sup>1</sup>H-NMR spectrum of compound 59 (D<sub>2</sub>O, 500 MHz, r.t.).

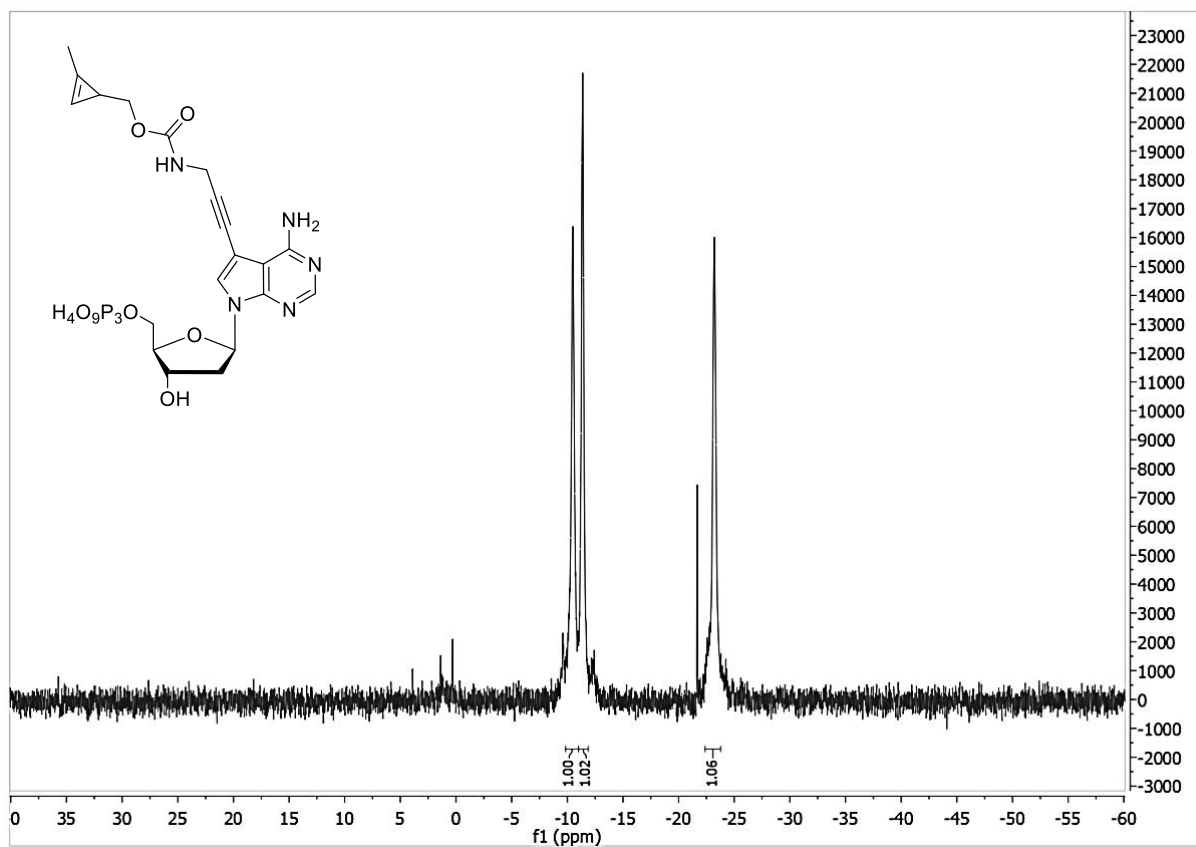


Figure S 95: <sup>31</sup>P-NMR spectrum of compound 59 (D<sub>2</sub>O, 243 MHz, r.t.).

## 7.3 Crystallography data

The single-crystal X-ray diffraction analysis was performed by Dr. Jörg-Martin Neudörfel.

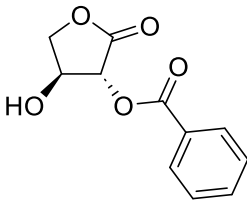
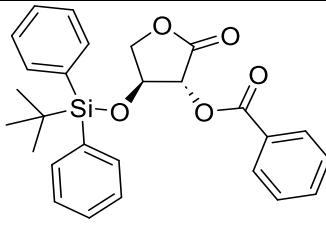
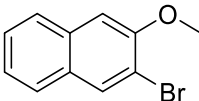
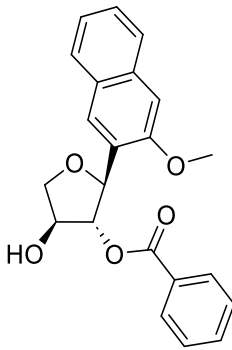
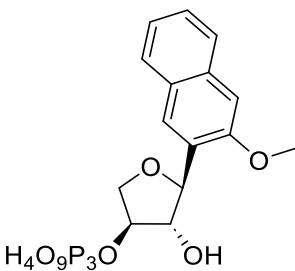
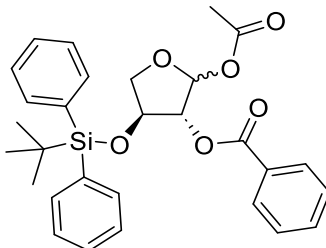
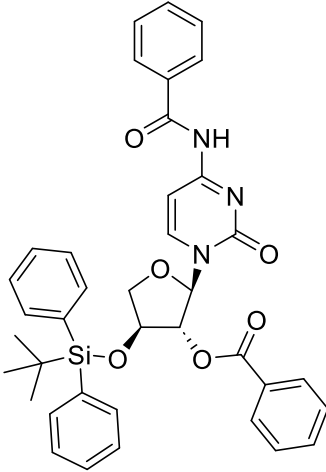
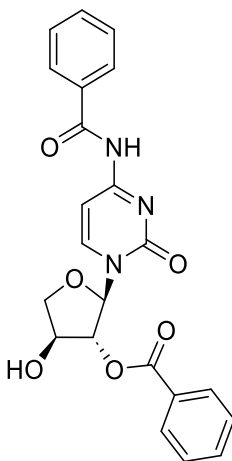
### Crystal data and structure refinement for X.

Identification code	hd28_1	
Empirical formula	C <sub>29</sub> H <sub>24</sub> N <sub>6</sub> O <sub>6</sub>	
Moiety formula	C <sub>29</sub> H <sub>24</sub> N <sub>6</sub> O <sub>6</sub>	
Formula weight	552.54	
Temperature	100(2) K	
Wavelength	1.54178 Å	
Crystal system	Monoclinic	
Space group	P2 <sub>1</sub>	
Unit cell dimensions	a = 10.9849(7) Å	a = 90°.
	b = 10.8904(6) Å	b = 94.624(4)°.
	c = 11.2938(7) Å	g = 90°.
Volume	1346.68(14) Å <sup>3</sup>	
Z	2	
Density (calculated)	1.363 Mg/m <sup>3</sup>	
Absorption coefficient	0.815 mm <sup>-1</sup>	
F(000)	576	
Crystal size	0.100 x 0.020 x 0.010 mm <sup>3</sup>	
Theta range for data collection	3.927 to 72.098°.	
Index ranges	-13 ≤ h ≤ 13, -13 ≤ k ≤ 11, -13 ≤ l ≤ 13	
Reflections collected	33937	
Independent reflections	5187 [R(int) = 0.0798]	
Completeness to theta = 67.679°	100.0 %	
Absorption correction	Semi-empirical from equivalents	
Max. and min. transmission	0.7536 and 0.6165	
Refinement method	Full-matrix least-squares on F <sup>2</sup>	
Data / restraints / parameters	5187 / 1 / 378	
Goodness-of-fit on F <sup>2</sup>	1.062	
Final R indices [I > 2σ(I)]	R1 = 0.0379, wR2 = 0.0953	
R indices (all data)	R1 = 0.0445, wR2 = 0.0996	
Absolute structure parameter	0.07(11)	
Extinction coefficient	n/a	
Largest diff. peak and hole	0.190 and -0.164 e.Å <sup>-3</sup>	

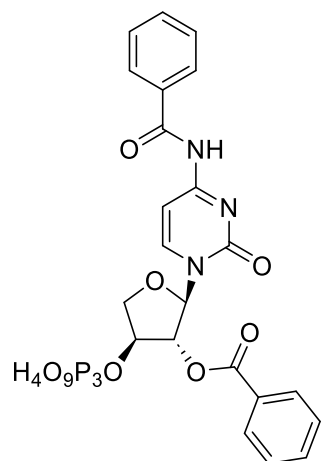
## 7.4 List of synthesized compounds

This chapter provides a list of all compounds that were synthesized within the framework of this thesis.

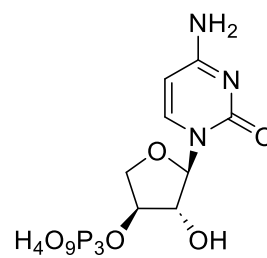
Table S 6: Compounds synthesized in this study.

#	Chemical structure	#	Chemical structure
14		12	
10		16	
6		18	
23		24	

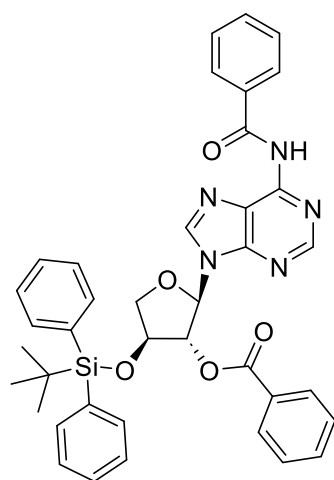
25



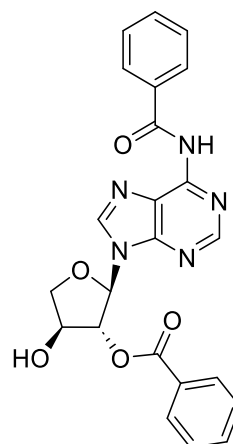
7



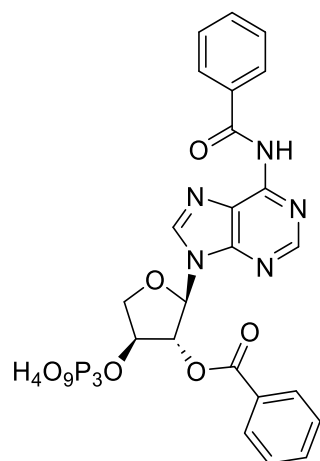
27



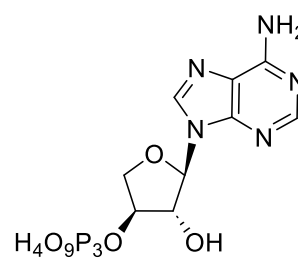
28



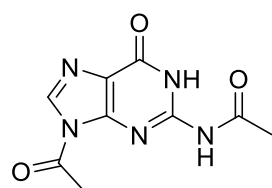
29



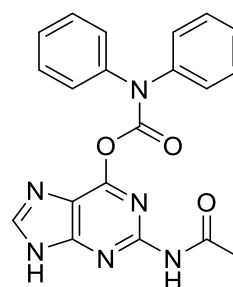
8



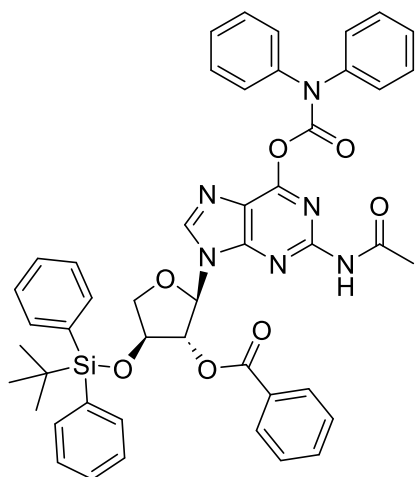
31



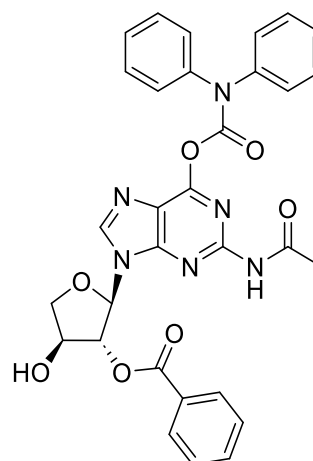
32



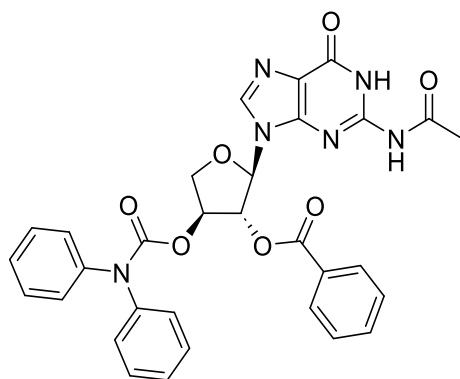
33



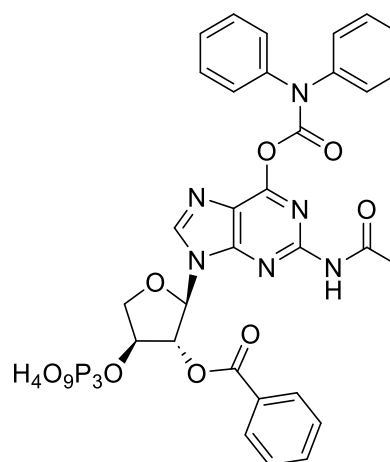
34



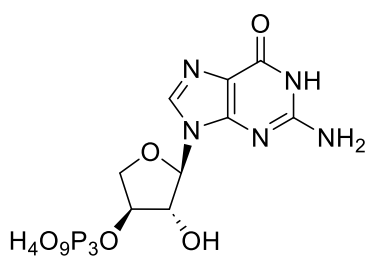
43



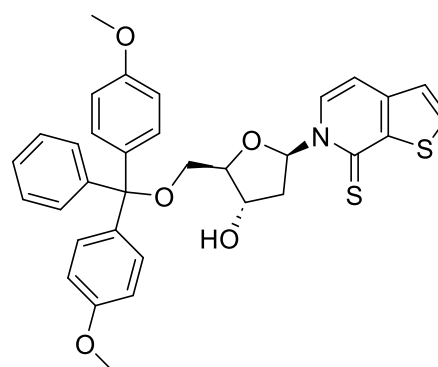
35



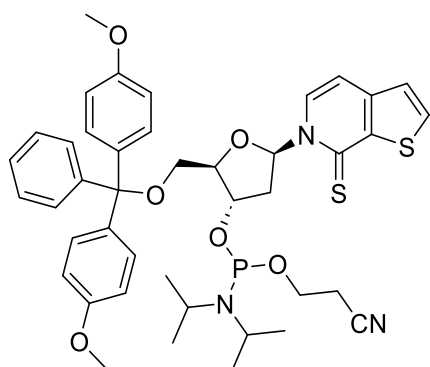
9



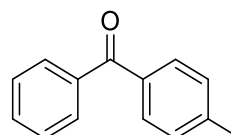
38



36

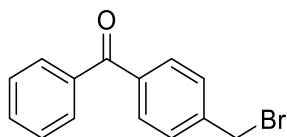


44

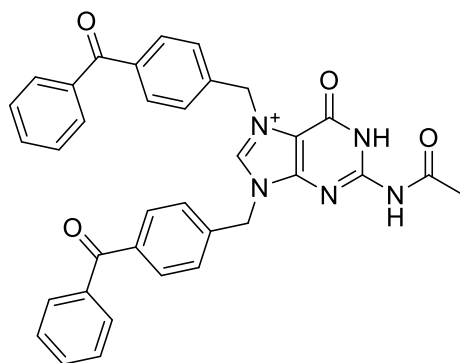




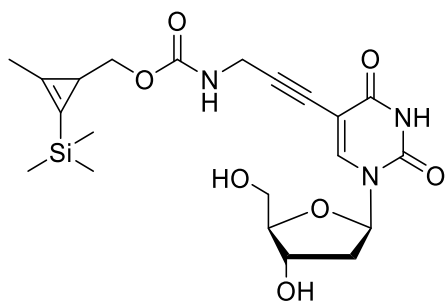
40



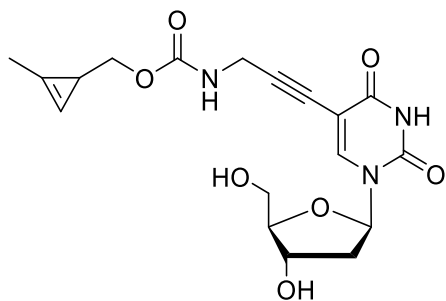
48



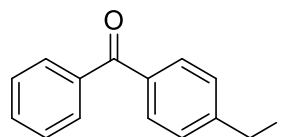
65



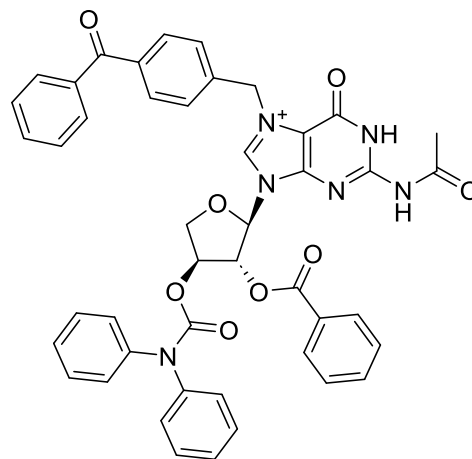
60



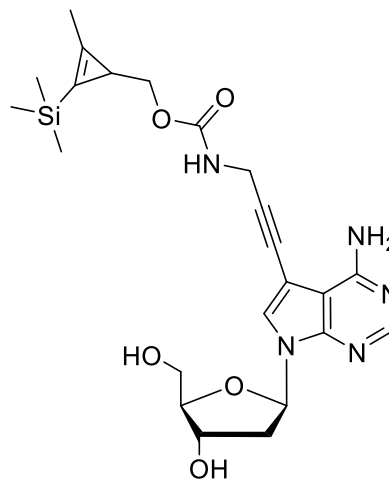
47



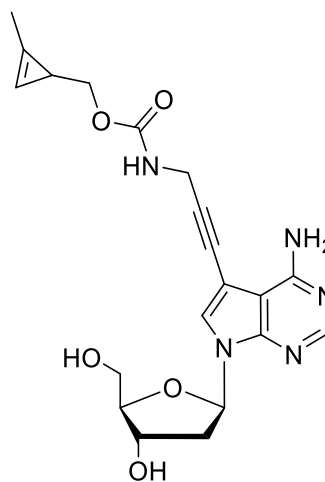
50



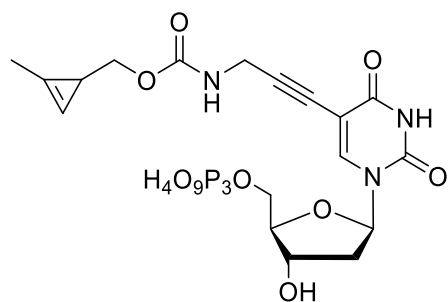
66



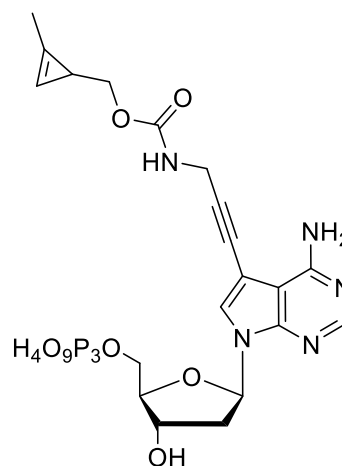
61



58



59



## 7.5 List of sequences

This chapter provides a list of all DNA, TNA, RNA, and amino acid sequences that were used or prepared in this study.

### 7.5.1 DNA primers

This chapter provides a list of all synthetic DNA primers utilized in cloning experiments, as well as in Sanger sequencing, and primer extension assays.

**Kod-RI\_ColonyPCR\_fw** (26 nt)

5'-ATGATTCTGGACACCGATTATATCAC-3'

**Kod-RI\_ColonyCPR\_rv** (21 nt)

5'-TTAGGTGCCTTTCGGTTTCAG-3'

**Seq\_pQE80HT\_1\_fw** (21 nt)

5'-GAGCGGATAACAATTTACACAC-3'

**Seq\_pQE80HT\_4\_rv** (18 nt)

5'-GTTCAGAACGCTCGGTTG-3'

**Seq\_pASG-IBA35\_1\_fw** (24 nt)

5'-GTTGACACTCTATCATTGATAGAG-3'

**Seq\_pASG-IBA35\_4\_rv** (24 nt)

5'-GAGTGAGCTAACTCACATTAATTG-3'

**Kod-RI\_ampli\_25bp\_pQE80HT\_fw** (25 nt)

5'-ATCCGAAAACCTGTATTTTCAGTCT-3'

**Kod-RI\_ampli\_25bp\_pQE80HT\_rv** (15 nt)

5'-CAGGTCGACCCGGGG-3'

**Kod-RI\_ampli\_15bp\_pQE80HT\_fw** (30 nt, 5'-overhang for In-Fusion cloning underlined)

5'-CTGTATTTTCAGTCTATGATTCTGGACACC-3'

**Kod-RI\_ampli\_15bp\_pQE80HT\_rv** (22 nt, 5'-overhang for In-Fusion cloning underlined)

5'-CGGGGTACCGAGCTCTTAGGTG-3'

**Kod-RI\_ampli\_REcloning\_pQE80HT\_fw** (46 nt, SacI recognition site underlined – asterisk marks the position of cleavage, start of Kod-RI coding sequence)

5'-TGTATTTTCAGTCTGAGCT\*CATGATTCTGGACACCGATTATATCAC-3'

**Kod-RI\_ampli\_REcloning\_pQE80HT\_rv** (41 nt, HindIII recognition site underlined – asterisk marks the position of cleavage, end of Kod-RI coding sequence)

5'-AAGCTCAGCTAATTAAAGCT\*TTTAGGTGCCTTTCGGTTTCAG-3'

**Kod-RI\_ampli\_Stargate\_pASG-IBA35\_fw** (51 nt, Esp3I recognition site underlined, **cleavage site in bold** – asterisk marks the position of cleavage, start of Kod-RI coding sequence)

5'-GACGCGTCTCC\*AATGATTCTGGACACCGATTATATCAC-3'

**Kod-RI\_ampli\_Stargate\_pASG-IBA35\_rv** (50 nt, Esp3I recognition site underlined, **cleavage site in bold** – asterisk marks the position of cleavage, end of Kod-RI coding sequence)

5'-GATGCGTCTCC\*TCCCTTAGGTGCCTTTCGGTTTCAG -3'

**Kod-RI\_ampli\_25bp\_pASG-IBA35\_fw** (51 nt, 5'-overhang for AQUA cloning underlined)

5'-GGCTAGCCATCACCATCACCATCACATGATTCTGGACACCGATTATATCAC-3'

**Kod-RI\_ampli\_25bp\_pASG-IBA35\_rv** (50 nt, 5'-overhang for AQUA cloning underlined)

5'-TATTGGGCTGCAGCGTCTCTCATTCCGGATTAGGTGCCTTTCGGTTTCAG-3'

**Kod-RI\_ampli\_15bp\_pASG-IBA35\_fw** (42 nt, 5'-overhang for In-Fusion cloning underlined)

5'-CACCATCACCATCACATGATTCTGGACACCGATTATATCAC-3'

**Kod-RI\_ampli\_15bp\_pASG-IBA35\_rv** (36 nt, 5'-overhang for In-Fusion cloning underlined)

5'-GTCTCTCATTCCGGATTAGGTGCCTTTCGGTTTCAG-3'

**pQE80HT\_PCRIlin\_fw** (16 nt)

5'-GAGCTCGGTACCCCGG-3'

**pQE80HT\_PCRIlin\_rv** (25 nt)

5'-AGACTGAAAATACAGGTTTTTCGGAT-3'

**pASG-IBA35\_PCRIlin\_fw** (20 nt)

5'-TCCGGAATGAGAGACGCTGC-3'

**pASG-IBA35\_PCRIlin\_rv** (23 nt)

5'-GTGATGGTGATGGTGATGGCTAG-3'

**Seq\_Kod-RI\_2\_fw** (22 nt)

5'-GTTGATCTGCCGTATGTTGATG-3'

**Seq\_Kod-RI\_3\_rv** (21 nt)

5'-GACGAACAATTTCCAGACCAC-3'

**Seq\_pASG-IBA35\_Stargate\_rv4** (18 nt)

5'-AGTGTAGCGGTCACGCTG-3'

**pQE80HT\_Kod-RI\_2313delA\_fw** (24 nt, insertion of A into the coding sequence of Kod-RI\_delA2313 underlined)

5'-GTGATCGTGCAATTCCGTTTGATG-3'

**pQE80HT\_Kod-RI\_2313delA\_rv** (22 nt)

5'-CAATGCGACCGCTACCTTTCAG-3'

**Mut1S\_RSGA\_fw** (23 nt)

5'-TGGCCAGCAGTTATTATGGCTAC-3'

**Mut1\_RSGA\_rv** (23 nt)

5'-GAATTTTGATACGACGCTGACGA-3'

**Mut2\_RSGA-fw** (25 nt)

5'-GAAATTGTTTCGTCGTGATTGGAGTG-3'

**Mut2G\_RSGA\_rv** (22 nt)

5'-CAGACCGCCGGTTGTAATTTTG-3'

**Mut3\_RSGA\_fw** (27 nt)

5'-ACCTGAAAGATTATAAAGCAACCGGTC-3'

**Mut3E\_RSGA\_rv** (28 nt)

5'-CACGGGTAATCTGTTTCATGAATAACCAG-3'

**Mut4\_RSGA\_fw** (30 nt)

5'-ATGCCGAGTATTATATCGAGAATCAGGTTC-3'

**Mut4A\_RSGA\_rv** (28 nt)

5'-CATATTTGTGTTTCGCCGGATCAAATTC-3'

**Seq\_RSGA\_2\_fw** (21 nt)

5'-TTGATCTGCCGTATGTTGATG-3'

**Seq\_RSGA\_3\_rv** (21 nt)

5'-CAATCACGACGAACAATTTCC-3'

**PEx\_primer** (20 nt)

5'-GACACTCGTATGCAGTAGCC-3'

**6-FAM\_PEx\_primer** (20 nt)

5'-(6)-FAM-GACACTCGTATGCAGTAGCC-3'

## 7.5.2 List of RNA, DNA and, TNA sequences

This chapter provides all unmodified and modified synthetic RNA sequences and DNA templates, along with the resulting complementary sequences of the DNA/TNA chimeras and the DNA and DNA<sup>CP</sup> sequences. The molecular weights of the resulting full-length oligonucleotides are given as references for those sequences that were analyzed by LC MS.

Note: The listed full-length DNA/TNA chimera were not always obtained. Please refer to Chapter 3.1 for further information.

**DNA\_Temp\_longATGC** (51 nt, primer binding site underlined)

5'-CGATGGACATTCGAGCATAAGATCCAGTCATGGCTACTGCATACGAGTGTC-3'

→ **resulting DNA/TNA chimera** (DNA primer sequence underlined, *TNA sequence italic*):

5'-GACACTCGTATGCAGTAGCCATGACTGGATCTTATGCTCGAATGTCCATCG-3'

**DNA\_Temp\_shortATGC** (29 nt, primer binding site underlined)

5'-AGATCAGTCGGCTACTGCATACGAGTGTC-3'

→ **resulting DNA/TNA chimera** (DNA primer sequence underlined, *TNA sequence italic*):

5'-GACACTCGTATGCAGTAGCCGACTGATCT-3'

**DNA\_Temp\_shortATGC<sup>NaM</sup>** (30 nt, primer binding site underlined)

5'-AGATCAXGTCGGCTACTGCATACGAGTGTC-3'

X= dNaM

→ **resulting DNA/TNA chimera** (DNA primer sequence underlined, *TNA sequence italic*):

5'-GACACTCGTATGCAGTAGCCGACZTGATCT-3'

Z = tTPT3

**DNA\_Temp\_midATGC** (40 nt, primer binding site underlined)

5'-TACGAATACATGCCAGTCATGGCTACTGCATACGAGTGTC-3'

→ **resulting DNA/TNA chimera** (DNA primer sequence underlined, *TNA sequence italic*):

5'-GACACTCGTATGCAGTAGCCATGACTGGCATGTATTCGTA-3'

**DNA\_Temp\_midATGC<sup>NaM</sup>** (41 nt, primer binding site underlined)

5'-TACGAATXACATGCCAGTCATGGCTACTGCATACGAGTGTC-3'

X= dNaM

→ **resulting DNA/TNA chimera** (DNA primer sequence underlined, *TNA sequence italic*):

5'-GACACTCGTATGCAGTAGCCATGACTGGCATGTZATTCGTA-3'

Z = tTPT3

**DNA\_pTemp\_mid2ATGC** (40 nt, primer binding site underlined)

5'-Phos-TACGAATACATACCAGTCATGGCTACTGCATACGAGTGTC-3'

→ **resulting DNA/TNA chimera** (DNA primer sequence underlined, *TNA sequence italic*):

5'-GACACTCGTATGCAGTAGCCATGACTGGTATGTATTCGTA-3'

(MW = 12583 g mol<sup>-1</sup>)

→ **resulting DNA sequence** (DNA primer sequence underlined, *DNA sequence italic*):

5'-GACACTCGTATGCAGTAGCCATGACTGGTATGTATTCGTA-3'

(MW = 12326 g mol<sup>-1</sup>)

→ **resulting DNA<sup>CP</sup> sequence** (DNA primer sequence underlined, *DNA sequence italic*):

5'-GACACTCGTATGCAGTAGCCA\*TGA\*CTGGTA\*TGTA\*TTCGTA\*-3' \* = dA<sup>CP</sup>  
 (MW = 13135 g mol<sup>-1</sup>)

**DNA\_pTemp\_mid2ATGC<sup>NaM</sup>** (41 nt, primer binding site underlined)

5'-Phos-TACGAATXACATACCAGTCATGGCTACTGCATACGAGTGTC-3' X = dNaM  
 → **resulting DNA/TNA chimera** (DNA primer sequence underlined, *TNA sequence italic*):  
 5'-GACACTCGTATGCAGTAGCCAATGACTGGTATGTZATTCGTA-3' Z = tTPT3  
 (MW = 12914 g mol<sup>-1</sup>)

**DNA\_pTemp\_mid2ATGC<sup>TPT3</sup>** (41 nt, primer binding site underlined)

5'-Phos-TACGAATZACATACCAGTCATGGCTACTGCATACGAGTGTC-3' Z = dTPT3  
 → **resulting DNA/TNA chimera** (DNA primer sequence underlined, *TNA sequence italic*):  
 5'-GACACTCGTATGCAGTAGCCAATGACTGGTATGTXATTCGTA-3' X = tNaM  
 (MW = 12905 g mol<sup>-1</sup>)

**DNA\_pTemp\_midATC** (37 nt, primer binding site underlined)

5'-Phos-TACAATACATCCATCATGGCTACTGCATACGAGTGTC-3'  
 → **resulting DNA/TNA chimera** (DNA primer sequence underlined, *TNA sequence italic*):  
 5'-GACACTCGTATGCAGTAGCCAATGATGGATGTATTGTA-3'  
 (MW: 11741 g mol<sup>-1</sup>)

**DNA\_pTemp\_midATC<sup>NaM</sup>** (38 nt, primer binding site underlined)

5'-Phos-TACAATXACATCCATCATGGCTACTGCATACGAGTGTC-3' X = dNaM  
 → **resulting DNA/TNA chimera** (DNA primer sequence underlined, *TNA sequence italic*):  
 5'-GACACTCGTATGCAGTAGCCAATGATGGATGTZATTGTA-3' Z = tTPT3  
 (MW: 12073 g mol<sup>-1</sup>)

**DNA\_pTemp\_midATC<sup>TPT3</sup>** (38 nt, primer binding site underlined)

5'-Phos-TACAATZACATCCATCATGGCTACTGCATACGAGTGTC-3' Z = dTPT3  
 → **resulting DNA/TNA chimera** (DNA primer sequence underlined, *TNA sequence italic*):  
 5'-GACACTCGTATGCAGTAGCCAATGATGGATGTXATTGTA-3' X = tNaM  
 (MW = 12064 g mol<sup>-1</sup>)

**DNA\_pTemp\_midATG** (36 nt, primer binding site underlined)

5'-Phos-TAGAATAGATGATGATGGCTACTGCATACGAGTGTC-3'  
 → **resulting DNA/TNA chimera** (DNA primer sequence underlined, *TNA sequence italic*):  
 5'-GACACTCGTATGCAGTAGCCAATCATCATCTATTCTA-3'

(MW = 11267 g mol<sup>-1</sup>)

**DNA\_pTemp\_midATG<sup>NaM</sup>** (37 nt, primer binding site underlined)

5'-Phos-TAGAATXAGATGATGATGGCTACTGCATACGAGTGTC-3' X = **dNaM**

→ **resulting DNA/TNA chimera** (DNA primer sequence underlined, *TNA sequence italic*):

5'-GACACTCGTATGCAGTAGCCATCATCATCTZATTCTA-3' Z = **tTPT3**

(MW = 11598 g mol<sup>-1</sup>)

**DNA\_pTemp\_midATG<sup>TPT3</sup>** (37 nt, primer binding site underlined)

5'-Phos-TAGAATZAGATGATGATGGCTACTGCATACGAGTGTC-3' Z = **dTPT3**

→ **resulting DNA/TNA chimera** (DNA primer sequence underlined, *TNA sequence italic*):

5'-GACACTCGTATGCAGTAGCCATCATCATCTXATTCTA-3' X = **tNaM**

(MW = 11589 g mol<sup>-1</sup>)

**Z-RNA<sup>m8G</sup>** (26 nt)

5'-Mod-ACYCYCYCYCGUUUUCYCYCYCYCGU-3' Y = **m<sup>8</sup>G**, Mod = abasic amine

**A-RNA<sup>WT</sup>** (26 nt)

5'-Mod-ACGCGCGCGCGUUUUCGCGCGCGCGU-3' Mod = abasic amine

### 7.5.3 Vector sequences

This chapter provides the sequences of vectors utilized in this study, along with the respective vector maps.

**pQE80HT (4556 bp)**

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAA  
 TTGTGAGCGGATAACAATTTACACAGAATTCATTAAGAGGAGAAATTAAGTATGAGAG  
 GATCGCATCACCATCACCATCACGGATCCGAAAACCTGTATTTTCAGTCTGAGCTCGGT  
 ACCCCGGGTCGACCTGCAGCCAAGCTTAATTAGCTGAGCTTGGACTCCTGTTGATAGAT  
 CCAGTAATGACCTCAGAACTCCATCTGGATTTGTTGAGAACGCTCGGTTGCCGCCGGGC  
 GTTTTTTATTGGTGAGAATCCAAGCTAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCT  
 AAAATGGAGAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAA  
 GAACATTTTGAGGCATTTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTTCAGCTG  
 GATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTT  
 ATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTTTCGTATGGCAATGAAAGAC  
 GGTGAGCTGGTGATATGGGATAGTGTTACCCCTTGTACACCGTTTTCCATGAGCAAAC



TGAAACGTTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACAT  
ATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTAT  
TGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAA  
CGTGGCCAATATGGACAACCTTCTTCGCCCCCGTTTTCCACCATGGGCAAATATTATACGC  
AAGGCGACAAGGTGCTGATGCCGCTGGCGATTACAGGTTTCATCATGCCGTTTGTGATGG  
CTTCCATGTGCGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCG  
GGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCCTTAAACGCCTGGGGTAATGACTCTCT  
AGCTTGAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTTCGTTTTAT  
CTGTTGTTTGTGCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCCTCTAGATTACGTG  
CAGTCGATGATAAGCTGTCAAACATGAGAATTGTGCCTAATGAGTGAGCTAACTTACATT  
AATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATT  
AATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTT  
TTTTTTTTACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGA  
GAGTTGCAGCAAGCGGTCCACGCTGGTTTGGCCAGCAGGCGAAAATCCTGTTTGATG  
GTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGA  
GATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGC  
CATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTACAGCATTGCA  
TGGTTTGTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGA  
ATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAG  
AACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTC  
CACGCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGT  
CAGAGACATCAAGAAATAACGCCGGAACATTAGTGACAGGCAGCTTCCACAGCAATGGC  
ATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGAT  
TGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCAC  
GCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCG  
TGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGGCCGCCAGTT  
GTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCC  
CGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAG  
AGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTTACATTCACCACCCTG  
AATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCACCATTGAT  
GGTGTGCGAATTTCCGGGCAGCGTTGGGTCTGGCCACGGGTGCGCATGATCTAGAGCT  
GCCTCGCGCGTTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGAC  
GGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTC  
AGCGGGTGTGGCGGGTGTGCGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGG  
AGTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTAAGTACTGAGAGTGCACCATAT  
GCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCC  
GCTTCCCTCGCTCACTGACTCGCTGCGCTCGGTGCTTCGGCTGCGGCGAGCGGTATCAG

CTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAAC  
ATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGT  
TTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGG  
TGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCG  
TGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCG  
GGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGT  
TCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTACGCCGACCGCTGCGCCTTA  
TCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGC  
AGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTG  
AAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCT  
GAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCG  
CTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCT  
CAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAACGAAACTCAG  
TTAAGGGATTTTGGTCATGGCGTTAAGGGATTTTGGTCATGAATTAATTCTTAGAAAAAC  
TCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTT  
GAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCATAGGATGGCA  
AGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTC  
CCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGG  
TGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACG  
CTCGTCATCAAATCACTCGCATCAACCAAACCGTTATTCAATTCGTGATTGCGCCTGAGC  
GAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACC  
GGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCT  
AATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGG  
AGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTC  
TGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAAC  
CTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTGCGACCTGATTGCCCGACATTA  
TCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTA  
GAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCTTGTATTACTGTTTATGTAA  
GCAGACAGTTTTATTGTTTCATGACCATGACATTAACCTATAAAAATAGGCGTATCACGAG  
GCCCTTTCGTCTTCAC

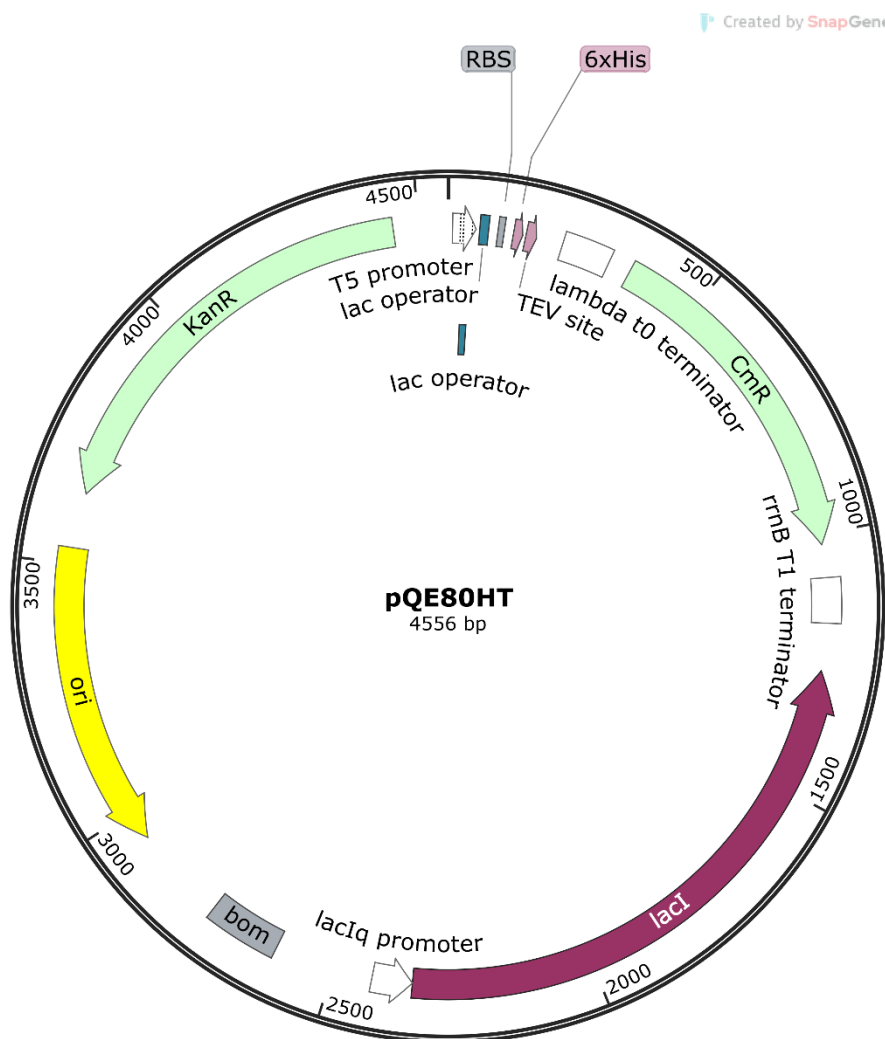


Figure S 96: Vector map of pQE80HT.

### pASG-IBA35 (3883 bp)

TCACGGATCTCCACGCGCCCTGTAGCGGCGCATTAAAGCGCGGCGGGTGTGGTGGTTA  
 CGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTT  
 CCCTTCCTTTCTCGCCACGTTTCGCCGGCTTTCCCGTCAAGCTCTAAATCGGGGGCTCC  
 CTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAATTGATTAGGGT  
 GATGGTTCACGTAGTGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGA  
 GTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTC  
 GGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTTCGGCCTATTGGTTAAAAAATGAG  
 CTGATTTAACAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTTCAGGTG  
 GCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAA  
 ATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGA  
 AGAGTATGAGTATTCAACATTTCCGTGTGCCCTTATTCCCTTTTTTTCGGGCATTTTGCC  
 TTCCTGTTTTTGTCTACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTG  
 GGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTT

TCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGG  
TATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTGCCGCATACACTATTCTCAG  
AATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGT  
AAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTC  
TGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCA  
TGTAACCTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAG  
CGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACCTATTAACCTGGCGA  
ACTACTTACTCTAGCTTCCCGGCAACAATTGATAGACTGGATGGAGGCGGATAAAGTTG  
CAGGACCACTTCTGCGCTCGGCCCTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGA  
GCCGGTGAGCGTGGCTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCT  
CCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGA  
CAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAGGAATTAATGATGTCTCG  
TTTAGATAAAAGTAAAGTGATTAACAGCGCATTAGAGCTGCTTAATGAGGTCCGGAATCGA  
AGGTTTAAACAACCCGTAAACTCGCCCAGAAGCTAGGTGTAGAGCAGCCTACATTGTATT  
GGCATGTAAAAATAAGCGGGCTTTGCTCGACGCCTTAGCCATTGAGATGTTAGATAGG  
CACCATACTCACTTTTGCCTTTAGAAGGGGAAAGCTGGCAAGATTTTTTACGTAATAAC  
GCTAAAAGTTTTAGATGTGCTTTACTAAGTCATCGCGATGGAGCAAAGTACATTTAGGT  
ACACGGCCTACAGAAAAACAGTATGAAACTCTCGAAAATCAATTAGCCTTTTTATGCCAA  
CAAGGTTTTTCACTAGAGAATGCATTATATGCACTCAGCGCAGTGGGGCATTTTACTTTA  
GGTTGCGTATTGGAAGATCAAGAGCATCAAGTGCCTAAAGAAGAAAGGGAAACACCTAC  
TACTGATAGTATGCCGCCATTATTACGACAAGCTATCGAATTATTTGATCACCAAGGTGC  
AGAGCCAGCCTTCTTATTCGGCCTTGAATTGATCATATGCGGATTAGAAAAACAACCTAA  
ATGTGAAAGTGGGTCTTAAAAGCAGCATAACCTTTTTCCGTGATGGTAACTTCACTAGTT  
TAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAATCCCTTAACGTGA  
GTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATC  
CTTTTTTCTGCGGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTG  
GTTTGTGTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACCTGGCTTACAGCAG  
AGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGA  
ACTCTGTAGCACCCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCC  
AGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGG  
CGCAGCGGTCGGGCTGAACGGGGGTTCTGTGCACACAGCCCAGCTTGGAGCGAACGA  
CCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGA  
AGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAC  
GAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCTGTGCGGTTTTCGCCAC  
CTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA  
ACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTTGCTGGCCTTTTTGCTCACATG  
ACCCGACACCATCGAATGGCCAGATGATTAATTCCTAATTTTTGTTGACACTCTATCATT

GATAGAGTTATTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAATGAATAGTTCGAC  
 AAAAATCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACAAATGGCTAGCCATCA  
 CCATCACCATCACTCCGGAATGAGAGACGCTGCAGCCCAATACGCAAACCGCCTCTCC  
 CCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGC  
 GGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTT  
 TACTTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTTACA  
 CAGGAAACAGCTATGACCATGATTACGCCAAGCTCGAAATTAACCCTCACTAAAGGGAA  
 CAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAAGTGTGGATCCCCCGGG  
 CTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCCGGTA  
 CCCAATTCGCCCTATAGTGAGTCGTATTACAATCACTGGCCGTCGTTTTACAACGTCGT  
 GACTGGGAAAACCCTGGCGTTACCCAACCTAATCGCCTTGCAGCACATCCCCCTTTCGC  
 CAGCTGGCGTAATAGCGAAGAGGCCCGCTCCTTTCGCTTTCCTTCCCTTCTTTCTCGCC  
 ACGTTCGCCGGCTTTCCCGTCAAGCTCTAAATCGGGGGCTCCCTTAGGGTTCCGATT  
 TAGTGCTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTCACCTCGAG  
 CGTCTCAGGGAGCTAAGGGAGCCACCCGCAAGCTTGACCTGTGAAGTAAAAATGGCG  
 CACATTGTGCGACATTTTTTTTTGTCTGCCGTTTACCGCTACTGCG

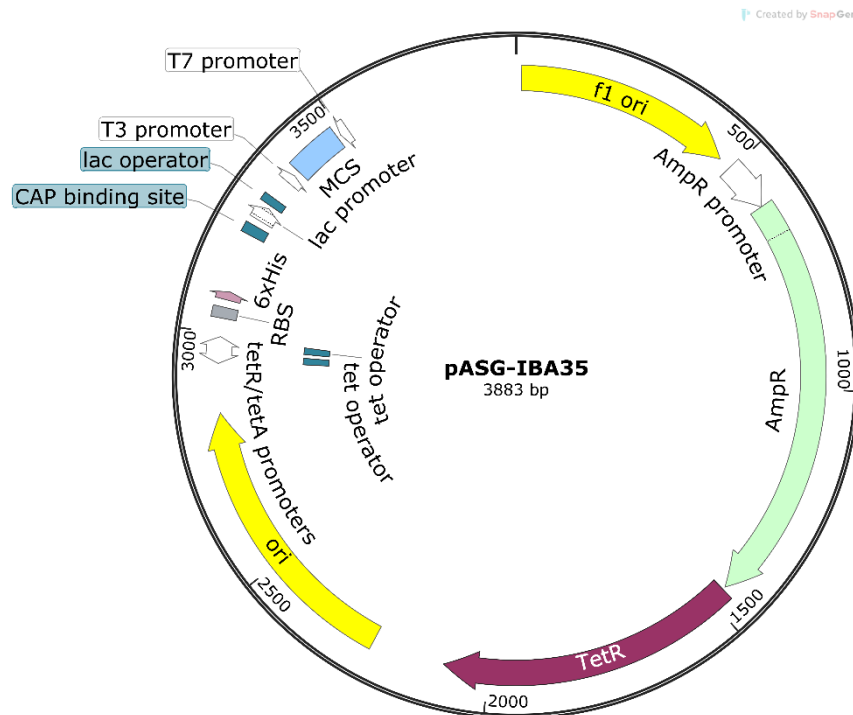


Figure S 97: Vector map of pASG-IBA35.

## 7.5.4 Gene and plasmid sequences

This chapter provides the sequences of all genes and plasmids including the plasmid maps used in this study.

### Kod-RI gene string (2375 bp)

**bold** = coding sequence

ATCCGAAAACCTGTATTTTCAGTCTATGATTCTGGACACCGATTATATCACCGAAGATG  
**GTAAACCGGTGATTTCGCATCTTCAAAAAGAAAACGGCGAGTTCAAATCGAGTATG**  
**ATCGTACCTTTGAGCCGTATTTCTATGCACTGCTGAAAGATGATAGCGCCATTGAAGA**  
**AGTAAAAAAATCACCGCAGAACGTCATGGCACCGTTGTTACCGTTAAACGTGTTGAA**  
**AAAGTGCAGAAAAAATTCCTGGGTCGTCCGGTTGAAGTTTGGAACTGTATTTACCC**  
**ATCCGCAGGATGTTCCGGCAATTCGTGATAAAATTCGTGAACATCCGGCAGTGATCGA**  
**TATCTATGAATATGATATTCGTTCCGCAACGCTACCTGATTGATAAAGGTCTGGTTC**  
**CGATGGAAGGTGATGAGGAACTGAAAATGCTGGCATTGTCATTGCAACCCTGTATCA**  
**TGAAGGTGAAGAATTTGCCGAAGGTCCGATTCTGATGATTAGCTATGCAGATGAAGAA**  
**GGTGCACGCGTTATTACCTGGAAAAATGTTGATCTGCCGTATGTTGATGTTGTTAGCA**  
**CCGAACGTGAAATGATTAACGTTTTCTGCGTGTGGTGAAGAAAAAGATCCGGATGT**  
**GCTGATTACCTATAACGGCGATAATTCGATTCGCCTATCTGAAAAACGCTGCGAA**  
**AACTGGGTATTAACCTTTGCACTGGGTCGTGATGGTAGCGAACCGAAAATTCAGCGTA**  
**TGGGTGATCGTTTTGCAGTTGAAGTTAAAGGTCGCATCCACTTTGATCTGTATCCGGTT**  
**ATTCGTGCTACCATTAATCTGCCGACCTATACACTGGAAGCAGTTTATGAAGCCGTTTT**  
**TGGTCAGCCGAAAGAAAAGGTTTATGCCGAAGAAATCACACCAGCATGGGAAACAGG**  
**TGAAAATCTGGAACGTGTTGCACGTTATAGCATGGAAGATGCAAAAGTTACCTATGAG**  
**CTGGGCAAAGAATTTCTGCCTATGGAAGCACAGCTGAGCCGTCTGATTGGTCAGAGC**  
**CTGTGGGATGTTAGCCGTAGCAGCACCGGTAATCTGGTTGAATGGTTTCTGCTGCGTA**  
**AAGCCTATGAACGTAATGAACTGGCACCGAACAAACCGGATGAAAAGAGCTGGCAC**  
**GTCGTCGTCAGAGCTATGAAGGTGGTTATGTTAAAGAACCGGAACGTGGTCTGTGGG**  
**AAAATATTGTTTATCTGGATTTCCGTAGCCTGTATCCGAGCATTATCATTACCCATAAT**  
**GTTAGTCCGGATACACTGAATCGTGAAGGCTGTAAAGAATATGATGTTGCACCGCAG**  
**GTTGGTCATCGTTTTGTAAAGATTTCCGGGTTTTATTCCGAGCCTGCTGGGTGATCT**  
**GCTGGAAGAACGTCAGAAAATCAAAAAAAGATGAAGGCCACCATCGATCCGATTGA**  
**ACGTAAACTGCTGGATTATCGTCAGCGTCGTATCAAATTCTGGCCAATAGTTATTATG**  
**GCTACTATGGTTATGCACGTGCCGTTGGTATTGCAAAGAATGTGCAGAAAGCGTTAC**  
**CGCCTGGGGTCGTGAATATATCACCATGACCATTAAAGAAATCGAAGAAAAATATGG**  
**CTTCAAAGTGATCTATAGCGATACCGATGGTTTTTTTTGCAACCATTCCGGGTGCAGAT**  
**GCCGAAACCGTTAAAAGAAAGCAATGGAATTCCTGAAGTATATCAACGCAAACTG**  
**CCTGGTGCACCTGGAACCTGGAATATGAAGGTTTCTATAAACGCGGTTTCTTCGTGACCA**

AGAAAAAATACGCAGTGATTGATGAAGAGGGCAAATTACAACCCGTGGTCTGGAAA  
TTGTTTCGTCGTGATTGGAGTGAAATTGCCAAAGAAACCCAGGCACGTGTTCTGGAAGC  
ACTGTTAAAAGATGGTGTGTGGAAAAAGCCGTGCGCATTGTGAAAGAAGTTACAGA  
AAAAGTGAAGTATGAAGTGCCTCCGGAAAAGCTGGTTATTCATATTCAGATTACC  
CGTGACCTGAAAGATTATAAAGCAACCGGTCCGCATGTTGCAGTTGCAAAACGTCTG  
GCAGCACGTGGTGTGAAAATTCGTCCGGGTACAGTTATTAGCTATATTGTTCTGAAAG  
GTAGCGGTTCGCATTGGTGTGATCGTGCAATTCCGTTTGATGAATTTGATCCGACCAAACA  
CAAATATGATGCCGAGTATTATATCGAGAATCAGGTTCTGCCTGCCGTTGAACGCATT  
CTGCGTGCATTTGGTTATCGTAAAGAAGATCTGCGTTATCAGAAAACCCGTCAGGTGG  
GTCTGAGCGCATGGCTGAAACCGAAAGGCACCTAAGAGCTCGGTACCCCGGGTTCGAC  
CTG

**pQE80HT\_Kod-RI (6881 bp)**

Underlined = His-Tag, underlined with dots = TEV-cleavage site, **bold** = coding sequence

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAA  
TTGTGAGCGGATAACAATTTACACAGAAATTCATTAAGAGGAGAAATTAAGTATGAGAG  
GATCGCATCACCATCACCATCACGGATCCGAAAACCTGTATTTTTCAGTCTATGATTCTGG  
**ACACCGATTATATCACCGAAGATGGTAAACCGGTGATTCCGCATCTTCAAAAAGAAAA**  
**CGGCGAGTTCAAATCGAGTATGATCGTACCTTTGAGCCGTATTTCTATGCACTGCTG**  
**AAAGATGATAGCGCCATTGAAGAAGTGAAAAAATCACCGCAGAACGTCATGGCACC**  
**GTTGTTACCGTTAAACGTGTTGAAAAAGTGCAGAAAAAATTCCTGGGTCGTCCGGTTG**  
**AAGTTTGGAACTGTATTTTACCCATCCGCAGGATGTTCCGGCAATTCGTGATAAAAT**  
**CGTGAACATCCGGCAGTGATCGATATCTATGAATATGATATTCCGTTCCGCAAACGCT**  
**ACCTGATTGATAAAGTCTGGTTCCGATGGAAGGTGATGAGGAACTGAAAATGCTGG**  
**CATTTGCCATTGCAACCCTGTATCATGAAGGTGAAGAATTTGCCGAAGGTCCGATTCT**  
**GATGATTAGCTATGCAGATGAAGAAGGTGCACGCGTTATTACCTGGAAAAATGTTGAT**  
**CTGCCGTATGTTGATGTTGTTAGCACCGAACGTGAAATGATTAACGTTTTTCTGCGTGT**  
**GGTGAAGAAAAAGATCCGGATGTGCTGATTACCTATAACGGCGATAATTTGATTTTC**  
**GCCTATCTGAAAAACGCTGCGAAAAACTGGGTATTAACCTTTGCACTGGGTCGTGATG**  
**GTAGCGAACCGAAAATTCAGCGTATGGGTGATCGTTTTGCAGTTGAAGTTAAAGGTCCG**  
**CATCCACTTTGATCTGTATCCGGTTATTCGTTCGTACCATTAATCTGCCGACCTATACAC**  
**TGGAAGCAGTTTATGAAGCCGTTTTTGGTCAGCCGAAAGAAAAGTTTATGCCGAAGA**  
**AATCACCAACCGCATGGGAAACAGGTGAAAATCTGGAACGTGTTGCACGTTATAGCAT**  
**GGAAGATGCAAAAGTTACCTATGAGCTGGGCAAAGAATTTCTGCCTATGGAAGCACA**  
**GCTGAGCCGTCTGATTGGTTCAGAGCCTGTGGGATGTTAGCCGTAGCAGCACCGGTAA**  
**TCTGGTTGAATGGTTTCTGCTGCGTAAAGCCTATGAACGTAATGAACTGGCACCGAAC**  
**AAACCGGATGAAAAAGAGCTGGCACGTCGTCGTCAGAGCTATGAAGGTGGTTATGTT**

AAAGAACCGGAACGTGGTCTGTGGGAAAATATTGTTTATCTGGATTTCCGTAGCCTGT  
ATCCGAGCATTATCATTACCCATAATGTTAGTCCGGATACACTGAATCGTGAAGGCTG  
TAAAGAATATGATGTTGCACCGCAGGTTGGTCATCGTTTTTTGTAAAGATTTTCCGGGT  
TTATTCCGAGCCTGCTGGGTGATCTGCTGGAAGAACGTCAGAAAATCAAAAAAAGAT  
GAAGGCCACCATCGATCCGATTGAACGTAAACTGCTGGATTATCGTCAGCGTCGTATC  
AAAATTCTGGCCAATAGTTATTATGGCTACTATGGTTATGCACGTGCCCGTTGGTATTG  
CAAAGAATGTGCAGAAAGCGTTACCGCCTGGGGTCGTGAATATATCACCATGACCATT  
AAAGAAATCGAAGAAAAATATGGCTTCAAAGTGATCTATAGCGATACCGATGGTTTTT  
TTGCAACCATTCCGGGTGCAGATGCCGAAACCGTTAAAAGAAAGCAATGGAATTCC  
TGAAGTATATCAACGCAAACTGCCTGGTGCCTGGAAGTGAATATGAAGGTTTCTA  
TAAACGCGGTTTCTTCGTGACCAAGAAAAAATACGCAGTGATTGATGAAGAGGGCAA  
AATTACAACCCGTGGTCTGGAAATTGTTTCGTGCTGATTGGAGTGAAATTGCCAAAGAA  
ACCCAGGCACGTGTTCTGGAAGCACTGTTAAAAGATGGTGATGTGGAAAAAGCCGTG  
CGCATTGTGAAAGAAGTTACAGAAAACTGAGCAAGTATGAAGTGCCTCCGGAAAAG  
CTGTTATTTCATATTCAGATTACCCGTGACCTGAAAGATTATAAAGCAACCGGTCCGC  
ATGTTGCAGTTGCAAAACGTCTGGCAGCACGTGGTGTGAAAATTCGTCCGGGTACAGT  
TATTAGCTATATTGTTCTGAAAGGTAGCGGTCGCATTGGTGATCGTGCAATTCCGTTTTG  
ATGAATTTGATCCGACCAACACAAATATGATGCCGAGTATTATATCGAGAATCAGGT  
TCTGCCTGCCGTTGAACGCATTCTGCGTGCATTTGGTTATCGTAAAGAAGATCTGCGTT  
ATCAGAAAACCCGTCAGGTGGGTCTGAGCGCATGGCTGAAACCGAAAGGCACCTAAG  
AGCTCGGTACCCCGGGTTCGACCTGCAGCCAAGCTTAATTAGCTGAGCTTGGACTCCTG  
TTGATAGATCCAGTAATGACCTCAGAACTCCATCTGGATTTGTTTCAGAACGCTCGGTTGC  
CGCCGGGCGTTTTTTATTGGTGAGAATCCAAGCTAGCTTGGCGAGATTTTCAGGAGCTA  
AGGAAGCTAAAATGGAGAAAAAATCACTGGATATAACCACCGTTGATATATCCCAATGG  
CATCGTAAAGAACATTTTGAGGCATTTTCAGTCAGTTGCTCAATGTACCTATAACCAGACC  
GTTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTAT  
CCGGCCTTTATTACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTTTCGTATGGC  
AATGAAAGACGGTGAGCTGGTGTATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCC  
ATGAGCAAACCTGAAACGTTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAG  
TTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCT  
AAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGT  
TTTGATTTAAACGTGGCCAATATGGACAACCTTCTTCGCCCCCGTTTTTACCATGGGCAAA  
TATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTTCAGGTTTCATCATGCCGT  
TTGTGATGGCTTCCATGTTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGT  
GGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCTTAAACGCCTGGGGT  
AATGACTCTTAGCTTGAGGCATCAAATAAACGAAAGGCTCAGTCGAAAGACTGGGCC  
TTTTCGTTTTATCTGTTGTTTGTTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCCTCT



AGATTACGTGCAGTCGATGATAAGCTGTCAAACATGAGAATTGTGCCTAATGAGTGAGC  
TAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTG  
CCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCATTGGGCG  
CCAGGGTGGTTTTCTTTTACCAGTGAGACGGGCAACAGCTGATTGCCCTTACCAGCC  
TGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAAT  
CCTGTTTGTATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTAT  
CCACTACCGAGATATCCGCACCAACGCGCAGCCCGACTCGGTAATGGCGCGCATTG  
CGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATT  
CAGCATTGTCATGGTTTGTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCG  
CTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGC  
GCCGAGACAGA ACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGA  
CCAGATGCTCCACGCCAGTCGCGTACCGTCTTCATGGGAGAAAATAACTGTTGATG  
GGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGACAGGCAGCTTCCA  
CAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGC  
GCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCG  
ACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTG  
CGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTT  
GCCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCT  
TCCACTTTTTCCCGCGTTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAAC  
GGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACAT  
TCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATAACCGCGAAAGGTTTTG  
CACCATTTCGATGGTGTGCGAATTTCCGGGCAGCGTTGGGTCTGGCCACGGGTGCGCAT  
GATCTAGAGCTGCCTCGCGCGTTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAG  
CTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGT  
CAGGGCGCGTCAGCGGGTGTGGCGGGTGTGGGGGCGCAGCCATGACCCAGTCACGT  
AGCGATAGCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTA CTGAGA  
GTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCA  
GGCGCTCTTCCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCCGGCTGCGGCG  
AGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACG  
CAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGC  
GTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCT  
CAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGG  
AAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCC  
TTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTC  
GGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTCCAGCCCGAC  
CGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATC  
GCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCT

ACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTAT  
CTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCA  
AACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGA  
AAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAA  
CGAAAACCTCACGTTAAGGGATTTTGGTCATGGCGTTAAGGGATTTTGGTCATGAATTAAT  
TCTTAGAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAA  
TACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCC  
ATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAA  
CCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACG  
ACTGAATCCGGTGAGAATGGCAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGC  
CAGCCATTACGCTCGTCATCAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGAT  
TGCGCCTGAGCGAGACGAAATACGCGATCGCTGTAAAAGGACAATTACAAACAGGAAT  
CGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTACCTGAATCAG  
GATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCAT  
GCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAG  
CCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTT  
CAGAAACAACCTCTGGCGCATCGGGCTTCCCATAACAATCGATAGATTGTCGCACCTGATT  
GCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTA  
ATCGCGGCCTAGAGCAAGACGTTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTA  
CTGTTTATGTAAGCAGACAGTTTTATTGTTTCATGACCATGACATTAACCTATAAAAAATAGG  
CGTATCACGAGGCCCTTTTCGTCTTCAC

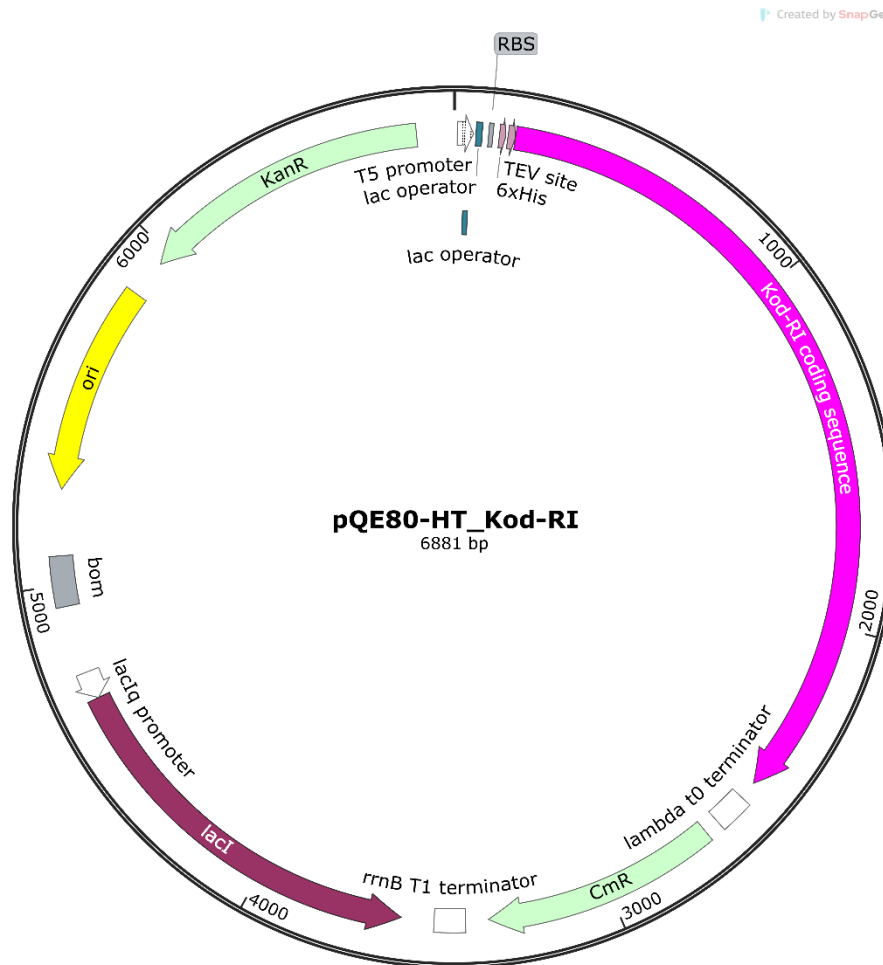


Figure S 98: Plasmid map of pQE80HT\_Kod-RI.

### pQE80HT\_Kod-RSGA (6881 bp)

Underlined = His-Tag, underlined with dots = TEV-cleavage site, **bold** = coding sequence

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAA  
 TTGTGAGCGGATAACAATTTACACACAGAATTCATTAAGAGGAGAAATTA ACTATGAGAG  
 GATCGCATCACCATCACCATCACGGATCCGAAAACCTGTATTTTCAGTCTATGATTCTGG  
**ACACCGATTATATCACCGAAGATGGTAAACCGGTGATTCGCATCTTCAAAAAAGAAAA**  
**CGGCGAGTTCAAAATCGAGTATGATCGTACCTTTGAGCCGTATTTCTATGCACTGCTG**  
**AAAGATGATAGCGCATTGAAGAAGTGAAAAAATCACCGCAGAACGTCATGGCACC**  
**GTTGTTACCGTTAAACGTGTTGAAAAAGTGCAGAAAAAATTCCTGGGTCGTCCGGTTG**  
**AAGTTTGGAACTGTATTTACCCATCCGCAGGATGTTCCGGCAATTCGTGATAAAATT**  
**CGTGAACATCCGGCAGTGATCGATATCTATGAATATGATATTCCGTTCCGCAAACGCT**  
**ACCTGATTGATAAAGGTCTGGTTCCGATGGAAGGTGATGAGGAACTGAAAATGCTGG**  
**CATTTGCCATTGCAACCCTGTATCATGAAGGTGAAGAATTTGCCGAAGGTCCGATTCT**  
**GATGATTAGCTATGCAGATGAAGAAGGTGCACGCGTTATTACCTGGAAAAATGTTGAT**  
**CTGCCGTATGTTGATGTTGTTAGCACCGAACGTGAAATGATTAACGTTTTTCTGCGTGT**

GGTGAAAGAAAAAGATCCGGATGTGCTGATTACCTATAACGGCGATAATTTTCGATTC  
GCCTATCTGAAAAACGCTGCGAAAACTGGGTATTAACTTTGCCTGGGTCTGATG  
GTAGCGAACCGAAAATTCAGCGTATGGGTGATCGTTTTGCAGTTGAAGTTAAAGGTGC  
CATCCACTTTGATCTGTATCCGGTTATTCGTCTGACCATTAATCTGCCGACCTATACAC  
TGGAAGCAGTTTATGAAGCCGTTTTTGGTCAGCCGAAAGAAAAGTTTTATGCCGAAGA  
AATCACCACCGCATGGGAAACAGGTGAAAATCTGGAACGTGTTGCACGTTATAGCAT  
GGAAGATGCAAAGTTACCTATGAGCTGGGCAAAGAATTTCTGCCTATGGAAGCACA  
GCTGAGCCGTCTGATTGGTCAGAGCCTGTGGGATGTTAGCCGTAGCAGCACCGGTAA  
TCTGGTTGAATGGTTTCTGCTGCGTAAAGCCTATGAACGTAATGAACTGGCACCGAAC  
AAACCGGATGAAAAGAGCTGGCACGTCTGTCAGAGCTATGAAGGTGGTTATGTT  
AAAGAACCGGAACGTGGTCTGTGGGAAAATATTGTTTATCTGGATTTCCGTAGCCTGT  
ATCCGAGCATTATCATTACCCATAATGTTAGTCCGGATACACTGAATCGTGAAGGCTG  
TAAAGAATATGATGTTGCACCGCAGGTTGGTCATCGTTTTTGTAAAGATTTTCCGGGT  
TTATTCCGAGCCTGCTGGGTGATCTGCTGGAAGAACGTCAGAAAATCAAAAAAAGAT  
GAAGGCCACCATCGATCCGATTGAACGTAACCTGCTGGATTATCGTCAGCGTCGTATC  
AAAATTCTGGCCAGCAGTTATTATGGCTACTATGGTTATGCACGTGCCCGTTGGTATTG  
CAAAGAATGTGCAGAAAGCGTTACCGCCTGGGGTCTGTAATATATCACCATGACCATT  
AAAGAAATCGAAGAAAAATATGGCTTCAAAGTGATCTATAGCGATACCGATGGTTTTT  
TTGCAACCATTCCGGGTGCAGATGCCGAAACCGTTAAAAGAAAGCAATGGAATTCC  
TGAAGTATATCAACGCAAAACTGCCTGGTGCCTGGAACCTGGAATATGAAGGTTTCTA  
TAAACGCGGTTTCTTCGTGACCAAGAAAAAATACGCAGTGATTGATGAAGAGGGCAA  
AATTACAACCGGCGGTCTGGAAATTGTTCTGTCGTGATTGGAGTGAATTGCCAAAGAA  
ACCCAGGCACGTGTTCTGGAAGCACTGTTAAAAGATGGTGATGTGGAAAAAGCCGTG  
CGCATTGTGAAAGAAGTTACAGAAAACTGAGCAAGTATGAAGTGCCTCCGAAAAG  
CTGGTTATTCATGAACAGATTACCCGTGACCTGAAAGATTATAAAGCAACCGGTCCGC  
ATGTTGCAGTTGCAAACGTCTGGCAGCACGTGGTGTGAAAATTCGTCCGGGTACAGT  
TATTAGCTATATTGTTCTGAAAGGTAGCGGTCCGATTGGTGATCGTGCAATTCCGTTTTG  
ATGAATTTGATCCGGCGAAACACAAATATGATGCCGAGTATTATATCGAGAATCAGGT  
TCTGCCTGCCGTTGAACGCATTCTGCGTGCATTTGGTTATCGTAAAGAAGATCTGCGTT  
ATCAGAAAACCCGTCAGGTGGGTCTGAGCGCATGGCTGAAACCGAAAGGCACCTAAG  
AGCTCGGTACCCCGGGTGCACCTGCAGCCAAGCTTAATTAGCTGAGCTTGGACTCCTG  
TTGATAGATCCAGTAATGACCTCAGAACTCCATCTGGATTTGTTTCAAGACGCTCGGTTGC  
CGCCGGGCGTTTTTTATTGGTGAGAATCCAAGCTAGCTTGGCGAGATTTTCAGGAGCTA  
AGGAAGCTAAAATGGAGAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGG  
CATCGTAAAGAACATTTTGGAGGCATTTTCAAGTCAAGTTGCTCAATGTACCTATAACCAGACC  
GTTTCAAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAAGTTTTAT  
CCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTTTCGTATGGC

AATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTACCCCTTGTTACACCGTTTTCC  
ATGAGCAAACCTGAAACGTTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAG  
TTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCT  
AAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGT  
TTTGATTTAAACGTGGCCAATATGGACAACCTTCTTCGCCCCCGTTTTACCATGGGCAAA  
TATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTTCATCATGCCGT  
TTGTGATGGCTTCCATGTCCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGT  
GGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCCTTAAACGCCTGGGGT  
AATGACTCTCTAGCTTGAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCC  
TTTTGTTTTATCTGTTGTTTGTCCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCCTCT  
AGATTACGTGCAGTCGATGATAAGCTGTCAAACATGAGAATTGTGCCTAATGAGTGAGC  
TAAC TTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTGCTG  
CCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCCTATTGGGCG  
CCAGGGTGGTTTTTCTTTTACCAGTGAGACGGGCAACAGCTGATTGCCCTTACCAGCC  
TGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGGCGAAAAT  
CCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTAT  
CCCCTACCAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTG  
CGCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATT  
CAGCATTGTCATGGTTTGTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCG  
CTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGC  
GCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGA  
CCAGATGCTCCACGCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATG  
GGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCA  
CAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCG  
GCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCG  
ACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTG  
CGACGGCGCGTGACAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTT  
GCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCT  
TCCACTTTTTCCCGCGTTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAAC  
GGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTTACAT  
TCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCAGCGAAAGGTTTTG  
CACCATTGATGGTGTCCGAATTTCCGGCAGCGTTGGGTCCCTGGCCACGGGTGCGCAT  
GATCTAGAGCTGCCTCGCGGTTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAG  
CTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGT  
CAGGGCGCGTCAGCGGGTGTGGCGGGTGTCCGGGCGCAGCCATGACCCAGTCACGT  
AGCGATAGCGGAGTGTATACTGGCTTAACCTATGCGGCATCAGAGCAGATTGTAAGTGA  
GTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCA

GGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTTCGTTCCGGCTGCGGCG  
AGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACG  
CAGGAAAGAACATGTGAGCAAAGGCCAGCAAAGGCCAGGAACCGTAAAAAGGCCG  
GTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCT  
CAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGG  
AAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCC  
TTTCTCCCTTCGGGAAGCGTGGCGCTTCTCATAGCTCACGCTGTAGGTATCTCAGTTC  
GGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGAC  
CGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATC  
GCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCT  
ACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTAT  
CTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCGGCA  
AACAAACCACCGCTGGTAGCGGTGGTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGA  
AAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAA  
CGAAAACCTCACGTTAAGGGATTTTGGTCATGGCGTTAAGGGATTTTGGTCATGAATTAAT  
TCTTAGAAAACTCATCGAGCATCAAATGAACTGCAATTTATTCATATCAGGATTATCAA  
TACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCC  
ATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAA  
CCTATTAATTTCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACG  
ACTGAATCCGGTGAGAATGGCAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGC  
CAGCCATTACGCTCGTCATCAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGAT  
TGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAAT  
CGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAG  
GATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCAT  
GCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAG  
CCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTT  
CAGAAACAACCTCTGGCGCATCGGGCTTCCCATAACAATCGATAGATTGTGCGCACCTGATT  
GCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTA  
ATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCTTGTATTA  
CTGTTTATGTAAGCAGACAGTTTTATTGTTTCATGACCATGACATTAACCTATAAAAAATAGG  
CGTATCACGAGGCCCTTTCGTCTTCAC

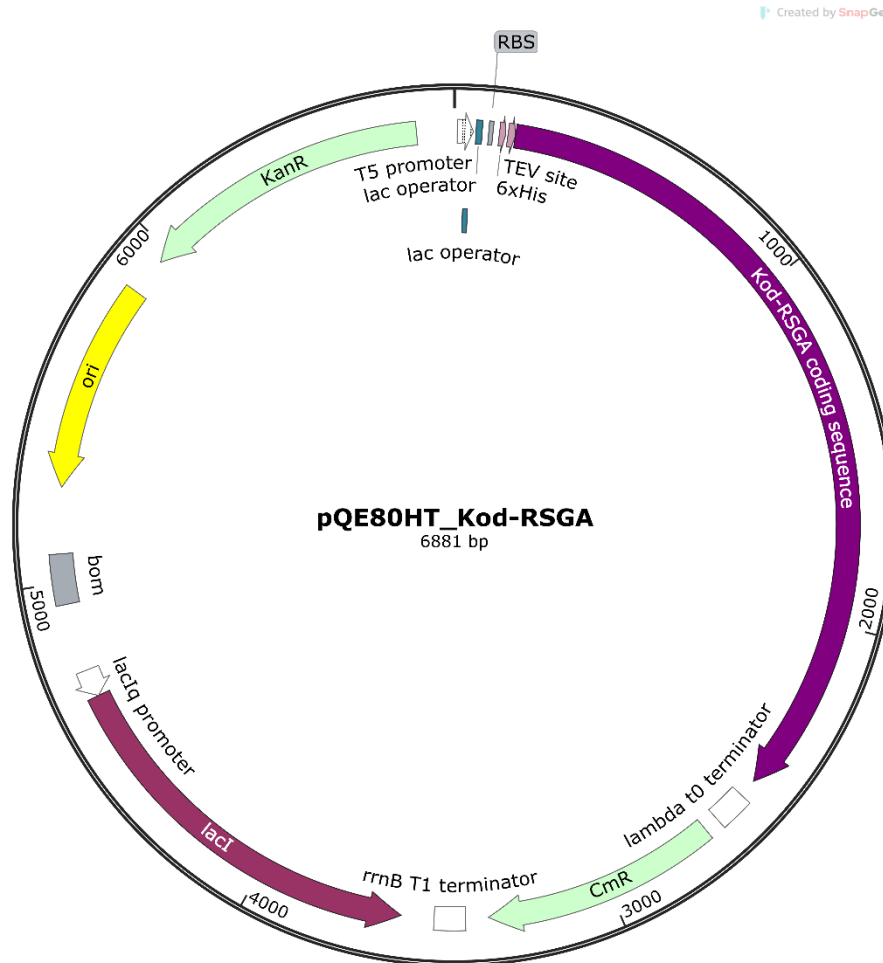


Figure S 99: Plasmid map of pQE80HT\_Kod-RSGA.

**pQE80HT\_Kod-RI\_2313delA (6880 bp)**

Underlined = His-Tag, underlined with dots = TEV-cleavage site, **bold** = coding sequence

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAA  
 TTGTGAGCGGATAACAATTTACACAGAAATTCATTAAGAGGAGAAATTA ACTATGAGAG  
 GATCGCATCACCATCACCATCACGGATCCGAAAACCTGTATTTTCAGTCTATGATTCTGG  
**ACACCGATTATATCACCGAAGATGGTAAACCGGTGATTTCGCATCTTCAAAAAAGAAAA**  
**CGGCGAGTTCAAAATCGAGTATGATCGTACCTTTGAGCCGTATTTCTATGCACTGCTG**  
**AAAGATGATAGCGCATTGAAGAAGTGAAAAAATCACCGCAGAACGTCATGGCACC**  
**GTTGTTACCGTTAAACGTGTTGAAAAAGTGCAAAAAAATTCCTGGGTCGTCCGGTTG**  
**AAGTTTGGAACTGTATTTACCCATCCGCAGGATGTTCCGGCAATTCGTGATAAAATT**  
**CGTGAACATCCGGCAGTGATCGATATCTATGAATATGATATTCCGTTCCGCAAACGCT**  
**ACCTGATTGATAAAGTCTGGTTCCGATGGAAGGTGATGAGGAACTGAAAATGCTGG**  
**CATTTGCCATTGCAACCCTGTATCATGAAGGTGAAGAATTTGCCGAAGGTCCGATTCT**  
**GATGATTAGCTATGCAGATGAAGAAGGTGCACGCGTTATTACCTGGAAAAATGTTGAT**  
**CTGCCGTATGTTGATGTTGTTAGCACCGAACGTGAAATGATTAACGTTTTTCTGCGTGT**

GGTGAAAGAAAAAGATCCGGATGTGCTGATTACCTATAACGGCGATAATTTTCGATTC  
GCCTATCTGAAAAACGCTGCGAAAACTGGGTATTAACTTTGCCTGGGTCTGATG  
GTAGCGAACCGAAAATTCAGCGTATGGGTGATCGTTTTGCAGTTGAAGTTAAAGGTGC  
CATCCACTTTGATCTGTATCCGGTTATTCGTCTGACCATTAATCTGCCGACCTATACAC  
TGGAAGCAGTTTATGAAGCCGTTTTTGGTCAGCCGAAAGAAAAGTTTATGCCGAAGA  
AATCACCACCGCATGGGAAACAGGTGAAAATCTGGAACGTGTTGCACGTTATAGCAT  
GGAAGATGCAAAGTTACCTATGAGCTGGGCAAAGAATTTCTGCCTATGGAAGCACA  
GCTGAGCCGTCTGATTGGTCAGAGCCTGTGGGATGTTAGCCGTAGCAGCACCGGTAA  
TCTGGTTGAATGGTTTCTGCTGCGTAAAGCCTATGAACGTAATGAACTGGCACCGAAC  
AAACCGGATGAAAAAGAGCTGGCACGTCGTCGTCAGAGCTATGAAGGTGGTTATGTT  
AAAGAACCGGAACGTGGTCTGTGGGAAAATATTGTTTATCTGGATTTCCGTAGCCTGT  
ATCCGAGCATTATCATTACCCATAATGTTAGTCCGGATACACTGAATCGTGAAGGCTG  
TAAAGAATATGATGTTGCACCGCAGGTTGGTCATCGTTTTTGTAAAGATTTTCCGGGT  
TTATTCCGAGCCTGCTGGGTGATCTGCTGGAAGAACGTCAGAAAATCAAAAAAAGAT  
GAAGGCCACCATCGATCCGATTGAACGTAACCTGCTGGATTATCGTCAGCGTCGTATC  
AAAATTCTGGCCAATAGTTATTATGGCTACTATGGTTATGCACGTGCCCGTTGGTATTG  
CAAAGAATGTGCAGAAAGCGTTACCGCCTGGGGTCTGTAATATATCACCATGACCATT  
AAAGAAATCGAAGAAAAATATGGCTTCAAAGTGATCTATAGCGATACCGATGGTTTTT  
TTGCAACCATTCCGGGTGCAGATGCCGAAACCGTTAAAAGAAAGCAATGGAATTCC  
TGAAGTATATCAACGCAAAACTGCCTGGTGCCTGGAACCTGGAATATGAAGGTTTCTA  
TAAACGCGGTTTCTTCGTGACCAAGAAAAAATACGCAGTGATTGATGAAGAGGGCAA  
AATTACAACCCGTGGTCTGGAAATTGTTCTGTCGTGATTGGAGTGAAATTGCCAAAGAA  
ACCCAGGCACGTGTTCTGGAAGCACTGTTAAAAGATGGTGATGTGGAAAAAGCCGTG  
CGCATTGTGAAAGAAGTTACAGAAAAACTGAGCAAGTATGAAGTGCCTCCGAAAAG  
CTGGTTATTCATATTCAGATTACCCGTGACCTGAAAGATTATAAAGCAACCGGTCCGC  
ATGTTGCAGTTGCAAACGTCTGGCAGCACGTGGTGTGAAAATTCGTCCGGGTACAGT  
TATTAGCTATATTGTTCTGAAAGGTAGCGGTCGCATTGGTGATCGTGCATTCCGTTTGA  
TGAATTTGATCCGACCAAACACAAATATGATGCCGAGTATTATATCGAGAATCAGGTT  
CTGCCTGCCGTTGAACGCATTCTGCGTGCATTTGGTTATCGTAAAGAAGATCTGCGTT  
ATCAGAAAACCCGTCAGGTGGGTCTGAGCGCATGGCTGAAACCGAAAGGCACCTAAG  
AGCTCGGTACCCCGGGTCCGACCTGCAGCCAAGCTTAATTAGCTGAGCTTGGACTCCT  
GTTGATAGATCCAGTAATGACCTCAGAACTCCATCTGGATTTGTTCAGAACGCTCGGT  
TGCCGCCGGGCGTTTTTTATTGGTGAGAATCCAAGCTAGCTTGGCGAGATTTTCAGGA  
GCTAAGGAAGCTAAAATGGAGAAAAAATCACTGGATATACCACCGTTGATATATCCC  
AATGGCATCGTAAAGAACATTTTGGAGGCATTTTCAGTCAGTTGCTCAATGTACCTATAAC  
CAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACA  
AGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTT



CGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCCACCCTTGTTACA  
CCGTTTTCCATGAGCAAACCTGAAACGTTTTTCATCGCTCTGGAGTGAATACCACGACGA  
TTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTG  
GCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGT  
GAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACCTTCTTCGCCCCCGTTT  
TCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTG  
AGGTTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACA  
ACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGC  
CCTTAAACGCCTGGGGTAATGACTCTCTAGCTTGAGGCATCAAATAAAACGAAAGGCT  
CAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTGCGGTGAACGCTCTCCTGAG  
TAGGACAAATCCGCCCTCTAGATTACGTGCAGTCGATGATAAGCTGTCAAACATGAGA  
ATTGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTC  
CAGTCGGGAAACCTGTGCTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGA  
GGCGGTTTGCATATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAAC  
AGCTGATTGCCCTTACCAGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTG  
GTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATG  
AGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCC  
GGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCAT  
CGCAGTGGGAACGATGCCCTCATTAGCATTGTCATGGTTTGTGAAAACCGGACATG  
GCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATT  
TATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAAC  
AGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCAGTCGCGTACCG  
TCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATA  
ACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCG  
GATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTT  
ACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTG  
ATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACT  
GGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCG  
GTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTTCGCAG  
AAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCAT  
ACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCT  
TCCGGGCGCTATCATGCCATACCGCGAAAGTTTTGCACCATTGATGGTGTGCGGAAT  
TTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGCATGATCTAGAGCTGCCTCGCGCG  
TTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTTCACAGC  
TTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTG  
TTGGCGGGTGTGCGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATA  
CTGGCTTAACTATGCGGCATCAGAGCAGATTGTAAGTACTGAGAGTGCACCATATGCGGTGT

GAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCGCTTCC  
TCGCTCACTGACTCGCTGCGCTCGGTCGTTCCGGCTGCGGCGAGCGGTATCAGCTCAC  
TCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATG  
TGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTT  
TTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAG  
GTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCTGGAAGCTCCCT  
CGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTT  
CGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGT  
CGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTACGCCGACCGCTGCGC  
CTTATCCGGTAACACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTG  
GCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGA  
GTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGC  
GCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAAC  
AAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAA  
AAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAC  
GAAACTCACGTTAAGGGATTTTGGTCATGGCGTTAAGGGATTTTGGTCATGAATTAA  
TTCTTAGAAAACTCATCGAGCATCAAATGAACTGCAATTTATTCATATCAGGATTAT  
CAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCA  
GTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCA  
ATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATG  
AGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGT  
TCAACAGGCCAGCCATTACGCTCGTCATCAAATCACTCGCATCAACCAAACCGTTAT  
TCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATT  
ACAAACAGGAATCGAATGCAACCCGGCGCAGGAACACTGCCAGCGCATCAACAATATT  
TTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAG  
TGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAG  
GCATAAATTCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAAC  
GCTACCTTTGCCATGTTTCAGAAACAACCTCTGGCGCATCGGGCTTCCCATACAATCGA  
TAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATC  
AGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGG  
CTCATAACACCCCTTGTACTGTTTATGTAAGCAGACAGTTTTATTGTTTCATGACCAT  
GACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTTCGTCTTAC

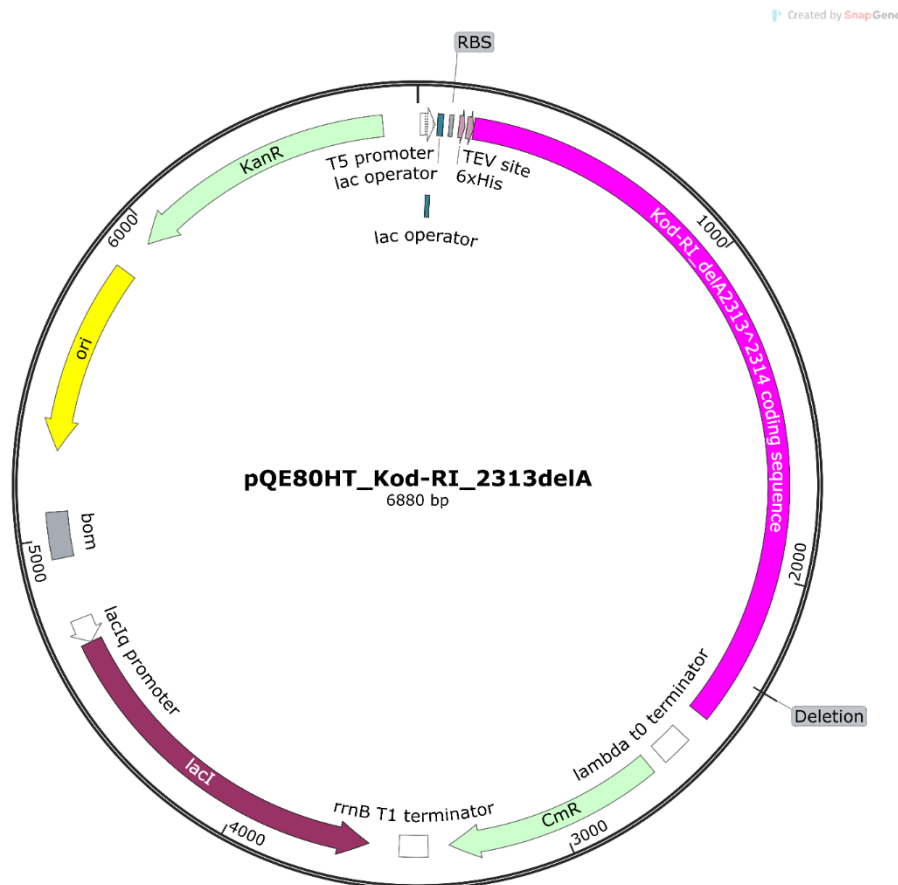


Figure S 100: Plasmid map of pQE80HT\_Kod-RI\_2313delA.

### pMA-RQ\_Kod-RI (4722 bp)

**bold** = coding sequence

```

CTAAATTGTAAGCGTTAATATTTTGTAAAATTCGCGTTAAATTTTTGTAAATCAGCTCAT
TTTTTAACCAATAGGCCGAAATCGGCAAATCCCTTATAAATCAAAGAATAGACCGAGA
TAGGGTTGAGTGCCGCTACAGGGCGCTCCCATTCGCCATTCAGGCTGCGCAACTGTT
GGGAAGGGCGTTTTCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGA
TGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCAGTCACGACGTTGTAA
AACGACGGCCAGTGAGCGCGACGTAATACGACTCACTATAGGGCGAATTGGCGGAAGG
CCGTCAAGGCCGCATTTAATTAATGATTTTTAGTCTGAGCTCATGATTCTGGATACCGA
TTATATCACCGAAGATGGTAAACCGGTGATTTCGCATCTTCAAAAAGAAAACGGCGA
GTTCAAATCGAGTATGATCGTACCTTTGAGCCGTATTTCTATGCACTGCTGAAAGATG
ATAGCGCCATTGAAGAAGTGAAAAAATCACCGCAGAACGTCATGGCACCGTTGTTA
CCGTTAAACGTGTTGAAAAGTGCAGAAAAAATTCCTGGGTCGTCCGGTTGAAGTTTG
GAACTGTATTTTACCCATCCGCAGGATGTTCCGGCAATTCGTGATAAAATTCGTGAA
CATCCGGCAGTGATCGATATCTATGAATATGATATTCCGTTCCGCAAACGCTACCTGA
TTGATAAAGGTCTGGTTCCGATGGAAGGTGATGAGGAAGTAAAATGCTGGCATTTCG
CATTGCAACCCTGTATCATGAAGGTGAAGAATTTGCCGAAGGTCCGATTCTGATGATT

```

AGCTATGCAGATGAAGAAGGTGCACGCGTTATTACCTGGAAAAATGTTGATCTGCCGT  
ATGTTGATGTTGTTAGCACCGAACGTGAAATGATTAAACGTTTTCTGCGTGTGGTGAA  
AGAAAAAGATCCGGATGTGCTGATTACCTATAACGGCGATAATTTTCGATTTGCGCTAT  
CTGAAAAACGCTGCGAAAACTGGGTATTAACCTTGCCTGGGTGCGTATGGTAGC  
GAACCGAAAATTCAGCGTATGGGTGATCGTTTTGCAGTTGAAGTTAAAGGTCGCATCC  
ACTTTGATCTGTATCCGGTTATTCGTGCTACCATTAATCTGCCGACCTATACACTGGAA  
GCAGTTTATGAAGCCGTTTTTGGTCAGCCGAAAGAAAAGGTTTATGCCGAAGAAATCA  
CCACCGCATGGGAAACAGGTGAAAATCTGGAACGTGTTGCACGTTATAGCATGGAAG  
ATGCAAAAGTTACCTATGAGCTGGGCAAAGAATTTCTGCCTATGGAAGCACAGCTGA  
GCCGTCTGATTGGTCAGAGCCTGTGGGATGTTAGCCGTAGCAGCACCGGTAATCTGG  
TTGAATGGTTTCTGCTGCGTAAAGCCTATGAACGTAATGAACTGGCACCGAACAAACC  
GGATGAAAAAGAGCTGGCACGTCGTCGTCAGAGCTATGAAGGTGGTTATGTTAAAGA  
ACCGGAACGTGGTCTGTGGGAAAATATTGTTTATCTGGATTTCCGTAGCCTGTATCCG  
AGCATTATCATTACCATAATGTTAGTCCGGATACACTGAATCGTGAAGGCTGTAAAG  
AATATGATGTTGCACCGCAGGTTGGTCATCGTTTTTGTAAAGATTTTCCGGGTTTTATT  
CCGAGCCTGCTGGGTGATCTGCTGGAAGAACGTCAGAAAATCAAAAAAAGATGAAG  
GCCACCATCGATCCGATTGAACGTAACCTGCTGGATTATCGTCAGCGTCGTATCAAAA  
TTCTGGCCAATAGTTATTATGGCTACTATGGTTATGCACGTGCCCGTTGGTATTGCAA  
GAATGTGCAGAAAGCGTTACCGCCTGGGGTCGTGAATATATCACCATGACCATTA  
GAAATCGAAGAAAAATATGGCTTCAAAGTGATCTATAGCGATACCGATGGTTTTTTTG  
CAACCATTCCGGGTGCAGATGCCGAAACCGTTAAAAAGAAAGCAATGGAATTCCTGA  
AGTATATCAACGCAAACTGCCTGGTGCCTGGAACCTGGAATATGAAGGTTTCTATA  
ACGCGGTTTCTTCGTGACCAAGAAAAAATACGCAGTGATTGATGAAGAGGGCAAAT  
TACAACCCGTGGTCTGGAAATTGTTGTCGTCGATTGGAGTGAAATTGCCAAAGAAACC  
CAGGCACGTGTTCTGGAAGCACTGTTAAAAGATGGTGATGTGGAAAAGCCGTGCGC  
ATTGTGAAAGAAGTTACAGAAAACTGAGCAAGTATGAAGTGCCTCCGGAAAAGCTG  
GTTATTCATATTCAGATTACCGTGACCTGAAAGATTATAAAGCAACCGGTCCGCATG  
TTGCAGTTGCAAAACGTCTGGCAGCACGTGGTGTGAAAATTCGTCCGGGTACAGTTAT  
TAGCTATATTGTTCTGAAAGGTAGCGGTCGCATTGGTGATCGTGCAATTCGTTTTGATG  
AATTTGATCCGACCAAACACAAATATGATGCCGAGTATTATATCGAGAATCAGGTTCT  
GCCTGCCGTTGAACGCATTCTGCGTGCATTTGGTTATCGTAAAGAAGATCTGCGTTAT  
CAGAAAACCCGTCAGGTGGGTCTGAGCGCATGGCTGAAACCGAAAGGCACCTAAAAG  
CTTAATTAGCTGAGCTTGGCGCGCCCTGGGCCTCATGGGCCTCCGCTCACTGCCCGC  
TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAACATGGTCATAGCTGTTTCCTTG  
CGTATTGGGCGCTCTCCGCTTCCCTCGCTCACTGACTCGCTGCGCTCGGTGTTCCGGT  
AAAGCCTGGGGTGCCTAATGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGG  
CCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGA

CGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCC  
CTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTC  
CGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCA  
GTTCCGGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCC  
CGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACT  
TATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGG  
TGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTG  
GTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCC  
GGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCG  
CAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGT  
GGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCT  
AGATCCTTTTAAATTAATAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAACTTG  
GTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCC  
TTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTAC  
CATCTGGCCCCAGTGCTGCAATGATACCGCGAGAACCACGCTCACCGGCTCCAGATTT  
ATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTA  
TCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGT  
TAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGT  
TTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCC  
ATGTTGTGCAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTGTGAGAAGTAAGTT  
GGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGC  
CATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGT  
GTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACA  
TAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACCTCTCAA  
GGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCT  
TCAGCATCTTTTACTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGC  
CGCAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCTTTTCA  
ATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATT  
TAGAAAAATAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

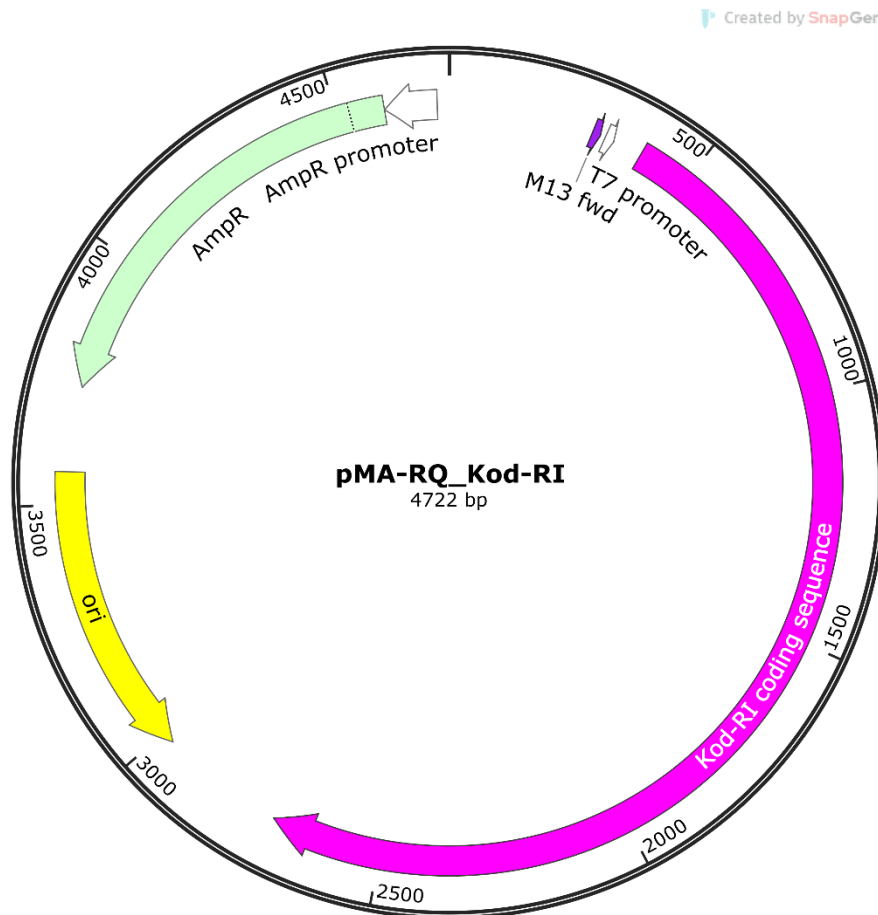


Figure S 101: Plasmid map of pMA-RQ\_Kod-RI.

**pET28a(+)\_mZBP1short<sup>WT</sup> (5850 bp)**

Underlined = His-Tags, underlined with dots = Thrombin cleavage site, **bold** = coding sequence

**GCAGAGGCGCCGGTGGATCTTAGTACTGGAGATAACTTGAACAGAAAATTTTACAA  
GTACTCTCGGACGATGGTGGGCCGGTCAAATCGGTCAATTAGTAAAGAAGTGTCAG  
GTACCAAAGAAGACTCTGAACCAGGTGTTATATAGACTCAAGAAAGAAGATCGCGTTT  
CGAGCCCTGAACCGGCCACCTGGAGTATTGGTGGTGCAGCCAGCGGCCGACGGTGCA  
CCGGCGATTCCCGAAAATAGCTCAGCGCAACCGTCTCTGGACGAGCGTATTCTGCGC  
TTTCTTGAGGCGAACGGCCCCATCGTGCTCTCCATATTGCGAAAGCCCTCGGTATGA  
CAACCGCGAAGGAAGTAAATCCTCTGTTGTACAGTATGCGGAACAAACATCTCTTGTC  
TTACGATGGTCAAACCTTGAAAATATATCATAGCAGACAAGAGGGTCAGGATATTGTG  
CTGCCTTGCTCTCAGGCTGCCCCCGAACGCATCACGTGGATCAGGCCGGGCTGGAG  
**CCGACCGAAATATTTCTGCTGCTTCCGATCAAGTTTTGGGACCTCGAGCACCACCACC**  
ACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGC  
CACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAACGGGTCTTGAGGGGT  
TTTTTGCTGAAAGGAGGAAGTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGG**

CGCATTAAGCGCGGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAG  
CGCCCTAGCGCCCGCTCCTTTGCTTTCTTCCCTTCTTTCTCGCCACGTTGCGCCGGCT  
TTCCCGGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGG  
CACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTG  
ATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTT  
CCAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTT  
GCCGATTTGCGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTT  
TAACAAAAATATTAACGCTTACAATTTAGGTGGCACTTTTCGGGGAAATGTGCGCGGAAC  
CCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAG  
AAAACTCATCGAGCATCAAATGAACTGCAATTTATTCATATCAGGATTATCAATACCAT  
ATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACCTCACCGAGGCAGTTCATAGG  
ATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTAT  
TAATTTCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGA  
ATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGC  
CATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCG  
CCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAA  
TGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATA  
TTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCAT  
CATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCGGTCAGCCAG  
TTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGA  
AACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCC  
GACATTATCGCGAGCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCG  
CGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTT  
TATGTAAGCAGACAGTTTTATTGTTTCATGACCAAATCCCTTAACGTGAGTTTTCGTTCCA  
CTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCG  
CGTAATCTGCTGCTTGCAAACAAAAAACCAACCGCTACCAGCGGTGGTTTGTGTTGCCGG  
ATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACCA  
AATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACC  
GCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGT  
CGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCCGG  
CTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTG  
AGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGG  
ACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAG  
GGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGT  
CGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGG  
CCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCTGCGTTAT  
CCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGC

AGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATG  
CGGTATTTTCTCCTTACGCATCTGTGCGGTATTTACACCGCAATGGTGCACCTCTCAGTA  
CAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACT  
GGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTT  
GTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTG  
TCAGAGGTTTTACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCA  
GCGTGGTCGTGAAGCGATTACAGATGTCTGCCTGTTTCATCCGCGTCCAGCTCGTTGA  
GTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTT  
TTTTCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTTCATGGGGGTAAT  
GATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCC  
CGGTTACTGGAACGTTGTGAGGGTAAACAACCTGGCGGTATGGATGCGGCGGGACCAGA  
GAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAG  
GGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACT  
TCCGCGTTTTCCAGACTTTACGAAACACGGAACCGAAGACCATTTCATGTTGTTGCTCAG  
GTCGCAGACGTTTTGCAGCAGCAGTCGCTTACGTTTCGCTCGCGTATCGGTGATTTCATT  
CTGCTAACCCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACG  
ATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAAC  
GTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATAC  
CGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCTCTCGCCGAAAAT  
GACCCAGAGCGCTGCCGGCACCTGTCTACGAGTTGCATGATAAAGAAGACAGTCATA  
AGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTGGAAG  
GCTCTCAAGGGCATCGGTGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAAT  
TGC GTT GCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAAT  
GAATCGGCCAACGCGCGGGGAGAGGCGGTTTGC GTATTGGGCGCCAGGGTGGTTTTT  
CTTTTACCAGTGAGACGGGCAACAGCTGATTGCCCTTACCGCCTGGCCCTGAGAGA  
GTTGCAGCAAGCGGTCCACGCTGGTTTGGCCAGCAGGCGAAAATCCTGTTTGATGGT  
GGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCCTACCGAGA  
TATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCA  
TCTGATCGTTGGCAACCAGCATCGCAGTGGAACGATGCCCTCATTTCAGCATTTGCATG  
GTTTGTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAAT  
TTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAA  
CTTAATGGGCCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCA  
CGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCA  
GAGACATCAAGAAATAACGCCGGAACATTAGTGACAGGCAGCTTCCACAGCAATGGC  
ATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGT  
GCACCGCCGCTTTACAGGCTTCGACGCGCTTCGTTCTACCATCGACACCACCACGCT  
GGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGC



AGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTT  
 GTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGC  
 GTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGA  
 CACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTTACATTACCACCCTGAATT  
 GACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTTCGATGGTG  
 TCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTA  
 GGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGC  
 CCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCAT  
 GAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCA  
 GCAACCGCACCTGTGGCGCCGGTGTATGCCGGCCACGATGCGTCCGGCGTAGAGGATC  
 GAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAAC  
 AATCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAG  
CCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATG

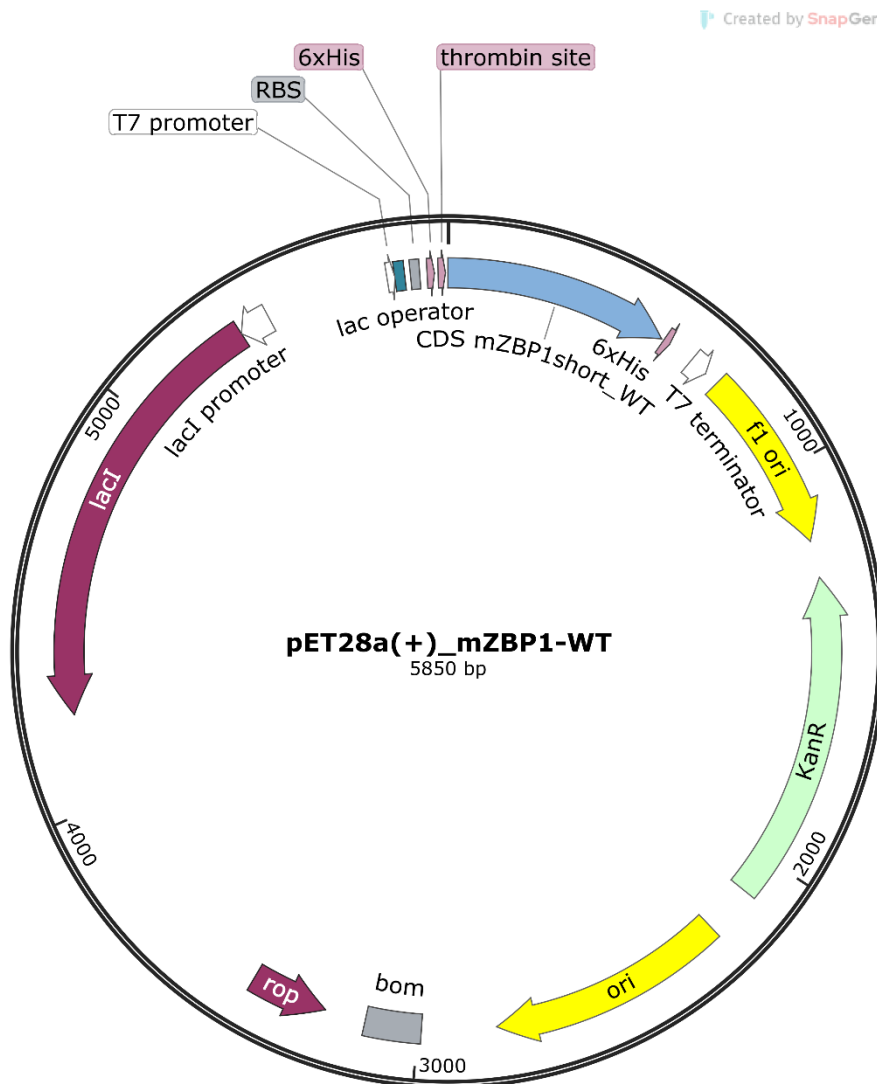


Figure S 102: Plasmid map of pET28a(+)\_mZBP1<sup>WT</sup>.

**pET28a(+)\_mZBP1short<sup>mZα1</sup> (5850 bp)**

Underlined = His-Tags, underlined...with...dots = Thrombin cleavage site, **bold** = coding sequence

**GCAGAGGCGCCGGTGGATCTTAGTACTGGAGATAACTTGGAACAGAAAATTTTACAA  
 GTACTCTCGGACGATGGTGGGCCGGTCAAATCGGTCAATTAGTAAAGAAGTGTGAG  
 GTACCAAAGAAGACTCTGGACCAGGTGTTAGCGAGACTCAAGAAAGAAGATCGCGTT  
 TCGAGCCCTGAACCGGCCACCTGGAGTATTGGTGGTGCAGCCAGCGGCGACGGTGC  
 ACCGGCGATTCCCGAAAATAGCTCAGCGCAACCGTCTCTGGACGAGCGTATTCTGCG  
 CTTTCTTGAGGCGAACGGCCCCATCGTGCTCTCCATATTGCGAAAGCCCTCGGTATG  
 ACAACCGCGAAGGAAGTAAATCCTCTGTTGTACAGTATGCGGAACAAACATCTTTGT  
 CTTACGATGGTCAAACCTTGAAAATATATCATAGCAGACAAGAGGGTCAGGATATTGT  
 GCTGCCTTGCTCTCCAGGCTGCCCCGAACGCATCACGTGGATCAGGCCGGGCTGGA  
 GCCGACCGAAATATTTCTGCTGCTTCCGATCAAGTTTTGGGACCTCGAGCACCACCAC  
CACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTG  
 CCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAACGGGTCTTGAGGGG  
 TTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCG  
 GCGCATTAAAGCGCGGGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCA  
 GCGCCCTAGCGCCCGCTCCTTTGCTTTCTTCCCTTCTTTCTCGCCACGTTGCGCGGC  
 TTTCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACG  
 GCACCTCGACCCCAAAAACCTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCT  
 GATAGACGGTTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGT  
 TCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTT  
 GCCGATTTGCGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTT  
 TAACAAAATATTAACGCTTACAATTTAGGTGGCACTTTTCGGGGAAATGTGCGCGGAAC  
 CCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAG  
 AAAAATCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCAT  
 ATTTTTGAAAAGCCGTTTCTGTAATGAAGGAGAAAACCTCACCGAGGCAGTTCCATAGG  
 ATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTAT  
 TAATTTCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGA  
 ATCCGGTGAGAATGGCAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGC  
 CATTACGCTCGTCATCAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCG  
 CCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAA  
 TGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATA  
 TTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCAT  
 CATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCGGTCAGCCAG  
 TTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGA  
 AACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCC**

GACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCG  
CGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTACTGT  
TATGTAAGCAGACAGTTTTATTGTTTCATGACCAAATCCCTTAACGTGAGTTTTCGTTCCA  
CTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCG  
CGTAATCTGCTGCTTGCAAACAAAAAACCACCGCTACCAGCGGTGGTTTGTGGCCGG  
ATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAGATACCA  
AATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACC  
GCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGT  
CGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCCGG  
CTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTG  
AGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGG  
ACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAG  
GGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGT  
CGATTTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGG  
CCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTAT  
CCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGC  
AGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATG  
CGGTATTTCTCCTTACGCATCTGTGCGGTATTTACACCCGCAATGGTGCCTCTCAGTA  
CAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACTCCGCTATCGCTACGTGACT  
GGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTT  
GTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTG  
TCAGAGGTTTTACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCA  
GCGTGGTCGTGAAGCGATTACAGATGTCTGCCTGTTTCATCCGCGTCCAGCTCGTTGA  
GTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTT  
TTTTCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTTCATGGGGGTAAT  
GATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCC  
CGGTTACTGGAACGTTGTGAGGGTAAACAACCTGGCGGTATGGATGCGGGCGGGACCAGA  
GAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAG  
GGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACT  
TCCGCGTTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTTCATGTTGTTGCTCAG  
GTCGCAGACGTTTTGCAGCAGCAGTCGCTTACGTTTCGCTCGCGTATCCGGTATTGATT  
CTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCTCAACGACAGGAGCAGC  
ATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAAC  
GTTTGGTGGCGGGACCAGTGACGAAGGCTTGAAGCGAGGGCGTGCAAGATTCCGAATAC  
CGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCTCGCCGAAAT  
GACCCAGAGCGCTGCCGGCACCTGTCTACGAGTTGCATGATAAAGAAGACAGTCATA  
AGTGCGGGCAGCAGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAG

GCTCTCAAGGGCATCGGTGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAAT  
TGC GTT GCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAAT  
GAATCGGCCAACGCGCGGGGAGAGGCGGTTTGC GTATTGGGCGCCAGGGTGGTTTTT  
CTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTACC GCCTGGCCCTGAGAGA  
GTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGT  
GGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCC ACTACCGAGA  
TATCCGCACCAACGCGCAGCCCCGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCA  
TCTGATCGTTGGCAACCAGCATCGCAGTGGAACGATGCCCTCATT CAGCATTTGCATG  
GTTTGTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAAT  
TTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAA  
CTTAATGGGCCC GCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCA  
CGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCA  
GAGACATCAAGAAATAACGCCGGAACATTAGTG CAGGCAGCTTCCACAGCAATGGCATC  
CTGGTCATCCAGCGGATAGTTAATGATCAGCCC ACTGACGCGTTGCGCGAGAAGATTGT  
GCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCT  
GGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGC  
AGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTT  
GTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGC  
GTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGA  
CACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATT  
GACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTTGCGCCATTCGATGGTG  
TCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTA  
GGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGC  
CCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCAT  
GAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCA  
GCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATC  
GAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAAC  
AATCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAG  
CCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATG

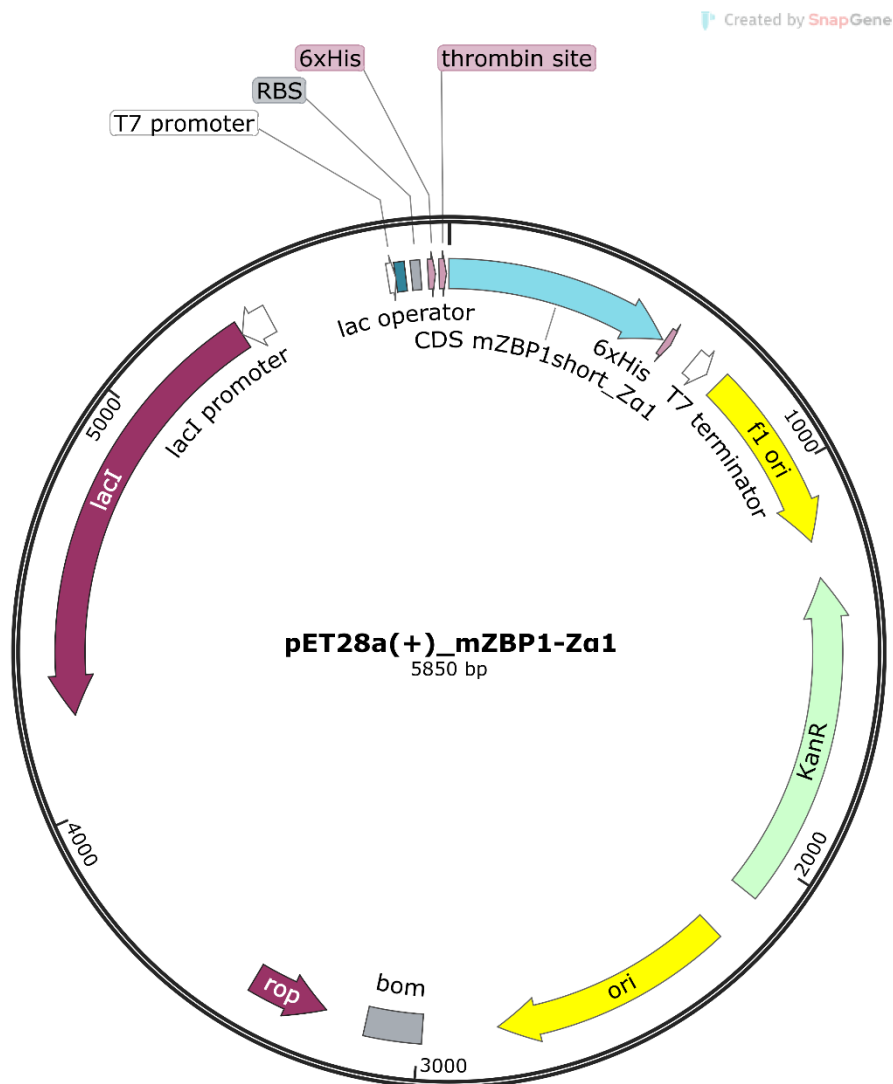


Figure S 103: Plasmid map of pET28a(+)\_mZBP1<sup>mZα1</sup>.

**pET28a(+)\_mZBP1short<sup>mZα2</sup> (5850 bp)**

Underlined = His-Tags, underlined with dots = Thrombin cleavage site, **bold** = coding sequence

**GCAGAGGCCCGGTGGATCTTAGTACTGGAGATAACTTGGAACAGAAAATTTTACAA  
GTA**  
**CTCTCGGACGATGGTGGGCCGGTCAAATCGGTCAATTAGTAAAGAAGTGTCAG  
GTACCAAAGAAGACTCTGAACCAGGTGTTATATAGACTCAAGAAAGAAGATCGCGTTT  
CGAGCCCTGAACCGGCCACCTGGAGTATTGGTGGTGCAGCCAGCGGCGACGGTGCA  
CCGGCGATTCCGAAAATAGCTCAGCGCAACCGTCTCTGGACGAGCGTATTCTGCGC  
TTTCTTGAGGCGAACGGCCCCATCGTGCTCTCCATATTGCGAAAGCCCTCGGTATGA  
CAACCGCGAAGGAAGTAGATCCTCTGTTGGCGAGTATGCGGAACAAACATCTTTGT  
CTTACGATGGTCAA**  
**ACTTGGAAAATATATCATAGCAGACAAGAGGGTCAGGATATTGT  
GCTGCCTTGCTCTCCAGGCTGCCCGAACGCATCACGTGGATCAGGCCGGGCTGGA  
GCCGACCGAAATATTTCTGCTGCTTCCGATCAAGTTTTGGGACCTCGAGCACCACCAC**

CACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTG  
CCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAACGGGTCTTGAGGGG  
TTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCG  
GCGCATTAAAGCGCGGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCA  
GCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCTTTCTCGCCACGTTGCGCGGC  
TTTCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACG  
GCACCTCGACCCCAAAAACTTGATTAGGGTGTGGTTCACGTAGTGGGCCATCGCCCT  
GATAGACGGTTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGT  
TCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTT  
GCCGATTTGCGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTT  
TAACAAAATATTAACGCTTACAATTTAGGTGGCACTTTTCGGGGAAATGTGCGCGGAAC  
CCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAG  
AAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCAT  
ATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACCTCACCGAGGCAGTTCCATAGG  
ATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTAT  
TAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGA  
ATCCGGTGAGAATGGCAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGC  
CATTACGCTCGTCATCAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCG  
CCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAA  
TGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATA  
TTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCAT  
CATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCGTCAGCCAG  
TTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGA  
AACAACTCTGGCGCATCGGGCTTCCATACAATCGATAGATTGTCGCACCTGATTGCCC  
GACATTATCGCGAGCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCG  
CGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGATTACTGTT  
TATGTAAGCAGACAGTTTTATTGTTTCATGACCAAAATCCCTTAACGTGAGTTTTTCGTTCCA  
CTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCG  
CGTAATCTGCTGCTTGCAAACAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGG  
ATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACCA  
AATACTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACC  
GCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGT  
CGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGCGG  
CTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTG  
AGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGG  
ACAGGTATCCGGTAAGCGGCAGGGTCCGAACAGGAGAGCGCACGAGGGAGCTTCCAG  
GGGGAAACGCCTGGTATCTTTATAGTCCTGTCCGGTTTTCGCCACCTCTGACTTGAGCGT

CGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGG  
CCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTAT  
CCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGC  
AGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATG  
CGGTATTTCTCCTTACGCATCTGTGCGGTATTTACACCGCAATGGTGCCTCTCAGTA  
CAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACT  
GGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTT  
GTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTG  
TCAGAGTTTTTACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCA  
GCGTGGTCGTGAAGCGATTACAGATGTCTGCCTGTTTCATCCGCGTCCAGCTCGTTGA  
GTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTT  
TTTTCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTTCATGGGGGTAAT  
GATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCC  
CGGTTACTGGAACGTTGTGAGGGTAAACAACACTGGCGGTATGGATGCGGCGGGACCAGA  
GAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAG  
GGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACT  
TCCGCGTTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTTCATGTTGTTGCTCAG  
GTCGCAGACGTTTTGCAGCAGCAGTCGCTTACGTTTCGCTCGCGTATCGGTGATTCAAT  
CTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCAGC  
ATCATGCGCACCCGTGGGGCCGCGCATGCCGCGGATAATGGCCTGCTTCTCGCCGAAAC  
GTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATAC  
CGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCTCGCCGAAAAT  
GACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATA  
AGTGCGGCGACGATAGTCATGCCCCGCGCCACCGGAAGGAGCTGACTGGGTTGAAG  
GCTCTCAAGGGCATCGGTGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAAT  
TGCGTTGCGCTCACTGCCCGCTTTCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAAT  
GAATCGGCCAACGCGCGGGGAGAGGCGGTTTTGCGTATTGGGCGCCAGGGTGGTTTTT  
CTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTACCGCCTGGCCCTGAGAGA  
GTTGCAGCAAGCGGTCCACGCTGGTTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGT  
GGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGA  
TATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCAGCGCCA  
TCTGATCGTTGGCAACCAGCATCGCAGTGGAACGATGCCCTCATTTCAGCATTTGCATG  
GTTTGTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAAT  
TTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAA  
CTTAATGGGCCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCA  
CGCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCA  
GAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATC

CTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGT  
 GCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCT  
 GGCACCCAGTTGATCGGGCGGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGC  
 AGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTT  
 GTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGC  
 GTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGA  
 CACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATT  
 GACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTG  
 TCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTA  
 GGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGC  
 CCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCAT  
 GAGCCCGAAGTGGCGAGCCCGATCTTCCCATCGGTGATGTCGGCGATATAGGCGCCA  
 GCAACCGCACCTGTGGCGCCGGTGTATGCCGGCCACGATGCGTCCGGCGTAGAGGATC  
 GAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAAC  
 AATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAG  
CCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATG

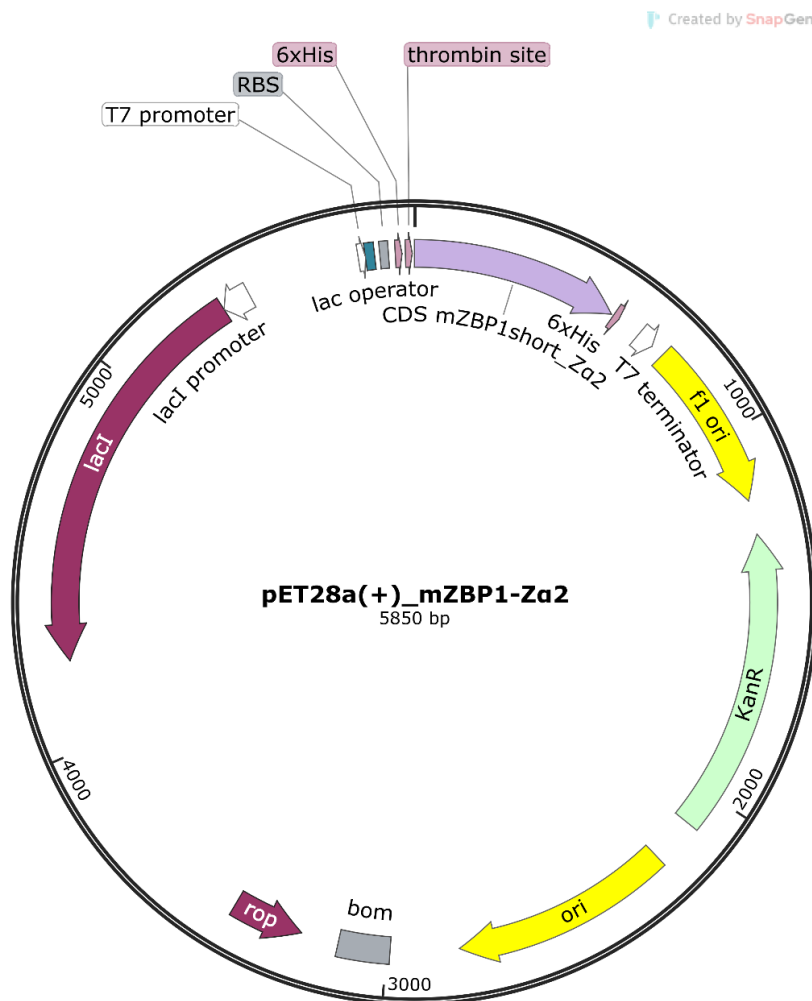


Figure S 104: Plasmid map of pET28a(+)\_mZBP1<sup>mZα2</sup>.



**pET28a(+)\_mZBP1short<sup>mZα1-2</sup> (5850 bp)**

Underlined = His-Tags, underlined...with...dots = Thrombin cleavage site, **bold** = coding sequence

**GCAGAGGCGCCGGTGGATCTTAGTACTGGAGATAACTTGGAACAGAAAATTTTACAA  
GTACTCTCGGACGATGGTGGGCCGGTCAAATCGGTCAATTAGTAAAGAAGTGTCAG  
GTACCAAAGAAGACTCTGGACCAGGTGTTAGCGAGACTCAAGAAAGAAGATCGCGTT  
TCGAGCCCTGAACCGGCCACCTGGAGTATTGGTGGTGCAGCCAGCGGCGACGGTGC  
ACCGGCGATTCCCGAAAATAGCTCAGCGCAACCGTCTCTGGACGAGCGTATTCTGCG  
CTTTCTTGAGGCGAACGGCCCCATCGTGCTCTCCATATTGCGAAAGCCCTCGGTATG  
ACAACCGCGAAGGAAGTAGATCCTCTGTTGGCGAGTATGCGGAACAAACATCTCTTG  
TCTTACGATGGTCAAACCTTGAAAAATATATCATAGCAGACAAGAGGGTCAGGATATTG  
TGCTGCCTTGCTCTCCAGGCTGCCCCCGAACGCATCACGTGGATCAGGCCGGGCTGG  
AGCCGACCGAAATATTTCTGCTGCTTCCGATCAAGTTTTGGGACCTCGAGCACCACCA  
CCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCT  
GCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGG  
GTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGC  
GGCGCATTAAGCGCGGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCC  
AGCGCCCTAGCGCCCGCTCCTTTGCTTTCTTCCCTTCTTTCTCGCCACGTTTCGCCGG  
CTTTCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTAC  
GGCACCTCGACCCCAAAAACCTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCC  
CTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTT  
GTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGAT  
TTTGCCGATTTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAA  
TTTTAACAAAATATTAACGCTTACAATTTAGGTGGCACTTTTCGGGGAAATGTGCGCGGA  
ACCCCTATTTGTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTT  
AGAAAACTCATCGAGCATCAAATGAACTGCAATTTATTCATATCAGGATTATCAATACC  
ATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACCTCACCGAGGCAGTTCCATAG  
GATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTA  
TTAATTTCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTG  
AATCCGGTGAGAATGGCAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGC  
CATTACGCTCGTCATCAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCG  
CCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAA  
TGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATA  
TTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCAT  
CATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAG  
TTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGA  
AACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCC**

GACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCG  
CGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTT  
TATGTAAGCAGACAGTTTTATTGTTTCATGACCAAATCCCTTAACGTGAGTTTTTCGTTCCA  
CTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCG  
CGTAATCTGCTGCTTGCAAACAAAAAACCACCGCTACCAGCGGTGGTTTGTGGCCGG  
ATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAGATACCA  
AATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACC  
GCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGT  
CGTGTCTTACCGGGTTGGAAGTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCCGG  
CTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTG  
AGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGG  
ACAGGTATCCGGTAAGCGGCAGGGTCCGAACAGGAGAGCGCACGAGGGAGCTTCCAG  
GGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGT  
CGATTTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGG  
CCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTAT  
CCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGC  
AGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATG  
CGGTATTTCTCCTTACGCATCTGTGCGGTATTTACACCCGCAATGGTGCCTCTCAGTA  
CAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACTCCGCTATCGCTACGTGACT  
GGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTT  
GTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTG  
TCAGAGGTTTTACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCA  
GCGTGGTCGTGAAGCGATTACAGATGTCTGCCTGTTTCATCCGCGTCCAGCTCGTTGA  
GTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTT  
TTTTCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTTCATGGGGGTAAT  
GATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCC  
CGGTTACTGGAACGTTGTGAGGGTAAACAACCTGGCGGTATGGATGCGGGCGGGACCAGA  
GAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAG  
GGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACT  
TCCGCGTTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTTCATGTTGTTGCTCAG  
GTCGCAGACGTTTTGCAGCAGCAGTCGCTTACGTTTCGCTCGCGTATCGGTGATTCATT  
CTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCAGC  
ATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAAC  
GTTTGGTGGCGGGACCAGTGACGAAGGCTTGAAGCGAGGGCGTGCAAGATTCCGAATAC  
CGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCTCGCCGAAAT  
GACCCAGAGCGCTGCCGGCACCTGTCTACGAGTTGCATGATAAAGAAGACAGTCATA  
AGTGCGGGCAGCAGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAG

GCTCTCAAGGGCATCGGTGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAAT  
TGC GTT GCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAAT  
GAATCGGCCAACGCGCGGGGAGAGGCGGTTTGC GTATTGGGCGCCAGGGTGGTTTTT  
CTTTTCACCAAGT GAGACGGGCAACAGCTGATTGCCCTTACCGCCTGGCCCTGAGAGA  
GTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGT  
GGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCCTACCGAGA  
TATCCGCACCAACGCGCAGCCCGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCA  
TCTGATCGTTGGCAACCAGCATCGCAGTGGAACGATGCCCTCATT CAGCATTTGCATG  
GTTTGTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAAT  
TTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAA  
CTTAATGGGCCC GCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCA  
CGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCA  
GAGACATCAAGAAATAACGCCGGAACATTAGTG CAGGCAGCTTCCACAGCAATGGCATC  
CTGGTCATCCAGCGGATAGTTAATGATCAGCCCCTGACGCGTTGCGCGAGAAGATTGT  
GCACCGCCGCTTTACAGGCTTCGACGCGCTTCGTTCTACCATCGACACCACCACGCT  
GGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTG  
AGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTT  
GTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGC  
GTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGA  
CACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATT  
GACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTG  
TCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTA  
GGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGC  
CCAACAGTCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCAT  
GAGCCCGAAGTGGCGAGCCCGATCTTCCCATCGGTGATGTCGGCGATATAGGCGCCA  
GCAACCGCACCTGTGGCGCCGGT GATGCCGGCCACGATGCGTCCGGCGTAGAGGATC  
GAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAAC  
AATCCCCTCTAGAAATAATTTTGT TTA ACTTTAAGAAGGAGATATACCATGGGCAGCAG  
CCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATG

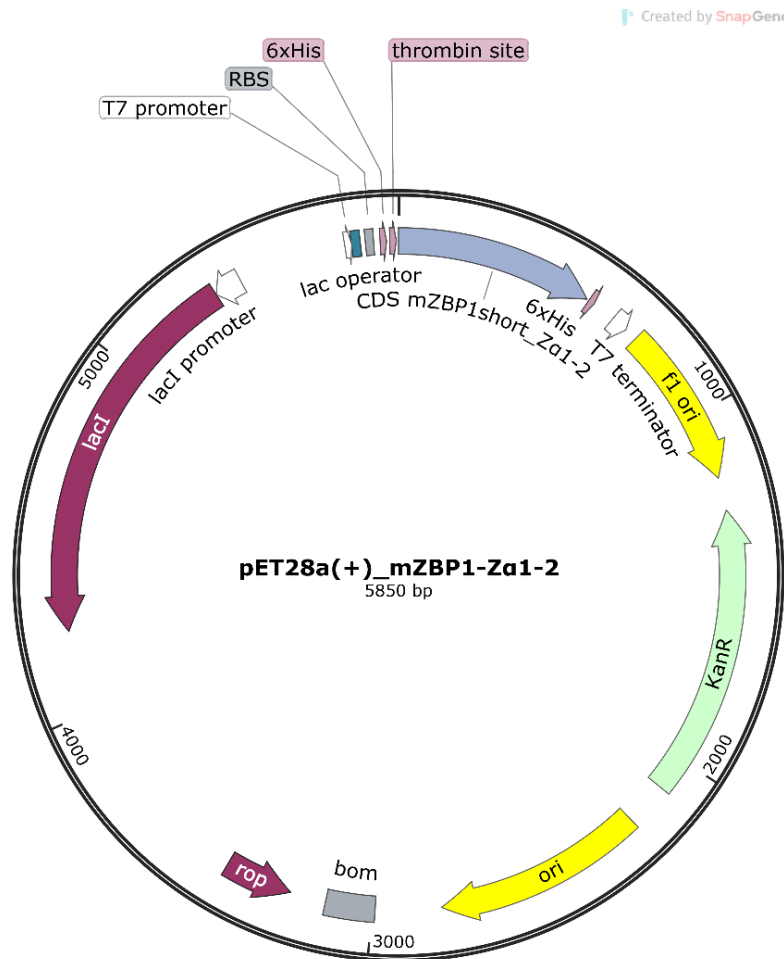


Figure S 105: Plasmid map of pET28a(+)\_mZBP1<sup>mZα1-2</sup>.

### 7.5.5 Amino acid sequences

This chapter provides the one-letter amino acid sequences of all proteins that were expressed and purified during this study.

#### Kod-RI polymerase<sup>151</sup>

MILDTDYITEDGKPVIRIFKKENGEFKIEYDRTFEPYFYALLKDDSAIEEVKKITAERHGTVVTV  
 KRVEKVQKKFLGRPVEVWKLYFTHPQDVPAIRDKIREHPAVIDIYEDIPFAKRYLIDKGLVPM  
 EGDEELKMLAFIAITLYHEGEEFAEGPILMISYADEEGARVITWKNVDLPYVDVSTEREMIK  
 RFLRVVKEKDPDVLITYNGDNDFAYLKKRCEKLGINFALGRDGSEPQIRMGDRFAVEVKG  
 RIHFDLYPVIRRTINLPTYTLEAVYEAVFGQPKEKVYAAEEITTAWETGENLERVARYSMEDAK  
 VTYELGKEFLPMEAQLSRLIGQSLWDVSRSSSTGNLVEWFLLRKAYERNELAPNKPDEKELA  
 RRRQSYEGGYVKEPERGLWENIVYLDFRSLYPSIIITHNVSPDTLNLREGCKEYDVAPQVGH  
 R FCKDFPGFIPSLLDLLEERQKIKKKMKATIDPIERKLLDYRQRRRIKILANSYYGYYGYARARW

YCKECAESVTAWGREYITMTIKEIEEKYGFKVIYSDDGFFATIPGADAETVKKKAMEFLKYIN  
AKLPGALELEYEGFYKRGFFVTKKKYAVIDEEGKITTRGLEIVRRDWSEIAKETQARVLEALL  
KGDVVEKAVRIVKEVTEKLSKYEVPEKLVIIHQITRDLKDYKATGPHVAVAKRLAARGVKIRP  
GTVISYIVLKGSGRIGDRAIPFDEFDPTKHKYDAEYYIENQVLP AVERILRAFGYRKEDLRYQK  
TRQVGLSAWLKPKGT\*

**Kod-RSGA polymerase<sup>152</sup>**

MILD TDYIT EDGKPVIRIFKKENG EFKIEYDRTFEPYFYALLKDDSAIEEVKKITAERHGTVVTV  
KRVEKVQKKFLGRPVEVWKLYFTHPQDVPAIRDKIREHPAVIDIYEYDIPFAKRYLIDKGLVPM  
EGDEELKMLAFIAIATLYHEGEEFAEGPILMISYADEEGARVITWKNVDLPYVDVSTEREMIK  
RFLRVVKEKDPDVLITYNGDNDFAYLKKRCEKLGINFALGRDGSEPKIQRMGDRFAVEVKG  
RIHFDLYPVIRRTINLPTYTLEAVYEAVFGQPKEKVYAEIITAWETGENLERVARYSMEDAK  
VTYELGKEFLPMEAQLSRLIGQSLWDVSRSSSTGNLVEWFLLRKAYERNELAPNKPDEKELA  
RRRQSYEGGYVKEPERGLWENIVYLDFRSLYPSIIITHNVSPDTLNREGCKEYDVAPQVGHR  
FCKDFPGFIPSLLDLLEERQKIKKKMKATIDPIERKLLDYRQRRIKILASSYYGYGYARARW  
YCKECAESVTAWGREYITMTIKEIEEKYGFKVIYSDDGFFATIPGADAETVKKKAMEFLKYIN  
AKLPGALELEYEGFYKRGFFVTKKKYAVIDEEGKITTRGLEIVRRDWSEIAKETQARVLEALL  
KGDVVEKAVRIVKEVTEKLSKYEVPEKLVIIHQITRDLKDYKATGPHVAVAKRLAARGVKIR  
PGTVISYIVLKGSGRIGDRAIPFDEFDPAKHKYDAEYYIENQVLP AVERILRAFGYRKEDLRYQ  
KTRQVGLSAWLKPKGT\*

**mZBP1<sup>WT</sup>**

AEAPVDLSTGDNLEQKILQVLSDDGGPVKIGQLVKKCQVPKKTLDQVLYRLKKEDRVSSPEP  
ATWSIGGAASGDGAPAIPENSSAQPSLDERILRFLEANGPHRALHIAKALGMTTAKEVNPLLY  
SMRNKHLLSYDGQTWKIYHSRQEGQDIVLPCSPGCPRTHHVDQAGLEPTEIFLLLPIKFWD

**mZBP1<sup>mZα1</sup>**

AEAPVDLSTGDNLEQKILQVLSDDGGPVKIGQLVKKCQVPKKTLDQVLYRLKKEDRVSSPEP  
ATWSIGGAASGDGAPAIPENSSAQPSLDERILRFLEANGPHRALHIAKALGMTTAKEVNPLLY  
SMRNKHLLSYDGQTWKIYHSRQEGQDIVLPCSPGCPRTHHVDQAGLEPTEIFLLLPIKFWD

**mZBP1<sup>mZα2</sup>**

AEAPVDLSTGDNLEQKILQVLSDDGGPVKIGQLVKKCQVPKKTLDQVLYRLKKEDRVSSPEP  
ATWSIGGAASGDGAPAIPENSSAQPSLDERILRFLEANGPHRALHIAKALGMTTAKEVDPLLA  
SMRNKHLLSYDGQTWKIYHSRQEGQDIVLPCSPGCPRTHHVDQAGLEPTEIFLLLPIKFWD

**mZBP1<sup>mZα1-1</sup>**

AEAPVDLSTGDNLEQKILQVLSDDGGPVKIGQLVKKCQVPKKTLDQVLARLKKEDRVSSPEP  
 ATWSIGGAASGDGAPAIPESSAQPSLDERILRFLEANGPHRALHIAKALGMTTAKEVDPLLA  
 SMRNKHLLSYDGQTWKIYHSRQEGQDIVLPCSPGCPRTHHVDQAGLEPTEIFLLLPIKFWD

## 7.6 List of abbreviations

Abbreviation	Definition
<sup>1</sup> Ψ	<i>N</i> 1-methylpseudouridine
2'F-RNA	2'-fluoro RNA
2'MOE RNA	2'- <i>O</i> -methoxyethyl RNA
2'OMe RNA	2'- <i>O</i> -methyl RNA
6Cl-G	6-chloropurine-2-amine
6-FAM	6-carboxyfluorescein
7dG	7-deazaguanosine
<b>A</b>	adenine
ADAR1	adenosine deaminase acting on RNA 1
AIBN	azobisisobutyronitrile
AmNA	amido-bridged nucleic acid
AMV	avian myeloblastosis virus
APS	ammoniumperoxodisulfate
aq.	aqueous
AQUA	Advanced Quick Assembly
ASO	antisense oligonucleotide
BNA	bridged nucleic acid
BODIPY	boron-dipyrromethene
BP	benzophenone
bp	base pairs
bpNA	boranophosphate nucleic acid
br	broad
BSA	<i>N,O</i> -bis(trimethylsilyl)acetamide
BSA	bovine serum albumin
Bst	<i>Bacillus stearothermophilus</i>
Bz-Cl	benzoyl chloride
<b>C</b>	cytosine
C5	complementary protein C5
CASP8	caspase-8
cDNA	complementary DNA
CeNA	cyclohexene nucleic acid
CEP	cyanoethyl phosphoramidite
CEP-Cl	bis(2-cyanoethyl)- <i>N,N</i> -diisopropylchlorophosphoramidite
CHARMM	chemistry at Harvard macromolecular mechanics
c-Hex	cyclohexane
cLNA	carbocyclic LNA
CP	methylcyclopropene

---

CuAAC	copper-catalyzed azide-alkyne cycloaddition
CV	column volume
CXCL9	transplant rejection biomarker
dA	deoxyadenosine
dA <sup>CPTP</sup>	C <sup>7</sup> -methylcyclopropene-modified 7-deaza-2'-deoxy-adenosine triphosphate
dC	deoxycytosine
DCC	<i>N,N'</i> -dicyclohexylcarbodiimid
DCE	1,2-dichloroethane
DCM	dichloromethane
DD	death domain
ddH <sub>2</sub> O	ultrapure water
DED	death effector domain
dG	deoxyguanosine
DIBAL-H	diisobutylaluminium hydride
DIPEA	diisopropylethylamine
DMAc	<i>N,N</i> -dimethylacetamide
DMAP	4-(dimethylamino)pyridine
DMF	<i>N,N</i> -dimethylformamide
DMF-DMA	<i>N,N'</i> -dimethylformamide dimethyl acetal
DMSO	dimethylsulfoxide
DMT-Cl	4,4'-dimethoxytrityl chloride
DNA	deoxyribonucleic acid
dPAGE	denaturing polyacrylamide gel electrophoresis
DPC-Cl	diphenylcarbonyl chloride
DTT	dithiothreitol
dU <sup>CPTP</sup>	C <sup>5</sup> -methylcyclopropene-modified 2'-deoxyuridine triphosphate
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
EdU	5-ethynyl-2'-deoxyuridine
<i>e.g.</i>	for example (Lat. <i>exempli gratia</i> )
EI	electron ionization
EMSA	electrophoretic mobility shift assays
eq.	equivalent(s)
ESI	electrospray ionisation
<i>et al.</i>	and other (Lat. <i>et alii</i> )
EtOH	ethanol
exo-	exonuclease deficient
exo+	exonuclease proficient
exSELEX	genetic alphabet expansion SELEX
exTNA	expanded genetic alphabet TNA
FADD	Fas-associated protein with death domain
FANA	2'-fluoro-arabinonucleic acid
FDA	U.S. Food and Drug Administration
fl	full-length
FNA	flexible nucleic acid
FNorm	normalized fluorescence
FT-IR	Fourier-transform infrared
FPLC	fast-protein liquid chromatography

---

---

G	guanine
GNA	glycol nucleic acid
GROMACS	Groningen machine for chemical simulations
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HFIP	Hexafluoroisopropanol
HIV	human immunodeficiency virus
HNA	hexitol nucleic acid
HMBC	heteronuclear Multiple Bond Correlation
HR	high resolution
HSQC	heteronuclear single quantum coherence
Hz	hertz
MHz	megahertz
<i>i.e.</i>	that is (Lat. <i>id est</i> )
iEDDA	inverse electron-demand Diels-Alder
IFN-I	interferon-I
IF- $\gamma$	cytokine interferon- $\gamma$
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl- $\beta$ -D-thiogalactoside
iso <b>C</b>	isocytosine
iso <b>G</b>	isoguanine
<i>J</i>	coupling constant
KD	kinase domain
$K_d$	dissociation constant
Kod	<i>Thermococcus kodakarensis</i>
LB	Luria-Bertani
LC-MS	liquid chromatography mass spectrometry
LNA	locked nucleic acid
m	murine
m <sup>5</sup> <b>C</b>	5-methylcytosine
m <sup>8</sup> <b>G</b>	8-methylguanosine
m <sup>8</sup> d <b>G</b>	8-methyl-2'-deoxyguanosine
m <sup>8</sup> m <b>G</b>	2'-O-methyl-8-methylguanosine
MD	molecular dynamics
MeCN	acetonitrile
MeOH	methanol
MLKL	RHIM-mediated recruitment of mixed lineage kinase domain-like pseudokinase
mRNA	messenger RNA
MST	microscale thermophoresis
NBS	<i>N</i> -bromosuccinimide
<i>n</i> -BuLi	<i>n</i> -butyllithium
ncRNA	non-coding RNA
NF- $\kappa$ B	nuclear factor kappa-light-chain-enhancer of activated B cells
NHS	<i>N</i> -hydroxysuccinimide
<i>N</i> -MeO-BNA	<i>N</i> -MeO-amino bridged nucleic acid
NMR	Nuclear magnetic resonance
nPAGE	native polyacrylamide gel electrophoresis
npNA	phosphoramidate nucleic acid

---



---

nt	nucleotide
OD <sub>600</sub>	optical density at 600 nm
OMe	methoxy
ON	oligonucleotide
o/n	overnight
OTA	ochratoxin A
PCR	polymerase chain reaction
PDB	protein data bank
PDL-1	programmed cell death ligand
PEX	primer extension assays
phNA	(S) <sub>p</sub> -P-alkyl-phosphonate nucleic acid
phT	phNA thymidine
PMG	phosphonomethylglycerol
PNA	peptide nucleic acid
Pol	polymerase
ppm	parts per million
proton sponge	1,8-bis( <i>N,N</i> -dimethylamino)naphthalene
psNA	phosphothioate nucleic acid
pTNA	phosphonomethyl threosyl nucleic acid
quant.	quantitative
RE	restriction enzyme
RHIM	receptor-interacting protein homotypic interaction motif
RIPK	receptor-interacting serine/threonine-protein kinase
RNA	ribonucleic acid
RNAi	RNA interference
ROS	reactive oxygen species
RP-HPLC	reversed-phase high-performance liquid chromatography
rRNA	ribosomal RNA
RT	reverse transcriptase
r.t.	room temperature
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
sat.	saturated
SDS	sodium dodecyl sulfate
SELEX	Systematic Evolution of Ligands by EXponential Enrichment
siRNA	small interfering RNA
SOMAmers	slow off-rate aptamers
SPAAC	strain-promoted azide-alkyne cycloaddition
ssDNA	single stranded DNA
SSII	SuperScript II
ssRNA	single stranded RNA
SVDPE	snake venom phosphodiesterase
T	thymine
Taq	<i>Thermus aquaticus</i>
tATP	α-L-threofuranosyl A nucleoside triphosphate
TB	terrific broth
TBAF	tetrabutylammonium fluoride
TBDPS-Cl	<i>tert</i> -butyldiphenylchlorosilane
TBE	Tris-borate-EDTA
<i>t</i> -BuLi	<i>tert</i> -butyllithium

---

tCTP	$\alpha$ -L-threofuranosyl C nucleoside triphosphate
TEA	triethylammonium
TEAB	triethylammoniumbicarbonate
TEMED	tetramethylenediamine
Tet-BDP	6-methyl-tetrazine-BODIPY
tG <sup>BPTP</sup>	benzophenone-modified $\alpha$ -L-threofuranosyl guanine nucleotide
Tgo	<i>Thermococcus gorgonarius</i>
tGTP	$\alpha$ -L-threofuranosyl guanosine triphosphate
THF	tetrahydrofuran
TLC	thin layer chromatography
TMS	trimethylsilyl
TMSOTf	trimethylsilyl trifluoromethanesulfonate
TNA	threose nucleic acid
tNaMTP	$\alpha$ -L-threofuranosyl <b>NaM</b> nucleoside triphosphate
TNF $\alpha$	tumor Nekrosis Factor alpha
tNTP	$\alpha$ -L-threofuranosyl nucleoside triphosphates
TP	triphosphate
TRIS	Tris-(hydroxymethyl)-aminomethane
tTPT3TP	$\alpha$ -L-threofuranosyl <b>TPT3</b> nucleoside triphosphate
tTTP	$\alpha$ -L-threofuranosyl thymidine triphosphate
UB	unnatural base
UBP	unnatural base pair
UV	ultraviolet
VEGF165	vascular endothelial growth factor isoform
v/v	volume per volume
VWF	von Willebrand Factor
WT	wild-type
w/v	weight per volume
x <sup>5</sup> C	C-5 halogenated cytosine
x <sup>8</sup> G	C-8 halogenated guanine
X-Gal	5-Bromo-4-chloro-3-indoyl-beta-D-galactopyranoside
XNA	xeno nucleic acid
Z $\alpha$	Z-DNA-binding domains
ZBP1	Z-DNA-binding protein 1
Z-NA	Z-nucleic acid
$\Psi$	pseudouridine

## 7.7 List of tables

Table 1: Comparison of the helical parameters of the different naturally occurring nucleic acid conformations.....	37
Table 2: Summary of the LC-MS results obtained from primer extension assays using DNA_pTemp_mid2ATGC <sup>NaM</sup> .....	82
Table 3: Summary of the LC-MS results obtained from primer extension assays using DNA_pTemp_midATG <sup>NaM</sup> .....	85

Table 4: Summary of the LC-MS results obtained from primer extension assays using DNA_pTemp_mid2ATGC <sup>TPT3</sup> .....	87
Table 5: Software applications used in this study. ....	133
Table 6: Databases used in this study. ....	134
Table 7: Equipment used in this study. ....	182
Table 8: Consumables used in this study. ....	184
Table 9: Reagents used in this study.....	185
Table 10: Buffers used in this study.....	187
Table 11: Enzymes used in this study .....	189
Table 12: Bacterial cells used in this study. ....	189
Table 13: Vectors used in this study. ....	190
Table 14: Plasmids and gene strings used in this study.....	190
Table 15: Kits used in this study. ....	190
Table 16: Composition of denaturing polyacrylamide gels (5 mL total volume) with different polyacrylamide concentrations. ....	191
Table 17: Composition of 8 % native polyacrylamide gels (5 mL total volume). ....	192
Table 18: Composition of denaturing SDS polyacrylamide separating gels (5 mL total volume) with different polyacrylamide concentrations. ....	192
Table 19: Overview of working concentrations for different antibiotics for bacterial cultivation used in this study.....	196
Table 20: General pipetting scheme for colony PCRs.....	200
Table 21: General cycling conditions for colony PCR employed in this study. ....	200
Table 22: General pipetting scheme for the amplification of the Kod-RI gene with Phusion Flash High Fidelity PCR Master Mix. ....	203
Table 23: PCR cycling conditions using Phusion Flash High Fidelity PCR Master Mix for the amplification of the Kod-RI gene with 25 bp overhangs for AQUA cloning and 15 bp overhangs for In-Fusion cloning into the pQE80HT vector. ....	203
Table 24: PCR cycling conditions using Phusion Flash High Fidelity PCR Master Mix for the amplification of the Kod-RI gene and simultaneous introduction of overhangs compatible with either restriction enzyme cloning into the pQE80HT vector or Stargate cloning, AQUA cloning and In-Fusion cloning into the pASG-IBA 35 vector. ....	204
Table 25: General pipetting scheme for the linearization of pQE80HT and pASG-IBA35 vectors with Phusion Flash High Fidelity PCR Master Mix. ....	205
Table 26: PCR cycling conditions for high fidelity PCR linearization of pQE80HT and pASG-IBA35 vectors using Phusion Flash High Fidelity PCR Master Mix. ....	206
Table 27: Calculated mass of DNA insert for the corresponding vector:insert ratios used in AQUA cloning with the pQE80HT vector. ....	207

Table 28: Calculated mass of DNA insert for the corresponding vector:insert ratios used in AQUA cloning with the pASG-IBA35 vector.....	207
Table 29: Calculated mass of DNA insert for the corresponding vector:insert ratios used in In-Fusion cloning with the pQE80HT vector.....	208
Table 30: Calculated mass of DNA insert for the corresponding vector:insert ratios used in In-Fusion cloning with the pASG-IBA35 vector. ....	208
Table 31: General pipetting scheme for the digestion and ligation reaction during Stargate Cloning.....	209
Table 32: Detailed pipetting scheme for the restriction enzyme digest of the pQE80HT vector with SacI-HF and HindIII-HF restriction enzymes. ....	210
Table 33: Detailed pipetting scheme for the restriction enzyme digest of the Kod-RI gene with SacI-HF and HindIII-HF restriction enzymes.....	210
Table 34: General pipetting scheme for the ligation reaction during restriction enzyme cloning. ....	211
Table 35: General pipetting scheme for the linearization and site-directed mutagenesis of pQE80HT_Kod-RI_2313delA with Phusion Flash High Fidelity PCR Master Mix.....	212
Table 36: PCR cycling conditions for the PCR linearization and site-directed mutagenesis of pQE80HT_Kod-RI_2313delA with Phusion Flash High Fidelity PCR Master Mix.....	212
Table 37: General pipetting scheme for the ligation reaction of the linearized and 5'-phosphorylated pQE80HT_Kod-RI_2313delA. ....	213
Table 38: General pipetting scheme for the linearization and targeted mutation of pQE80HT_Kod-RI with Phusion Flash High Fidelity PCR Master Mix.....	214
Table 39: PCR cycling conditions for the PCR-based linearization and targeted mutation of pQE80HT_Kod-RI using Phusion Flash High Fidelity PCR Master Mix. ....	215
Table 40: General pipetting scheme for the ligation reaction of linearized and 5'-phosphorylated plasmid.....	216
Table 41: Annealing temperatures for the primer sets 2-4 for PCR-based linearization and targeted mutation of the plasmid from the previous round using Phusion Flash High Fidelity PCR Master Mix. ....	216
Table 42: General pipetting scheme for the TNA polymerase activity assay.....	219
Table 43: Temperature profile for TNA polymerase activity assays. ....	220
Table 44: Pipetting scheme for a small-scale TNA primer extension assay using Terminator polymerase with tTPT3TP at 50 °C and 70 °C.....	220
Table 45: General temperature profile for a small-scale TNA primer extension assay using Terminator polymerase with tTPT3TP at 50 °C and 70 °C.....	221
Table 46: Pipetting scheme for small-scale TNA primer extension assays using Terminator polymerase with tTPT3TP and different additives.....	222

Table 47: Combination of additives used in small-scale primer extension assays.....	222
Table 48: General temperature profile for a small-scale TNA primer extension assay using Therminator polymerase with tTPT3TP and different combinations of additives. ....	223
Table 49: Pipetting scheme for small-scale TNA primer extension assays using Therminator polymerase with varying tTPT3TP concentrations. ....	223
Table 50: General temperature profile used for a small-scale TNA primer extension assay using Therminator polymerase with varying tTPT3TP concentrations.....	224
Table 51: Pipetting scheme for a large-scale TNA primer extension assay with tTPT3TP using Therminator polymerase.....	225
Table 52: Pipetting scheme for a large-scale TNA primer extension assay with tNaMTP using Therminator polymerase.....	225
Table 53: Pipetting scheme for a large-scale TNA primer extension assay with tTPT3TP using Kod-RI polymerase.....	226
Table 54: Pipetting scheme for a large-scale TNA primer extension assay with tNaMTP using Kod-RI polymerase.....	226
Table 55: Pipetting scheme for a large-scale TNA primer extension assay with tTPT3TP using Kod-RSGA polymerase. ....	227
Table 56: Pipetting scheme for a large-scale TNA primer extension assay with tNaMTP using Kod-RSGA polymerase. ....	227
Table 57: General temperature profile for large scale TNA primer extension assay.....	228
Table 58: Residue topology entries TTP3, TNAM, TG2, TA2, TC2, TT2 and TTP from the modified “merged.rtp” of the Gromacs – Charmm 36 force field. ....	229
Table 59: Additional force field parameters used for threose-nucleic acid building blocks. .	242
Table 60: General pipetting scheme for DNA primer extension assays with the cyclopropene- modified dA <sup>CPTP</sup> .....	247
Table 61: General pipetting scheme for the positive control of DNA primer extension assays using only canonical nucleotides. ....	247
Table 62: Temperature profile for DNA primer extension assays with dA <sup>CPTP</sup> .....	248
Table 63: FPLC gradient of the IMAC purification of the mZBP1 variants.....	250
Table 64: General pipetting scheme for the NHS-ester click reaction of 5'-amino modified RNA with AF488 NHS-ester.....	251
Table 65: General pipetting scheme for EMSA with Z-RNA <sup>8mG</sup> and mZBP1 <sup>WT</sup> .....	252
Table 66: General pipetting scheme for EMSA with A-RNA <sup>WT</sup> and mZBP1 <sup>WT</sup> .....	253
Table 67: mZBP1 <sup>WT</sup> ligand concentrations used for MST measurements with Z-RNA <sup>m8G</sup> . ...	254
Table S 1: Protein concentration of the Kod-RI polymerase stock. ....	290
Table S 2: Protein concentration of the Kod-RSGA polymerase stock.....	295

Table S 3: Area under the curve (AUC) analysis of the crude reaction mixture of the <i>N</i> -benzophenone-functionalization of threoguanosine 43 after 24 h reaction time. ....	304
Table S 4: Area under the curve (AUC) analysis of the crude reaction mixture of the <i>N</i> -benzophenone-functionalization of threoguanosine 43 after 48 h reaction time. ....	305
Table S 5: Protein concentrations of the distinct mZBP1 variant stocks. ....	310
Table S 6: Compounds synthesized in this study. ....	342

## 7.8 List of figures

Figure 1: Schematic representation of the unidirectional flow of biological information according to the central dogma of molecular biology. <sup>1</sup> .....	1
Figure 2: Possible RNA progenitors within a pre-RNA world scenario. ....	2
Figure 3: Possible modification sites in the tripartite structure of nucleic acids. ....	3
Figure 4: Examples of repeating units of phosphate backbone-modified artificial nucleic acids. ....	4
Figure 5: Examples of repeating units of artificial nucleic acids with minor modifications at the 2'-position of the sugar moiety. ....	6
Figure 6: Examples of repeating units of xeno nucleic acids compared to RNA. ....	7
Figure 7: Schematic comparison between the canonical nucleic acids (RNA/DNA) and TNA. ....	10
Figure 8: C:G Watson-Crick base pair and G/G Hoogsteen base pair. ....	12
Figure 9: Attractive sites for modification of the canonical pyrimidine and purine nucleobases. ....	15
Figure 10: Overview of the first generation UBPs to expand the genetic alphabet. ....	16
Figure 11: Overview of the second generation UBPs to expand the genetic alphabet. ....	18
Figure 12: Examples of repeating units of artificial nucleic acids with a dual modification pattern. ....	21
Figure 13: TNA with an increased chemical density developed by the Chaput research group. ....	23
Figure 14: Click reactions used in bioorthogonal applications. ....	25
Figure 15: Commonly used dienes and dienophiles in inverse electron-demand Diels-Alder reactions in order of reactivity. ....	26
Figure 16: Reaction cascade of the bioorthogonal inverse electron-demand Diels-Alder reaction between strained alkenes and 1,2,4,5-tetrazines. ....	26
Figure 17: TPT3 nucleoside triphosphates with reactive handles used in the Kath-Schorr research group for downstream nucleic acid functionalization via the iEDDA click reaction. .	27

Figure 18: Schematic representation of Systematic Evolution of Ligands by Exponential Enrichment (SELEX). .....	30
Figure 19: Schematic representation of one click-SELEX cycle. ....	33
Figure 20: Different conformations adopted by nucleic acids. ....	36
Figure 21: Nucleoside conformations and sugar pucker found in Z-nucleic acids. ....	37
Figure 22: Nucleobase modifications at the C <sup>5</sup> position of cytosine and the C <sup>8</sup> position of guanine stabilize the Z-conformation of nucleic acids. ....	39
Figure 23: Three dimensional structure of the Z-DNA binding protein 1 from <i>mus musculus</i> . ....	41
Figure 24: ZBP1-mediated cell death pathways upon Z-nucleic acid binding: necroptosis and apoptosis. ....	41
Figure 25: Required TNA nucleotides for the enzymatic preparation of TNA with an expanded genetic alphabet. ....	51
Figure 26: Schematic representation of small-scale TNA primer extension assays to prepare tTPT3-modified exTNA with Therminator polymerase. ....	61
Figure 27: Schematic representation of the optimized large-scale primer extension assay for the preparation of tTPT3-modified exTNA. ....	62
Figure 28: dPAGE results of Therminator polymerase-mediated synthesis towards tTPT3-modified exTNA using dNaM-modified DNA templates containing either only dA, dT, and dG or dA, dT, and dC. ....	63
Figure 29: MS results of Therminator polymerase-mediated synthesis towards tTPT3-modified exTNA using dNaM-modified DNA templates containing either only dA, dT, and dG or dA, dT, and dC. ....	64
Figure 30: Schematic representation of large-scale primer extension assays for the preparation of tNaM-modified exTNA. ....	65
Figure 31: dPAGE results of Therminator polymerase-mediated synthesis towards tNaM-modified exTNA using dTPT3-modified DNA templates containing either only dA, dT, and dG or dA, dT, and dC. ....	66
Figure 32: MS results of Therminator polymerase-mediated synthesis towards tNaM-modified exTNA using dTPT3-modified DNA templates containing either only dA, dT, and dG or dA, dT, and dC. ....	67
Figure 33: Overview of the employed cloning methods for the preparation of the expression plasmid harboring the Kod-RI gene. ....	69
Figure 34: Nickel-based IMAC purification of Kod-RI polymerase. ....	71
Figure 35: dPAGE results of Kod-RI polymerase-mediated synthesis towards tTPT3-modified exTNA using dNaM-modified DNA templates containing either only dA, dT, and dG or dA, dT, and dC. ....	72

Figure 36: MS results of Kod-RI polymerase-mediated synthesis towards tTPT3-modified exTNA using dNaM-modified DNA templates containing either only dA, dT, and dG or dA, dT, and dC. ....	73
Figure 37: dPAGE results of Kod-RI polymerase-mediated synthesis towards tNaM-modified exTNA using dTPT3-modified DNA templates containing either only dA, dT, and dG or dA, dT, and dC. ....	74
Figure 38: MS results of Kod-RI polymerase-mediated synthesis towards tNaM-modified exTNA using dTPT3-modified DNA templates containing either only dA, dT, and dG or dA, dT, and dC. ....	75
Figure 39: Nickel-based IMAC purification of Kod-RSGA polymerase. ....	78
Figure 40: dPAGE results of Kod-RSGA polymerase-mediated synthesis of tTPT3-modified exTNA using a dNaM-modified DNA template. ....	79
Figure 41: Comparison of mass spectra obtained from Kod-RSGA polymerase-mediated synthesis of tTPT3-modified exTNA using DNA_pTemp_midATGC <sup>NaM</sup> with and without MnCl <sub>2</sub> supplementation. ....	80
Figure 42: dPAGE results of Kod-RSGA polymerase-mediated synthesis of tTPT3-modified exTNA using a dNaM-modified DNA template containing only dA, dT, and dG. ....	83
Figure 43: Comparison of mass spectra obtained from Kod-RSGA polymerase-mediated synthesis of tTPT3-modified exTNA using DNA_pTemp_midATG <sup>NaM</sup> with and without MnCl <sub>2</sub> supplementation. ....	84
Figure 44: Kod-RSGA polymerase-mediated synthesis of tNaM-modified exTNA using a dTPT3-modified DNA template. ....	86
Figure 45: Representative snapshot (cluster analysis) of a 200 ns MD simulation of the Kod-RSGA polymerase in complex with a DNA template containing either dNaM or dTPT3 and a nascent TNA oligonucleotide containing the respective counterpart tTPT3 or tNaM. ....	88
Figure 46: Average distance of the 2'-OH nucleophile of either tTPT3 or tNaM in a nascent TNA to the 3' α-phosphate of the incoming natural TNA triphosphate. ....	89
Figure 47: Proposed strategy for a traceless approach to suppress G/G Hoogsteen base pair formation during enzymatic TNA synthesis. ....	91
Figure 48: Schematic representation of the envisioned iEDDA-based click-SELEX process. ....	103
Figure 49: Analysis of the crude reaction mixture of dU <sup>CPTP</sup> (58) obtained after triphosphate synthesis. ....	107
Figure 50: Analysis of the crude reaction mixture of dA <sup>CPTP</sup> (59) after triphosphate synthesis. ....	108
Figure 51: Schematic representation of preliminary experiments to test the polymerase processing of dA <sup>CPTP</sup> and its downstream functionalization via iEDDA. ....	112



Figure 52: 20 % dPAGE analysis of primer extension assays with dA <sup>CP</sup> TP using Taq polymerase and further downstream functionalization of DNA <sup>CP</sup> with Tet-BDP via iEDDA. .114	
Figure 53: Structure of dA <sup>CP</sup> TP (59) with the assigned <sup>1</sup> H NMR chemical shifts of the protons from the cyclopropene moiety.....116	
Figure 54: Stability evaluation of dA <sup>CP</sup> TP (59) by LC-MS analysis. ....117	
Figure 55: Comparison of carbamate- and amide-substituted cyclopropenes: kinetics ( $k_2$ ) of the iEDDA reaction with differently substituted CPs and their stability. <sup>364,403</sup> .....118	
Figure 56: mZBP1 <sup>WT</sup> and the three mutants mZBP1 <sup>mZα1</sup> , mZBP1 <sup>mZα2</sup> , and mZBP1 <sup>mZα1-2</sup> . ....122	
Figure 57: Cobalt-based IMAC purification of mZBP1 variants. ....122	
Figure 58: Fluorescently-labeled GC-rich RNA hairpin containing m <sup>8</sup> G modifications to stabilize the Z-conformation. ....124	
Figure 59: 8 % nPAGE of comparative interaction studies of mZBP1 <sup>WT</sup> with Z-RNA <sup>m8G</sup> and A-RNA <sup>WT</sup> . ....124	
Figure 60: Microscale thermophoresis binding curves for AF-488 labeled Z-RNA <sup>m8G</sup> with mZBP1 <sup>WT</sup> .....126	
Figure 61: Overview of the four research projects on artificial nucleic acids with enhanced functionalities that were conducted within the framework of this thesis.....127	
Figure 62: Structure of the L-threose nucleic acid residues named “TTP3, TNAM, TG2, TA2, TC2, TT2 and TTP” in the modified Gromacs <sup>378</sup> – Charmm36 <sup>379</sup> force field (see also Table 58). .....229	
Figure S 1: dPAGE results of Therminator and Kod-RI polymerase-mediated synthesis towards tTPT3-modified exTNA at different temperatures.....283	
Figure S 2: dPAGE results of Therminator polymerase-mediated synthesis towards tTPT3-modified exTNA using different additives.....283	
Figure S 3: dPAGE results of Therminator polymerase-mediated synthesis towards TPT3-modified exTNA using different tTPT3TP concentrations ( <i>i.e.</i> 0.25 mM, 0.5 mM, 0.75 mM, 1.0 mM). ....284	
Figure S 4: Complete gel of the 20 % dPAGE of PEx with Therminator and Kod-RI polymerases using DNA_pTemp_midATG <sup>NaM</sup> and DNA_pTemp_midATC <sup>NaM</sup> . ....284	
Figure S 5: UV chromatograms and deconvoluted ESI <sup>-</sup> mass spectra of purified primer extension assays using Therminator polymerase employing DNA template DNA_pTemp_midATG <sup>NaM</sup> and tTPT3TP. ....285	
Figure S 6: UV chromatograms and deconvoluted ESI <sup>-</sup> mass spectra of purified primer extension assays using Therminator polymerase employing DNA template DNA_pTemp_midATC <sup>NaM</sup> and tTPT3TP. ....286	

Figure S 7: UV chromatograms and deconvoluted ESI <sup>-</sup> mass spectra of purified primer extension assays using Therminator polymerase employing DNA template DNA_pTemp_midATG <sup>TPT3</sup> and tNaMTP.....	287
Figure S 8: UV chromatograms and deconvoluted ESI <sup>-</sup> mass spectra of purified primer extension assays using Therminator polymerase employing DNA template DNA_pTemp_midATC <sup>TPT3</sup> and tNaMTP.....	288
Figure S 9: Exemplary SDS-PAGE (10 %) of the pooled fractions from one affinity chromatography purification run of His-tagged Kod-RI polymerase. ....	289
Figure S 10: Denaturing 15 % PAGE analysis of the Kod-RI polymerase activity assay with 3 hours incubation at 55 °C. ....	289
Figure S 11: Bovine Serum albumin standard curve prepared with 8 different concentrations using the Pierce 660 nm Protein assay kit. ....	290
Figure S 12: UV chromatograms and deconvoluted ESI <sup>-</sup> mass spectra of purified primer extension assays using Kod-RI polymerase employing DNA template DNA_pTemp_midATG <sup>NaM</sup> and tTPT3TP. ....	290
Figure S 13: UV chromatograms and deconvoluted ESI <sup>-</sup> mass spectra of purified primer extension assays using Kod-RI polymerase employing DNA template DNA_pTemp_midATC <sup>NaM</sup> and tTPT3TP. ....	291
Figure S 14: UV chromatograms and deconvoluted ESI <sup>-</sup> mass spectra of purified primer extension assays using Kod-RI polymerase employing DNA template DNA_pTemp_midATG <sup>TPT3</sup> and tNaMTP.....	292
Figure S 15: UV chromatograms and deconvoluted ESI <sup>-</sup> mass spectra of purified primer extension assays using Kod-RI polymerase employing DNA template DNA_pTemp_midATC <sup>TPT3</sup> and tNaMTP.....	293
Figure S 16: Exemplary SDS-PAGE (10 %) of the pooled fractions from one affinity chromatography purification run of His-tagged Kod-RSGA polymerase.....	294
Figure S 17: Denaturing 15 % PAGE analysis of the Kod-RSGA polymerase activity assay with 3 hours incubation at 55 °C. ....	294
Figure S 18: UV chromatograms and deconvoluted ESI <sup>-</sup> mass spectra of purified primer extension assays using Kod-RSGA polymerase employing DNA template DNA_pTemp_mid2ATGC <sup>NaM</sup> and tTPT3TP. ....	296
Figure S 19: UV chromatograms and deconvoluted ESI <sup>-</sup> mass spectra of purified primer extension assays using Kod-RSGA polymerase employing DNA template DNA_pTemp_mid2ATGC and canonical tNTPs.....	297
Figure S 20: UV chromatograms and deconvoluted ESI <sup>-</sup> mass spectra of purified primer extension assays using Kod-RSGA polymerase employing DNA template DNA_pTemp_midATG <sup>NaM</sup> and tTPT3TP.....	298

Figure S 21: UV chromatograms and deconvoluted ESI <sup>+</sup> mass spectra of purified primer extension assays using Kod-RSGA polymerase employing DNA template DNA_pTemp_midATG and canonical tNTPs. ....	299
Figure S 22: UV chromatograms and deconvoluted ESI <sup>+</sup> mass spectra of purified primer extension assays using Kod-RSGA polymerase employing DNA template DNA_pTemp_mid2ATGC <sup>TPT3</sup> and tNaMTP. ....	300
Figure S 23: UV chromatograms and deconvoluted ESI <sup>+</sup> mass spectra of purified primer extension assays using Kod-RSGA polymerase employing DNA template DNA_pTemp_mid2ATGC and canonical tNTPs. ....	301
Figure S 24: Representative snapshot (cluster analysis) of 200 ns MD simulation of Kod-RSGA polymerase in complex with a DNA template containing dNaM and a nascent TNA oligonucleotide containing tTPT3. ....	301
Figure S 25: LC-MS analysis of the <i>N</i> <sup>7</sup> -functionalization of <i>N</i> <sup>2,9</sup> -diacetylated guanine 31 using 5.0 equivalents of BP-Br (40) as a simplified test system. ....	302
Figure S 26: LC-MS analysis of the <i>N</i> <sup>7</sup> -functionalization of threoguanosine 33 with different amounts of equivalents of BP-Br(40). ....	303
Figure S 27: RP-HPLC analysis of the crude reaction mixture of the <i>N</i> <sup>7</sup> -functionalization of threoguanosine 43. ....	304
Figure S 28: UV traces ( $\lambda = 254$ nm) of the crude products obtained after triphosphate synthesis of dU <sup>CP</sup> TP (58, left) and dA <sup>CP</sup> TP (59, right). ....	305
Figure S 29: Full image of the 20 % dPAGE analysis of primer extension assays with dA <sup>CP</sup> TP using Taq polymerase and further downstream functionalization of DNA <sup>CP</sup> with Tet-BDP...	306
Figure S 30: ESI <sup>+</sup> mass spectra of DNA <sup>WT</sup> and DNA <sup>CP</sup> . ....	306
Figure S 31: 20 % dPAGE analysis of primer extension assays with dA <sup>CP</sup> TP using OneTaq polymerase and further downstream functionalization of DNA <sup>CP</sup> with Tet-BDP. ....	307
Figure S 32: Evaluation of the stability of A <sup>CP</sup> TP (59) by NMR. ....	308
Figure S 33: SDS-PAGE (15 %) of the expression and IMAC purification of His-tagged mZBP1 <sup>WT</sup> . ....	308
Figure S 34: SDS-PAGE (15 %) of the expression and IMAC purification of His-tagged mZBP1 <sup>mZa1</sup> . ....	309
Figure S 35: SDS-PAGE (15 %) of the expression and IMAC purification of His-tagged mZBP1 <sup>mZa2</sup> . ....	309
Figure S 36: SDS-PAGE (15 %) of the expression and IMAC purification of His-tagged mZBP1 <sup>mZa1-2</sup> . ....	310
Figure S 37: Complete 15 % SDS-PAGE of the different stocks of expressed and purified mZBP1 variants. Ladder: ....	310
Figure S 38: 8 % nPAGE of titrating amounts of AF88-labeled Z-RNA <sup>m8G</sup> and A-RNA <sup>WT</sup> . ....	311

Figure S 39: Complete 8 % nPAGE of comparative interaction studies of mZBP1 <sup>WT</sup> with Z-RNA <sup>m8G</sup> and A-RNA <sup>WT</sup> .....	311
Figure S 40: Microscale thermophoresis pretests .....	312
Figure S 41: <sup>1</sup> H-NMR spectrum of compound 14 (CD <sub>3</sub> OD, 400 MHz, r.t.) .....	313
Figure S 42: <sup>13</sup> C-NMR spectrum of compound 14 (CD <sub>3</sub> OD, 126 MHz, r.t.) .....	314
Figure S 43: <sup>1</sup> H-NMR spectrum of compound 12 (CDCl <sub>3</sub> , 400 MHz, r.t.) .....	314
Figure S 44: <sup>13</sup> C-NMR spectrum of compound 12 (CDCl <sub>3</sub> , 126 MHz, r.t.) .....	315
Figure S 45: <sup>1</sup> H-NMR spectrum of compound 10 (CDCl <sub>3</sub> , 300 MHz, r.t.) .....	315
Figure S 46: <sup>1</sup> H-NMR spectrum of compound 16 (CDCl <sub>3</sub> , 600 MHz, r.t.) .....	316
Figure S 47: <sup>13</sup> C-NMR spectrum of compound 16 (CDCl <sub>3</sub> , 151 MHz, r.t.) .....	316
Figure S 48: <sup>1</sup> H-NMR spectrum of compound 6 (D <sub>2</sub> O, 500 MHz, r.t.) .....	317
Figure S 49: <sup>31</sup> P-NMR spectrum of compound 6 (D <sub>2</sub> O, 202 MHz, r.t.) .....	317
Figure S 50: <sup>1</sup> H-NMR spectrum of compound 18 (CDCl <sub>3</sub> , 400 MHz, r.t.) .....	318
Figure S 51: <sup>13</sup> C-NMR spectrum of compound 18 (CDCl <sub>3</sub> , 126 MHz, r.t.) .....	318
Figure S 52: <sup>1</sup> H-NMR spectrum of compound 24 (DMSO- <i>d</i> <sub>6</sub> , 500 MHz, r.t.) .....	319
Figure S 53: <sup>13</sup> C-NMR spectrum of compound 24 (DMSO- <i>d</i> <sub>6</sub> , 126 MHz, r.t.) .....	319
Figure S 54: <sup>1</sup> H-NMR spectrum of compound 25 (D <sub>2</sub> O, 400 MHz, r.t.) .....	320
Figure S 55: <sup>31</sup> P-NMR spectrum of compound 25 (D <sub>2</sub> O, 162 MHz, r.t.) .....	320
Figure S 56: <sup>1</sup> H-NMR Spectrum of compound 7 (D <sub>2</sub> O, 500 MHz, r.t.) .....	321
Figure S 57: <sup>31</sup> P-NMR Spectrum of compound 7 (D <sub>2</sub> O, 202 MHz, r.t.) .....	321
Figure S 58: <sup>1</sup> H-NMR spectrum of compound 28 (DMSO- <i>d</i> <sub>6</sub> , 400 MHz, r.t.) .....	322
Figure S 59: <sup>13</sup> C-NMR spectrum of compound 28 (DMSO- <i>d</i> <sub>6</sub> , 126 MHz, r.t.) .....	322
Figure S 60: <sup>1</sup> H-NMR spectrum of compound 29 (D <sub>2</sub> O, 400 Hz, r.t.) .....	323
Figure S 61: <sup>31</sup> P-NMR spectrum of compound 29 (D <sub>2</sub> O, 126 Hz, r.t.) .....	323
Figure S 62: <sup>1</sup> H-NMR spectrum of compound 8 (D <sub>2</sub> O, 500 MHz, r.t.) .....	324
Figure S 63: <sup>31</sup> P-NMR spectrum of compound 8 (D <sub>2</sub> O, 202 MHz, r.t.) .....	324
Figure S 64: <sup>1</sup> H-NMR spectrum of compound 31 (DMSO- <i>d</i> <sub>6</sub> , 499 MHz, r.t.) .....	325
Figure S 65: <sup>13</sup> C-NMR spectrum of compound 31 (DMSO- <i>d</i> <sub>6</sub> , 75 MHz, r.t.) .....	325
Figure S 66: <sup>1</sup> H-NMR spectrum of compound 32 (DMSO- <i>d</i> <sub>6</sub> , 500 MHz, r.t.) .....	326
Figure S 67: <sup>1</sup> H-NMR spectrum of compound 33 (DMF- <i>d</i> <sub>7</sub> , 500 MHz, r.t.) .....	326
Figure S 68: <sup>13</sup> C-NMR spectrum of compound 33 (DMF- <i>d</i> <sub>7</sub> , 126 MHz, r.t.) .....	327
Figure S 69: <sup>1</sup> H-NMR spectrum of compound 34 (CD <sub>3</sub> OD, 500 MHz, r.t.) .....	327
Figure S 70: <sup>13</sup> C-NMR spectrum of compound 34 (CD <sub>3</sub> OD, 126 MHz, r.t.) .....	328
Figure S 71: <sup>1</sup> H-NMR spectrum of compound 35 (D <sub>2</sub> O, 400 MHz, r.t.) .....	328
Figure S 72: <sup>31</sup> P-NMR spectrum of compound 35 (D <sub>2</sub> O, 162 MHz, r.t.) .....	329
Figure S 73: <sup>1</sup> H-NMR spectrum of compound 9 (D <sub>2</sub> O, 400 MHz, r.t.) .....	329
Figure S 74: <sup>31</sup> P-NMR spectrum of compound 9 (D <sub>2</sub> O, 162 MHz, r.t.) .....	330

Figure S 75: $^1\text{H}$ -NMR spectrum of compound 38 ( $\text{CDCl}_3$ , 600 MHz, r.t.).....	330
Figure S 76: $^{13}\text{C}$ -NMR spectrum of compound 38 ( $\text{CDCl}_3$ , 151 MHz, r.t.).....	331
Figure S 77: $^1\text{H}$ -NMR spectrum of compound 36 ( $\text{CDCl}_3$ , 400 MHz, r.t.).....	331
Figure S 78: $^{31}\text{P}$ -NMR spectrum of compound 36 ( $\text{CDCl}_3$ , 162 MHz, r.t.).....	332
Figure S 79: $^1\text{H}$ -NMR spectrum of compound 44 ( $\text{CDCl}_3$ , 300 MHz, r.t.).....	332
Figure S 80: $^{13}\text{C}$ -NMR spectrum of compound 44 ( $\text{CDCl}_3$ , 75 MHz, r.t.).....	333
Figure S 81: $^1\text{H}$ -NMR spectrum of compound 40 ( $\text{CDCl}_3$ , 300 MHz, r.t.).....	333
Figure S 82: $^{13}\text{C}$ -NMR spectrum of compound 40 ( $\text{CDCl}_3$ , 75 MHz, r.t.).....	334
Figure S 83: $^1\text{H}$ -NMR spectrum of compound 47 ( $\text{CDCl}_3$ , 300 MHz, r.t.).....	334
Figure S 84: $^{13}\text{C}$ -NMR spectrum of compound 47 ( $\text{CDCl}_3$ , 75 MHz, r.t.).....	335
Figure S 85: $^1\text{H}$ -NMR spectrum of compound 43 ( $\text{DMSO}-d_6$ , 600 MHz, r.t.).....	335
Figure S 86: $^1\text{H}$ -NMR spectrum of compound 65 ( $\text{CD}_3\text{OD}$ , 500 MHz, r.t.).....	336
Figure S 87: $^{13}\text{C}$ -NMR spectrum of compound 65 ( $\text{CD}_3\text{OD}$ , 126 MHz, r.t.).....	336
Figure S 88: $^1\text{H}$ -NMR spectrum of compound 66 ( $\text{CD}_3\text{OD}$ , 500 MHz, r.t.).....	337
Figure S 89: $^{13}\text{C}$ -NMR spectrum of compound 66 ( $\text{CD}_3\text{OD}$ , 126 MHz, r.t.).....	337
Figure S 90: $^1\text{H}$ -NMR spectrum of compound 60 ( $\text{CD}_3\text{OD}$ , 600 MHz, r.t.).....	338
Figure S 91: $^{13}\text{C}$ -NMR spectrum of compound 60 ( $\text{CD}_3\text{OD}$ , 151 MHz, r.t.).....	338
Figure S 92: $^1\text{H}$ -NMR spectrum of compound 61 ( $\text{CD}_3\text{OD}$ , 600 MHz, r.t.).....	339
Figure S 93: $^{13}\text{C}$ -NMR spectrum of compound 61 ( $\text{CD}_3\text{OD}$ , 151 MHz, r.t.).....	339
Figure S 94: $^1\text{H}$ -NMR spectrum of compound 59 ( $\text{D}_2\text{O}$ , 500 MHz, r.t.).....	340
Figure S 95: $^{31}\text{P}$ -NMR spectrum of compound 59 ( $\text{D}_2\text{O}$ , 243 MHz, r.t.).....	340
Figure S 96: Vector map of pQE80HT.....	355
Figure S 97: Vector map of pASG-IBA35.....	357
Figure S 98: Plasmid map of pQE80HT_Kod-RI.....	363
Figure S 99: Plasmid map of pQE80HT_Kod-RSGA.....	367
Figure S 100: Plasmid map of pQE80HT_Kod-RI_2313delA.....	371
Figure S 101: Plasmid map of pMA-RQ_Kod-RI.....	374
Figure S 102: Plasmid map of pET28a(+)_mZBP1 <sup>WT</sup> .....	377
Figure S 103: Plasmid map of pET28a(+)_mZBP1 <sup>mZ<math>\alpha</math>1</sup> .....	381
Figure S 104: Plasmid map of pET28a(+)_mZBP1 <sup>mZ<math>\alpha</math>2</sup> .....	384
Figure S 105: Plasmid map of pET28a(+)_mZBP1 <sup>mZ<math>\alpha</math>1-2</sup> .....	388

## 7.9 List of schemes

Scheme 1: Retrosynthetic considerations providing access to tNaMTP (6).....	51
Scheme 2: Synthesis of glycosyl donor 12.....	52
Scheme 3: Synthesis of tNaMTP (6).....	52

Scheme 4: Two-step one-pot synthesis of the universal glycosyl donor 18 starting from glycosyl donor 12.....	54
Scheme 5: Synthesis of tCTP (7).....	55
Scheme 6: Neighboring group participation causes high selectivity for the $\alpha$ -stereoisomer during the Vorbrüggen glycosylation.....	55
Scheme 7: Synthesis of tATP (8).....	56
Scheme 8: Synthesis of tGTP (9). ....	57
Scheme 9: Synthesis of dTPT3-CEP (36).....	59
Scheme 10: Proposed synthetic strategy toward the benzophenone-photocaged threoguanosine triphosphate 39. ....	92
Scheme 11: Two-step synthesis of threoguanosine 34 and the formed byproduct 43.....	93
Scheme 12: Vorbrüggen glycosylation of protected guanine 32 with universal glycosyl donor 18.....	94
Scheme 13: Two-step synthesis of 4-(bromomethyl)benzophenone (BP-Br, 40). ....	95
Scheme 14: First attempt at the $N^7$ -functionalization of threoguanosine 34 with one equivalent of BP-Br (40). ....	95
Scheme 15: Synthesis of 4-(iodomethyl)benzophenone (BP-I, 47) via the Finkelstein reaction. ....	96
Scheme 16: Second attempt at the $N^7$ -functionalization of threoguanosine 34 with one equivalent of BP-I (47).....	96
Scheme 17: Preliminary study on the $N^7$ -functionalization of $N^{2,9}$ -diacetyl-protected guanine 31 as a simplified test system with 1.0 and 5.0 equivalents of BP-Br (40). ....	97
Scheme 18: Preliminary study on the $N^7$ -functionalization of threoguanosine 33 with different amounts of equivalents of BP-Br (40). ....	98
Scheme 19: Products obtained by the benzophenone-functionalization of threoguanosine 43 without the $O^6$ diphenylcarbamoyl protecting group at the guanine.....	99
Scheme 20: Proposed synthetic pathway to synthesize tG <sup>BP</sup> TP (39) with the $N^7$ -functionalization performed on threoguanosine 56 without the diphenylcarbamoyl protecting group at the $O^6$ oxygen of guanine. ....	100
Scheme 21: Retrosynthetic considerations providing access to the target compounds dU <sup>CP</sup> TP (58) and dA <sup>CP</sup> TP (59).....	104
Scheme 22: Synthetic route towards cyclopropene-modified nucleoside triphosphates dU <sup>CP</sup> TP (58) and dA <sup>CP</sup> TP (59).....	105
Scheme 23: Schematic representation of the Yoshikawa and Ludwig method for the generation of nucleoside 5'-triphosphates. <sup>389</sup> ....	106
Scheme 24: Schematic representation of the popular Ludwig-Eckstein nucleoside triphosphate synthesis approach. <sup>400</sup> .....	110

Scheme 25: Schematic representation of the nucleoside triphosphate synthesis approach using the cyclic pyrophosphoryl P-amidite approach developed by the Jessen research group.<sup>398</sup> .....111

Scheme 26: Proposed synthetic route towards the cyclopropene-modified deazaA nucleoside 80 containing an amide linker. ....119





## 8 Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation mit dem Titel „Artificial nucleic acids with enhanced functionalities“ selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Teilpublikationen:

Depmeier, H.; Kath-Schorr, S., Expanding the Horizon of the Xeno Nucleic Acid Space: Threose Nucleic Acids with Increased Information Storage, *J. Am. Chem. Soc.* **2024**, *146* (11), 7743–7751 (doi: 10.1021/jacs.3c14626)

Köln, den 30.06.2024

---

Hannah Depmeier