technische universität dortmund

MAX PLANCK INSTITUTE OF MOLECULAR PHYSIOLOGY

# Incorporation of modified genetic building blocks into nucleic acids and their potential biotechnological application

**Dissertation**

For the achievement of the academic degree

of the Doctor in Natural Sciences (Dr. rer. nat.)

Submitted to

Department of Chemistry and Chemical Biology

Technical University Dortmund

by

**Eric Constantin Ogel**

Born on 20.01.1998 in Mainz

Dortmund 2025

Die vorliegende Dissertation wurde im Zeitraum von April 2022 bis Oktober 2025 unter der Anleitung von Dr. Sidney Becker am Max-Planck-Institut für Molekulare Physiologie und an der Fakultät für Chemie und Chemische Biologie der Technischen Universität Dortmund angefertigt.

Dekan: Prof. Dr. Stefan M. Kast

Erstgutachter: Dr. Sidney Becker

Zweitgutachter: Prof. Dr. Dr. h.c. Herbert Waldmann

"The science of today will be the technology of tomorrow."

*Edward Teller*

Two master's students participated in the project “New labeling techniques for 5-Formylcytosine“ presented in this dissertation. Therefore, some of the findings of this dissertation are previously reported in the following master's theses:

Nguyen, Susanne. “Chemical labelling of 5-formylcytosine in DNA and its potential use for DNA sequencing”, Master thesis, TU Dortmund, **2024**.[1]

Schöne, Anna. “Development of novel chemical labeling strategies for the detection of 5-formylcytosine in DNA”, Master thesis, TU Dortmund, **2025**.[2]

Select parts of this work, particularly in Chapter 6 (2-Amino-DDP as a pH-responsive nucleobase), were published verbatim in the following publication:

Eric Ogel, Sidney Becker, “Synthesis and Incorporation of a pH-Responsive Nucleoside into DNA Sequences”, *Chembiochem* **2025**, e70086.[3] https://doi.org/10.1002/cbic.202500650

At the time of publication of this thesis, the article is an Online Version of Record before inclusion in an issue (Early View).[3]

# Acknowledgements

First and foremost, I would like to express my sincere gratitude to Dr. Sidney Becker for the opportunity to start the Becker Lab as his first PhD student, as well as for his great supervision. His consistent support and constructive feedback have significantly contributed to my professional growth and personal development.

I would like to express my deepest gratitude to Prof. Dr. Dr. h.c. Herbert Waldmann for providing an excellent research infrastructure in Department 4, for insightful feedback on my project, and for taking on the role of second examiner.

I want to thank my TAC members Prof. Dr. Daniel Summerer and Prof. Dr. Peter 't Hart for their valuable guidance and scientific advice.

I am also truly grateful for the opportunity to be a part of the International Max Planck Research School for Living Matter. The diverse workshops, talks, retreats, and numerous social events have all played a significant role in my growth as a researcher while allowing me to form new friendships. I am particularly thankful for the chance to attend international conferences, such as the Gordon Research Conference on Nucleosides, Nucleotides and Oligonucleotides, which were really rewarding. In addition to new scientific perspectives, I also gained cultural insights. None of this would have been possible without the dedication and expertise of Dr. Lucia Sironi and Christa Hornemann, whose passion and organizational skills made this excellent graduate program possible. Both of them offered consistent support and always have an open door to listen.

When I began my doctoral studies, I enjoyed a warm welcome in Dortmund with great support from my colleagues from Department 4 and the CGC, especially Lara, Aylin, Kai and Rachel, for which I am very grateful.

Over time, the Becker Group continued to grow and became a great interdisciplinary environment with motivated and appreciative colleagues, for which I am very grateful. I would like to thank the entire Becker Group for their support, both in the laboratory and outside. Thank you for the pleasant working atmosphere and enjoyable group activities, including laser tag and bowling. I will particularly remember our nice group retreat to the Eifel.

I especially acknowledge Raveena and Cedrik for chemistry discussions and their contributions to the positive working atmosphere in our chemistry lab A3.15, and for their acceptance of my selection of music. Special thanks go to Bhavana for her assistance with biochemistry-related questions and for her expert

guidance in preparing gels. I want to thank Shanice for her support with the Orbitrap, and Marlon for his assistance in all questions regarding sequencing.

I am really grateful for my two master's students, Susanne and Anna, for their fantastic support with the 5fC project. For the successful research collaboration in the field of enzymatic triphosphate synthesis, I would like to express my gratitude to Dr. Maryke Fehlau and Dr. Nicolas Cornelissen. In addition, I like to thank the staff in the analytical department for their reliable processing of my samples.

I would also like to thank my colleagues from Department 4 and the CGC groups for the constructive scientific discussions we have had, as well as the enjoyable lunch and coffee breaks. Here, I want to highlight Aylin, Lara, Celine, Jana, Kai, Rachel, Gian, Jan, and Niko. I would like to thank Jana, Celine and Tobi for the lovely game evenings.

I want to thank all my colleagues and friends who have supported and motivated me throughout all my studies and beyond. I would like to express my heartfelt gratitude to all the wonderful people I met during my studies in Mainz. A very special thank you goes to Nino, Seb, Diana, and Julius, who have been by my side since our very first semester. Our circle of friends was greatly enriched by Moritz, Franjo, Jannis, Alina, Siska and Simone. Together, we formed a close group that supported, encouraged, and inspired one another through every challenge and success. I really value the strong connection we've built, and I'm so grateful that, no matter the physical distance between us now, our friendship keeps growing and brings so much joy.

My sincere and deepest gratitude goes to my whole family, especially to my parents and siblings. Mama, Papa, Philipp, Denise: Ihr habt mich während meines gesamten Lebens und meiner akademischen Laufbahn, vom Bachelor- und Masterstudium bis hin zur Promotion, immer unterstützt, motiviert und an mich geglaubt. Eure Unterstützung, euer Vertrauen und eure Liebe haben mich so weit gebracht. Dafür bin ich euch unendlich dankbar.

Finally, I would like to thank my partner for always supporting me in all situations. Your kindness, patience, and understanding mean the world to me.

## Table of Contents

# 1. Abstract

Natural DNA modifications, particularly cytosine modifications, have been demonstrated to play a central role in various biological processes. Their alterations have been linked to a multitude of diseases, including cancer, neurological disorders, and developmental disorders.[4-6] By conducting a detailed analysis of these modifications, we can gain a more profound understanding of biological processes and develop potential therapeutic approaches.[6-8] While significant progress in 5-methylcytosine (5mC) research has been made, the less abundant 5-formylcytosine (5fC) remains a relatively underexplored modification.[8] Conventional malononitrile-based labeling has been shown to convert 5fC to a base read as T/U by polymerases, thereby differentiating it from cytosine.[9] However, this method is limited in its conversion efficiency and can lead to DNA damage and degradation. In this work, novel labeling reagents were evaluated and a new method for the selective labeling of 5fC was developed. Acetylacetone was used to induce 5-formylcytosine-to-thymine conversion sequencing in single-base resolution. We termed this novel technique, which exhibits high labeling efficiency and selectivity, AceF-Seq. Under optimal conditions, a quantitative conversion could be achieved with no observable DNA degradation. AceF-Seq is thereby significantly enhancing the precision and reliability of 5fC detection in DNA and supporting advanced epigenetic studies.

The following research project focused on the development of a programmable, pH-responsive nucleoside to build DNA nanostructures. The modified nucleobase 2-Amino-dideaza purine (2-Amino-DDP) was synthesized as triphosphate (dNTP) and as phosphoramidite and incorporated into DNA enzymatically and by solid phase oligonucleotide synthesis. As indicated by the measured melting points, there is a strong destabilization of the base pair 2-Amino-DDP:T under slightly acidic conditions that is not observed when pairing with cytosine. Additionally, the pH-dependent incorporation of 2-Amino-DDP triphosphate by different DNA polymerases was observed. They exhibited a clear preference for cytosine under lower pH and for thymine under higher pH conditions, suggesting that 2-Amino-DDP has the capacity to switch between paired and unpaired states based on pH. This innovation advances the potential for constructing DNA-based nanotechnology and may also enhance acid-base catalysis in nucleic acid enzymes, analogous to histidine's function in proteins.

The third project explored the design and synthesis of the unnatural nucleobase 2,6-Diamino-dideaza purine (2,6-Diamino-DDP). In its protonated form, it is believed to form a novel base pair with three hydrogen bonds with labeled 5fC-bases, thereby expanding the genetic alphabet. The nucleoside was synthesized using an optimized Buchwald-Hartwig amidation and the 2,6-Diamino-DDP triphosphate was produced enzymatically, using a modified polyphosphate kinase. During primer extension experiments, it could be fully incorporated opposite a labeled version of 5fC, therefore forming a novel base pair. This approach has the potential to optimize 5fC-sequencing with direct readout and could support further applications such as DNA data storage and aptamer development.

# 2. Zusammenfassung

Natürliche DNA-Modifikationen, insbesondere Veränderungen an Cytosin, spielen eine zentrale Rolle in verschiedenen biologischen Prozessen. Ihre Veränderungen wurden mit einer Vielzahl von Krankheiten in Verbindung gebracht, darunter Krebs, neurologische Störungen und Entwicklungsstörungen.[4-6] Durch eine detaillierte Analyse dieser Modifikationen können wir ein tieferes Verständnis biologischer Prozesse erlangen und potenzielle therapeutische Ansätze entwickeln.[6-8] Während bei der Erforschung von 5-Methylcytosin (5mC) bedeutende Fortschritte erzielt wurden, ist das seltenere 5-Formylcytosin (5fC) relativ wenig erforscht.[8] Die Reaktion mit Malononitril wandelt 5fC in eine Base um, die von Polymerasen als T/U gelesen wird, wodurch es sich von Cytosin unterscheidet.[9] Diese Methode ist jedoch in ihrer Effizienz begrenzt und kann zu DNA-Schäden führen. In dieser Arbeit wurden verschiedene Reagenzien evaluiert, um eine neue Methode zur selektiven Markierung von 5fC zu entwickeln. Acetylaceton wurde verwendet, um die Umwandlung von 5-Formylcytosin zu Thymin bei der Sequenzierung zu induzieren. Wir haben diese neuartige Technik, die eine hohe Effizienz und Selektivität aufweist, AceF-Seq genannt. Unter optimalen Bedingungen konnte eine quantitative Umwandlung ohne DNA-Schäden erreicht werden. AceF-Seq verbessert damit die Präzision und Zuverlässigkeit der 5fC-Detektion in der DNA erheblich und unterstützt weitere epigenetische Studien.

Im zweiten Forschungsprojekt wurde ein programmierbares, pH-responsives Nukleosid entwickelt. Die modifizierte Nukleobase 2-Amino-dideaza-purin (2-Amino-DDP) wurde als Triphosphat (dNTP) und als Phosphoramidit synthetisiert und enzymatisch sowie durch Festphasen-Oligonukleotidsynthese in die DNA eingebaut. Wie die gemessenen Schmelzpunkte zeigen, kommt es unter leicht sauren Bedingungen zu einer starken Destabilisierung des Basenpaares 2-Amino-DDP:T, die bei der Paarung mit Cytosin nicht zu beobachten ist. Zusätzlich wurde der pH-abhängige Einbau von 2-Amino-DDP-Triphosphat durch verschiedene DNA-Polymerasen beobachtet. Sie zeigten eine klare Präferenz für Cytosin bei niedrigerem pH und für Thymin bei höherem pH, was darauf hindeutet, dass 2-Amino-DDP die Fähigkeit besitzt, je nach pH-Wert zwischen gepaarten und ungepaarten Zuständen zu wechseln. Auf Basis dieser Entdeckung könnten DNA-basierte Nanotechnologien entwickelt und Säure-Base-Katalyse in Nukleinsäureenzymen verbessert werden, analog zur Funktion von Histidin in Proteinen.

Das dritte Projekt befasste sich mit dem Design und der Synthese der unnatürlichen Nukleobase 2,6-Diamino-dideaza-Purin (2,6-Diamino-DDP). In seiner protonierten Form bildet es ein neues Basenpaar mit drei Wasserstoffbrückenbindungen mit markierten 5fC-Basen, wodurch das genetische Alphabet erweitert wird. Das Nukleosid wurde unter Verwendung einer optimierten Buchwald-Hartwig-Amidierung synthetisiert und das 2,6-Diamino-DDP-Triphosphat wurde enzymatisch unter Verwendung einer modifizierten Polyphosphatkinase hergestellt. Bei Primer-Verlängerungs-Experimenten konnte es vollständig gegenüber einer markierten Version von 5fC eingebaut werden und bildete so ein neues Basenpaar. Dieser Ansatz hat das Potenzial, die 5fC-Sequenzierung mit direkter Auslesung zu optimieren und könnte weitere Anwendungen wie die Speicherung von Daten in DNA und die Entwicklung von Aptameren unterstützen.

# 3. Introduction

## 3.1 DNA structure

The genetic code, written in the DNA, is based on the four nucleobases adenine (A), guanine (G), thymine (T) and cytosine (C).[10] The chemical composition of nucleic acids makes them ideal for use as carriers and transmitters of genetic information.[10] Nucleic acids are linear polymers, while each monomer is comprised of three moieties: a sugar unit, a phosphate, and a nucleobase.[10] The sequence of bases within the polymer is unique for each nucleic acid and forms a linear structure of organized information.[10] X-ray diffraction images from Watson, Crick and Franklin revealed the double-helical structure of the DNA in 1953.[11-12] Under physiological conditions, the majority of DNA exists in the B-form consisting of two helical, antiparallel polynucleotide strands that are wound in a clockwise direction around a common axis.[10] The sugar-phosphate backbone is positioned externally, with the purine (A/G) and pyrimidine (T/C) bases situated internally within the helix, while two neighboring bases are 0.34 nm apart.[10] Every 3.4 nm, the helical structure repeats, so 10.4 bases make up one helical turn, while the diameter of the helix is 2 nm.[10] Less hydrated DNA strands form A-DNA, where the double helix displays a wider diameter and shorter helical repeats compared to the B-DNA helix.[10] Its base pairs are exhibiting a greater inclination towards the helix axis.[10] In the helix of the A-form, fewer water molecules bind to the phosphate backbone than in B-DNA, making it the preferred form in conditions of dehydration.[10] However, double-stranded RNA regions and some RNA-DNA hybrids can adopt a form of double helix that is highly similar to A-DNA under hydrated conditions.[10]

The DNA double helix molecule features two distinctive grooves, known as major and minor grooves.[10] This is a consequence of the relative orientation of the glycosidic bonds of a base pair, which are not diametrically opposed.[10] The minor groove contains the O-2 of a pyrimidine and the N-3 of a purine, which can act as hydrogen bond acceptors.[13] In the major groove, N-7 of a adenine or guanine, O-4 of thymine and O-6 of guanine are potential hydrogen bond acceptors, while the exocyclic amines N-6 of adenine and N-4 of cytosine can act as hydrogen donors.[10] This allows the binding of proteins, but also natural products to DNA.[14] In B-DNA, the major groove is wider (1.2 nm vs. 0.6 nm) and deeper (0.85 nm vs. 0.75 nm) than the minor groove.[10] It also has a greater number of features that distinguish one base pair from another, making the major groove more accessible for interactions with proteins that recognize specific DNA sequences.[10] A third type of double helix, Z-DNA, is left-handed, while the phosphate backbone forms a “zigzag” line, which led to the name Z-DNA.[10] The biological significance of Z-DNA is less pronounced than that of the A- and B-forms.[10]

DNA can vary not only in the form of its double helix or its sequence, but also through chemical modifications to the nucleobases themselves.[15] To date, more than 40 DNA modifications have already been

characterized in all three domains of life.[15] These play a crucial role in central cellular processes, such as the regulation of gene expression or the organization of chromatin structure.[15-18] In addition to the genetic sequence, DNA modifications represent the epigenetic code as another level of information.[15, 18-21] Further information on these points is provided in chapter 3.3.2.

## 3.2 DNA Polymerases

DNA polymerases catalyze the addition of dNTPs to a DNA strand.[10] Most DNA polymerases are template-dependent enzymes and the base of the incorporated dNTP needs to be complementary to the base in the template strand to be incorporated effectively.[10] The DNA chain elongation occurs in the 5' to 3' direction, making it necessary to have a primer with a free 3'-OH group in place in order to initiate the synthesis process.[10] In addition, DNA polymerases need a Lewis acid, such as $Mg^{2+}$, as cofactor.[10] A nucleophilic attack of the 3'-OH primer end on the innermost phosphorus center of the dNTP forms the new phosphodiester bond and releases pyrophosphate.[10] A significant number of DNA polymerases contain a 3' to 5' exonuclease domain, which identifies base pair mismatches and removes the incorrect nucleotide to be replaced by the correct one.[10] This function is also known as proofreading and contributes to the remarkably high accuracy of DNA replication, which has an error rate of less than $10^{-8}$ per base pair.[10]

For most of the polymerases used in this work, this exonuclease domain has been removed to prevent the removal of unnatural nucleosides after their incorporation.

## 3.3 Epigenetic regulation

There are heritable changes in gene function that occur without a change in the underlying DNA sequence.[17] These epigenetic changes can affect gene expression without changing the DNA code itself.[17] Environmental factors and cellular processes influence gene expression and can thus lead to changes in the phenotype of an organism.[17] Epigenetic regulation involves the addition or removal of chemical modifications to DNA, but also on histone proteins, which are an essential building block of chromatin.[17] The following sections will introduce the mechanisms of epigenetic regulation.

### 3.3.1 Histone modifications

Histones are basic proteins around which DNA is wrapped.[17] The high content of lysine and arginine in histones leads to their basic properties, which in turn results in a strong binding affinity for the polyphosphate backbone of DNA.[10] Covalent histone modifications such as acetylation, methylation or phosphorylation can either loosen or compact the chromatin structure and thus influence gene expression.[17] Histone methylation is dynamically regulated at specific sites and is essential for proper gene expression during

developmental processes.[22] Disruptions in this regulation can cause developmental defects and are linked to genetic disorders that can affect the skeleton and nervous system.[22] Histone modifications also regulate chromatin remodeling.[22] The packaging of chromosomal DNA into nucleosomes facilitates genome organization and compaction, but restricts access to regulatory DNA elements.[23] Chromatin components engage in transcriptional regulation, chromosome segregation, DNA replication, and DNA repair processes.[23] Specialized chromatin remodeling complexes (remodelers) execute a range of nucleosome-modifying activities, including nucleosome sliding, destabilization, ejection, and reorganization.[23] With these regulatory mechanisms, access to transcription factors and other regulatory proteins can be enabled or prevented, thereby influencing gene expression.[23]

### 3.3.2 Modifications on the DNA bases

There are naturally occurring, chemically modified variants of all four canonical nucleobases in the DNA.[15, 24] The following figure is a brief overview of DNA modifications that play important roles in gene regulation[9], development[25], and cell differentiation[26].

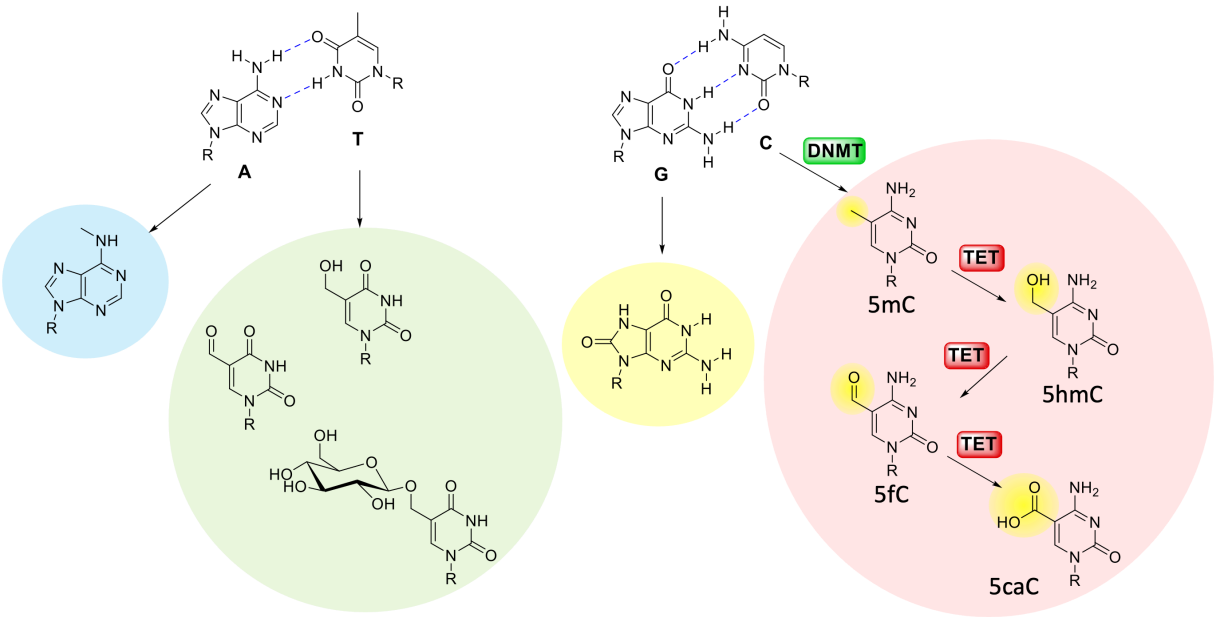



*Figure 1: A variety of chemical modifications on all four nucleobases that are naturally occurring in eukaryotic DNA.*

Especially the modifications at the 5-position of certain cytosines in DNA, typically in CpG islands, are crucial to many cellular functions[27] and are therefore discussed more detailed in chapters 3.4 to 3.8. The determination of these modifications by sequencing is discussed in chapter 3.9.3.

In addition to direct epigenetic influences that can result in changes in gene expression, there are also indirect influences. For instance, the presence of oxidative DNA damage, specifically 8-oxoguanine, has been observed to result in an approximately 300% increase in gene expression.[24] This phenomenon occurs through the regulation of gene transcription via base excision repair.[24]

### 3.4 Cytosine methylation

Historically, 5mC was the first described nucleobase modification, when a small amount of a methylated cytosine derivative was identified as a hydrolysis product of tuberculinic acid gained from bacteria in 1925.[28] But it was only decades later that its great relevance for a variety of regulatory mechanisms was recognized.[20] In 1948, R. Hotchkiss analyzed calf thymus DNA and observed an additional band near cytosine using paper chromatography.[29] The retardation factor was similar to that of cytosine, with a slight shift, indicating a modification of cytosine, which was later attributed to 5mC.[29] After it has been demonstrated that bacteriophage DNA can be selectively degraded by the host bacterium[30], the discovery of strain-specific methyltransferase activity in bacteria indicated a possible role of 5mC in the defense against bacteriophages.[30] It was later discovered that eukaryotes possess DNA methyltransferases that function as „writers“ and that they are also „reader“ proteins that are capable of recognizing 5-methylcytosine.[20]

In mammals, 5-methylcytosine (5mC) is the primary form of epigenetic DNA modification.[17] There is a lack of methylation in CpG islands, which are often located in close proximity to gene promoters,[31] while over 80% of CpG dinucleotides located outside of CpG islands are methylated.[17] Cytosine methylation plays a critical role in development and disease, and its functions include regulating gene expression[17] and silencing[26], and X-chromosome inactivation.[18, 32] The methyl group in 5mC is attached to the cytosine by a carbon-carbon bond, making it chemically stable and difficult to remove directly.[32] It can be installed by *de novo* methyltransferases DNMT3A and DNMT3B[33] and is genetically maintained by a robust mechanism involving DNMT1, which recognizes half-methylated CpG sites and ensures that the newly synthesized DNA strand is methylated after replication.[32, 34] The maintenance process also involves the protein UHRF1 (ubiquitin-like, containing PHD and RING finger domains 1), which co-localizes with DNMT1 during S-phase.[35] Correct 5mC-maintenance is essential for the proper regulation of gene expression and the prevention of genetic instability.[32]

Although 5mC is chemically stable, it can be reverted to its unmodified form through different mechanisms.[32] One mechanism is passive DNA demethylation. Because 5mC is read as cytosine by polymerases, a guanine is incorporated on the opposite strand. Therefore, the absence of effective DNA methylation maintenance leads to the gradual loss of 5mC during DNA replication.[32] Active demethylation,

on the other hand, occurs when specific enzymes gradually oxidize 5mC.[32, 35] This process is discussed in the following chapter 3.5.

### 3.5 5mC oxidation by TET enzymes

In mammals, the process of DNA methylation can be actively reversed by a multi-step biochemical reaction that starts with TET enzymes, a family of ten-eleven translocation dioxygenases.[36] TET enzymes oxidize 5-methylcytosine (5mC) stepwise to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC), as depicted in Figure 1.[32, 36] The latter two can be removed from DNA by base excision repair, mediated by enzymes such as thymine DNA glycosylase (TDG).[32]

The involvement of TET enzymes in active regulatory regions of the genome has been thoroughly investigated in embryonic stem cells.[37] However, these enzymes are also broadly expressed across differentiated tissues.[37] The existence of multiple TET isoforms is a consequence of the utilization of alternative regulatory elements in distinct tissues.[37] It has been observed that certain isoforms, including TET2, are deficient in the CXXC domain, a CpG-binding protein domain.[37] This has been implicated in the selective recruitment of these enzymes to specific genomic locations.[37] The TET3 protein exists in three primary isoforms, including a full-length variant with an N-terminal CXXC domain (Tet3FL).[38] While this CXXC domain is known to bind unmethylated CpG sites, it also exhibits a great affinity for 5-carboxylcytosine (5caC).[38]

### 3.6 5-hydroxymethylcytosine (5hmC)

5-hydroxymethylcytosine (5hmC) is the first product of the oxidation of 5mC by TET-enzymes and is found in many cell types, especially in the brain[16] and embryonic stem cells.[39] 5hmC is not just a temporary intermediate in the DNA demethylation pathway, but a stable epigenetic mark with a distinct genomic distribution and specific regulatory functions during development, aging, and cancer.[16, 26, 32, 40] In mammals, 5hmC has been observed to be present in elevated levels in certain tissues, particularly in neurons.[4, 40-41] It can make up to 0.6% of all nucleotides in Purkinje cells and 0.2% in granule cells.[16] The high abundance of 5hmC in neurons is thought to play important roles in gene regulation, neuronal development, and synaptic plasticity.[4, 40-41]

### 3.7 5-formylcytosine (5fC)

5-formylcytosine (5fC) is found in the DNA of many cell types[9, 19, 42-43], but also in RNA.[44] The first discovery of 5fC in embryonic stem cell DNA was published by the Carell group in 2011.[45] As a naturally occurring

modified base, 5fC plays a crucial role in epigenetic regulation[44], influencing gene expression[42] and cellular differentiation.[9] Some cellular functions of 5fC are depicted in Figure 2.

It has been shown that 5fC also plays a significant role in zygotic reprogramming by recruiting RNA Pol III in *Xenopus* and mouse embryos.[46] Because the tRNA expression *in vivo* is enhanced, 5fC is an activating epigenetic mark.[46] Although 5fC seems to have many biological functions, its prevalence is significantly lower than that of 5mC and 5hmC, with variable concentrations of around 20 to 200 ppm of cytosines[43], depending on the specific cell type.[19] The levels of 5fC can differ significantly between different cell types, suggesting that it may play distinct roles in different biological contexts.[19] In 2017, 5fC has been observed to form reversible conjugates with histone proteins in cells via Schiff-base formation between its formyl group and lysine side chains of proteins (Figure 2 bottom right).[47] It is hypothesized that the resulting reversible DNA-protein conjugates modify the chromatin structure and contribute to the epigenetic control of gene expression.[47]

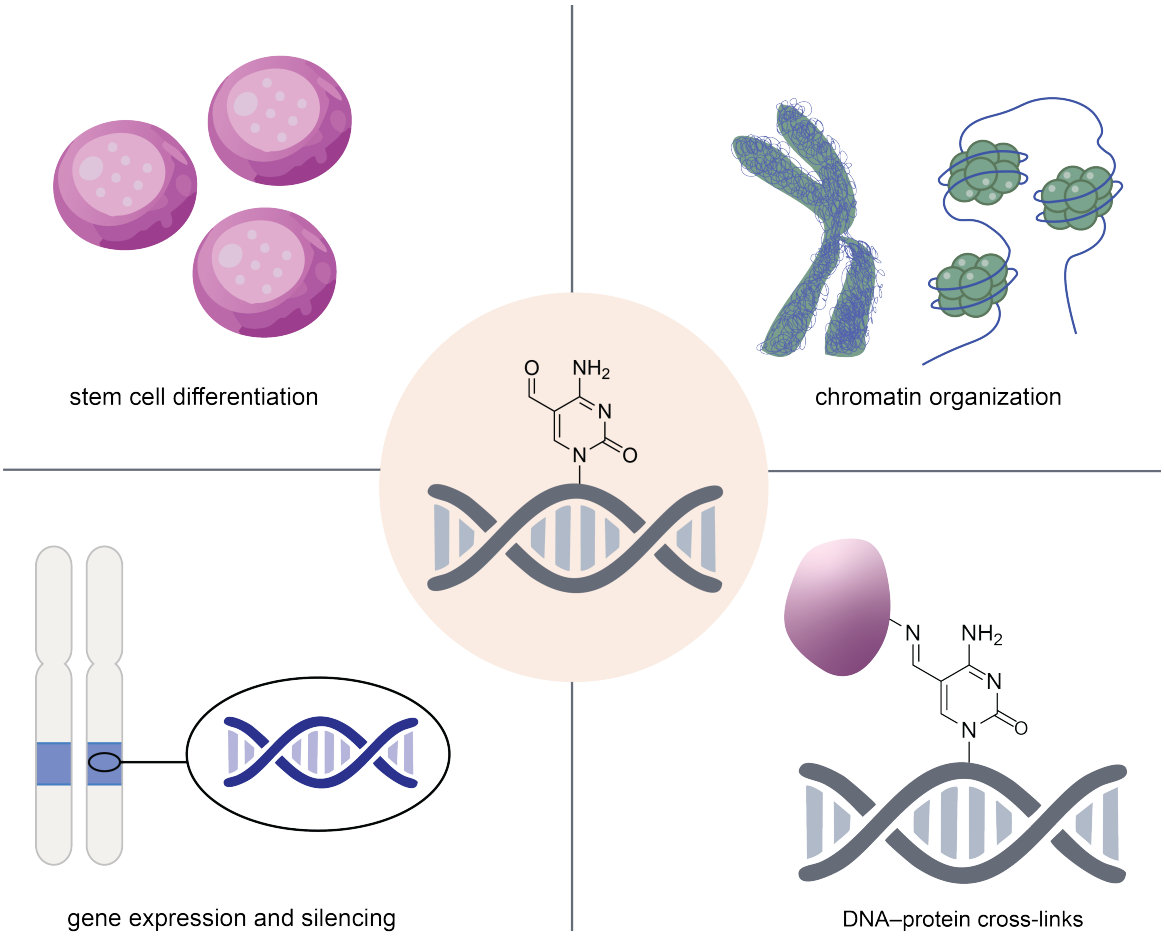



*Figure 2: Overview over a selection of cellular functions of 5fC.*

### 3.8 5-Carboxycytosine (5caC)

In the last oxidation step catalyzed by TET enzymes, 5-carboxylcytosine (5caC) is produced.[48] This process could be observed in vitro and in cultured cells.[48] Thymine-DNA glycosylase (TDG) is an enzyme that usually removes thymine moieties from G/T mismatches.[48] In 2011, it was discovered that human thymine DNA glycosylase has the capacity to efficiently recognize and remove 5caC.[48] In the absence of TDG in mouse embryonic stem cells, 5caC accumulates.[48] These findings suggest that TET-mediated oxidation of 5mC, followed by TDG-driven excision of 5caC, establishes a pathway for active DNA demethylation.[48] 5caC has been shown to directly interact with RNA polymerase II, thereby slowing transcription elongation.[49] This interaction, which involves hydrogen bonds between the conserved epi-DNA recognition loop in the polymerase and the carboxyl group of 5caC, has a direct and significant impact on gene expression.[49]

It has also been demonstrated that the methyl-CpG-binding domain (MBD) can be engineered to selectively recognize 5caC marks in CpG dyads, revealing the scaffold's evolutionary adaptability and potential for custom epigenetic reader design.[50]

### 3.9 DNA Sequencing: relevance and applications

The DNA sequence contains essential information that enables living organisms to survive and reproduce.[51] Determining this sequence is remarkably valuable for basic research, but also for various applications.[15] In the medical field, DNA sequencing is a valuable tool that enables the identification and diagnosis of diseases and genetic disorders.[52] Genomic sequence data has emerged as a foundational element in the field of personalized medicine, particularly in the context of cancer therapies, where it serves as a guide for therapeutic intervention.[52]

Prior to sequencing, the DNA needs to be isolated, for example from a blood sample.[6] The presence of cell-free nucleic acids (cfNAs) in blood plasma, derived from a variety of tissues and cell types, makes them collectible through minimally invasive routine blood draws.[6] The collection of cfNAs offers valuable and comprehensive insight into human physiology.[6] A scheme showing the concept of a liquid biopsy, after which the cfNAs can be isolated and sequenced, is shown in Figure 3.

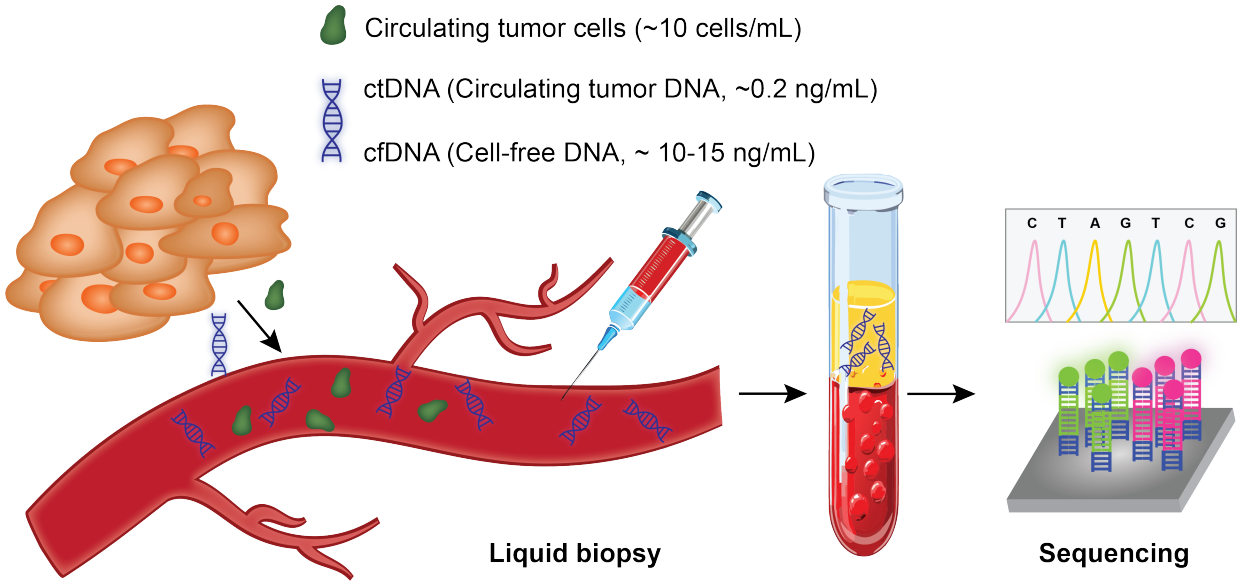



*Figure 3: Schematic figure showing the concept of a liquid biopsy, in which the cell-free nucleic acids, such as cell-free DNA and circulating tumor DNA (ctDNA) can be isolated and sequenced.*

Liquid biopsy assays of tumor-derived factors have several applications in cancer care.[53] In addition to the detection of cancer in an early stage, the success of cancer treatment can be monitored.[53] Advances in next-generation sequencing and molecular biology tools have enabled liquid biopsy technologies that use cfNAs for noninvasive diagnostics.[6] These technologies are used for prenatal testing, cancer monitoring, organ transplant surveillance, and infectious disease detection.[6] cfDNA is present in low concentrations of around 7 ng/mL in plasma, though this level can vary widely among individuals.[6] The cfDNA concentrations are often higher in older individuals and in those with inflammation, infection, obesity, or metastatic cancer.[6] These quantitative limitations cause significant challenges, particularly for applications requiring high analytical sensitivity.[6] The integration of molecular profiling with modern sequencing methods enables the identification of crucial genetic mutations and biomarkers, thereby contributing to a more precise diagnosis and treatment of cancer.[54]

Nevertheless, challenges such as cost, quality, and the preservation of tissue samples must be addressed.[54] This underscores the necessity of very sensitive techniques that enable accurate sequencing even with a limited quantity of DNA material.

### 3.9.1 Sanger sequencing

One of the earliest sequencing techniques is the dideoxy method or Sanger method, named after Frederick Sanger. He published his first chain-termination method in 1975[55] and the dideoxy nucleoside method in 1977.[56] In the latter publication, four polymerase reactions were used, all containing the four nucleotides, one of which is radioactively labeled.[56] In addition to the four dNTPs, each reaction mixture contains a

small percentage of the 2',3'-dideoxy analogue of a nucleotide (ddNTPs).[56] The chain-terminating ddNTPs lack a 3'-hydroxyl group, preventing the DNA polymerase from extending the DNA.[56] Consequently, DNA fragments of varying lengths are generated and separated by polyacrylamide gel electrophoresis (PAGE).[56] By comparing the four batches, the sequence can be read after exposure of the radioactive gel to a photographic film.[56] In more recent approaches, the ddNTPs are each coupled to a different fluorescent label at the base, thus avoiding the need to split into separate batches and to handle radioisotopes.[10] The generated fragments are separated according to size by capillary electrophoresis, and the fluorescence at each of the four wavelengths reveals the complementary sequence of the original DNA template.[10]

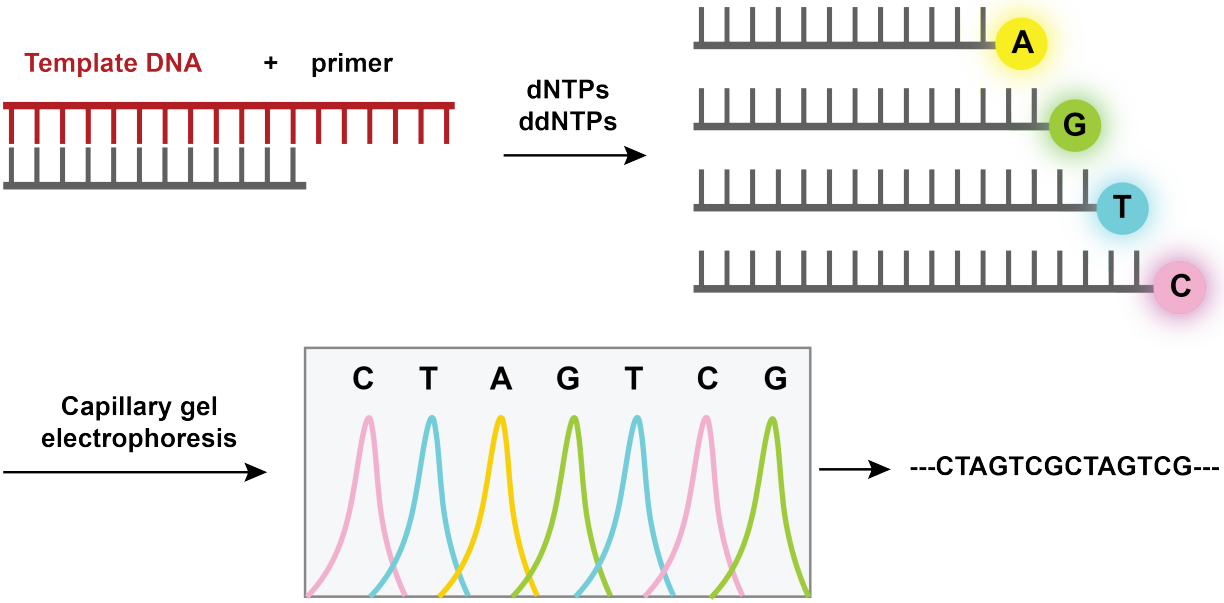



*Figure 4: Schematic workflow of Sanger sequencing.*

It is currently possible for automatic Sanger DNA sequencers to determine over one million bases per day using this method.[10] The major limitation of this method is the maximum length of the sequences of around 1,000 bases and the relatively low throughput compared to Next-Generation Sequencing (NGS) technologies.[10, 57]

### 3.9.2 Next-Generation Sequencing (NGS)

Next-Generation sequencing (NGS) techniques enable the examination of billions of sequence fragments within a single experimental procedure.[10, 58] NGS facilitates the parallel sequencing of multiple DNA templates (a process known as multiplexing) by immobilizing them on an array surface and subjecting them to the same reagents. [10, 51, 58] DNA fragments are amplified by PCR on a solid support and function as templates for a DNA polymerase, whereby the incorporation of a dNTP creates a signal that is detected

differently, depending on the specific NGS method used.[10] The signal can be created using either a luciferase (for pyrosequencing) or a fluorescent tag (for Illumina sequencing).[51, 59] This process is referred to as sequencing-by-synthesis (SBS).[58]

In pyrosequencing, which was first described in 1993[59], streptavidin coated magnetic beads and a DNA polymerase without 3´to 5´exonuclease activity are used. The nucleotides are added one by one, and every time one nucleotide is incorporated, one molecule of pyrophosphate is released.[59] This is used by ATP sulfurylase together with adenyl sulfate to form ATP, which in turn is used by luciferase to convert luciferin to oxyluciferin.[59] The light emitted in this process is then detected.[59] The intensity of the light signal is indicative of the number of nucleotides that have been incorporated.[59] Subsequently, the nucleotide mixture is removed prior to the addition of the next nucleotide.[59] This process is repeated for each of the four nucleotides until the DNA sequence of the single-stranded template is determined.[59]

In the reversible termination method, each nucleotide has its own fluorescent label and a reversibly blocked 3' end.[10] Once the nucleotide has been incorporated into the growing chain, it is identified by the fluorescent label and the blocking group is removed, allowing the process to be repeated.[10] One example for the reversible termination method that has been published by David R. Bentley and Shankar Balasubramanian.[51] This technique is using nano-wells that contain oligonucleotides as “anchoring points” on a chip. Thousands of copies of each DNA fragment are generated by bridge amplification PCR.[51] The use of 3′-OH-modified nucleotides enables an efficient incorporation while allowing the simultaneous addition of all four nucleotides and avoiding over-incorporation.[51] Following each round of synthesis, a camera takes a picture of the chip and a computer determines the added base by the wavelength of the fluorescent tag for each spot on the chip.[51] After washing the non-incorporated molecules away, the 3’-OH-protecting group is removed chemically.[51] The process is continued until the entire DNA molecule has been sequenced.[51] The schematic workflow of Illumina sequencing is shown in Figure 5.

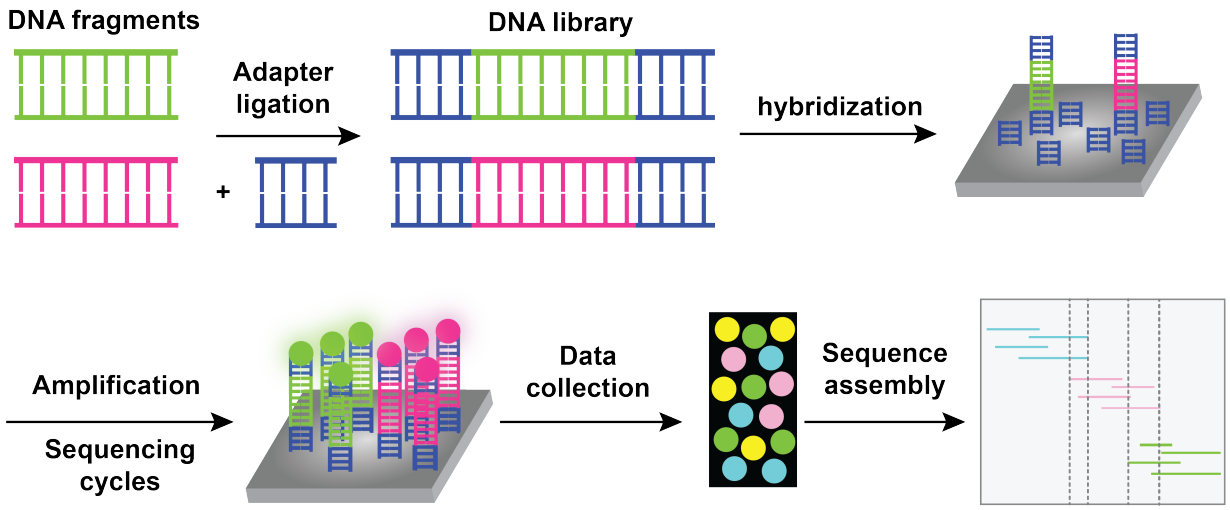



*Figure 5: Schematic workflow of Illumina sequencing.*

Another method published in 2011 uses semiconductor chips to directly sequence DNA by Ion Torrent, which also involves sequencing-by-synthesis.[60] The Ion Torrent technology converts chemical DNA codes into digital data on a semiconductor chip with over one million wells where DNA polymerase reactions occur.[61] When a specific dNTP is incorporated into the growing DNA strand, $H^+$ is released and a pH change is detected by a transistor.[61] The $H^+$ ions are detected and the signal is converted into a corresponding DNA sequence read by an algorithm.[61] Each reaction is detected by its own sensor, allowing for parallel and simultaneous detection of multiple sequences.[60-61]

### 3.9.3 Third-Generation Sequencing

A majority of the sequencing techniques require template amplification.[58] However, the sample amplification has some drawbacks, including the potential for errors in replication and sequence-dependent biases.[58] In addition, during the process of replication via PCR, epigenetic modifications on the nucleobases such as methylation get “diluted” and the additional information is lost.[58]

In 2003, a pioneering real-time observation of DNA synthesis in single molecules was made.[62] This observation was facilitated by the use of fluorescently labeled nucleotides and a DNA polymerase anchored in nanoscopic holes in metal films, called zero-mode waveguides.[62] The fluorescent excitation is limited by the zero-mode waveguide, which is small enough that just one growing DNA chain fits and causes signals.[58] Therefore, background fluorescence is strongly reduced and only the active site of the polymerase is visible.[25] Subsequent advancements in this field led to the development of the PacBio platform.[58] This technique achieves consensus sequence accuracy of 99.3% at 15-fold coverage, with error rates mainly dependent on the types of fluorophores used.[63] This technology has the capacity to

produce very long read lengths, often ranging from 10,000 to 100,000 base pairs, making it useful for *de novo* genome assembly, detecting structural variants, and resolving repetitive regions.[58]

Another approach is the (Oxford) Nanopore sequencing, where the DNA is passing through a thin protein pore, which is causing changes in electrical current that can be measured.[64] While the idea was already hypothesized in the 1980s[58], the efficient nanopore sequencing of long, natural DNA strands was published in 2014.[64] A key distinction of nanopore devices compared to other sequencing technologies is their ability to be exceptionally portable, with the capacity to fit as compactly as a USB stick.[64] This portability is a result of their detection method, which utilizes electronic signals instead of optical detection.[64]

### 3.9.4 Sequencing of epigenetic bases

As previously discussed, epigenetic base modifications, especially cytosine modifications in the 5-position, play a crucial role in fundamental cellular processes, such as the regulation of gene expression and the organization of the chromatin structure.[15, 20, 27] Consequently, these DNA modifications form a second level of information in addition to the genetic sequence. However, it is not possible to solve this code without the use of advanced methods, as most conventional methods are unable to read modifications.[65] In Sanger sequencing, or after amplifying a sample via PCR, guanine is incorporated opposite to modified cytosines.[25] The reason is that modifications on the 5-position in cytosine do not change the hydrogen bonding pattern. Therefore, G is incorporated opposite 5mC, 5hmC, 5fC and 5caC, causing a loss of the epigenetic information.[41] Consequently, numerous methods have been developed over time to make these modifications visible in sequencing.[9, 16, 25-26, 66]

To simplify the sequencing of epigenetic bases, one strategy is to enrich DNA fragments with modification-specific antibodies and subsequently sequence them.[67] Non-base-resolution approaches generally involve enriching oxidized forms of 5mC prior to high-throughput sequencing.[32] For instance, DNA immunoprecipitation (DIP) can be used to enrich 5hmC, 5fC, and 5caC through antibodies that specifically bind to these DNA modifications.[32] This approach (DIP-Seq) allows genome-wide mapping of epigenetic bases with a resolution of approximately 400 base pairs.[68] Nevertheless, this technique has the risk of off-target binding by the antibodies, which can lead to false-positive results.[68]

Recently developed methods avoid this issue and can determine cytosine modifications at the base level. These methods involve the use of either chemical or enzymatic techniques to induce a selective mutation of cytosine to thymine (C-to-T transition). Following the DNA amplification using PCR, a thymine is detected through DNA sequencing. This principle is employed in the Bisulfite sequencing method, which was first described in 1992[27] and is regarded as the gold standard for achieving single base resolution in 5mC sequencing.[25] When DNA is treated with sodium bisulfite, the resulting deamination reaction converts

unmodified cytosine to uracil, which is read as T after amplification, while 5mC remains as C due to its higher resistance to deamination.[27] This method enabled the first genome-scale methylation maps (methylomes).[20] However, bisulfite sequencing cannot distinguish 5hmC from 5mC[69], nor can it differentiate between 5fC and 5caC.[25] This limitation may have resulted in misinterpretation of epigenetic marks with different regulatory functions.[25] Other limitations of this method include degradation of the DNA[70] and incomplete conversion[71].

In order to address these limitations, several advanced methods have been developed.[25] TET-assisted bisulfite sequencing of 5-hmC (TAB-seq)[72] as well as oxidative bisulfite sequencing (oxBS-seq)[4] can distinguish between 5mC and 5hmC. In 2013, the method fCAB-seq[19] was developed. It combines chemical reduction with biotin tagging to detect 5fC. RedBS-seq[73] is another method to detect 5fC, based on chemical reduction of 5fC to 5hmC and subsequent bisulfite treatment. For the detection of 5caC, CAB-Seq[74] can be used to convert the carboxylic acid into an amide which can survive the bisulfite treatment without undergoing deamination.

However, all bisulfite-based methods have inherent disadvantages, as harsh treatment conditions cause sample degradation by creating abasic sites (more than 1 in every 200 bases)[70], which promote strand breaks and limit read length.[25] This degradation is particularly problematic in highly repetitive genomic regions, where the resulting short reads complicate the alignment to a reference sequence.[25]

The following section will provide an overview of bisulfite-free 5fC-detection methods using small-molecule labeling, due to their particular relevance to the research of this thesis. Some methods work at single-base resolution and involve the reaction of a nucleophilic labeling reagent with the formyl-group of 5fC to construct a new base. After cyclization, this “labeled” 5fC has a different hydrogen bonding pattern. In the following PCR amplification, the labeled 5fC therefore gets read as T and A is incorporated opposite to that base, allowing its detection.

One example is fC-CET[42] (cyclization-enabled C-to-T transition of 5fC), which is based on the reaction of an azide-containing derivative of 1,3-indandione with 5fC. This allows the selective enrichment of 5fC-containing DNA via click chemistry and biotin pulldown.[42]

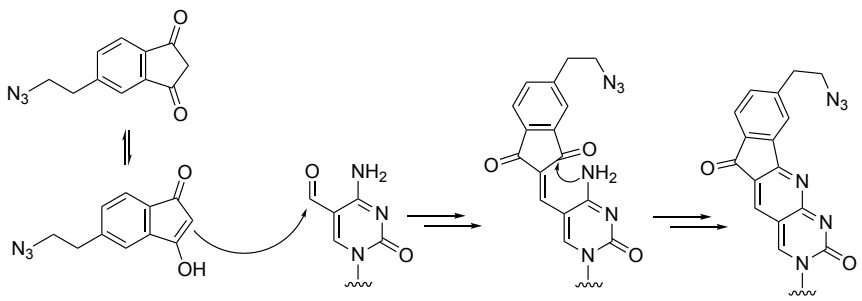


*Figure 6: Reaction of 5fC with a azide-modified derivative of indanedione used in fC-CET labeling.*

A notable disadvantage of this method is the requirement for multiple conversion and purification steps, which can result in significant degradation and loss of genomic DNA.[9] This poses a significant challenge to single-cell sequencing.[9] Consequently, genome-wide, base-resolution profiling of 5fC typically requires starting material from tens of thousands to millions of cells.[9]

In CLEVER-Seq[9] (chemical-labeling-enabled C-to-T conversion sequencing), malononitrile is used as labeling reagent. It causes the formation of the base 5fC-M, which is read as T, as shown in Figure 7.

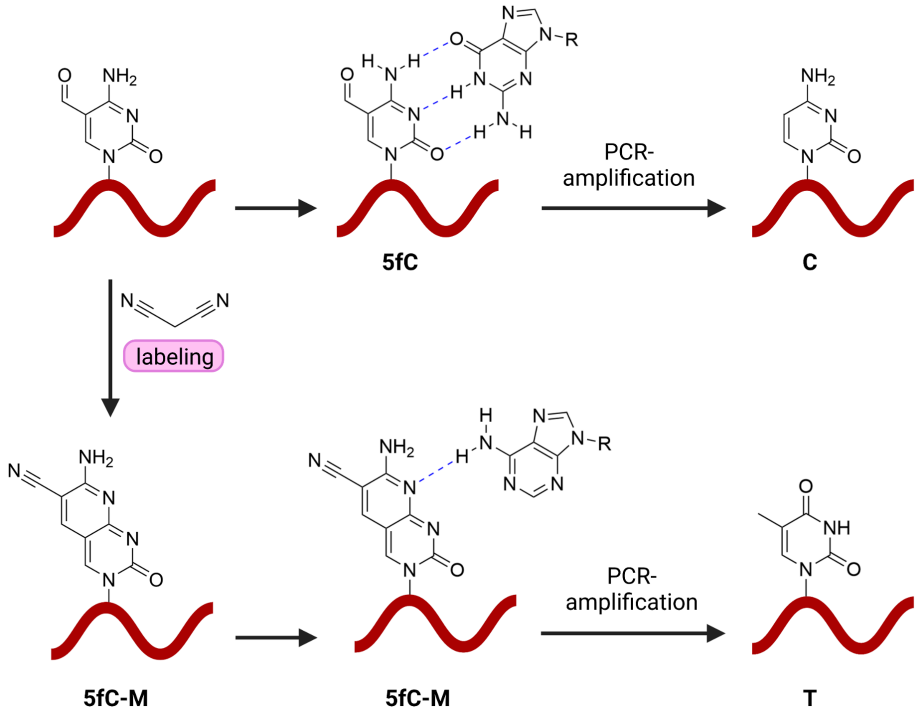



*Figure 7: Schematic representation of the behavior of 5fC during sequencing, on the top without labeling, on the bottom after malononitrile treatment.*

The reaction of malononitrile with 5fC is a Knoevenagel-type condensation, followed by an intramolecular reaction between the 6-amine and the nitrile moiety. It finally creates a base with three hydrogen bond acceptors, which is read as T by polymerases, as depicted in Figure 8.

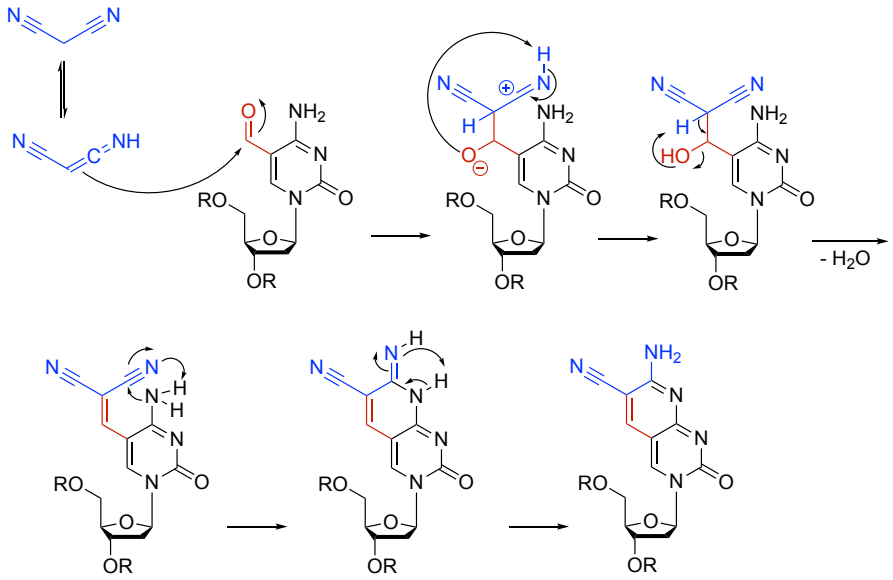


*Figure 8: Reaction mechanism of the labeling reaction of 5fC with malononitrile.*

After this chemical modification, polymerases read the modified 5fC as T/U and prefer to incorporate an A in the opposite strand.[44] This way, 5fC can be distinguished from the native cytosine. This approach also works for RNA labeling.[44] One issue is the weak interaction between 5fC-M and A with just one hydrogen bond, because 5fC-M is not protonated at physiological conditions.[75] Additionally, G can also be incorporated opposite to 5fC-M, because a possible wobble pair involving the exocyclic amine can be formed, as depicted in Figure 9.

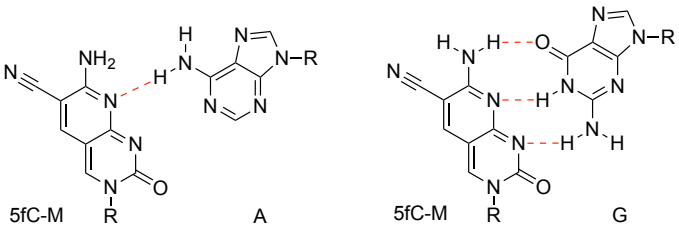



*Figure 9: Base pair of 5fC-M with A and wobble-pair of 5fC-M with G.*

Overall, the current methods help in understanding the important roles of naturally modified nucleobases, but are limited in terms of sensitivity and resolution.

### 3.10 Genetic alphabet expansion

The expansion of the genetic alphabet from the four-letter code to a larger alphabet results in an exponential increase in possible sequences, causing a substantial growth in information density. Some groups have already developed unnatural base pairs that could act as a third base pair in replication[76], transcription[77], and/or translation[78].

The group of Steven Benner described the incorporation of isoguanosine into an oligonucleotide opposite isocytidine in a DNA template with an appropriate polymerase[79] as well as other artificial nucleobases (2,6-diaminopyrimidine and 2,6-diaminopyridine)[80] more than 30 years ago. The group also described a new technology for incorporating non-canonical amino acids into peptides by ribosome-based translation.[81] They expanded the genetic code through the creation of a new codon-anticodon pair using nucleobases that have different hydrogen-bonding patterns.[81-82] With this technique, it was possible to generate peptides containing a non-proteinogenic amino acid via *in vitro* translation.[81] In 2019, DNA- and RNA-like systems built from eight nucleotide letters called “hachimoji” were introduced (Figure 10).

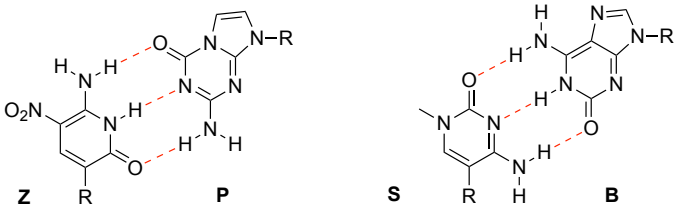



*Figure 10: Z-P and S-B base pairs in hachimoji.[77]*

This hachimoji DNA could also be transcribed to hachimoji RNA in the form of a functioning fluorescent aptamer.[77]

The group of Floyd E. Romesberg described a different approach to create new base pairs. Instead of inserting minor changes into the existing nucleic acids to build an alternating hydrogen bonding pattern, they are using hydrophobic and packing forces, including ring-stacking, to create non-polar base pairs.[83] Their new base pairs could be PCR-amplified and transcribed *in vitro*, and the mechanism of their replication has been characterized.[83] In 2017, the Romesberg group reported a semi-synthetic organism that encodes and retrieves increased information.[84]

The group of Ichiro Hirao created an unnatural base pair based on a purine and a pyridine analog, similar to the Benner group. They used 2-amino-6-(2-thienyl)purine and pyridin-2-one to expand the genetic code.[85] In addition to that, they created base pairs connected through an H-bonding-pattern which was used to create high-affinity DNA aptamers[86] that could be applicable to cancer cell imaging[87], biomarker

discovery[86], cancer cell profiling[87], anti-cancer therapies[87], and drug delivery systems.[76] DNA aptamers are short DNA molecules that have the ability to bind with high specificity to target molecules, making them potentially useful as diagnostic and therapeutic agents.[86] These aptamers are typically produced through *in vitro* selection techniques, such as systematic evolution of ligands by exponential enrichment (SELEX).[88] SELEX involves multiple rounds of selection and amplification from a diverse nucleic acid library, while each round applies greater selection pressure to enrich for aptamers with the desired characteristics.[76] However, even after extensive selection and amplification, it is often challenging to isolate nucleic acids that possess both high affinity and specificity.[86] To address this, it has been suggested that introducing nonstandard nucleotides could enhance the functional capabilities of nucleic acids.[86] One improved method for generating aptamers has significantly enhanced the affinity and stability of DNA aptamers by combining genetic alphabet expansion with mini-hairpin DNA technology.[86] The additional, unnatural bases can enhance the chemical and structural diversity of DNA aptamers. [86, 89] A mini-hairpin stabilizes DNA aptamers by protection against nuclease degradation and thermal denaturation.[86] A simplified illustration of this approach is depicted in Figure 11.

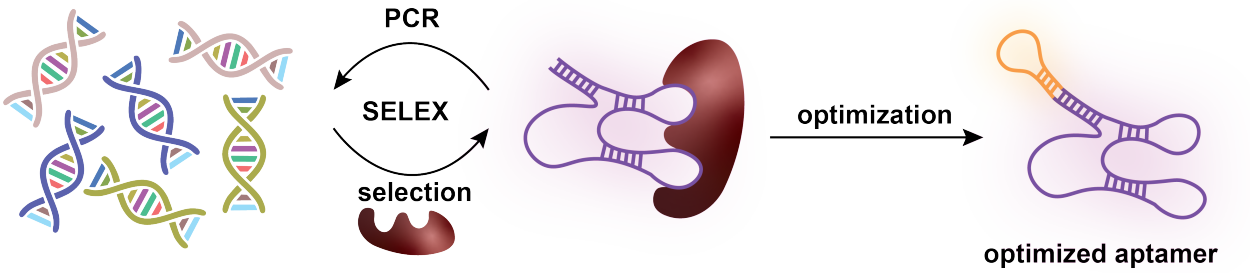



*Figure 11: Aptamer generation via SELEX, followed by optimization. Inspired by Ichiro Hirao.[89]*

With the incorporation of additional modified building blocks, the binding affinity can be over 100-fold improved compared to aptamers containing only natural bases.[86] The resulting stable, high-affinity DNA aptamers could be used for diagnostic and therapeutic applications.[86]

An additional potential application for an extended genetic alphabet is DNA data storage.[90] Molecular data storage has emerged as a promising solution for high-density and long-lasting information storage, a matter of increasing importance as the gap between data generation and current storage capabilities widens.[90] DNA already serves as an efficient medium for molecular archival data storage.[90] When adding unnatural base pairs, the information density increases even further. A potential workflow for DNA data storage is depicted in Figure 12.

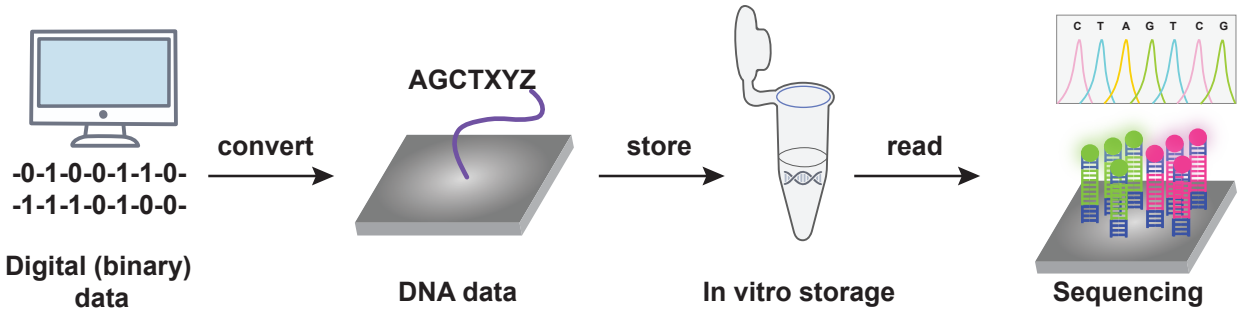



*Figure 12: Simplified workflow of DNA data storage, inspired by Luis Ceze.*[90]

Digital data, usually binary code, is converted into DNA sequences by a computer algorithm.[90] These sequences are subsequently synthesized by machines, resulting in multiple physical copies of each DNA strand.[90] Solid-phase oligonucleotide synthesis can be carried out on columns or on array-based solid supports, which allow for higher throughput.[90] The DNA can be stored *in vitro*, either frozen in solution or dried, or the DNA can be cloned and preserved within living cells.[90] When specific data is required, it can be selectively retrieved from the pool of DNA sequences through random access. This can be achieved via PCR-based amplification using primer pairs that correspond to targeted data segments created during the encoding step.[90] Then, automated sequencing technologies can read the detectable DNA. Common sequencing methods include Sanger sequencing, which exhibits a low throughput capacity[91], and sequencing-by-synthesis platforms such as Illumina[51], which demonstrates a higher throughput capacity. In recent developments, nanopore sequencing (e.g., those provided by Oxford Nanopore Technologies, ONT) has enabled real-time data reading.[25]

### 3.11 DNA nanotechnology and molecular machines

DNA nanotechnology harnesses the unique properties of DNA to build precise, nanoscale structures with programmable functions.[92] In nature, the cellular machinery is essential for the functioning of a variety of biological processes, from ATP synthase at the nanoscale to cell recognition at the microscale.[93] Artificial molecular machines have found wide-ranging applications, from basic research to biomedicine.[93] An early publication of artificial molecular machines in the form of molecular shuttles occurred in 1991.[94] In their work, a system in which a cationic bead moves back and forth like a shuttle between two identical hydroquinol stations was presented.[94] J. Fraser Stoddart, co-author of this study, shared the Nobel Prize in Chemistry together with Ben Feringa and Jean-Pierre Sauvage in 2016 for the design and synthesis of molecular machines.[95] Recent advances in harnessing unique physical and chemical properties of DNA have led to the development of DNA-based artificial molecular machines.[93]

While many polymers can be used to create molecular machines, DNA has emerged as a versatile and programmable building block for the development of complex materials at the nanoscale.[96] The

programmable nature of oligonucleotides in combination with high-yielding solid-phase synthesis enables the precise design of structures and allowing for control over their shape, size, and functionality.[96] The development of DNA origami, a platform that was first introduced in 2006[97], has been a significant advancement in the field of nanotechnology due to its programmable nature and precise nanoscale structure.[97-98] This platform has gained widespread adoption in diverse scientific disciplines, including chemistry, physics, and biology, due to its versatility and potential for advanced applications.[97-98] Their predictable mechanical properties, including elasticity and stability, enable their response to external forces to be calculated.[98] While it is now also possible to use deep learning-based methods to predict DNA shape features[99], recent advances have led to the development of sophisticated DNA-based artificial molecular machines.[93]

To construct a molecular machine that can achieve directional motion, it is necessary to overcome random thermal forces, which is impossible in a thermodynamic equilibrium without violating physical laws.[100] However, when out of equilibrium, directional movement can be achieved through Brownian ratchet mechanisms, which disrupt inversion symmetry.[100] This fundamental principle underlies the functionality of both biological motors and synthetic systems.[101] Hendrik Dietz's group built upon advances in DNA nanotechnology to develop a nanoscale rotary motor from DNA origami.[101] This motor is driven by a ratcheting mechanism and exhibits mechanical capabilities similar to those of biological motors, such as the $F_1F_0$-ATPase.[101-102] The concept is shown in Figure 13.

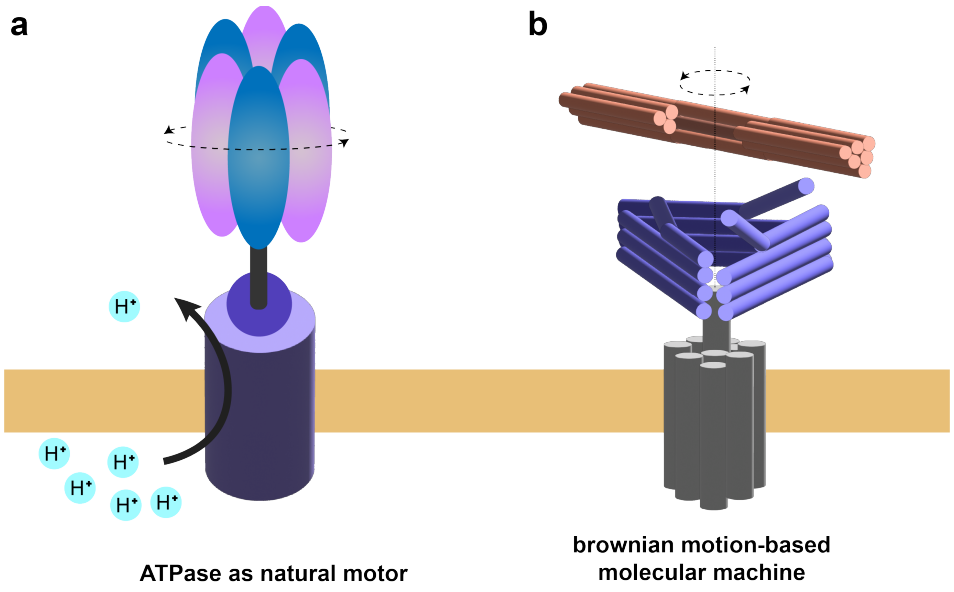



*Figure 13: Schematic representation of an ATPase and the DNA origami rotary ratchet motor[101].*

The construction is fixed through biotin-neutravidin anchors / linkers, and a non-rotating electric alternating current is applied to produce a kinetic asymmetry, which causes the rotation of the motor.[101]

### 3.12 pH-dependence of DNA structures

Around neutral pH, nucleic acid molecules are negatively charged due to the phosphate groups in their backbone, making them polyanions.[96] The natural base components of the nucleic acids remain electrically neutral and allow the correct formation of Watson-Crick base pairs.[103] However, when exposed to basic conditions, G, T, and U, can be deprotonated and therefore gain a negative charge.[103] Under acidic conditions, DNA can also be protonated at different positions. When single-stranded DNA is progressively acidified, C residues are protonated first, followed by A and G.[104] The preferred protonation sites in single-stranded DNA are the same as those found in isolated nucleosides: the N3 position of C, the N1 position of A, and the N7 position of G.[104] However, studies on double-stranded DNA show a different pattern: protonation begins at the N7 position of G and either the N7 or N3 position of A residues, while protonation of C residues at the N3 position only occurs after the DNA is unwound.[104] This change in charge significantly impacts not only the hydrogen bonding properties of the nucleic acids[105], but also their interaction with ions.[106]

For oligonucleotides, a specific relationship was observed between the p$K_a$ values of the bases and the pH of the aqueous solution.[107] Near neutral pH, and when the difference between a group's p$K_a$ value and the surrounding pH is less than 2, the group is likely to be involved in catalytic functions.[107] When the difference is greater than 2, the group is more likely to contribute to the formation and stabilization of the biomolecule's structure.[107] But there are exceptions to this general trend. Certain modified nucleobases, such as 7-deaza guanine (p$K_a \approx 10.3$) and 3-deaza guanine (p$K_a \approx 12.3$), have p$K_a$ values above 10 and exhibit decreased duplex stability, despite having a p$K_a$-pH > 2.[108] 5-aza- and 6-aza cytosine, which have p$K_a$ values of 2.6 and 2.8, respectively, also show decreased duplex stability.[109] This suggests that when the p$K_a$ of a nucleobase is either below 3 or above 10, its ability to form stable base pairs with its complementary partner is compromised.[107] This may be due to an increase in the hydrophilic character of the heterocycles, which is supported by the decreased lipophilic behavior of these modified nucleosides compared to their canonical counterparts.[107]

The p$K_a$ of a base can be modified through various chemical alterations, such as the introduction of electron-donating or electron-withdrawing groups. But to avoid perturbing the structure and other interactions, such as π-π interactions, additional groups should not be too bulky. An elegant way to increase the p$K_a$ in purines is to increase the electron density of the aromatic system by removing nitrogen. Replacing N3 or N7 in the purine ring with CH causes the N1 to be protonated already under mildly acidic conditions.[107] While adenine has a p$K_a$ of 3.7, 7-Deaza adenine has 5.3, and 3,7-Dideaza adenine has 8.3[107], so that the latter is protonated at physiological pH.

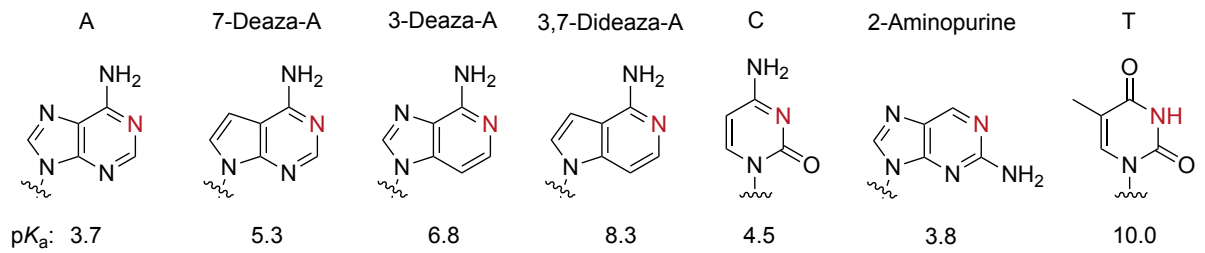



*Figure 14*: *$pK_a$ values*[107] *of canonical nucleobases and close structural analogues showing higher basicity in deaza- and dideaza-derivatives.*

When using 7-deazapurine nucleosides with amino groups in the 6-position, 2-position, or both positions, each of them could form a stable base pair with thymine.[110] Depending on the pH, 7-deaza-2′-deoxyadenosine and its 2-amino derivative can form stable base pairs not only with thymidine (dT), but also with cytidine (dC).[110] Using such deaza-derivatives of nucleobases can therefore help to design pH-responsive molecular switches.[110] Additionally, they could be used in an extended genetic alphabet and for aptamer generation.[86]

### 3.13 DNA nanotechnology for biomedical applications

DNA nanotechnology offers significant value in biomedicine.[92, 111] Over the past decade, advancements in this field have enabled the design of biocompatible DNA nanostructures for targeted drug delivery[92, 111], biosensing[112], and diagnostics[113]. DNA nanostructures have notably been used to improve the delivery of anticancer drugs, reducing toxicity and enhancing cellular uptake.[92, 114] Additionally, the synthesis of polymer-stabilized DNA fibers allows for the control of their lifetimes and the exploration of their potential applications in delivery systems and smart materials.[111] This offers significant potential for targeted, precise and biocompatible delivery systems.[111]

One example is the use of pH-responsive multifunctional DNA nano-micelles (DNMs) as delivery vehicles for a controlled release of the chemotherapeutic drug doxorubicin and anaplastic lymphoma kinase (ALK)-specific siRNA.[114] These pH-responsive DNA nano-micelles have been engineered using pH-sensitive triplex DNA structures.[92] The DNMs remain stable and monodispersed at various pH levels; however, under acidic conditions characteristic of the tumor microenvironment, the formation of C-G·$C^+$ triplexes triggers the release of therapeutic agents.[92] This controlled release mechanism allows DNMs to deliver the chemotherapeutic drug doxorubicin and specific small interfering RNA (siRNA) effectively, enhancing treatment efficacy against cancer, specifically against anaplastic large cell lymphoma.[92]

One example of an application of a DNA nanomachine as a diagnostic tool is the I-Switch, which acts as a pH sensor inside living cells using fluorescence resonance energy transfer (FRET).[113] The I-switch closes at a low pH and efficiently detects pH changes between 5.5 and 6.8.[113] Its ability to successfully track pH

during endosome maturation demonstrates the potential of DNA nanodevices for advanced sensing and diagnostics within living systems.[113]

# 4. Aim of the thesis

The investigation of modified nucleic acids enhances our knowledge of epigenetic mechanisms and could help to develop innovative diagnostic and therapeutic tools. In this dissertation, the research on modified DNA building blocks should lay the groundwork to enable biotechnological applications in the future. The research is centered on 5-formylcytosine, a natural DNA modification, and the two artificial nucleosides 2-Amino-DDP and 2,6-Diamino-DDP. The following figure presents a schematic overview of the three distinct projects.

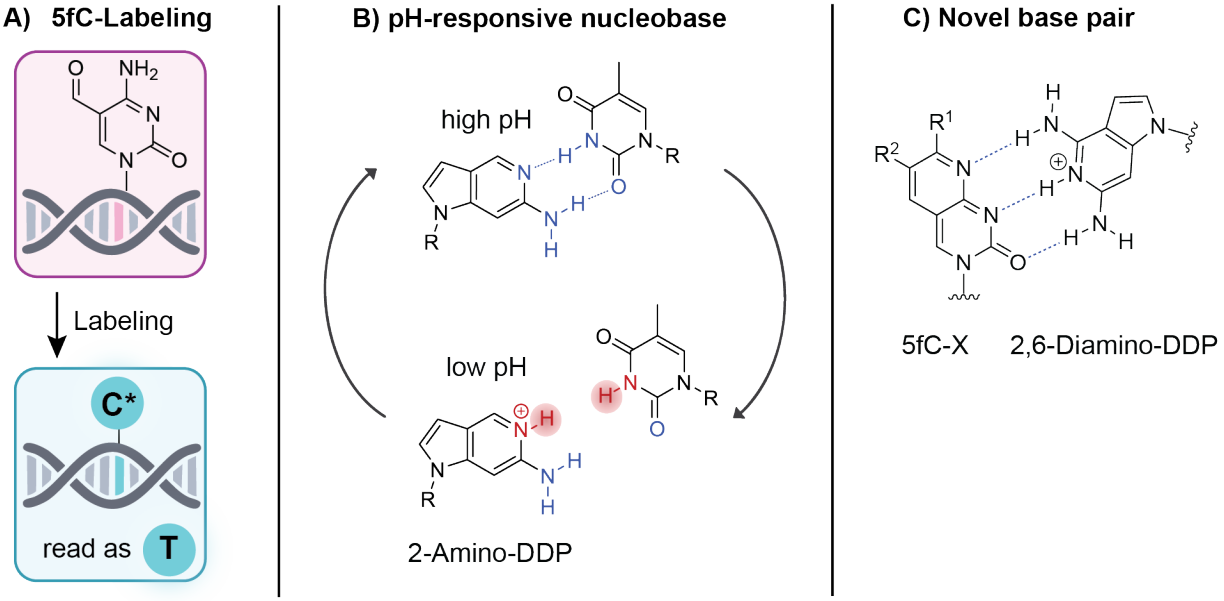



*Figure 15: Three research topics discussed in this dissertation. A) Development of a new labeling method for 5fC-detection. B) Synthesis of a pH-responsive nucleobase and its incorporation into DNA. C) Synthesis of a nucleobase with three H-bond-donors to create an artificial base pair with labeled 5fC-bases.*

In the first research project, a new method to detect 5-formylcytosine (5fC) should be developed to allow for single-base resolution epigenome sequencing. The resulting 5fC landscapes could enhance our understanding of the biological impact of 5fC. Conventional malononitrile-based CLEVER-Seq has been demonstrated to convert 5fC to a base read as T/U by polymerases, thereby differentiating it from cytosine.[9] However, the CLEVER-Seq method is limited in its conversion efficiency and can potentially lead to DNA degradation. In this work, a series of potential 5fC-labeling reagents is examined. The labeling should be selective for 5fC and compatible with DNA, without causing DNA degradation. The ultimate goal of this project is the discovery of a novel method for epigenetic research that could contribute to the understanding of 5fC's biological functions and its potential as a biomarker for disease-related processes. Moreover, novel diagnostic tools could detect changes in 5fC levels associated with diseases, including cancer, and monitor the effectiveness of treatments.

The objective of the second research project is to develop a programmable, pH-responsive nucleoside. This approach aims to mimic the dynamic behavior of biological molecular machines driven by proton gradients. The modified nucleobase 2-Amino-dideaza purine (2-Amino-DDP) should be synthesized as triphosphate (dNTP) and as phosphoramidite. The latter should be used to synthesize oligonucleotides that contain a base pair of 2-Amino-DDP with C or T, and their physicochemical properties should be determined. Melting point measurements are expected to reveal whether 2-Amino-DDP has the ability to cause a transition between paired and unpaired DNA states based on the pH. Consequently, this could enable DNA origami structures to respond dynamically to environmental changes. Furthermore, the pH-dependent incorporation of 2-Amino-DDP triphosphate by multiple DNA polymerases against C or T templates should be examined. It should be investigated whether the polymerases show a pH-dependent primer extension. Overall, the objective is to evaluate the potential of 2-Amino-DDP as new, pH-responsive building block for diverse applications, such as DNA-based molecular machines or nucleic acid enzymes.

The third research project is focused on the investigation of a new base pair with a three hydrogen-bond-donor base to expand the genetic alphabet. For this approach, the aim is to synthesize the unnatural nucleobase 2,6-Diamino-dideaza purine (2,6-Diamino-DDP), which was designed to have a low p$K_a$, enabling protonation at neutral pH. In its protonated form, it is hypothesized to form a novel base pair with labeled 5fC-bases, such as 5fC-M, as they have three hydrogen bond acceptors. 2,6-Diamino-DDP should be synthesized as triphosphate and as phosphoramidite, to test both the enzymatic and chemical incorporation into DNA strands. Ultimately, this approach could have potential applications in DNA data storage, sequencing, aptamer development, and PCR diagnostics.

# 5. New labeling techniques for 5-Formylcytosine

As a naturally occurring base modification, 5fC plays a crucial role in multiple cellular mechanisms, including epigenetic regulation[44] and gene expression[42], as explained in chapter 3.7. One common method to detect 5fC is CLEVER-Sequencing, which is described in detail in chapter 3.9.4. It contains a Knoevenagel condensation of 5fC with malononitrile followed by an intramolecular ring closure reaction between the amine and the nitrile. The newly created base, 5fC-M (malononitrile-modified 5fC) is read as T/U by polymerases and can be distinguished from the native cytosine.[44, 75] This approach works for DNA and RNA labeling, but is limited in terms of sensitivity and resolution, with conversion rates under 90% and observable DNA degradation.[44]

The primary problem encountered in 5fC sequencing is the incorporation of G opposite labeled 5fC bases, leading to the loss of distinction between C and 5fC. Undesired G incorporation can be caused by tautomerization or protonation of 5fC-M, which results in a changed hydrogen bonding pattern, or by the formation of a wobble pair including the exocyclic amine of 5fC-M. The different base pairs that can possibly form with 5fC-M are depicted below.

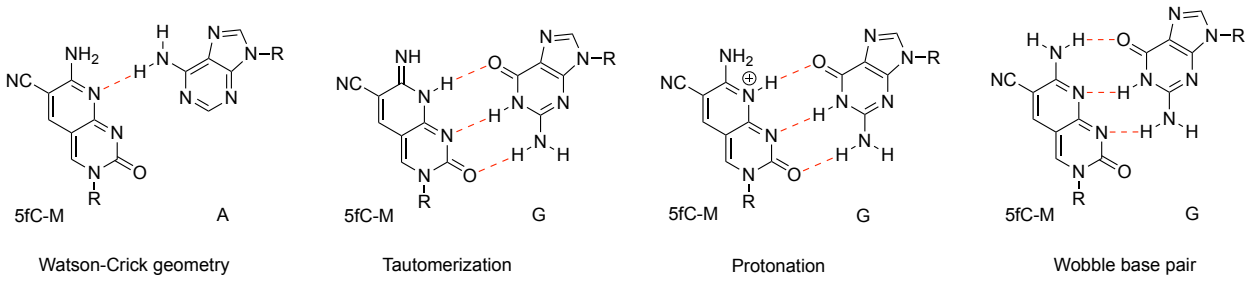



*Figure 16: Different possibilities for base pairs of 5fC-M with A or G caused by tautomerization, protonation, or by the formation of a wobble pair.*

To avoid the formation of a base pair with G, new labeling strategies should therefore reduce or avoid tautomerization and protonation of 5fC-M as well as wobble base pair formation. In this chapter, different labeling strategies are discussed with the goal to improve 5fC labeling and overcome current limitations. In the future, this could improve the understanding of 5fC's biological roles. The creation of new diagnostic tools could improve early detection of diseases and treatment monitoring.

## 5.1 Labeling of 5fC with Nitroacetonitrile (NAN)

For an improved 5fC labeling strategy compared to CLEVER-sequencing, a variety of reagents were tested. To achieve a Knoevenagel condensation, electron-withdrawing groups next to a C-H-acidic position are needed. A stronger electron withdrawing group would lower the p$K_a$ of labeled 5fC and prevent potential

mispairing with G caused by protonation. Additionally, a high solubility in aqueous buffers is advantageous to allow using high concentrations of the reagent. In a first screening, different nitrile-containing small molecules were used to react with 5fC-nucleoside as well as a 10-mer oligonucleotide containing one 5fC (Figure 17).

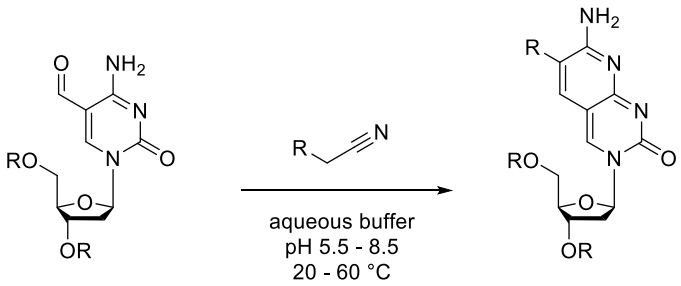



*Figure 17: Setup for testing different nitrile-based 5fC-labeling reagents.*

Malononitrile led to good labeling yields, confirming results of other groups[9, 43-44], but also hydrolyzed partially to an amide. Fluoroacetonitrile, *p*-Nitro-phenylacetonitrile and pentafluorophenylacetonitrile were not soluble in aqueous buffer, giving a suspension, and achieved no observable product formation.

A promising candidate is nitroacetonitrile, given its enhanced reactivity compared to malononitrile while being highly soluble in water. The strong electron-withdrawing nitro group is expected to cause a lower p$K_a$ in the labeled 5fC and therefore prevent potential mispairing with G caused by protonation. Consequently, nitroacetonitrile labeling is a potentially superior alternative to the currently employed malononitrile labeling approach. Pure nitroacetonitrile is known to be unstable and explosive.[115] Therefore, a more stable form was synthesized, the potassium salt of nitroacetonitrile (NAN). The synthesis was performed as described in the literature[115] and is shown in Figure 18.

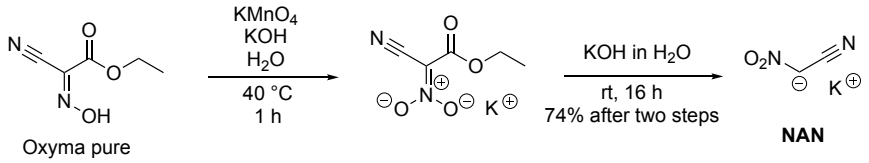



*Figure 18: Synthesis of NAN in two steps.*

Ethyl cyanohydroxyiminoacetate, commercially available as *Oxyma Pure*, is a reagent for peptide synthesis. In this case, the substance undergoes oxidation with potassium permanganate, resulting in the formation of a nitro group. Treatment with potassium hydroxide leads to the hydrolysis of the ester, resulting in decarboxylation and the formation of the potassium salt of nitroacetonitrile (**NAN**).

One disadvantage of this method is the difficult purification of NAN after the synthesis. Especially the separation from the excess of potassium hydroxide is challenging, because KOH and NAN have a similar solubility, as both are highly soluble in water and methanol. Therefore, a derivative that contains an additional carboxyl was synthesized by using the strategy shown in Figure 19. The carboxy-analogue of potassium nitroacetonitrile (**C-NAN**), could be produced with a high degree of purity due to an optimized synthesis route with a stable precursor[116] based on synthetic procedures described in the literature.[116-117] C-NAN has also been described to be used as a synthetic equivalent to NAN.[118]

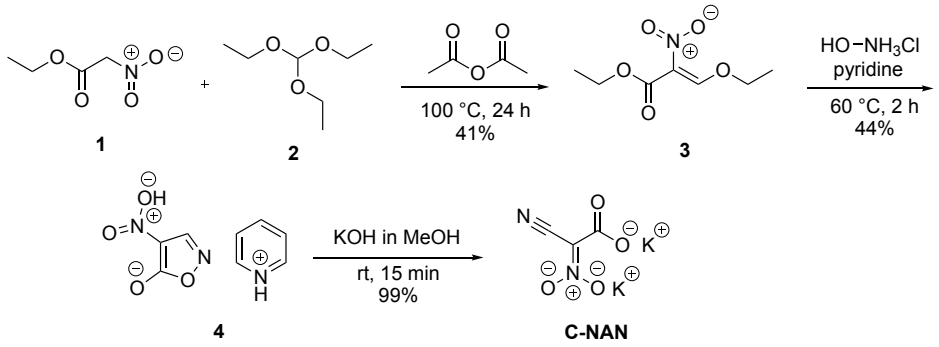



*Figure 19: Synthetic route for C-NAN in three steps.*

First, commercially available reagents **1** and **2** are heated with acetic anhydride to get **3**, which is treated with hydroxylamine in pyridine to form the nitroisoxazolone pyridinium salt **4**. This precursor demonstrates a higher level of stability in comparison to NAN and, when frozen, can be stored for years. Recrystallization from ethanol is an effective purification method that results in the formation of well-formed crystals that are of a remarkably high degree of purity, as shown in the HMBC-NMR spectrum in Figure 20.

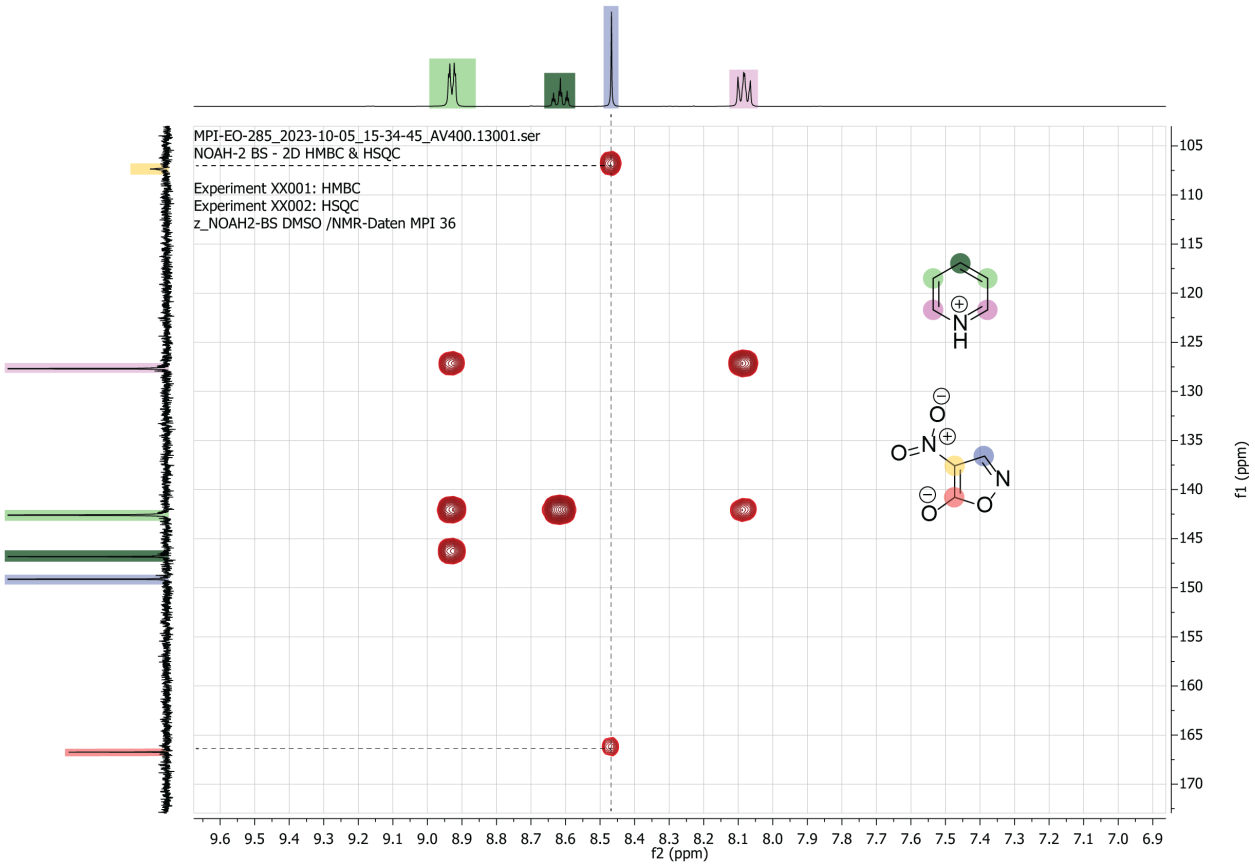



*Figure 20: $^{1}H$-$^{13}C$-HMBC-NMR-spectrum from the nitroisoxazolone pyridinium salt **4**.*

In the final step of the synthesis, the nitroisoxazolone pyridinium salt is added to a methanolic potassium hydroxide solution. Subsequently, the precipitation of white crystals of C-NAN can be observed immediately. Since the other components remain in the methanolic solution, the pure product can be isolated by simple filtration. Consequently, the synthesis of C-NAN is notably more efficient than that of NAN.

The reaction mechanism of the labeling of 5fC with NAN is shown in Figure 21. C-NAN undergoes spontaneous decarboxylation in solution[118], so the reaction of 5fC with C-NAN follows the same mechanism.

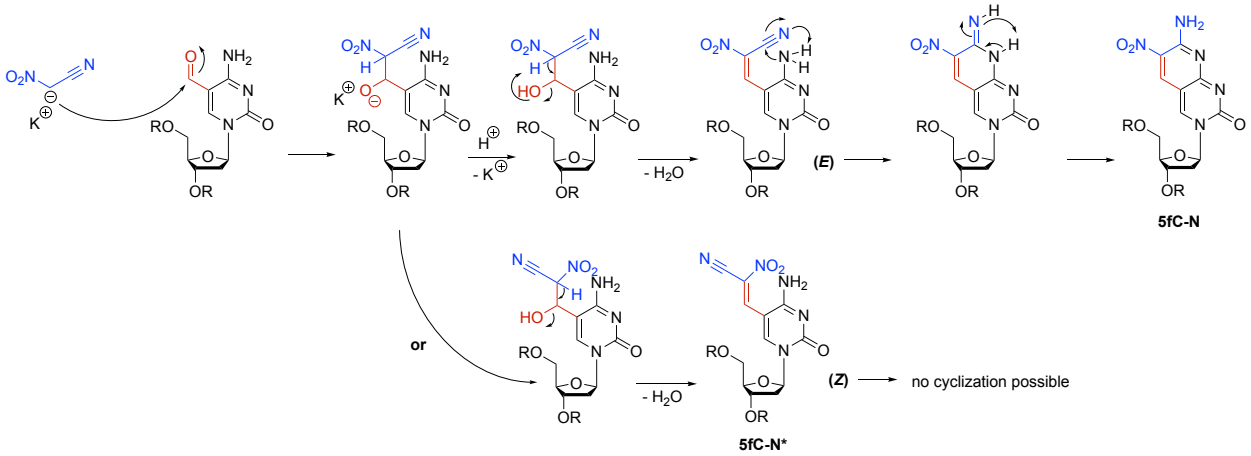



*Figure 21: Reaction mechanism of the labeling of 5fC with NAN.*

During the reaction with 5fC, the carbonyl group is attacked by the nucleophilic labeling reagent in a Knoevenagel-type reaction. After the release of a water molecule by condensation, there are two possible conformations. In the (*E*) conformation, the exocyclic amine is facing in the direction of the nitrile, so it can attack the nitrile in a nucleophilic reaction, as shown in the upper row of Figure 21. This reaction leads to an intramolecular 6-*exo*-dig cyclization to a compound that finally forms a bigger aromatic system, 5fC-N.

In contrast to the reaction with the symmetrical molecule malononitrile, a different conformer can be formed when using NAN or C-NAN. In the (*Z*) conformer, the nitrile points away from the exocyclic amine while the nitro group is adjacent to it. The double bond prevents free rotation and the lone pair of the exocyclic amine cannot overlap with the $\pi^*$ orbital of the carbon of the nitrile. This prevents cyclization, and the hydrogen-bond pattern of the 5fC-N* molecule remains unchanged. Therefore, the formation of this conformer is undesired for the labeling of 5fC, since it does not lead to a C-to-T transition during sequencing. Since cyclized and uncyclized 5fC-N* have identical masses, they cannot be distinguished by HRMS.

During her master's thesis, Susanne Nguyen supported me in optimizing the conditions for the 5fC-labeling. All tested conditions (buffers, pH, concentration of the reagents and reaction temperature) can be found in her thesis. The optimal conditions for an oligonucleotide concentration of 10 µM (5fC-25mer) were found to be: 500 mM NaOAc at pH 5.0, 100 mM labeling agent, 40 °C, 16 h. Under these conditions, almost quantitative conversion (>99%) was possible, according to LC-MS measurements shown in Figure 22.

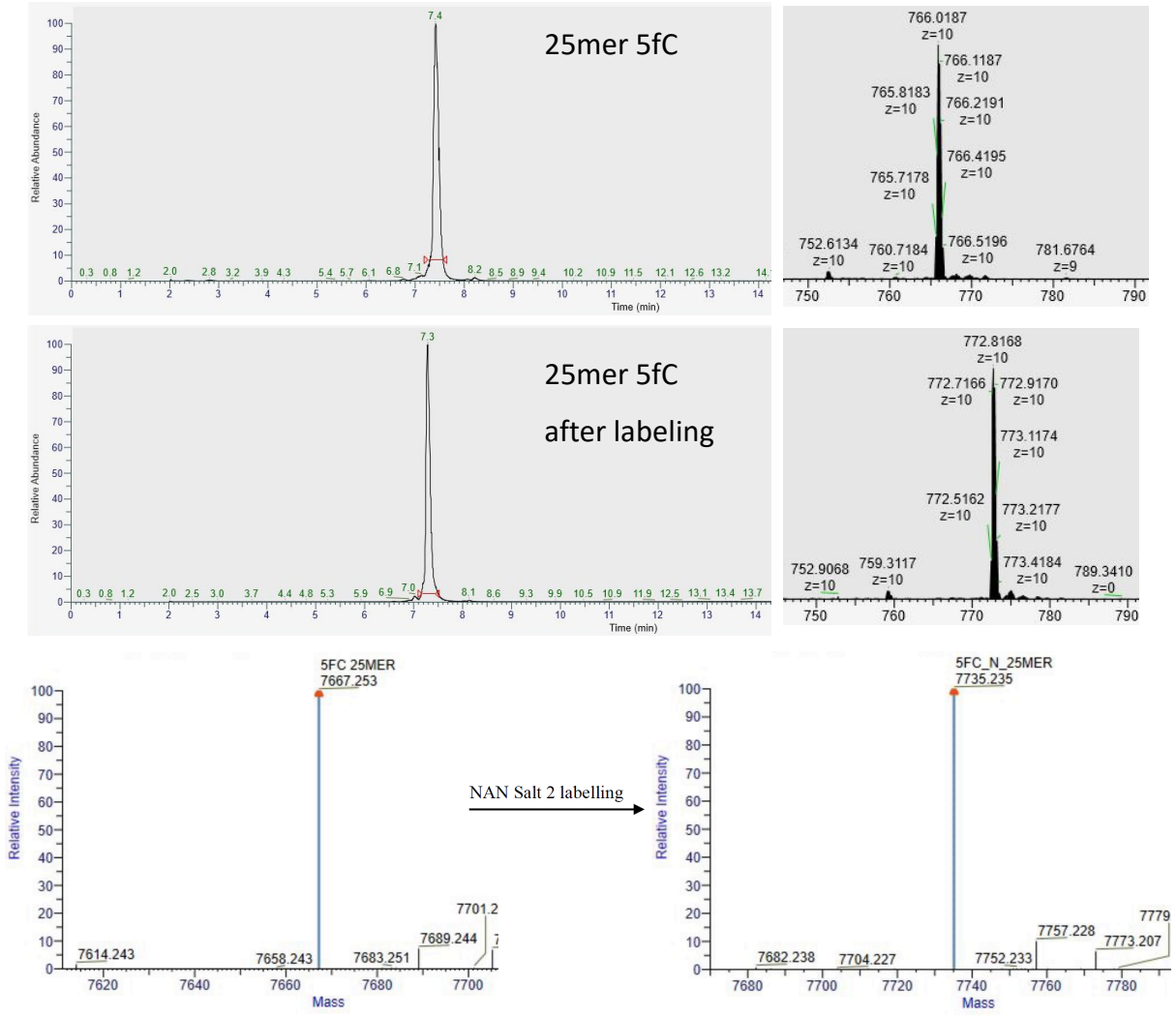



*Figure 22: Top and middle: LC-MS measurements of the 25mer 5fC oligonucleotide before (top) and after (middle) labeling with C-NAN. Bottom: Deconvoluted mass of the 25mer before (left) and after (right) the labeling reaction.*

After a successful condensation, the label should add 68.00 Da to the molecule's mass. The deconvoluted mass of the 25mer is 7735.235 Da after labeling, which is 67.982 Da higher than before labeling. Therefore, successful conversion of 5fC could be confirmed by HRMS. However, as mentioned above, the cyclization state is not visible due to the same mass of 5fC-N and 5fC-N*.

In order to determine whether the NAN-labelling also causes a C-to-T transition, Susanne Nguyen established a molecular cloning protocol. She could successfully label a 70-mer oligonucleotide containing one 5fC at position 36. Primers were designed not only to extend the oligonucleotide, but also to incorporate restriction enzyme recognition sites and a polymerase chain reaction (PCR) was performed to elongate the insert. The expected length of the PCR-amplified product was 127 base pairs. After amplification, the PCR product underwent restriction enzyme digestion and was then ligated into a plasmid vector. The vector was subsequently transformed into *E. coli* DH5α cells. Transformants were selected via antibiotic resistance on

agar plates. Plasmid DNA was extracted from the resulting colonies and submitted for Sanger sequencing to determine if the labeling process had induced a C-to-T conversion in position 456. A section of the sequencing results is shown in Figure 23.

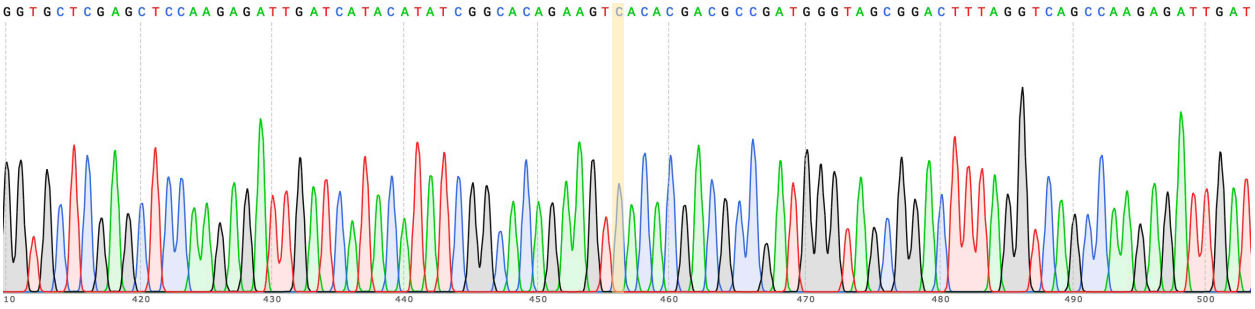



*Figure 23: Results of the Sanger-Sequencing of the recombinant plasmid containing labeled 5fC-N, showing a C in position 456.*

Sanger sequencing did not indicate successful C-to-T-transformation. A potential explanation could be an incomplete labeling or problems with the cyclization. If uncyclized 5fC-N* is present, polymerases recognize it as C, resulting in G incorporation. The polymerase may be incapable of reading over the base that contains a nitro group, leading to preferential amplification of unlabeled 5fC. Additionally, even in the cyclized form, a wobble pair may form between 5fC-N and G, as seen with 5fC-M, and depicted in Figure 24.

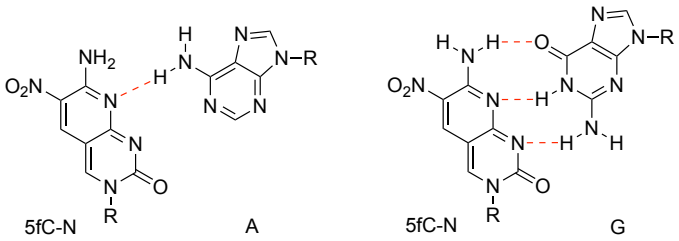



*Figure 24: Base pair of 5fC-N with A and wobble-pair of 5fC-N with G.*

The stability of the labeled product and the possibility of label detachment should be evaluated. To accomplish this objective, Anna Schöne, a master's student under my supervision, conducted a series of heat treatments on previously labeled 5fC-oligonucleotides. We could observe a majority of the labeled product detaching after the reaction, indicating that 5fC-N* was the major species present.

A possible mechanism for the label detachment is that water can nucleophilically attack the exocyclic double bond of 5fC-N*, resulting in its hydrolysis back to 5fC. Conversely, the aromatized heterocycle of 5fC-N is expected to exhibit significantly enhanced stability, thereby reducing its hydrolysis tendency, as explained in Figure 25.

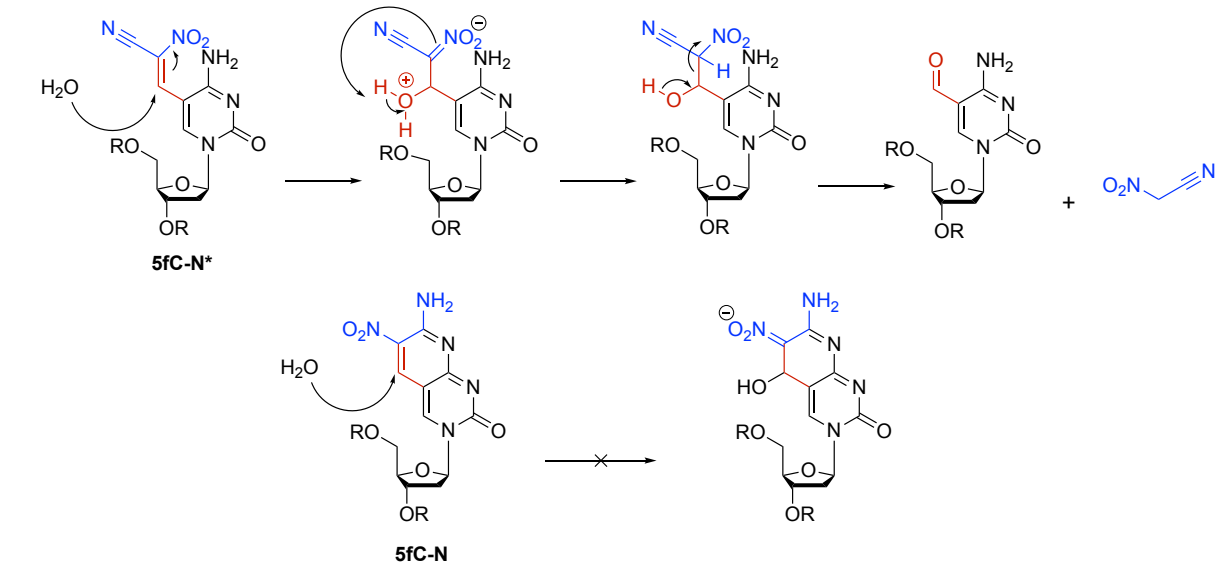



*Figure 25: Proposed reaction mechanism of the hydrolysis of uncyclized (Z) 5fC-N* while cyclized 5fC-N is less prone to reacting with nucleophiles.*

In the proposed mechanism, the lone electron pair of water can overlap with the π* molecular orbital of the C=C double bond in 5fC-N*, resulting in the formation of a new σ-bond and the breaking of the π-bond. In 5fC-N, the double bond is part of a larger aromatic system that is less prone to reacting with nucleophiles.

Anna Schöne and I tried to change the conformation from *Z* to *E* through UV light irradiation, based on a method described in 2022[119]. After 5 hours of irradiation with 350 nm UV light, 76% 5fC-N remained, while 14% lost the label and recovered unmodified 5fC. In addition, a species with the mass of a double-labeled adduct appeared. A possible reaction for its formation from hydrolyzed 5fC-N* is shown in Figure 26.

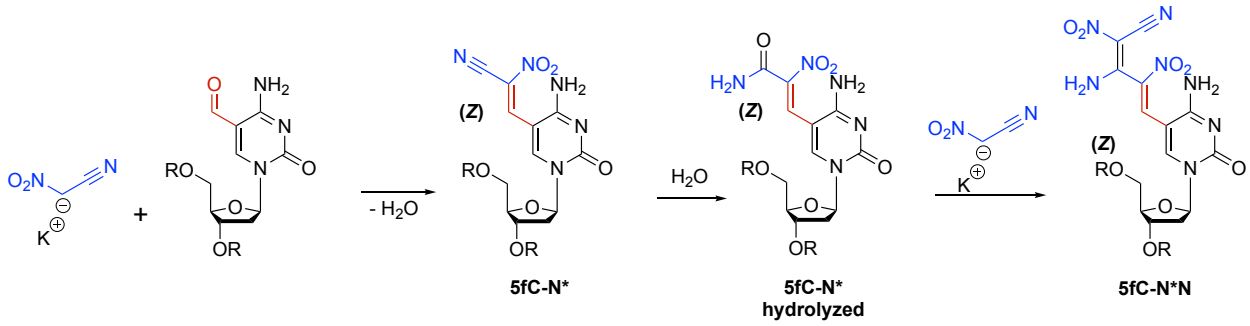



*Figure 26: Proposed reaction for the formation of the double-labeled adduct (5fC-N*N) from hydrolyzed 5fC-N*.*

Subsequent heat treatment led to a higher fraction of unmodified 5fC and hydrolyzed 5fC-N*. The LC-MS chromatograms are shown in Figure 27, and an overview of the species formed in each step is presented in Table 1.

*Table 1: Relative percentages of the species formed after labeling (10 µM 5fC-oligo T1, 100 mM NAN, 500 mM NaOAc buffer pH 5.0, 40 °C, 16 h), after UV irradiation (350 nm, 5 h), and after subsequent heating of the sample (80 °C, 2 h).*

| Species | After Labeling (%) | After UV irradiation (%) | After heat treatment (%) |
|---|---|---|---|
| 5fC | 1 | 14 | 30 |
| 5fC-N or 5fC-N* | 99 | 76 | 51 |
| 5fC-N* hydrolyzed | 0 | 0 | 10 |
| 5fC-N*N | 0 | 10 | 0 |

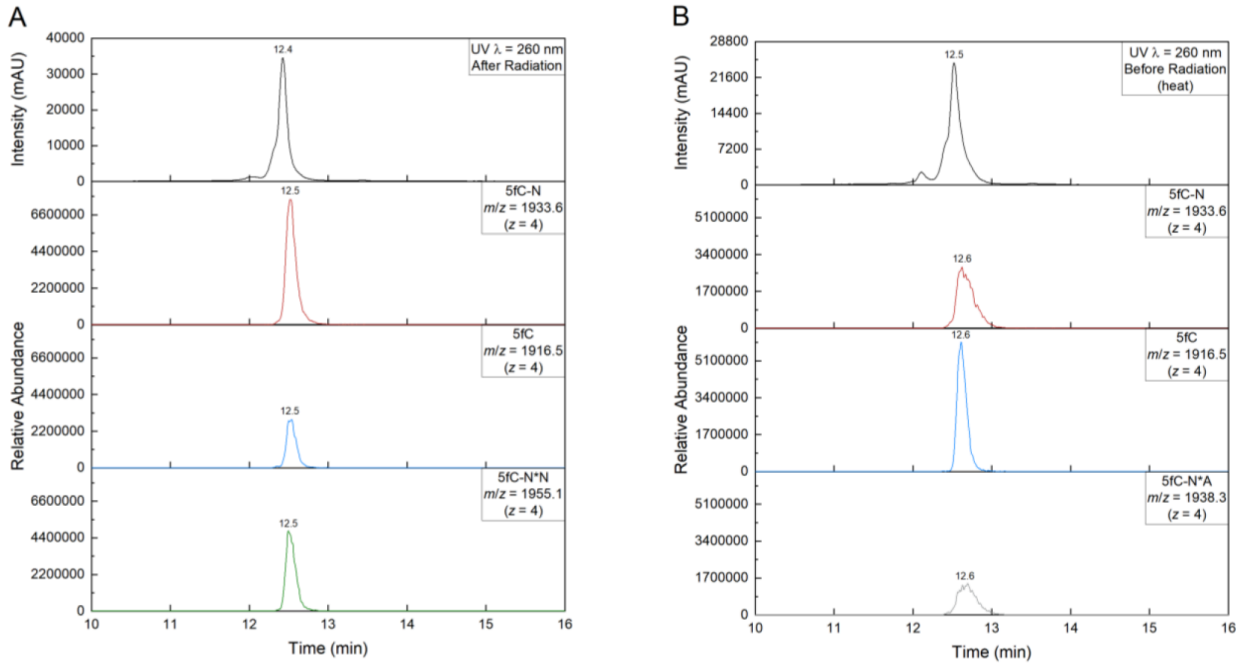



*Figure 27: LC-MS chromatograms of NAN-labeling (10 µM 5fC-oligo T1, 100 mM NAN, 500 mM NaOAc buffer pH 5.0 at 40 °C for 16 h) after irradiation at 350 nm for 5 h (A) and after heat treatment at 80 °C for 2 h (B). The reaction and measurements were performed with Anna Schöne as part of her master's thesis.[2]*

As no successful isomerization could be observed, we tested photocatalysts in the following step, as they have been shown to facilitate isomerization reactions, for example in alkenes[120]. Xanthone and riboflavin caused DNA degradation, while acetophenone caused label detachment, and fluorenone did not show any effect. The addition of benzophenone to the purified oligonucleotide caused partial DNA degradation. The remaining DNA consisted of 75% 5fC-N, 13% unmodified 5fC and 12% 5fC-N*N.[2] After heat treatment, 22% of 5fC-N remained, while 60% unmodified 5fC and 18% hydrolyzed 5fC-N were observed.[2] Heat treatment without previous UV treatment caused an increase in the formation of the double-labeled adduct (5fC-N*N) of up to 30%, along with DNA degradation. Further details are described in the master's thesis of Anna Schöne.[2]

Overall, no satisfactory isomerization could be achieved. Since UV irradiation can severely damage DNA itself[121] and DNA degradation was observed after heat treatment or UV irradiation, the NAN-labeling remains inefficient and an alternative strategy was necessary. Therefore, alternative labeling reagents were evaluated.

## 5.2 Labeling of 5fC with acetylacetone (Ace)

For the NAN-labeling, it was observed that a molecule that can react with 5fC in different conformations (*E* or *Z*) leads to complications with the following cyclization. Attempts to produce the pure *E* stereoisomer directly, but also by converting the *Z*- to the *E*-isomer, remained unsuccessful. Consequently, we focused on symmetrical molecules. A small, symmetrical, water-soluble molecule (analogous to malononitrile) has several advantages over asymmetrical labeling reagents, including the ability to avoid side reactions and undesired conformations.

As described in Figure 16, the main problem in 5fC labeling is the incorporation of G opposite labeled 5fC bases caused by tautomerization or protonation of 5fC-M or by the formation of a wobble pair. A labeling reagent that does not contain a nitrile will not lead to the formation of an exocyclic amine and therefore could avoid tautomerization or wobble pair formation. Acetylacetone is a promising candidate to label 5fC without causing an exocyclic amine, as shown in Figure 28.

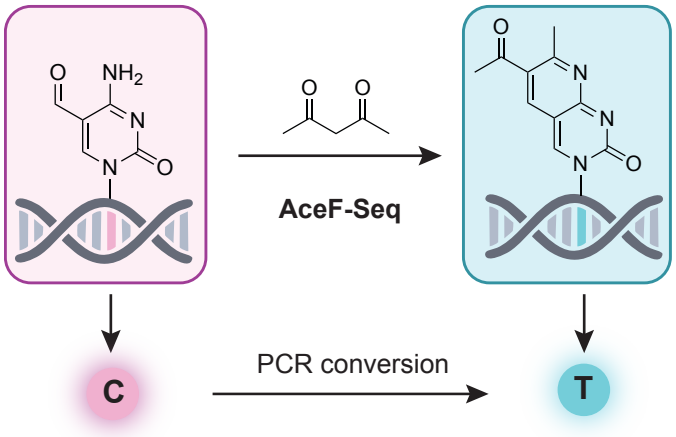



*Figure 28: Schematic workflow of acetylacetone-induced 5-formylcytosine-to-thymine conversion sequencing (AceF-Seq). We termed the acetylacetone-labeled 5fC shown on the right "5fC-A".*

The exocyclic methyl group does not offer a hydrogen bond donor and could avoid many possibilities of forming a base pair with G, as shown in Figure 29.

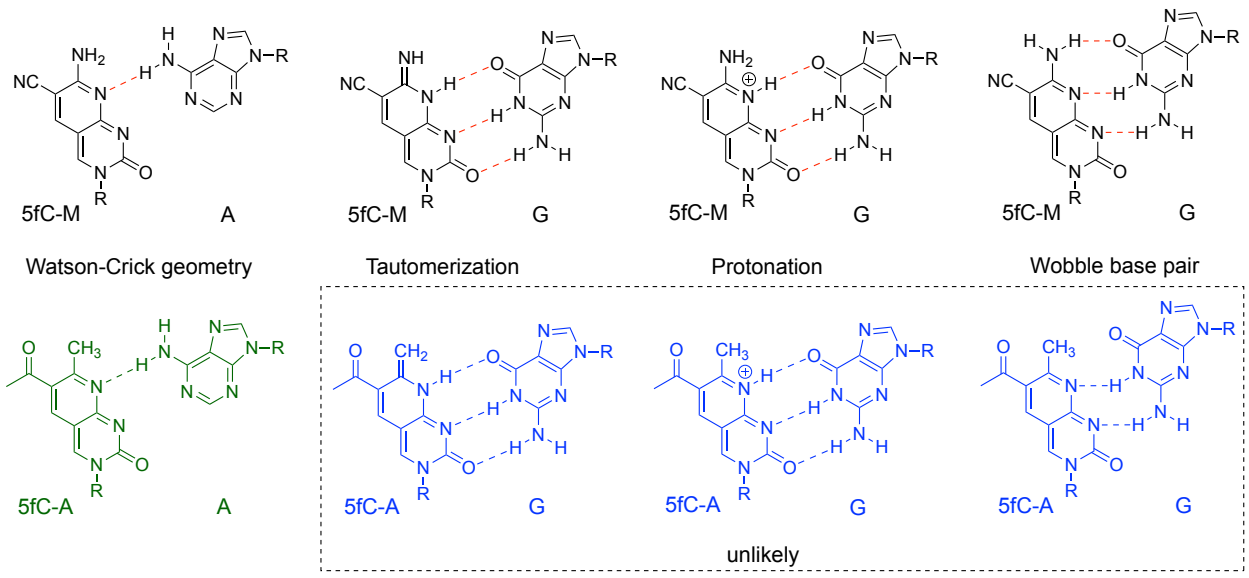



*Figure 29: Different possibilities for base pairs of 5fC-M and the acetylacetone-labeled 5fC-A with A or G caused by tautomerization, protonation, or by the formation of a wobble pair.*

Consequently, acetylacetone was examined for its potential to react with 5fC. The research group that published the fC-CET labeling method, which also relies on the Friedländer reaction, already suggested using acetylacetone for 5fC labeling.[42] However, they observed no detectable product formation using MALDI-TOF mass spectrometry.[42] Later, a reaction between 2-aminoacetophenone and acetylacetone, catalyzed by iron(III) chloride, was reported.[122] Due to the similarity between 2-aminoacetophenone and 2-aminobenzaldehyde, this type of reaction was identified as a promising candidate for 5fC labeling. Two possible mechanistic pathways have been proposed for this transformation.[123]

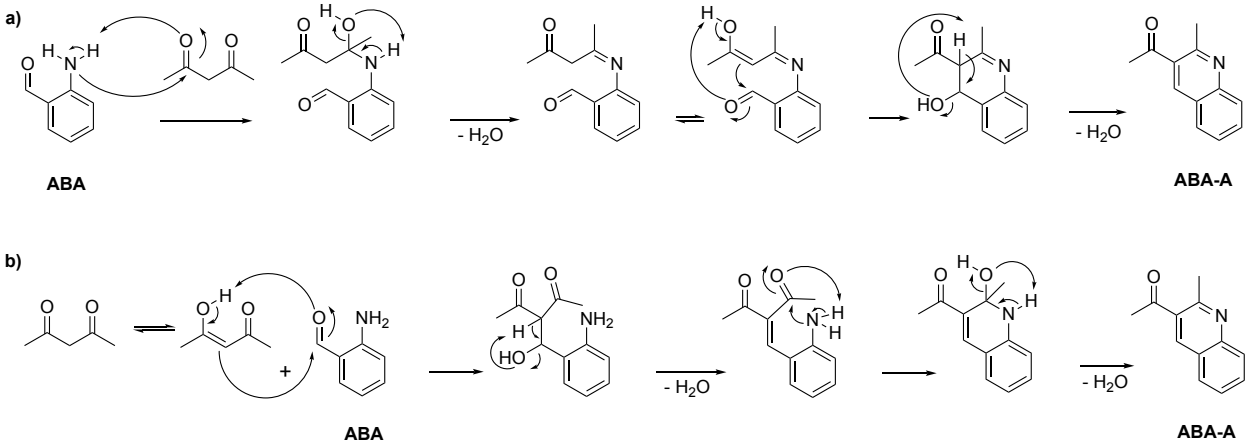



*Figure 30: Two possible reaction mechanisms of the reaction of 2-aminobenzaldehyde with acetylacetone.*

In the first proposed mechanism (Figure 30a), the amino group of 2-aminobenzaldehyde (ABA) initiates a nucleophilic attack on the carbonyl group of acetylacetone, resulting in the formation of an imine / Schiff base. After the first condensation, an intramolecular aldol condensation yields a cyclic intermediate, which subsequently undergoes an intramolecular condensation to give the final aromatic quinoline derivative ABA-A.

In the second pathway (Figure 30b), acetylacetone functions as the nucleophile, engaging in an attack on the aldehyde of 2-aminobenzaldehyde. Subsequently, an intermolecular condensation occurs. The intermediate then undergoes cyclization as a result of a nucleophilic attack by the exocyclic amine on the carbonyl group. This results in the formation of quinoline derivative ABA-A. While the proportion of reactions that proceed according to one mechanism or another remains unclear[123], both mechanisms result in the same cyclized product, ABA-A.

The following reactions and measurements discussed in this chapter (labeling of 5fC with acetylacetone) were performed together with Anna Schöne, as part of her master's thesis. She optimized the conditions for the labeling reaction of ABA to ultimately achieve full conversion (Figure 31).

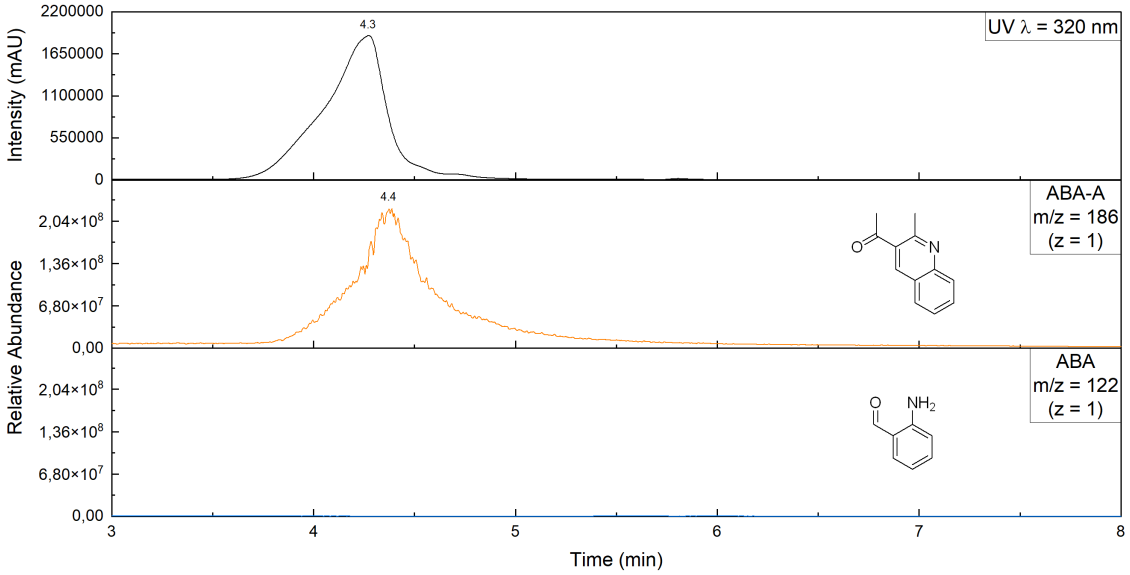



*Figure 31: LC/MS chromatograms of the labeling reaction of ABA (150 mM) with acetylacetone (1.5 M) and $FeCl_3$ (3.00 eq.) in water at room temperature for 16 h. The reaction and measurements were performed with Anna Schöne as part of her master's thesis.[2]*

## 5.3 Acetylacetone (Ace)-labeling on 5fC nucleoside

We performed the labeling of 5fC nucleoside under optimized conditions using 1 M acetylacetone in 500 mM NaOAc buffer at pH 5.5, with and without iron chloride. The UV chromatogram shown in Figure 32 indicated formation of a single main product.

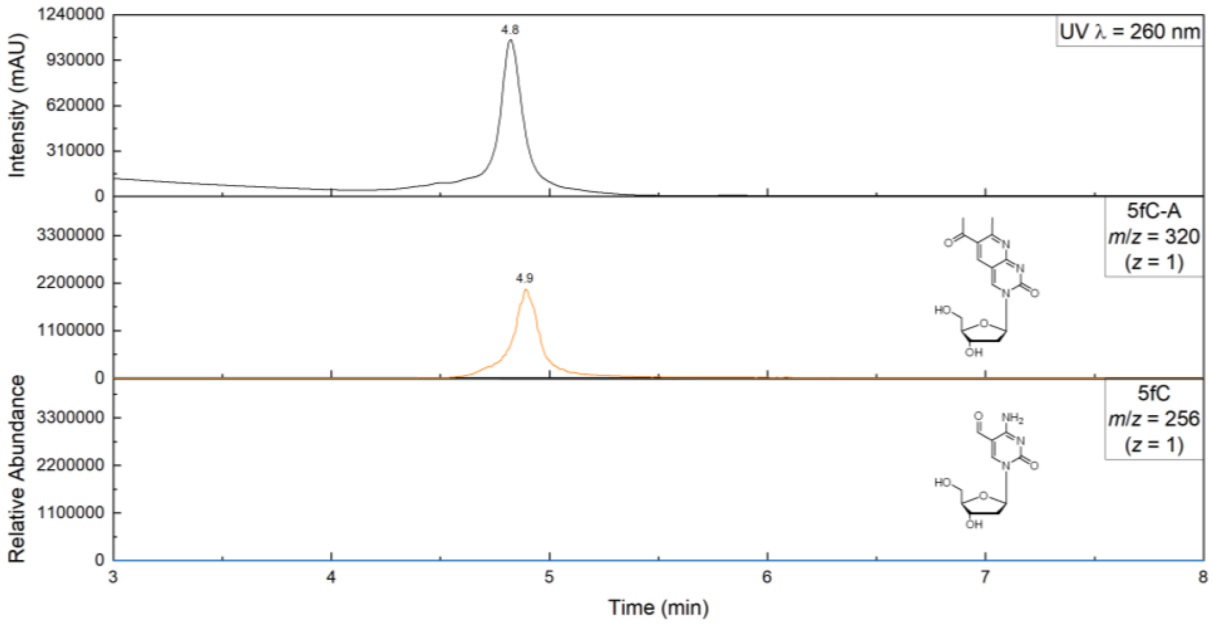



*Figure 32: LC/MS chromatograms of the reaction of 5fC-nucleoside (10 mM) with acetylacetone (1.0 M) in 500 mM NaOAc buffer at pH 5.5 at 60 °C for 16 h. The reaction and measurements were performed with Anna Schöne as part of her master's thesis.*[2]

Mass spectrometric analysis confirmed the presence of the labeled 5fC-A nucleoside with a mass signal of 320.03 (z = 1), which corresponds to the expected product.

### 5.4 Acetylacetone (Ace)-labeling on 5fC-containing ssDNA

In order to label 5fC-containing oligonucleotides, we tested various conditions. The buffer, reaction temperature, and concentrations of the individual components were optimized using a 25mer DNA-strand containing one 5fC. The sequence is AA**5**TGAAGACACGGCTATACATACT (with **5** = 5fC). Using the optimal conditions (10 µM oligo, 1.0 M acetylacetone, 1.0 mM $FeCl_3$, 500 mM NaOAc buffer at pH 5.5, 60 °C for 16 hours), over 99% of the 5fC was successfully labeled without causing detectable DNA degradation.

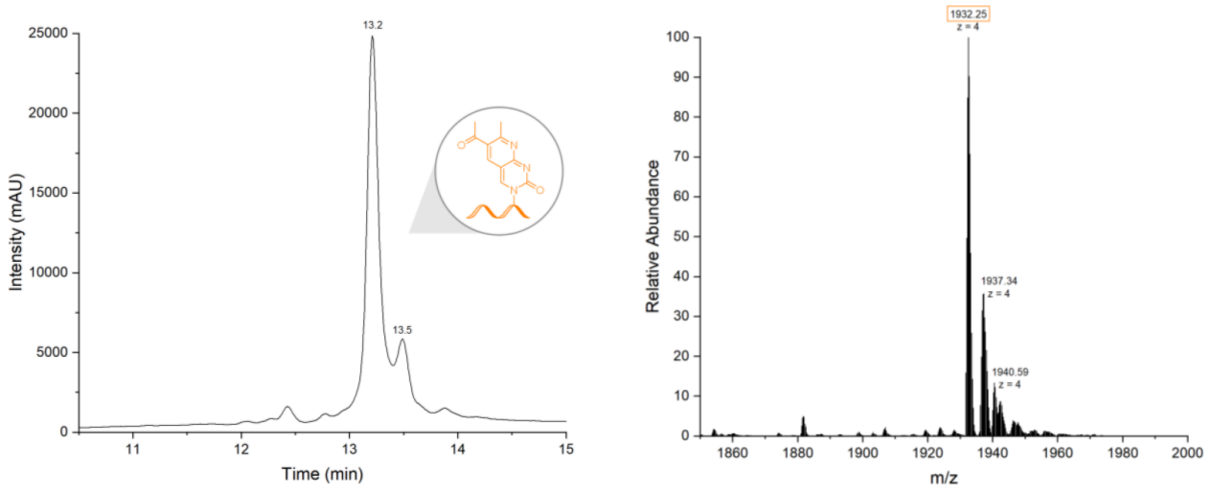



*Figure 33: LC/MS spectra of the reaction of a 5fC-containing 25-mer oligonucleotide (10 µM) with acetylacetone (1.0 M) and $FeCl_3$ (1 mM) 500 mM NaOAc buffer (pH 5.5) at 60 °C for 16 h. Monoisotopic masses: 5fC = 7667.34 Da (1916.57 Da at z = 4) and 5fC-A = 7731.34 Da (1932.25 Da at z = 4). The reaction and measurements were performed with Anna Schöne and this figure is part of her master's thesis.*[2]

Based on mass spectrometric data, acetylacetone-induced 5-formylcytosine-to-thymine conversion sequencing (AceF-Seq) can lead to full conversion, but to verify the expected C-to-T transition, additional experiments are needed. In order to directly compare the results with CLEVER-Seq, a 70-mer DNA oligonucleotide containing one 5fC was labeled with either acetylacetone or malononitrile.

Both labeling reactions achieved conversion rates of over 99%, determined by LC/MS. However, labeling with malononitrile resulted in partial degradation of the DNA. Despite testing various conditions, the optimal malononitrile labeling protocol still led to 16% DNA degradation, and degradation products could be observed (Figure 34, 0-3 min).[2]

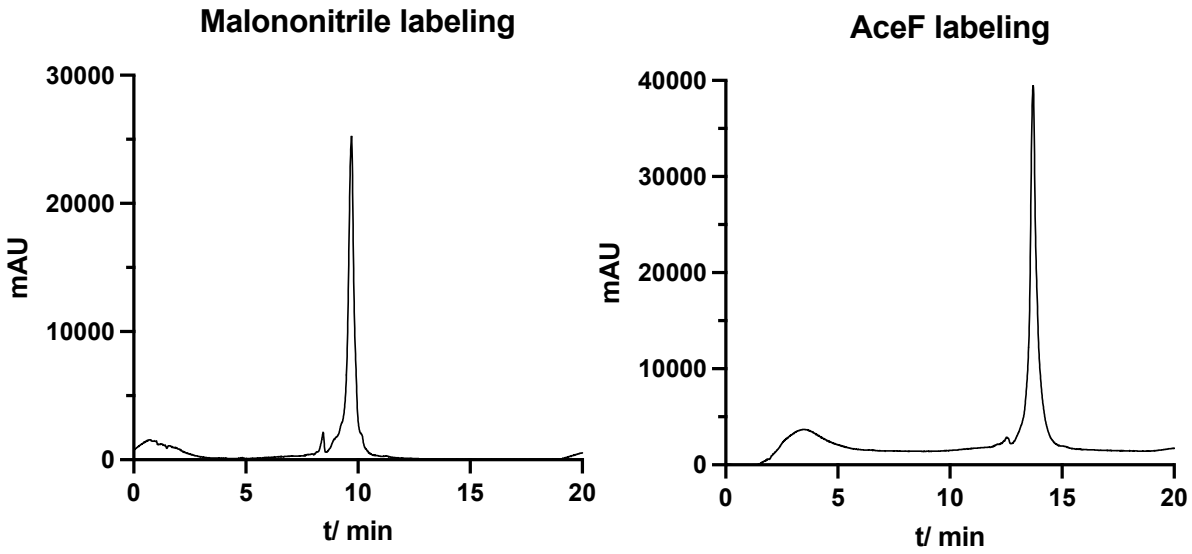



*Figure 34: LC-MS spectra of the labeling reaction of 5fC (70mer, 10 µM) with malononitrile (150 mM) in 200 mM $NH_4OAc$ buffer (pH 7.0) at 20 °C for 72 h and AceF labeling (1 M acetylacetone in 500 mM NaOAc buffer (pH 5.5) at 60 °C for 16 h). The reaction and measurements were performed with Anna Schöne as part of her master's thesis.[2]*

In contrast, oligonucleotides exhibited greater thermal stability with no detectable DNA degradation under the acetylacetone labeling conditions.

Primer extension experiments were performed to test how efficient polymerases can incorporate A opposite labeled the 5fC bases 5fC-M and 5fC-A. Primer **P1** and template **T1** were used; the sequences are listed in chapter 9.5.1. A first reaction was carried out with 0.5 µM labeled DNA template, 0.25 µM primer, 200 µM dNTPs, and 1 U Klenow Fragment (exo-) DNA polymerase in NEBufferTM (1X), incubated for 30 minutes at 37 °C. LC-MS analysis showed that 11% of the strands remained unextended, while 12% had one A incorporated, and 74% got extended further. 3% of the strands had a G instead.

Other polymerases were tested to determine the most efficient polymerase in terms of highest C-to-T transition. For these tests, only dATP and dGTP were added. The malononitrile labeling strategy was included as a reference under the same conditions to compare the performance of the acetylacetone labeling. A total of eight polymerases were tested. An overview of the results is displayed in Table 2 and Figure 35.

*Table 2: Results of the primer extension experiments with Ace-labeling and Malononitrile-labeling. Reactions were performed using 0.5 µM labeled DNA template, 0.25 µM primer, 200 µM each of dATP and dGTP, and one unit DNA polymerase, with an incubation time of 30 minutes. The relative percentages of A and G incorporation as well as unextended primer P are shown. The reactions and measurements were performed with Anna Schöne as part of her master's thesis.[2]*

| | Acetylacetone-Labeling | | | Malononitrile-Labeling | | |
|---|---|---|---|---|---|---|
| Polymerase | P (%) | A (%) | G (%) | P (%) | A (%) | G (%) |
| Bsu | 17.6 | 74.4 | 8.0 | 30.5 | 57.9 | 11.6 |
| Klenow (exo-) | 13.8 | 77.7 | 8.5 | 17.0 | 69.0 | 13.9 |
| Deep Vent (exo-) | 46.3 | 48.5 | 5.2 | 59.8 | 34.7 | 5.4 |
| Taq | 66.8 | 28.9 | 4.3 | 68.1 | 25.6 | 6.3 |
| Sulfolobus IV | 17.3 | 71.4 | 11.3 | 20.3 | 58.9 | 20.8 |
| Therminator | 0.1 | 83.6 | 16.3 | 0.9 | 76.3 | 22.8 |
| Bst | 0.8 | 80.7 | 18.6 | 2.1 | 71 | 26.9 |
| Bst 3.0 | 0.0 | 80.0 | 20.0 | 1.1 | 65.4 | 33.5 |

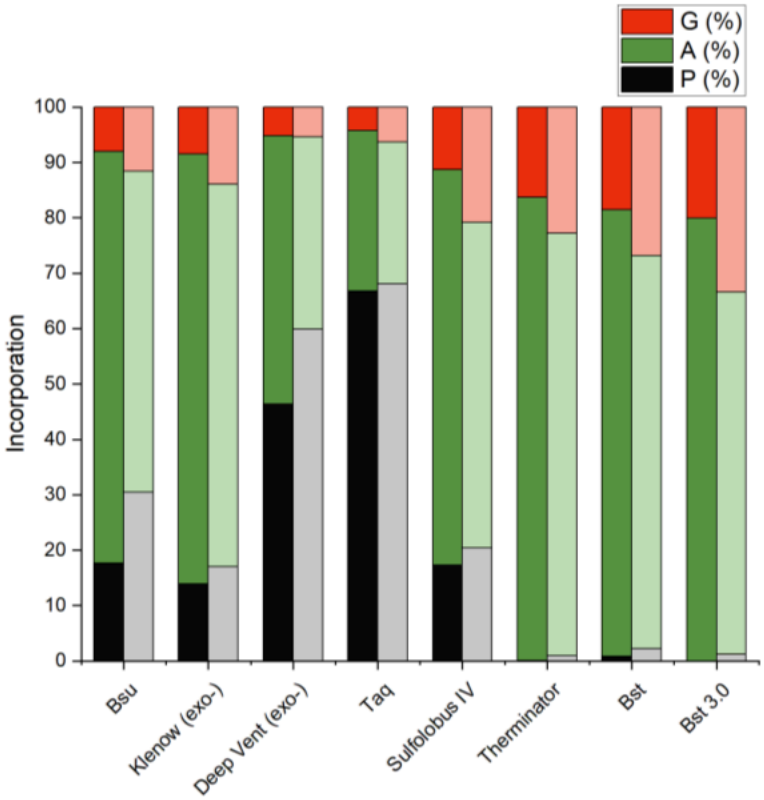



*Figure 35: Results of the primer extension experiments with Ace-labeling (left, darker bars) and Malo-labeling (right, lighter bars). Reactions were performed using 0.5 µM labeled DNA template, 0.25 µM primer, 200 µM each of dATP and dGTP, and one unit DNA polymerase, with an incubation time of 30 minutes. The relative percentages of A and G incorporation as well as unextended primer P are shown. The reaction and measurements were performed with Anna Schöne and this figure is part of her master's thesis.[2]*

Ace-Labeling demonstrated superior performance with a higher C-to-T conversion and lower stalling compared to malononitrile labeling in all tested polymerases.

Afterwards, primer overhang PCR was used to generate a double-stranded 148mer DNA. The sequences are listed in chapter 9.5.1. In the case of adenine incorporation opposite to the labeled 5fC in the initial polymerase reaction, thymine is incorporated in the opposite strand. This process enables the detection of a C-to-T transition by comparison to the reference sequence. The PCR products were then analyzed after Sanger sequencing.

*Table 3: Results from Sanger sequencing, including percentages of thymine (T) and cytosine (C) detected. The percentages are average values of two replicates per approach.*

| DNA Polymerase | Labeling | T (%) | C (%) |
|---|---|---|---|
| AmpliTaq Gold 360™ | None (reference) | 4 | 96 |
| | Acetylacetone | 91 | 9 |
| | Malononitrile | 84 | 16 |
| Deep Vent | Acetylacetone | 52 | 48 |
| | Malononitrile | 37 | 63 |
| Deep Vent exo- | Acetylacetone | 21 | 79 |
| | Malononitrile | 13 | 87 |
| KOD | Acetylacetone | 88 | 12 |
| | Malononitrile | 79 | 21 |
| OneTaq | Acetylacetone | 15 | 85 |
| | Malononitrile | 11 | 89 |
| Q5 Hot Start | Acetylacetone | 83 | 17 |
| | Malononitrile | 63 | 37 |

For Ace-labeling, the Sanger sequencing chromatograms verified that the specific cytosine 43 (5fC) was effectively converted to a thymine. Deep Vent (exo-) and OneTaq polymerases did not achieve a C-to-T conversion, neither with malononitrile nor with Ace-Labeling. Consequently, these polymerases have not been investigated further.

When using the AmpliTaq Gold™ 360 DNA polymerase, the malononitrile-labeled sample (right side in Figure 36) displays a blue cytosine-band with 16% relative size compared to the red thymine-band with 84%. As Ace-Labeling achieved 91% C-to-T conversion, it is more effective compared to malononitrile.

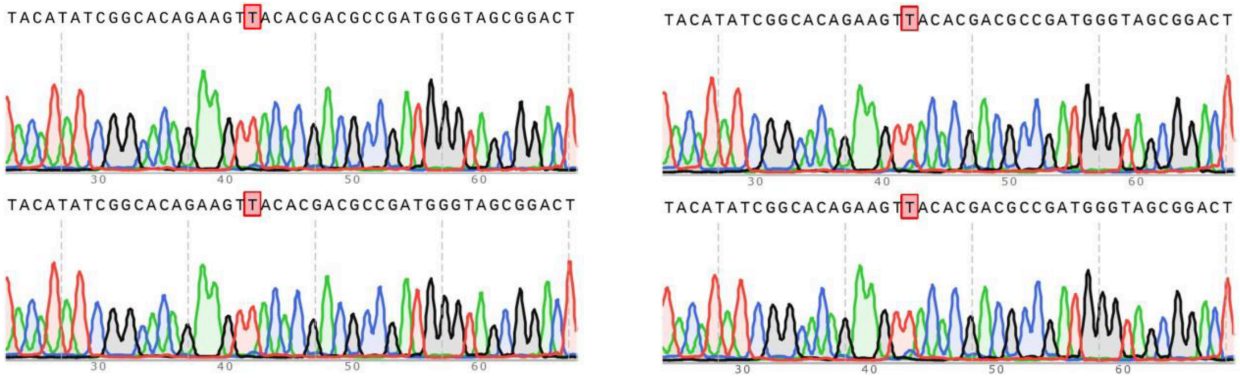



*Figure 36: Sanger sequencing results from PCR with AmpliTaq GoldTM 360 DNA polymerase (left: Ace-Labeling, right: Malo-Labeling).*

Using the Deep Vent DNA polymerase (Figure 37), the overall C-to-T-conversion was not efficient, as seen by the overlap of the C- and T- bands in position 43. After Ace-Labeling, the T band has a higher intensity of 52%, which leads to the interpretation of the base as T. Conversely, malononitrile-labeling had 37% T and 63% C and therefore did not achieve the desired C-to-T conversion.

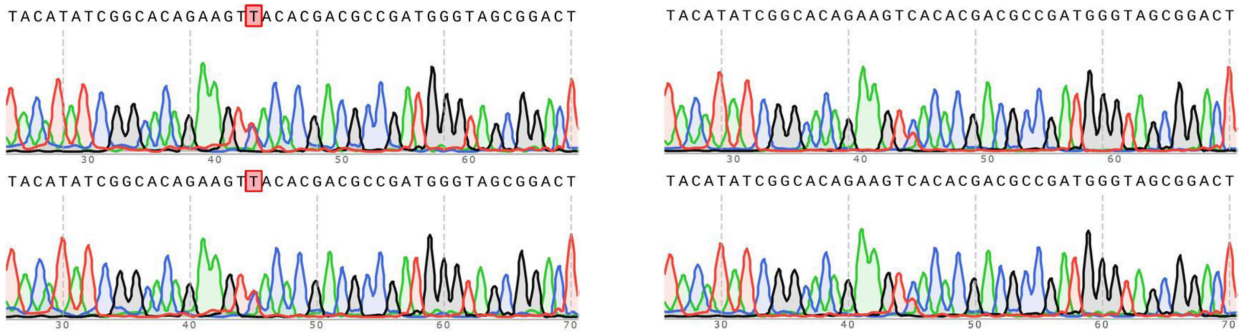



*Figure 37: Sanger sequencing results from PCR with Deep Vent® DNA polymerase (left: Ace-Labeling, right: Malo-Labeling).*

For the KOD DNA polymerase and Q5® Hot Start DNA polymerase, both labeling approaches were found to be effective. However, AceF-Seq was again superior (88% and 83%, respectively) to malononitrile (79% and 63%, respectively) as seen in Figure 38 and Figure 39.

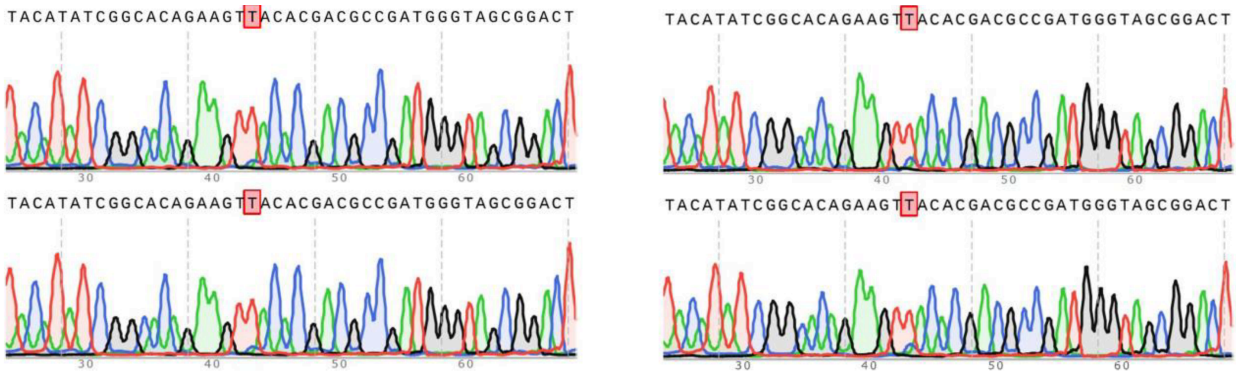



*Figure 38: Sanger sequencing results from PCR with KOD DNA polymerase (left: Ace-Labeling, right: Malo-Labeling).*

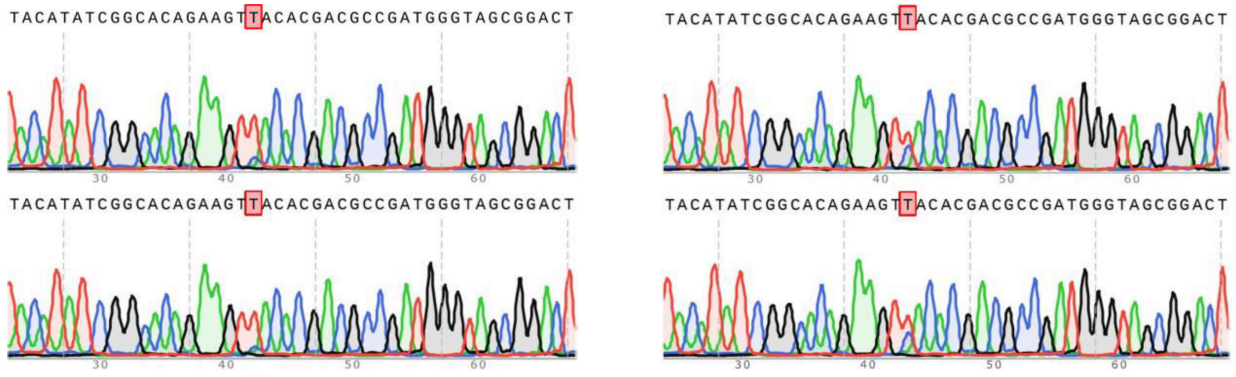



*Figure 39 Sanger sequencing results from PCR with Q5® Hot Start DNA polymerase (left: Ace-Labeling, right: Malo-Labeling).*

Overall, acetylacetone-labeling outperformed the malononitrile-labeling approach, illustrating the superiority of the AceF-Seq method in comparison to previous techniques.

## 5.5 New labeling techniques for 5fC - Conclusion and future perspectives

In this chapter, different small-molecule-based labeling strategies for 5-formylcytosine (5fC) in DNA were discussed. Conventional malononitrile-based labeling has been shown to convert 5fC to a base read as T/U by polymerases, thereby differentiating it from cytosine. However, this method is limited in its conversion efficiency and can lead to DNA degradation. Undesired G incorporation can be caused by tautomerization or protonation of 5fC-M or by the formation of a wobble pair including the exocyclic amine of 5fC-M. A robust detection of base modifications is important for single-base resolution epigenome sequencing to distinguish between different DNA modifications and to avoid false negative results. Consequently, novel labeling strategies were evaluated.

In this work, the potassium salt of nitroacetonitrile (NAN) was evaluated as a novel labeling reagent, initially demonstrating efficient conversion. Nevertheless, the formation of the (*Z*)-isomer which cannot cyclize results in a low C-to-T conversion. Attempts to change the conformation through UV light irradiation or heat to allow cyclization led to partial conversion of the (*Z*) to the (*E*) isomer, while treatment over a longer period of time led to the detachment of a majority of labeled product. Despite adjustments to the reaction conditions and the addition of photocatalysts, satisfactory results could not be obtained.

Consequently, a novel method for the selective labeling of 5fC was developed: Acetylacetone-induced 5-formylcytosine-to-thymine conversion sequencing, termed AceF-Seq. The method demonstrates high labeling efficiency and selectivity. Following an initial series of tests on 5fC nucleoside, AceF-Seq was further optimized for 25-mer and 70-mer DNA strands. In optimal conditions, a quantitative conversion can be achieved without observable DNA degradation. Sanger sequencing confirmed an effective C-to-T conversion while achieving higher conversion rates than malononitrile-labeling. AceF-Seq has been demonstrated to enhance the precision and reliability of 5fC detection.

In the near future, it may be possible to sequence entire genomes using AceF-Seq and detect the occurrence of 5fC. This could provide insight into the dynamics of active DNA demethylation and could potentially reveal entirely new biological functions of 5fC. Therefore, this research lays the groundwork for advanced epigenetic mapping and biomarker research.

# 6. 2-Amino-DDP as a pH-responsive nucleobase

Select parts of this chapter were published verbatim in the following publication: Eric Ogel, Sidney Becker, “Synthesis and Incorporation of a pH-Responsive Nucleoside into DNA Sequences”, *Chembiochem* **2025**, e70086. https://doi.org/10.1002/cbic.202500650. At the time of publication of this thesis, the article is an Online Version of Record before inclusion in an issue (Early View).[3]

Life on Earth depends on proton gradients to drive molecular machines that are part of essential biological processes, from photosynthesis to cellular respiration.[93] Nanotechnology, particularly DNA origami, is often used to mimic complex molecular machines. The structural versatility of DNA, coupled with its programmable thermodynamics, makes it a remarkable macromolecule. It has been demonstrated to be an excellent candidate for engineering molecular devices, also due to the efficiency of its synthesis.[124] DNA origami, however, is inherently static.[125-126] Even though it can simulate machine-like properties, it cannot harness proton gradients to transform into a dynamic system.[96] In this chapter, the development of a programmable nucleoside with pH-sensitive base pairing will be discussed. It should enable a reversible switch between paired and unpaired states near physiological pH. This design should allow for programmable sequences that transition between duplex and single-stranded DNA in response to pH changes, while natural sequences should not be affected.

The design of pH-responsive, orthogonal DNA motifs that maintain the integrity of canonical sequences requires careful consideration of p$K_a$ values. Cytosine undergoes protonation at pH levels below 4.5, while thymine deprotonates above pH 10. Therefore, too acidic or too basic solutions result in the destabilization of the DNA duplex, which subsequently leads to its denaturation. Consequently, the optimal p$K_a$ for a pH-sensitive nucleoside should ideally lie between these values, achieving responsiveness near physiological pH (~7.4). In order to maintain duplex stability and ensure enzymatic compatibility, the pH-responsive nucleobase must closely mimic natural bases. The modification of the acid-base properties of heterocyclic compounds can be achieved through the strategic adjustment of their electron density. It has been demonstrated that the removal of nitrogen atoms from the aromatic ring structure increases electron density, thereby raising the nucleobase's p$K_a$.**[107]** This approach offers a promising strategy for the regulation of nucleobase properties.

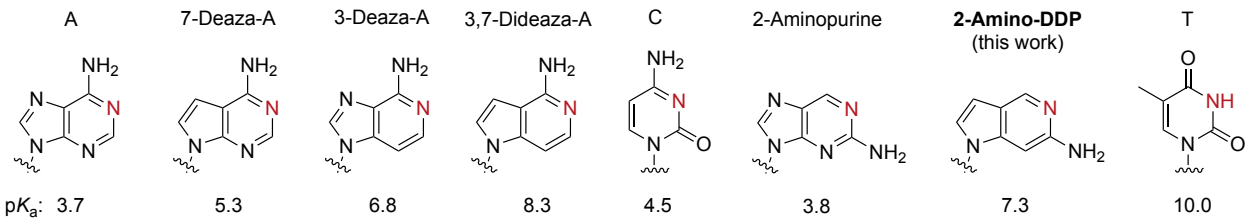



*Figure 40: $pK_a$ values[107] of canonical nucleobases and close structural analogues showing higher basicity in deaza- and dideaza-derivatives.*

2-Amino-dideaza purine (2-Amino-DDP) was chosen as novel building block because the p$K_a$ of this modified nucleobase was expected to be similar to its structural isomer 3,7-Dideaza adenine, which is around 8.3.[107] The base should therefore be protonated at acidic to neutral pH and uncharged at basic pH. This should result in a base pair analogous to the AT pair in basic environments, while the protonated form should build a base pair with C through the new hydrogen bond caused by the protonated N1. This behavior enables the construction of a molecular switch from a single strand of DNA that exhibits a reversible binding (on state) or non-binding (off state) property to the complementary strand, depending on the pH level, as illustrated in Figure 41.

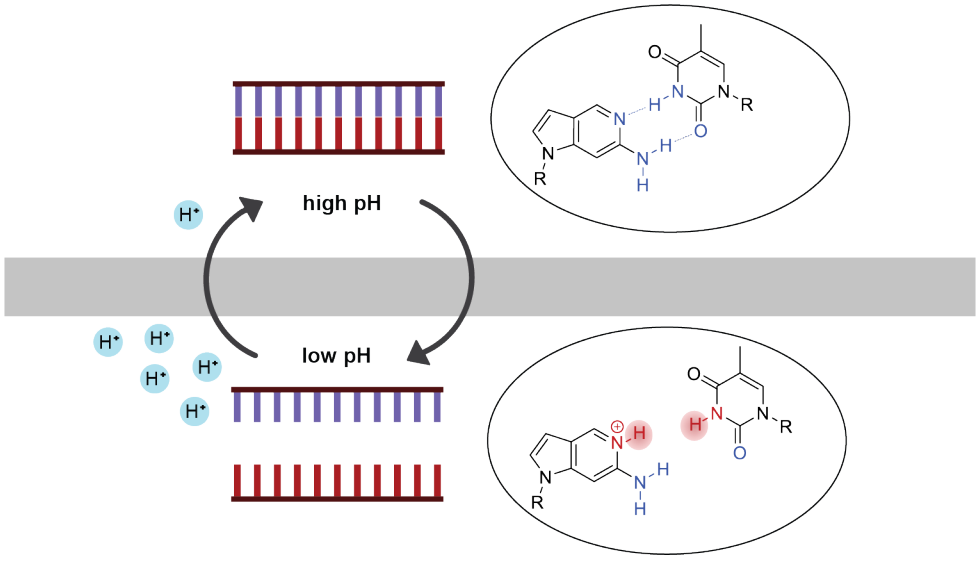



*Figure 41: Schematic representation of the pH-responsive base 2-Amino-DDP causing denaturation in an environment with low pH. This figure was modified from: Eric Ogel, Sidney Becker, "Synthesis and Incorporation of a pH-Responsive Nucleoside into DNA Sequences", Chembiochem **2025**, e70086, under a CC-BY 4.0 license.[3]*

## 6.1 Chemical triphosphate Synthesis

To further investigate its pH-responsive character and to test whether the triphosphate of 2-Amino-DDP could be incorporated by polymerases selectively opposite C or T, depending on the pH, the triphosphate was synthesized. The synthesis scheme of the triphosphate is shown in Figure 42.

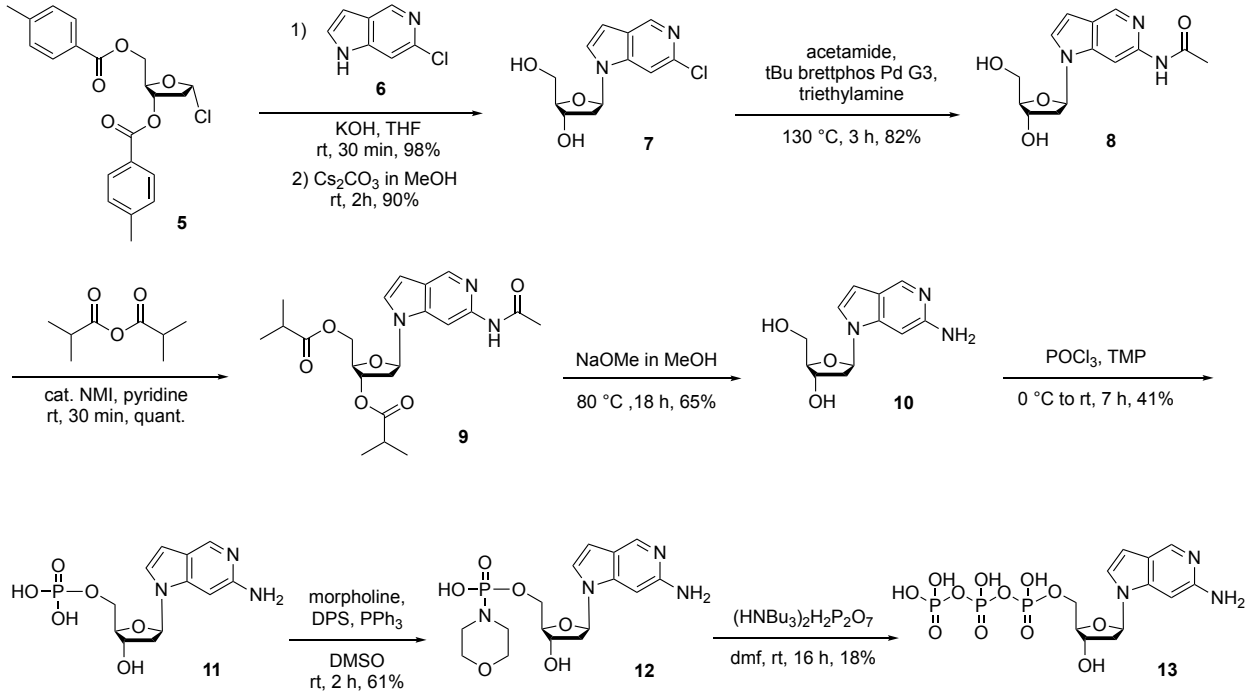



*Figure 42: Synthesis of 2-Amino-DDP triphosphate. This figure was modified from: Eric Ogel, Sidney Becker, "Synthesis and Incorporation of a pH-Responsive Nucleoside into DNA Sequences", Chembiochem **2025**, e70086, under a CC-BY 4.0 license.*

The triphosphate was synthesized in eight steps, starting from Hoffer's Chlorosugar **5** and 3,7-dideaza-2-chloropurine **6**, which are both commercially available. The glycosylation and the following deprotection of the hydroxyl groups were conducted as described in the literature[127] to give the 2-Chloro-DDP nucleoside **7**. The chlorine residue was then converted to amide **8** using a Buchwald-Hartwig amidation. Initially, the protocol documented in the literature[128] was employed, utilizing *tBu brettphos Pd G3* as pre-activated catalyst, $Cs_2CO_3$ as the base and *tert*-butanol as the solvent.[128] With 6.0 eq. of acetamide, the acetyl-protected nucleoside **8** could be isolated with 10% yield. To optimize the conversion rate, varying temperatures, bases and stoichiometric relations were tested for acetamide and for trifluoroacetamide, as shown in Table 4.

*Table 4: UV-determined yields of the Buchwald-Hartwig-Amidation of 7 with acetamide and trifluoroacetamide.*

| UV-yield (unisolated) | 2,6-lutidine, 120 °C for 16 h | Triethylamine, 120 °C, 16 h | KOtBu, 120 °C, 16 h | Triethylamine, 130 °C, 3 h |
|---|---|---|---|---|
| 6.00 eq. acetamide in tBuOH | 10% | 12% | 8% | 14% |
| 100 eq. acetamide, no solvent | 83% | 87% | 65% | 95% |
| 100 eq. trifluoroacetamide, no solvent | 0% | 0% | 0% | 0% |

The optimal conditions involve the use of acetamide as both the solvent and reagent, with the exclusion of any additional solvents and led to 82% isolated yield. The method resulted in large amounts of acetamide that cannot be removed by evaporation. A method for the separation of the nucleoside from the acetamide excess involved the protection of the 5'- and 3'-hydroxyl groups using isobutyric anhydride. The resulting protected nucleoside **9** was significantly less polar, allowing its transfer to the organic layer during extraction with dichloromethane (DCM), while the majority of the acetamide remained in the aqueous layer. Subsequent column chromatography allowed the isolation of the protected nucleoside in a highly pure form. The deprotection process required the use of a strong base, since ammonia, methylamine or a mixture of them (AMA) cleave the esters but are not able to hydrolyze the amide, even at 80 °C. The exceptional stability of the amide is particularly remarkable because the acetyl-, isobutyl-, or benzoyl protecting groups on the natural building blocks A and G can be deprotected much more rapidly.[129] Using ammonia at 65 °C, complete deprotection of A and G is achieved in less than 6 hours.[130] This observation highlights the significant impact of substituting nitrogen with CH in the aromatic system. The higher electron density of the system leads not only to higher p$K_a$ values, but also to an enhanced stability of the protective amide. The full deprotection of the nucleoside could be achieved by using a 0.5 M solution of KOH in MeOH at 80 °C for 18 h, accompanied by traces of depurination. The deprotected nucleoside **10** could be isolated in 65% yield. The conversion to the 5'-monophosphate **11** was then achieved through the reaction with $POCl_3$ in trimethyl phosphate, a method that has been described in the literature for a variety of nucleosides.[131] Subsequently, the nucleoside monophosphate was converted to the corresponding morpholidate **12**, using a procedure similar to the original publication from 1965.[132] **12** then reacted with pyrophosphate to yield the triphosphate **13**.

The initial purification process, carried out using high-performance liquid chromatography (HPLC), involved the use of an ion exchange column with a Tris-HCl buffer and an aqueous, 1.25-molar NaCl solution. The process enabled the removal of byproducts, such as the nucleoside monophosphate, while the nucleoside diphosphate remained unseparated from the triphosphate. Following this initial purification, an additional

HPLC purification was conducted, using a C18 column and a triethyl ammonium acetate (TEAA) buffer as an ion pairing agent.

### 6.2 Enzymatic triphosphate synthesis

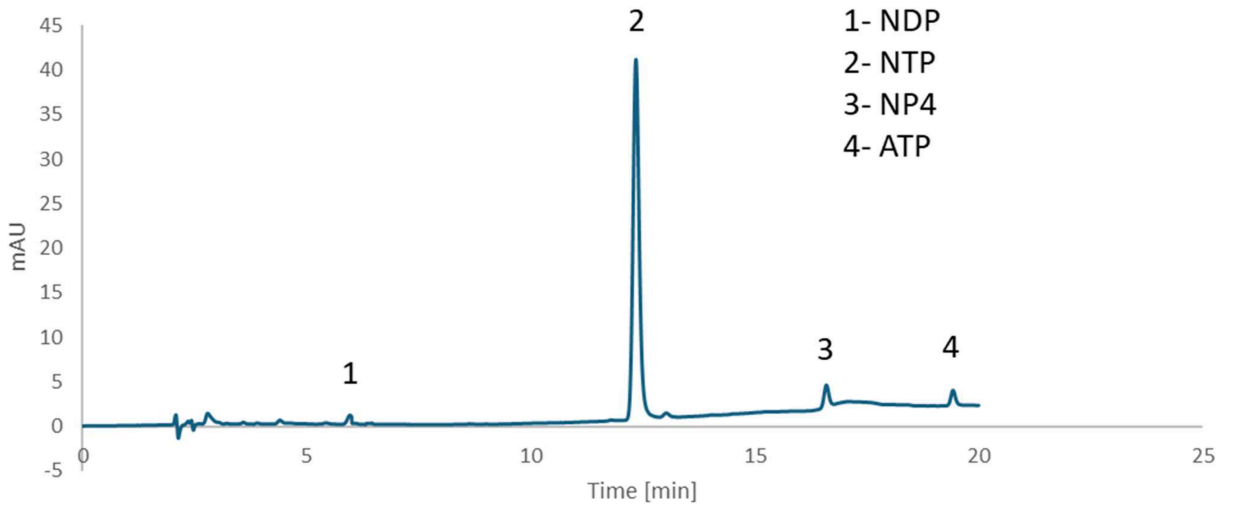
In order to achieve improved triphosphate yields with a reduced number of steps, the enzymatic triphosphate synthesis technique was evaluated. For this purpose, polyphosphate kinases (PPKs) were examined, as they catalyze the addition of phosphate groups to nucleoside monophosphates using various phosphate donors. Two PPKs (PPK04 and PPK05, both from BioNukleo) were tested under the conditions listed in Table 5.

*Table 5: Tested conditions for the enzymatic triphosphate reaction. SH= Sodium hexametaphosphate, SP = Sodium polyphosphate, MP = Monophosphate, DP = Diphosphate, TP = Triphosphate. Final concentrations: 0.1 µg / µL enzyme, 5 mM Monophosphate, 6 mM Polyphosphate, 10 mM $MgCl_2$, 70 mM Tris-HCl (pH = 7.5). Total volume per reaction: 100 µL.*

| Kinase | Phosphate source | Monophosphate | Ratio MP:DP:TP after 24 h | Ratio MP:DP:TP after 3 d |
|---|---|---|---|---|
| PPK 04 | SH | dAMP | 40 : 10 : 1 | - |
| | SH | 2-Amino-DDP | 75 : 8 : 1 | 33 : 5 : 1 |
| | SP | dAMP | 150 : 5 : 1 | - |
| | SP | 2-Amino-DDP | 100 : 4 : 1 | 300 : 8 : 1 |
| PPK 05 | SH | dAMP | 30 : 10 : 1 | - |
| | SH | 2-Amino-DDP | 75 : 10 : 1 | 25 : 8 : 1 |
| | SP | dAMP | 100 : 30 : 1 | - |
| | SP | 2-Amino-DDP | 150 : 15 : 1 | 60 : 10 : 1 |

The optimal conditions for achieving satisfactory results for PPK05 involved the use of sodium hexametaphosphate as the phosphate donor and a reaction temperature of 50 °C. The highest observed concentration of triphosphate was recorded after four days of reaction; however, some of the triphosphate began to hydrolyze back to di- and monophosphates. The process of hydrolysis may be accelerated by the $Mg^{2+}$ ions necessary for the activity of the enzyme, which are required to be present in the reaction.

To evaluate the potential of triphosphate synthesis when starting from the nucleoside rather than the monophosphate, a collaborative effort was undertaken with the company BioNukleo, who optimized a protocol with 2-Amino-DDP nucleoside to convert it to the triphosphate. In optimal conditions, 92.5% nucleoside triphosphate is generated, along with 4.3% nucleoside tetraphosphate and 2.2% nucleoside diphosphate, as shown in Figure 43.




| Verbindung | NDP | NTP | NP4 | ATP |
|---|---|---|---|---|
| Anteil | 2,2% | 92,5% | 4,3% | 1% |

*Figure 43: HPLC-chromatogram of the enzymatic 2-Amino-DDP triphosphate synthesis provided by BioNukleo. NP4 is the nucleoside tetraphosphate.*

In contrast to the higher yield, the drawback of enzymatic triphosphate synthesis is the high cost of PPKs. Consequently, chemical synthesis may be the preferred method for triphosphate production on a larger scale, while enzymatic triphosphate synthesis is the more efficient method of choice for smaller quantities. The following polymerase reactions were carried out using chemically and enzymatically synthesized nucleoside triphosphate.

### 6.3 Single nucleotide incorporation

To test the pH-responsive properties of 2-Amino-DDP, polymerase incorporation experiments were performed using single-nucleotide primer extension reactions. A 24-mer primer, which had the fluorophore 6-FAM attached at the 5'-end, was annealed to a 25-mer template with a C or T overhang at its 5'-end, as illustrated in Figure 44.

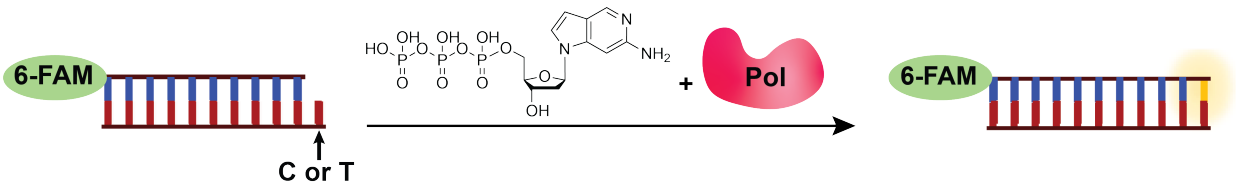



*Figure 44: Schematic illustration of the single nucleotide incorporation with a DNA polymerase. This figure was modified from: Eric Ogel, Sidney Becker, "Synthesis and Incorporation of a pH-Responsive Nucleoside into DNA Sequences", Chembiochem **2025**, e70086, under a CC-BY 4.0 license.[3]*

Given the absence of the N3 atom in 2-Amino-DDP, which results in a reduced electron density in the minor groove, the first objective was to investigate if exonuclease-deficient polymerases can incorporate 2-Amino-DDP triphosphate under their standard buffer conditions. In an initial screening, five different polymerases, which all lack exonuclease activity, were tested: Bsu (*Bsu* DNA Polymerase, Large Fragment), Kle (Klenow-Fragment), Bst (*Bst* DNA Polymerase, Large Fragment), The (Therminator™ DNA Polymerase), Sul (*Sulfolobus* DNA Polymerase IV). The resulting oligonucleotide strands were separated using Urea-PAGE and detected by the fluorescence signal caused by 6-FAM (depicted in Figure 45).

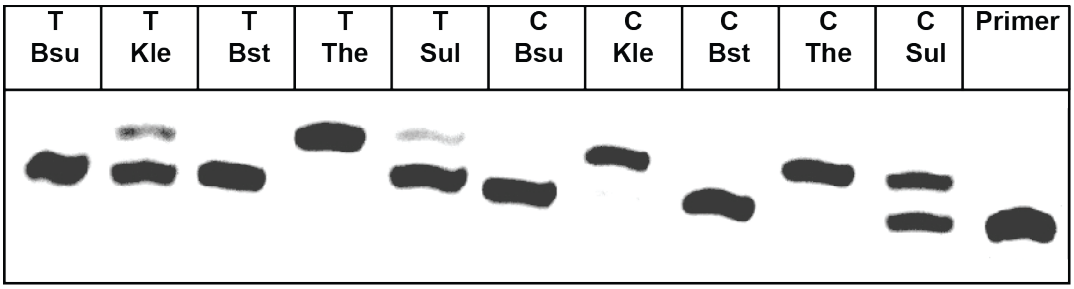



*Figure 45: Urea-PAGE of the enzymatic incorporation of 2-Amino-DDP with different polymerases. Conditions: 0.5 µM template, 0.25 µM primer, 100 µM dNTP, 0.05 units/µL polymerase, 10 min. Temp: 60 °C for Bst, Kle and Sul, 37 °C for Bsu and Kle. This figure was modified from: Eric Ogel, Sidney Becker, "Synthesis and Incorporation of a pH-Responsive Nucleoside into DNA Sequences", Chembiochem **2025**, e70086, under a CC-BY 4.0 license.[3]*

While Bsu and Bst polymerases showed no activity, the Therminator polymerase completely elongated the primer, both with T and with C as overhanging bases of the template strand. Kle and Sul showed partial incorporation. Further investigations included more polymerases and focused on the pH-dependency. To ensure proper pH-dependent incorporation, we required a polymerase that is able to accept modified dNTPs while remaining selective enough to distinguish between protonated and unprotonated 2-Amino-DDP. We found out that the Thermo Sequenase, a polymerase that was originally designed for sequencing[133], shows pH-dependence by a clear preference for C incorporation at lower pH values and a preference for T at higher pH values, as shown in Figure 46.

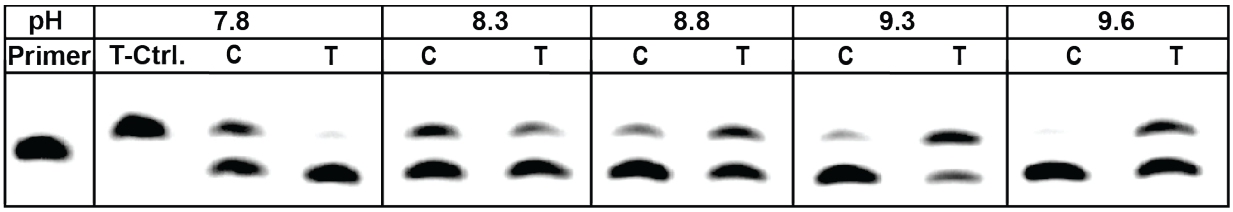



*Figure 46: Urea-PAGE of the enzymatic and pH-dependent incorporation of 2-Amino-DDP with Thermo Sequenase™ DNA Polymerase (Cytiva), conditions: 0.5 µM template, 0.25 µM primer, 25 µM dNTP, 0.05 units/µL polymerase, 30 mM Tris-HCl, 7.5 mM $MgSO_4$, 60 °C, 10 min. This figure was modified from: Eric Ogel, Sidney Becker, "Synthesis and Incorporation of a pH-Responsive Nucleoside into DNA Sequences", Chembiochem **2025**, e70086, under a CC-BY 4.0 license.[3]*

At pH 7.8, the Thermo Sequenase incorporated 2-Amino-DDP preferentially opposite cytosine (42% after 10 min), while less incorporation opposite T occurred (14% after 10 min). At a higher pH of 9.6, incorporation shifted to favor thymine in the template strand (39% opposite T, 11% opposite C). When doubling the Thermo Sequenase polymerase concentration to 0.10 units/µL at pH 8.3, 2-Amino-DDP is already fully incorporated after 5 min opposite to C, while only 17% incorporation occurred opposite T (Figure 47).

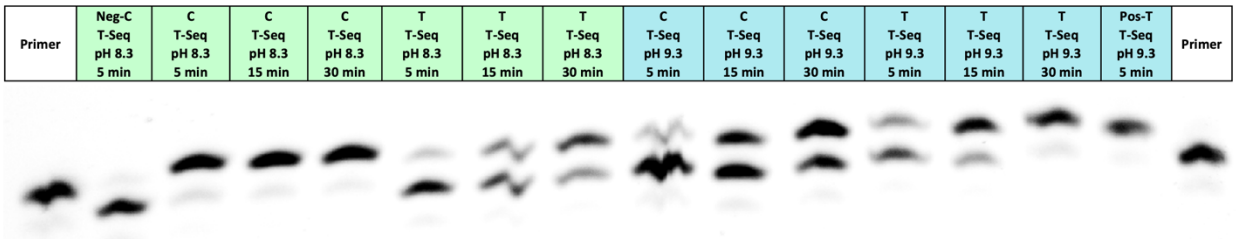



*Figure 47: Urea-PAGE of the incorporation of 2-Amino-DDP triphosphate with Thermo Sequenase™ DNA Polymerase (Cytiva), conditions: 0.5 µM template, 0.25 µM primer, 25 µM dNTP, 0.10 units/µL polymerase, 30 mM Tris-HCl, 7.5 mM $MgSO_4$, 60 °C. Neg-C (negative control): no dNTP added. Pos-C (positive control): 25 µM dATP as the only dNTP. This figure was modified from: Eric Ogel, Sidney Becker, "Synthesis and Incorporation of a pH-Responsive Nucleoside into DNA Sequences", Chembiochem **2025**, e70086, under a CC-BY 4.0 license.[3]*

The incorporation opposite T occurs at a significantly slower rate with only 17% incorporation after 5 min and 64% incorporation after 30 min. Conversely, at a pH of 9.3, there is a nearly complete primer extension (91%) opposite T after 30 minutes, while the extension opposite C is less efficient with 68%.

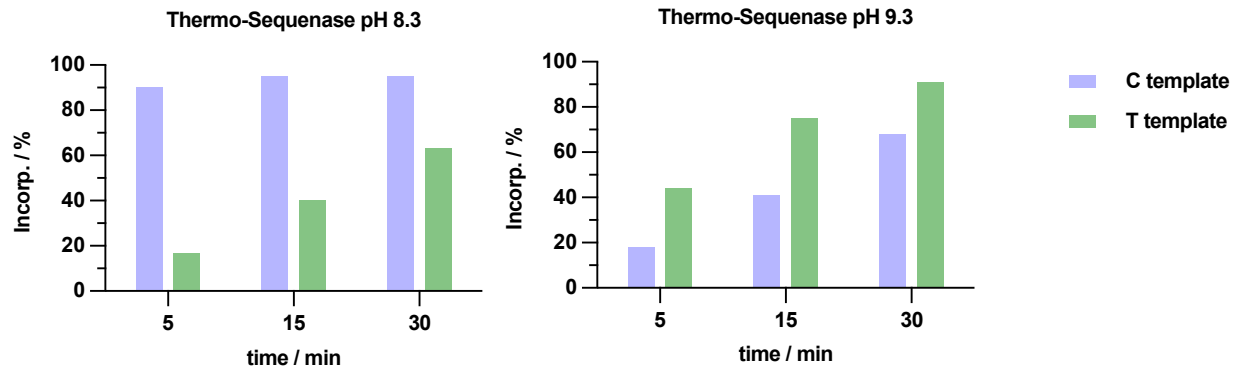



*Figure 48: Time-dependent incorporation of 2-Amino-DDP triphosphate with Thermo Sequenase™ DNA Polymerase (Cytiva), conditions: 0.5 µM template, 0.25 µM primer, 25 µM dNTP, 0.10 units/µL polymerase, 30 mM Tris-HCl, 7.5 mM $MgSO_4$, 60 °C.*

To investigate whether other polymerases can also incorporate 2-Amino-DDP in a pH-dependent manner, we tested KlenTaq. The KlenTaq polymerase is known to incorporate unnatural base pairs by inducing a Watson-Crick geometry[134] and also shows pH-responsive behavior in our experiments, as shown in Figure 49 and Figure 50.

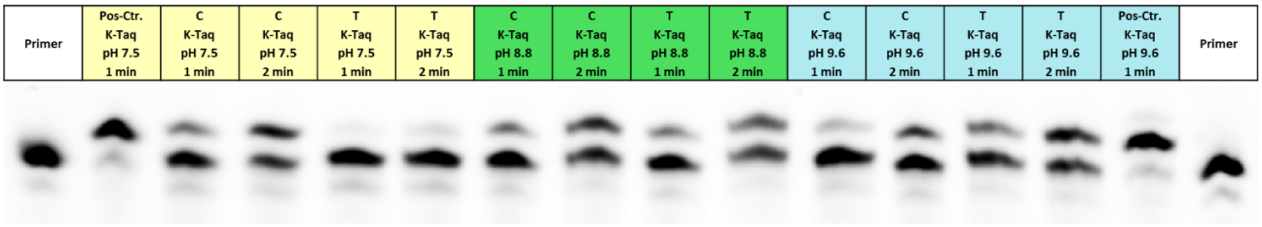



*Figure 49: Urea-PAGE of the incorporation of 2-Amino-DDP triphosphate with KlenTaq. Conditions: 0.5 µM template, 0.25 µM primer, 25 µM dNTP, 0.05 units/µL polymerase, 25 mM Tris-HCl, 40 mM KCl, 5 mM $MgSO_4$, 60 °C. This figure was modified from: Eric Ogel, Sidney Becker, "Synthesis and Incorporation of a pH-Responsive Nucleoside into DNA Sequences", Chembiochem **2025**, e70086, under a CC-BY 4.0 license.[3]*

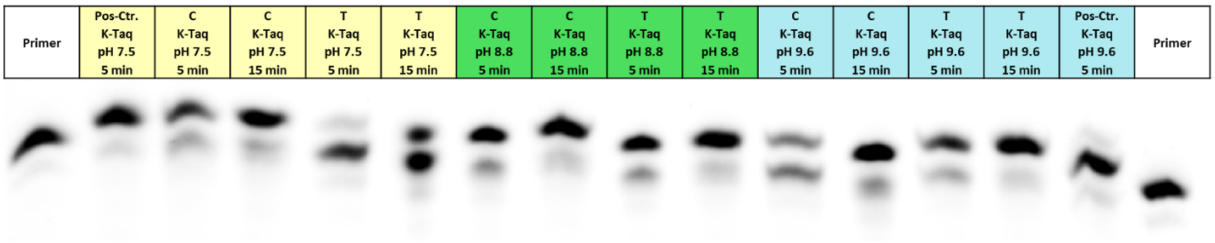



*Figure 50: Urea-PAGE of the incorporation of 2-Amino-DDP triphosphate with KlenTaq. Conditions: 0.5 µM template, 0.25 µM primer, 25 µM dNTP, 0.05 units/µL polymerase, 25 mM Tris-HCl, 40 mM KCl, 5 mM $MgSO_4$, 60 °C. This figure was modified from: Eric Ogel, Sidney Becker, "Synthesis and Incorporation of a pH-Responsive Nucleoside into DNA Sequences", Chembiochem **2025**, e70086, under a CC-BY 4.0 license.[3]*

At a pH of 7.5, 26-45% incorporation opposite C is observable after 1-2 minutes, while only minimal incorporation (4-8%) occurs opposite T. After 15 min, 73% incorporation appeared opposite C and 38% opposite T. At a pH of 8.8, the amount of incorporated nucleotide is very similar opposite C and T: 22-24% after 1 min, 43-44% after 2 min, 74% after 5 min and 79% incorporation after 15 min incubation time. At a pH of 9.6, incorporation opposite T is favored with 72% (versus 45% opposite C) after 5 min. After 15 minutes, incorporation is almost complete opposite both C (80%) and T (91%).

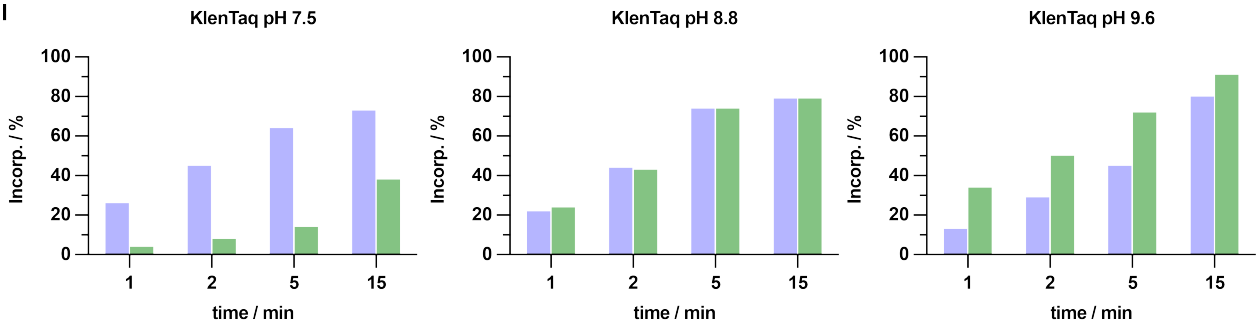



*Figure 51: Time-dependent incorporation of 2-Amino-DDP triphosphate with KlenTaq. Conditions: 0.5 µM template, 0.25 µM primer, 25 µM dNTP, 0.05 units/µL polymerase, 25 mM Tris-HCl, 40 mM KCl, 5 mM $MgSO_4$, 60 °C.*

This suggests that incorporation opposite C is strongly favored at lower pH values, while incorporation opposite T is slightly favored in more basic buffers. The reduced incorporation at lower pH values can be partially attributed to the enhanced activity of polymerases at higher pH values, leading to higher activity, irrespective of the type of triphosphate. This observation lends further support to the hypothesis that 2-Amino-DDP undergoes protonation at lower to neutral pH, thereby forming a more effective base pair with C. In contrast, the deprotonated form, which is predominant at higher pH, establishes a more robust base pair with T.

As a control, the same conditions were tested with 2-Aminopurine triphosphate instead of its dideaza-derivative, shown in Figure 52.

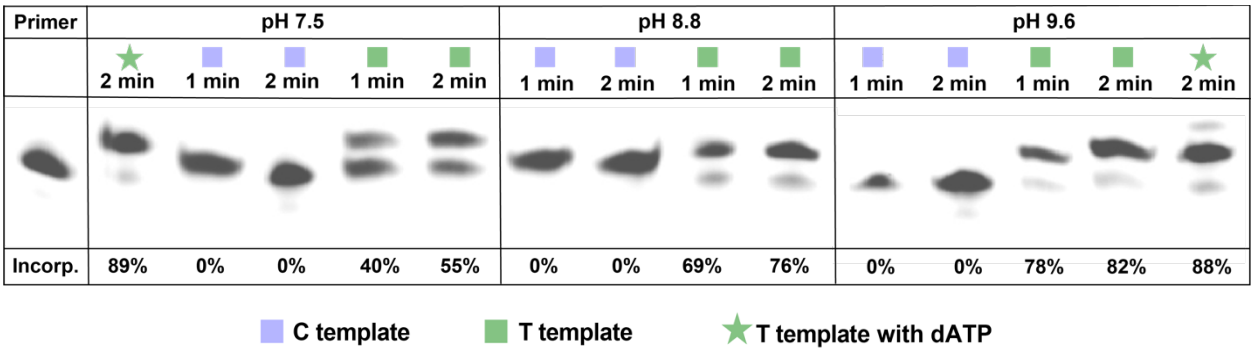



*Figure 52: Urea-PAGE of the incorporation of 2-Aminopurine with KlenTaq. Conditions: 0.5 µM template, 0.25 µM primer, 25 µM dNTP, 0.05 units/µL polymerase, 25 mM Tris-HCl, 40 mM KCl, 5 mM $MgSO_4$, 60 °C. The experiments where dATP was incorporated opposite a T template were conducted as positive control. This figure was modified from: Eric Ogel, Sidney Becker, "Synthesis and Incorporation of a pH-Responsive Nucleoside into DNA Sequences", Chembiochem **2025**, e70086, under a CC-BY 4.0 license.*[3]

As expected, there is no detectable incorporation of 2-Aminopurine opposite C at any tested pH. The incorporation opposite T is favored at basic pH, caused by the higher activity of KlenTaq at higher pH-values. This further supports the hypothesis that 2-Amino-DDP has a different base pairing pattern than 2-aminopurine caused by protonation in the N1 position.

To test whether the unnatural base pair induces polymerase stalling, a primer extension experiment with a template with 19 nucleotides overhang was used. In the first step, the 30-mer primer **P5** and the 49-mer DNA template **T6** were annealed. When 2-Amino-DDP triphosphate and Therminator polymerase were added, two incorporations opposite T occurred, as expected. The calculated product mass of 10,296.849 Da is in agreement with the deconvoluted experimental mass of 10,296.861 Da (Table 6 and Figure 55).

*Table 6: Calculated and deconvoluted masses of the oligonucleotides expected in this primer extension experiment.*

| Oligonucleotide | Calculated mass | Deconvoluted mass |
|---|---|---|
| 30-mer primer **P5** | 9676.731 | Not detected after extension |
| 31-mer primer (**P5**+X) | 9985.782 | Not detected after extension |
| 32-mer primer (**P5**+XX) | 10296.849 | 10296.861 |
| 49-mer Template **T6** | 14981.430 | Not detected after extension |
| 50-mer Template (**T6**+A) | 15292.472 | 15292.432 |
| 49-mer primer (fully extended) | 15591.7284 | Not detected after extension |
| 50-mer primer (extended + A-Tail) | 15904.786 | 15904.762 |

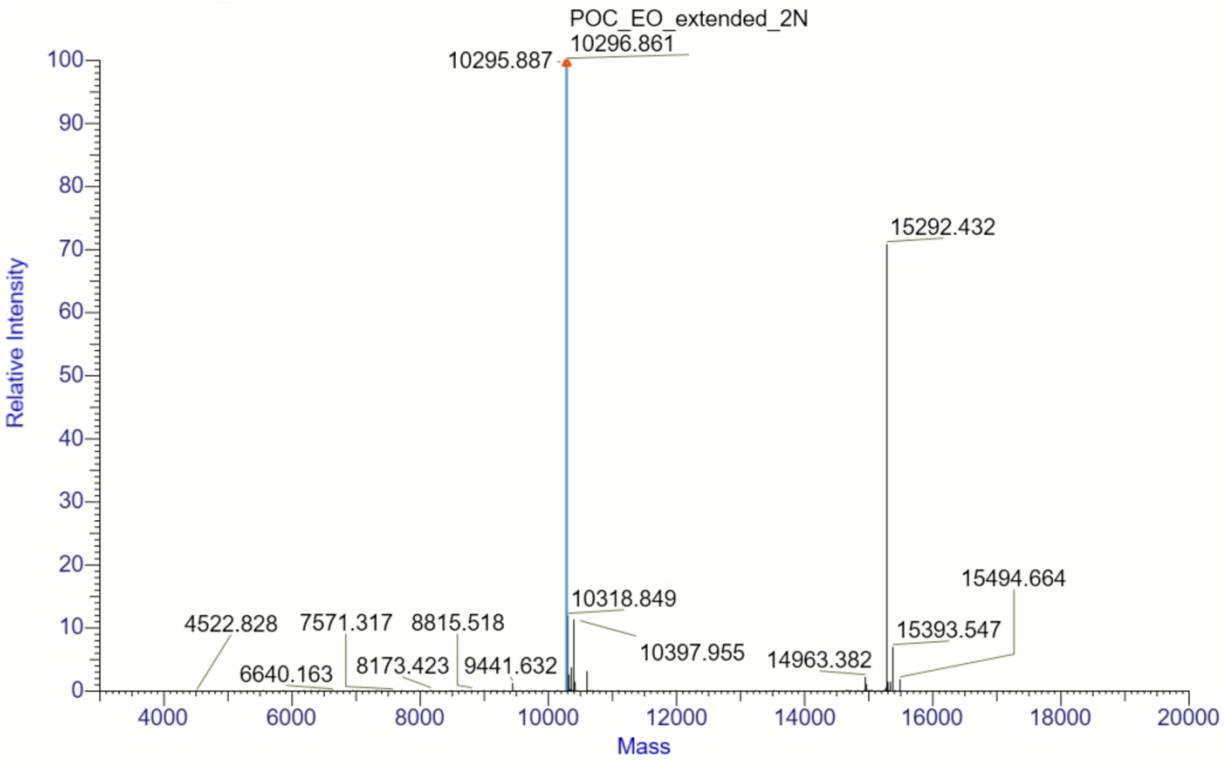



*Figure 53: HRMS-Spectrum of the mixture of elongated (Btn)-primer CGG GCG GAC CAG AAC CCT TGA GCA CAG AAA **XX** (with X = 2-Amino-DDP) and the template (**T6** incl. A-Tail). The calculated mass is 10296.849, the deconvoluted mass is 10296.861. The calculated mass of the template with single A-tailing is 15292.472, the deconvoluted mass is 15292.432. This figure was modified from: Eric Ogel, Sidney Becker, "Synthesis and Incorporation of a pH-Responsive Nucleoside into DNA Sequences", Chembiochem **2025**, e70086, under a CC-BY 4.0 license.*[3]

Additionally, 3'-A-tailing of the template was observed. In the next step, all remaining dNTPs were added. This enabled the full elongation of the primer without stalling, plus an additional 3'-A-tailing event at the end. The expected mass of 15904.786 Da was confirmed by the deconvoluted experimental mass of 15904.762 Da (Table 6 and Figure 54).

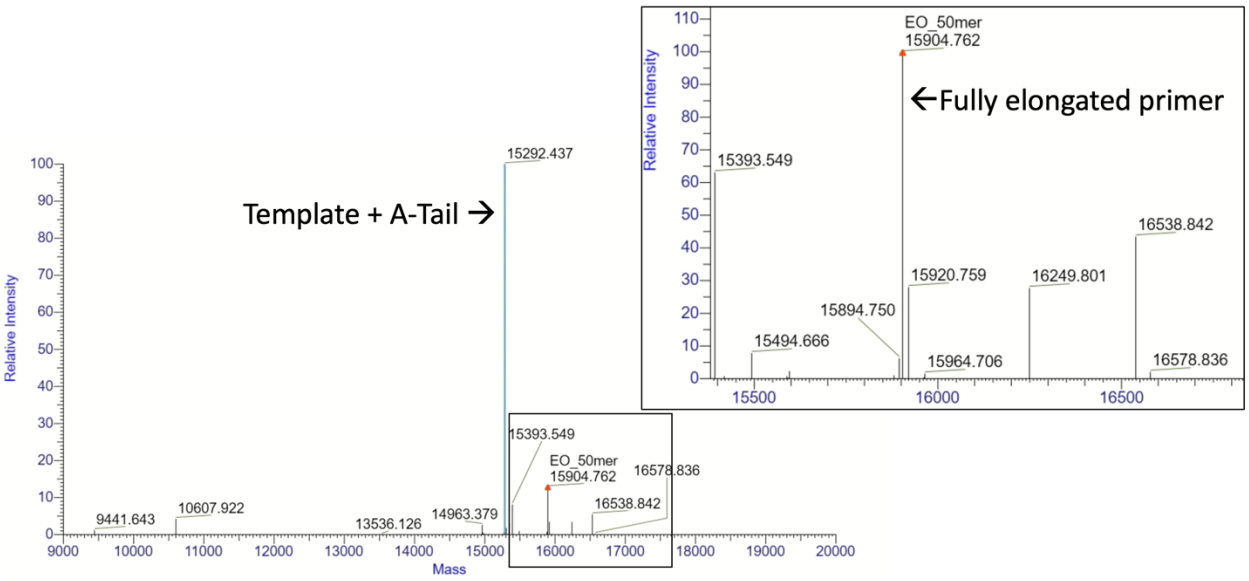



*Figure 54: Deconvoluted HRMS-Spectrum of the elongated 50-mer (Btn)-primer CGG GCG GAC CAG AAC CCT TGA GCA CAG AAA **XX**C GTC GAG TTA GCC GAA GA (with X = 2-Amino-DDP). The calculated mass is 15904.786, the deconvoluted mass we detected is 15904.762. This figure was modified from: Eric Ogel, Sidney Becker, "Synthesis and Incorporation of a pH-Responsive Nucleoside into DNA Sequences", Chembiochem **2025**, e70086, under a CC-BY 4.0 license.[3]*

As 2-Amino-DDP is readily accepted by polymerases it may be used for *in vitro* evolution studies to create DNAzymes. The p$K_a$ of our nucleobase near the physiological range might also allow for general acid-base catalysis analogous to histidine in protein enzymes.

### 6.4 Determination of the p$K_a$ of 2-Amino-DDP

To assess whether the pH-dependent behavior of 2-Amino-DDP observed during polymerase incorporation can be explained by its p$K_a$, a method described by Dardonville and colleagues[135] was used to determine the p$K_a$. For compounds that have a chromophore in close proximity to the ionization center, the extinction coefficient changes as a function of ionization.[135] This enables the measurement of their p$K_a$ values by UV spectrophotometry, as demonstrated for other compounds.[135] To determine the p$K_a$ of 2-Amino-DDP, UV/Vis-absorption of the nucleoside was measured in aqueous buffers at pH values ranging from 5.0 to 9.5 and at a wavelength of 200 to 600 nm. A section of the graph is shown in Figure 55. As the nucleoside is protonated at the aromatic nitrogen N1, the absorption spectrum changes with decreasing pH.

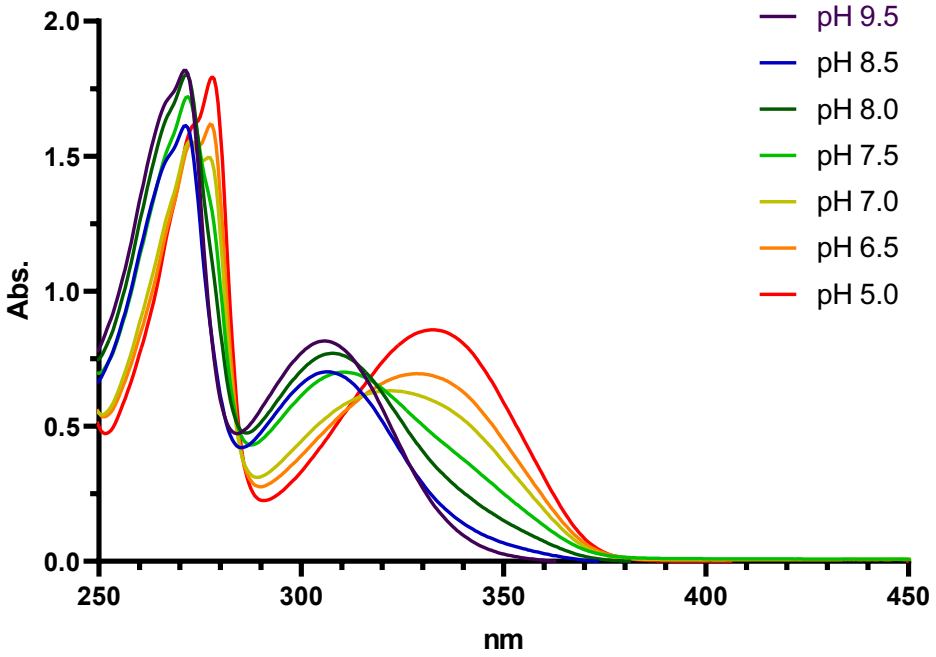



*Figure 55: Absorption spectrum of 2-Amino-DDP nucleoside in aqueous solution at pH values ranging from 5.0 to 9.5. The absorption at the wavelengths from 250 to 450 nm is shown. This figure was modified from: Eric Ogel, Sidney Becker, "Synthesis and Incorporation of a pH-Responsive Nucleoside into DNA Sequences", Chembiochem **2025**, e70086, under a CC-BY 4.0 license.*[3]

In addition to an absorption maximum at 260-270 nm, which is characteristic of aromatic compounds, a second peak emerges at around 310 nm within the basic pH range. As the pH of the buffer decreases, this second peak experiences a shift toward longer wavelengths, reaching 335 nm at pH 5.0. A plot of the total difference in absorption versus pH was constructed in order to determine the p$K_a$. The difference in absorption at 340 nm was used to calculate the p$K_a$ value using nonlinear regression, with a previously published method[135] to provide the p$K_a$ (2-Amino-DDP) = 7.3±0.2.

### 6.5 Phosphoramidite synthesis

The synthesis of oligonucleotides containing 2-Amino-DDP using a solid phase DNA synthesizer is possible after the synthesis of its protected phosphoramidite. Therefore, the selection of appropriate protective groups for the exocyclic amine is crucial. The most efficient strategy would be the use of *N*-acetyl-protected 2-Amino-DDP, because it can be produced in three steps, as illustrated in Figure 42. However, following the phosphoramidite synthesis, it is essential to remove the protecting groups from the DNA under conditions that do not cause DNA damage. As previously discussed, the electron-rich aromatic group results in a highly stable amide bond, thereby making the acetyl protecting group just as incompatible as the isobutyl or benzoyl groups in this case. Nevertheless, amides have been demonstrated to provide useful protection groups in oligonucleotide synthesis. They reduce electron density within the aromatic system,

thereby preventing side reactions. Therefore, it was hypothesized that an amide with strong electron withdrawing groups and thus greater reactivity would be more suitable for 2-Amino-DDP.

Pentafluorobenzoyl chloride was identified as a promising candidate as a protective group reagent due to its commercial availability and enhanced reactivity resulting from its electron-withdrawing fluorine substituents. When unprotected 2-Amino-DDP nucleoside reacts with pentafluorobenzoyl (pfp) chloride, both the amine and the hydroxyl groups get pfp-protected. Two possible synthetic strategies may be considered. In the first strategy, the hydroxyl groups are protected first, for example with silyl protecting groups. Then, the amine is protected with pfp, and finally the hydroxyl groups are deprotected. However, this approach results in yield losses and a more complex synthesis process involving three purification steps. An alternative approach is to fully protect 2-Amino-DDP with pfp and then selectively deprotect the pfp-esters. Given the fact that esters are significantly more reactive than amides, the selective cleavage should be possible, particularly due to the enhanced stability of the amide in 2-Amino-DDP. The synthesis scheme using the latter approach is shown in Figure 56.

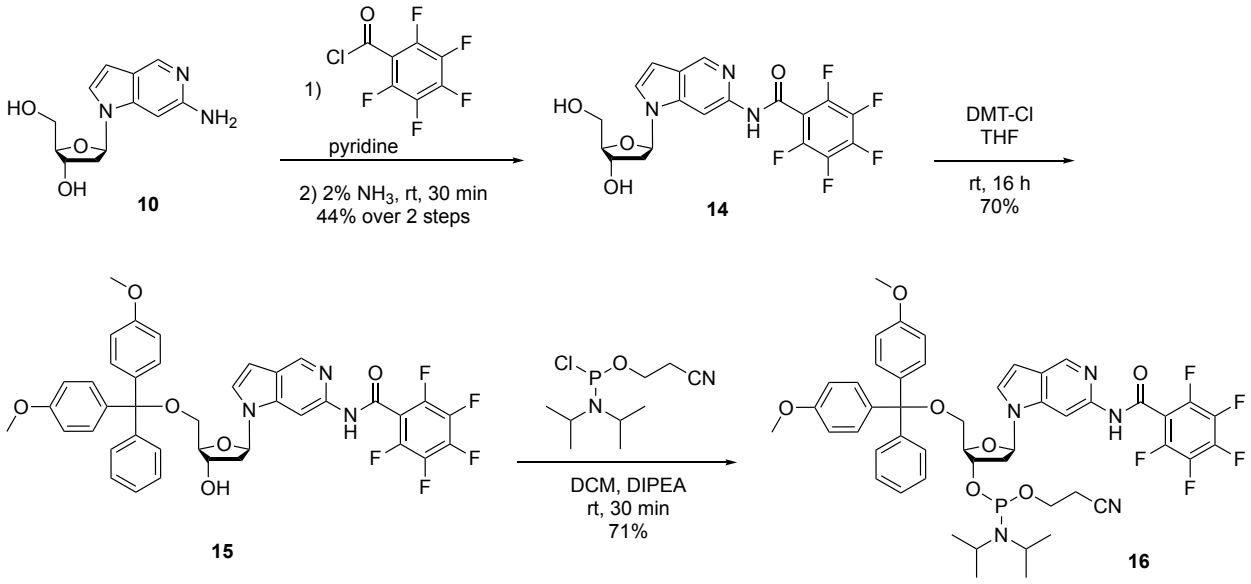



*Figure 56: Chemical synthesis of the 2-Amino-DDP Phosphoramidite with using a pentafluorobenzoyl (pfb) protecting group.*

The pentafluorobenzoyl-protection with the subsequent selective cleavage of the esters was successful and the pfb-protected 2-Amino-DDP **14** could be isolated with 44% purified yield over two steps. The DMT-protection to **15** and conversion to the phosphoramidite **16** worked with yields of 70% and 71% respectively. The phosphoramidite **16** was used to synthesize the 10mer DNA TCAGXGTAAG (X = 2-Amino-DDP). Notably, while the pfp-protecting group could be cleaved from the free nucleoside with ammonia, its deprotection from the oligonucleotide was highly inefficient with a maximum of 10-15% deprotection. Higher temperatures and the use of 100 mM NaOH solution led to DNA degradation. It is suspected that enhanced

base stacking within the oligonucleotide increases the stability of the protecting group relative to the nucleoside.

To test other electron-withdrawing groups, dichloroacetyl- and trichloroacetyl-protected 2-Amino-DDP were synthesized (Compounds **10c** and **10d** in the experimental part). During deprotection tests on the nucleoside, it was observed that the stability is still too high. The dichloroacetyl- and trichloroacetyl-protected 2-Amino-DDP were both fully stable in $NaHCO_3$-solution and solutions of ammonia in water or methanol at room temperature for 16 h. Treatment with 25% ammonia at 60 °C for 6 h led to minimal deprotection. Using 1% HCl in MeOH at room temperature for 16 h resulted in 40% deprotection. Nevertheless, there is still a need for a protective group that can be readily cleaved with ammonia, as this enables a simultaneous cleavage of the other protective groups without damaging the DNA.

To overcome this limitation, an acetal-based protecting strategy was employed to synthesize an amidine. Amidines have been described to be useful for solid-phase DNA synthesis[136] and are cleavable under basic conditions, independently of the aromatic electron density. It was shown that *N*,*N*-dibutylformamide dimethylacetal is a suitable protecting group for 2,6-Diamino purine[137], so it was synthesized in two steps and used to protect 2-Amino-DDP. The synthesis scheme is shown in Figure 57.

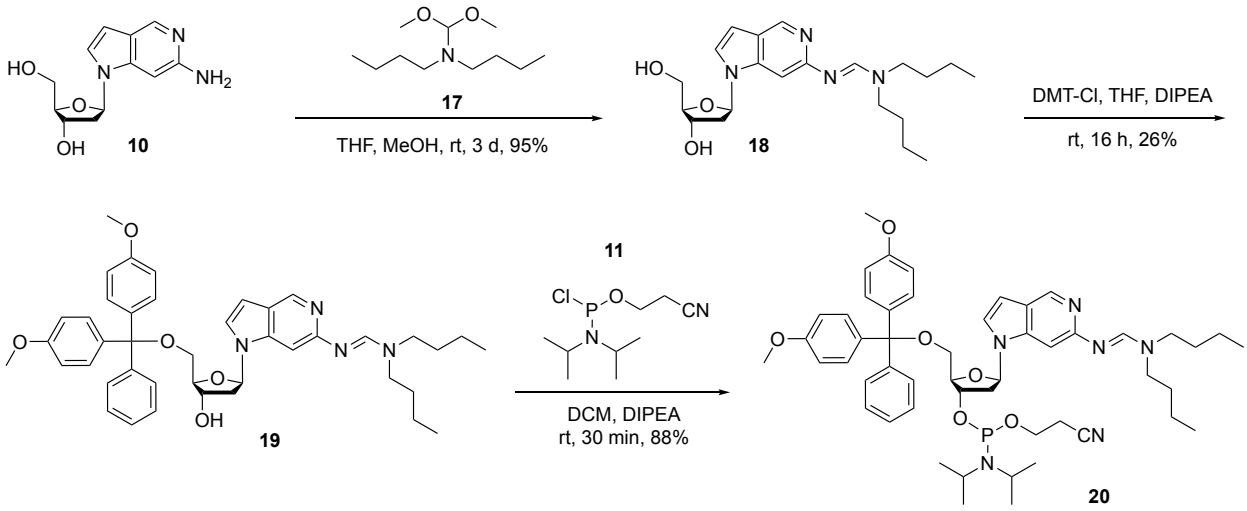



*Figure 57: Chemical synthesis of the 2-Amino-DDP phosphoramidite using dbf-dma protecting group. This figure was modified from: Eric Ogel, Sidney Becker, "Synthesis and Incorporation of a pH-Responsive Nucleoside into DNA Sequences", Chembiochem **2025**, e70086, under a CC-BY 4.0 license.[3]*

To produce *N*,*N*-dibutylformamide dimethylacetal **17**, which was used as protecting agent, *N*,*N*-dibutylformamide was first methylated with dimethylsulfate and then nucleophilically attacked with methanolate.[138] The acetal was purified by distillation through a fractionating column in a high vacuum and used to protect the 2-Amino-DDP nucleoside selectively at the exocyclic nitrogen as amidine **18**. Subsequently, the 5'-hydroxyl group was DMT-protected to get **19**.[139] In the last step, the protected

nucleoside was treated with a phosphorochloridite to convert the 3' end to phosphoramidite **20** [140], before it was immediately purified by column chromatography and directly incorporated into DNA oligonucleotides on a solid-phase DNA synthesizer. The phosphoramidite could be obtained in over 90% purity, determined by $^{31}P$ NMR and in the $^{1}H$ NMR (Experimental part).

The 10mer DNA TCAGXGTAAG (X = 2-Amino-DDP) was synthesized via Solid Phase Oligonucleotide Synthesis (SPOS). A simplified representation of the process is shown in Figure 58 and discussed below.

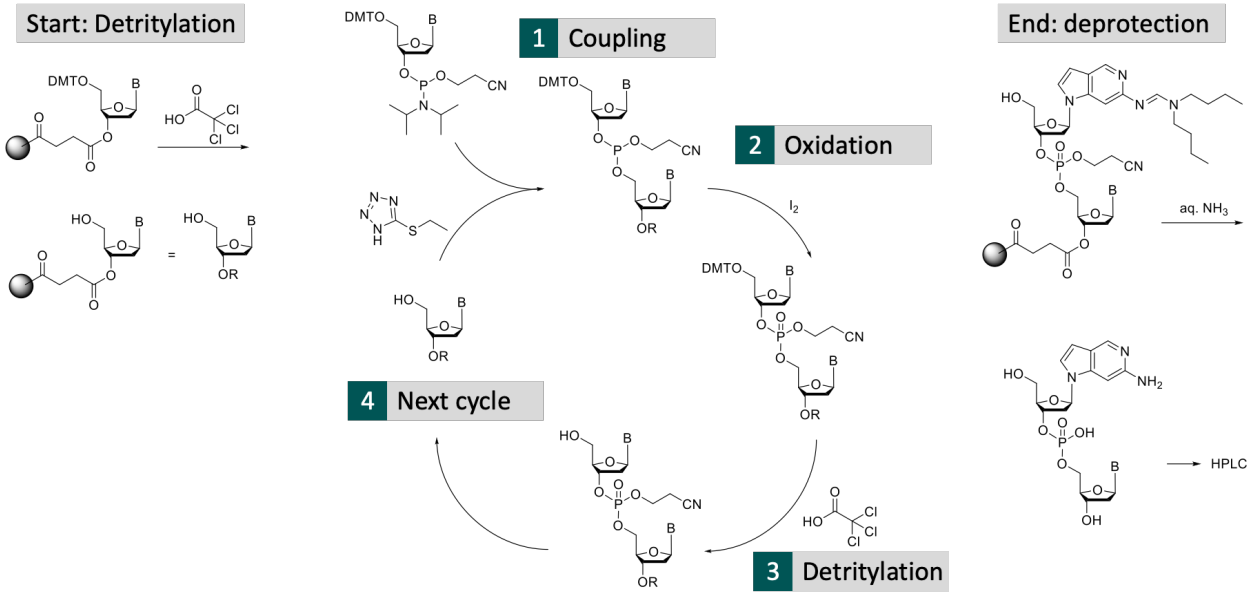



*Figure 58: Schematic representation of the steps involved in Solid Phase Oligonucleotide Synthesis (SPOS).*

Solid phase oligonucleotide synthesis (SPOS) is the standard method for the chemical synthesis of short oligonucleotides. In this process, synthesis occurs on a solid support, in this case CPG (controlled pore glass). The first nucleoside is connected to CPG with an ester linker attached to its 3' end. The phosphoramidite synthesis process involves a series of cyclically repeated steps, with the addition of one single nucleotide to the 5' end of the growing chain at a time.

The initial step in a synthesis cycle is the acidic removal of the 5'-DMT group from the nucleotide, which is termed detritylation. Consequently, the 5' hydroxyl group becomes available for the next reaction. In the following coupling step, a nucleotide phosphoramidite reacts with the free 5'-hydroxyl group of the growing oligonucleotide. To accelerate the reaction, the phosphoramidite is activated by making it more electrophilic using an activator. In this work, 5-(Ethylthio)-1*H*-tetrazole is used. The coupling time is 25 seconds for standard phosphoramidite, but for the modified 2-Amino-DDP phosphoramidite, it was allowed to couple for 600 seconds.

While the conversion rate is usually very high (typically over 98%), not all oligonucleotide chains undergo successful coupling. Therefore, a following capping step is performed through the acetylation of the free hydroxyl groups with a highly reactive mixture of acetic anhydride and *N*-methylimidazole. Capping prevents the reaction of free hydroxyl groups with other reagents, reducing error products forming. In the following oxidation step, the P(III) phosphite-triesters are oxidized to stable P(V) phosphate triesters. The cyanoethyl group prevents undesirable reactions at phosphorus during the synthesis. After the oxidation step, the detritylation of the 5'-DMT-group follows and the next cycle can begin.

The cycles of coupling, capping, oxidation and detritylation are repeated until the desired oligonucleotide sequence is obtained. In a final step, all protective groups are removed in ammonia solution. Under the used conditions, the oligonucleotide is also detached from its solid carrier by hydrolysis of the ester bond, as shown in Figure 58 on the right side.

Using the described Solid Phase Oligonucleotide Synthesis method, the 10mer DNA TCAG**X**GTAAG (X = 2-Amino-DDP) was synthesized. 128 µg (41.6 nmol) were isolated after HPLC. The deconvoluted HRMS-spectrum showing the 10mer product is shown in Figure 59.

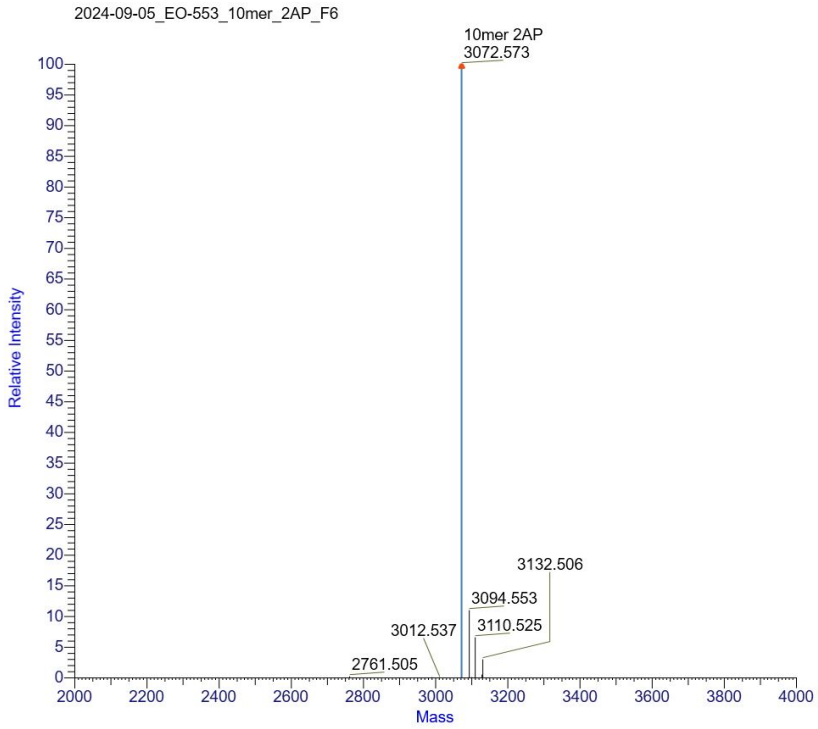



*Figure 59: Deconvoluted HRMS-spectrum of the 10mer DNA TCAGXGTAAG (X = 2-Amino-DDP). The calculated mass is 3072.580, the detected deconvoluted mass is 3072.573.*

## 6.6 CD Spectroscopy

The 10mer TCAGXGTAAG (X = 2-Amino-DDP) was then used for CD measurements. It was annealed to the corresponding template (CTTACCCTGA or CTTACTCTGA) at 3 µM oligonucleotide concentration in 10 mM phosphate buffer with pH from 6.0 to 9.5 and a total of 250 mM NaCl. The CD spectrum of double-stranded DNA functions as a tool for the identification of the secondary structure of DNA, while the presence of a strong positive peak at 275-280 nm and a negative peak at 245-250 nm is indicative of a B-DNA structure.[141] The CD spectra of the 10mer with C- or T-template are shown in Figure 60.

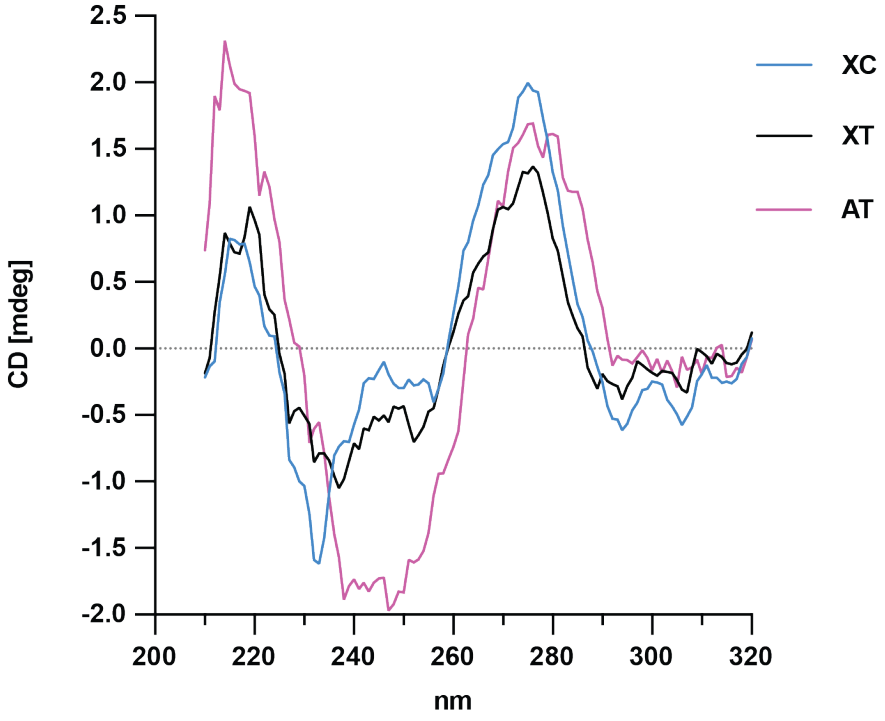



*Figure 60: The CD spectra of the canonical 10mer duplex with AT and the 2-Amino-DDP containing 10mer duplexes with C-template (XC) or T-template (XT) at a pH of 6.5 and 20 °C are presented. The displayed results are the mean of three consecutive measurements. This figure was modified from: Eric Ogel, Sidney Becker, "Synthesis and Incorporation of a pH-Responsive Nucleoside into DNA Sequences", Chembiochem **2025**, e70086, under a CC-BY 4.0 license.[3]*

All three spectra best match that of a B-DNA in terms of the relative position of the maxima and minima. The intensity of the spectrum for the base pair with T is slightly lower than that for the base pair with C. This hints towards the partial protonation of 2-Amino-DDP and the resulting repulsion disrupts the duplex structure.

## 6.7 Melting point measurements

The 10mer used for melting point measurements had the sequence TCAGXGTAAG. It was annealed to the corresponding template CTTACCCTGA or CTTACTCTGA at 3 µM oligonucleotide concentration in 10 mM phosphate buffer with pH from 6.0 to 9.5 and a total of 250 mM NaCl. The three melting curves shown in Figure 61 are presented as examples.

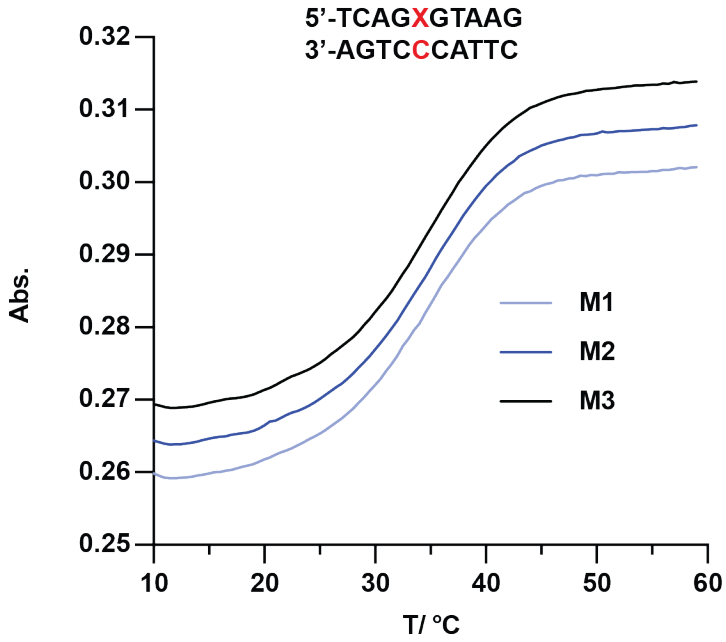



*Figure 61: The spectra display three melting curves of the 10mer duplex 'XC' at pH 6.5, recorded in succession and presented as illustrative examples. This figure was modified from: Eric Ogel, Sidney Becker, "Synthesis and Incorporation of a pH-Responsive Nucleoside into DNA Sequences", Chembiochem **2025**, e70086, under a CC-BY 4.0 license.*[3]

The melting curves were plotted using nonlinear regression, based on the difference of the absorption at 260 nm and 420 nm, to give the corresponding melting points. It has been shown that the melting point, and consequently, the stability of the new base pair are influenced by the pH value.

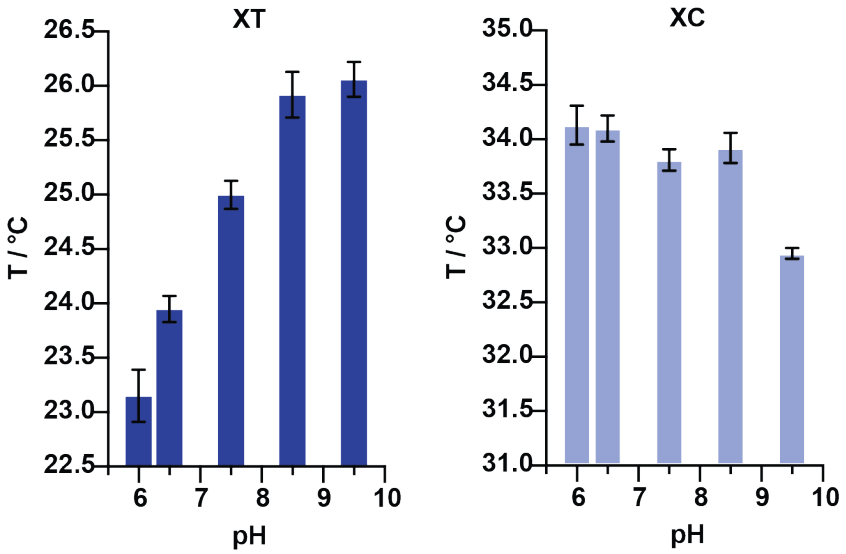



*Figure 62: The representation of melting points at five different pH values has been measured by UV absorption change at 260 nm. The results from the duplex are shown on the left, in which 2-Amino-DDP forms a base pair with T, and on the right for the base pair with C. This figure was modified from: Eric Ogel, Sidney Becker, "Synthesis and Incorporation of a pH-Responsive Nucleoside into DNA Sequences", Chembiochem **2025**, e70086, under a CC-BY 4.0 license.[3]*

It was observed that the 2-Amino-DDP building block does not demonstrate large pH-dependent changes in duplex stability when paired with cytosine. The thermal stability exhibited a minimal decrease, from 34.1 °C at pH 6.0 to 32.9 °C at pH 9.5 (Figure 62). This observation may be explained by the formation of a wobble base pair with cytosine following deprotonation.

The 2-Amino-DDP:T base pair, in contrast, demonstrates clear pH sensitivity, with melting temperatures rising from 23.2 °C at pH 6.0 to 26.1 °C at pH 9.5 (Figure 62). This phenomenon can be explained by the repulsion between the N1-H of 2,6-Diamino-DDP and the N3-H of T (Figure 63).

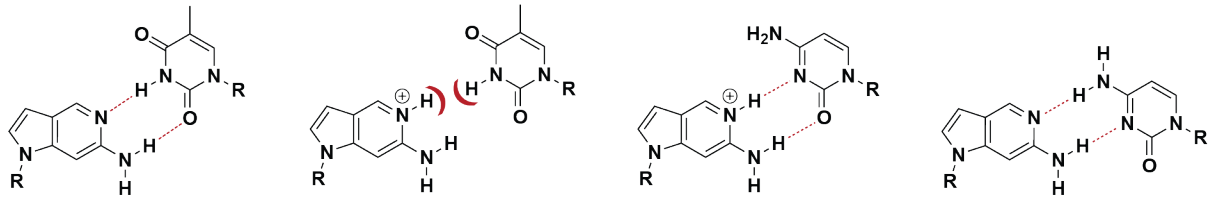



*Figure 63: Watson-Crick base pairing of 2-Amino-DDP pair with T and C, and potential wobble pair with C.*

The deprotonation of N1 at higher pH values results in the formation of a new hydrogen bond. In contrast to the base pair with cytosine, this base pair is incapable of forming a stabilizing wobble base pair upon protonation. Furthermore, the melting temperatures of the 2-Amino-DDP:T pair are considerably lower than those of the 2-Amino-DDP:C pair. This reduced thermal stability may be attributed to changes in duplex

dynamics and structure, similar to the effects observed with the 2-aminopurine:T pair. Although 2AP does not induce premelting, it is known to influence the stability of several adjacent base pairs and to decrease the overall melting temperature of the duplex.[142] A similar effect could explain the low melting point of the 2-Amino-DDP:T pair.

For comparison, we examined an oligonucleotide featuring a canonical A:T base pair at the same position, which had a slightly higher and nearly constant melting temperature, ranging from 34.8 °C at pH 6.0 to 35.5 °C at pH 9.5, consistent with the expected lack of pH sensitivity.

Notably, the significantly lower melting temperature of the 2-Amino-DDP:T pair offers a unique advantage. It enables selective sequence targeting while leaving canonical sequences which already have low melting temperatures unaffected. These properties enable the development of sequence-independent, pH-responsive DNA motifs that function orthogonally to natural DNA sequences, thereby creating new opportunities for dynamic applications in DNA nanotechnology.

## 6.8 2-Amino-DDP: Conclusion and future perspectives

This chapter discussed the development of a programmable nucleoside with pH-responsive base pairing to enable applications in DNA nanotechnology. This approach aimed to mimic the dynamic behavior of biological molecular machines driven by proton gradients. Achieving this objective required careful tuning of the p$K_a$ value. With a p$K_a$ of 7.3, 2-Amino-dideaza purine (2-Amino-DDP) should have the capacity to switch its pairing partners based on pH.

The modified nucleobase 2-Amino-DDP was synthesized as triphosphate and as phosphoramidite. Oligonucleotides which included 2-Amino-DDP were synthesized to enable the investigation of its base pair with C and T. Melting point measurements revealed a clear pH-dependence and a strong destabilization with T at a pH under 7. Consequently, this could enable DNA origami structures to respond dynamically to environmental changes. Furthermore, the pH-dependent incorporation of 2-Amino-DDP triphosphate opposite C and T templates was confirmed by using multiple DNA polymerases.

These findings have the potential for constructing DNA-based molecular machines that utilize proton gradients, and may also enhance acid-base catalysis in nucleic acid enzymes, analogous to histidine's function in proteins.

In future experiments, it may be advantageous to synthesize a series of oligonucleotides, each containing a varying number of 2-Amino-DDP, under different sequence contexts. Melting point measurements could be used to investigate the destabilizing influence of protonation in more detail. Consequently, 2-Amino-DDP phosphoramidite is a promising candidate for employment as a novel building block for a wide variety of DNA nanotechnology applications, ranging from molecular machines to nanoparticles and DNAzymes.

# 7. 2,6-Diamino-dideaza purine (2,6-Diamino-DDP)

The third research project is focused on the investigation of a new base pair with a three hydrogen-bond-donor base to expand the genetic alphabet. For this approach, the aim is to synthesize a nucleobase that can form a base pair with labeled 5fC-bases, such as 5fC-M or 5fC-A, which have been described in chapter 5. As described in chapter 5 and depicted in Figure 64 on the left side, labeled 5fC-bases form just one hydrogen bond with adenine.[75] Nevertheless, the majority of polymerases demonstrate a preference for incorporating adenine in the opposite position to 5fC-M.[75] This behavior occurs because the base pair forms a Watson-Crick geometry inside the polymerase that is not observed when paired with guanine.[75] In most methods for 5fC detection, the modification is identified by sequencing the sample before and after a (bio)chemical conversion and comparison to detect a C-to-T transition.[8-9, 19, 43] A problem with methods that rely on C-to-T conversion is the potential information loss due to the occurrence of natural C-to-T mutations[143], which can be challenging to distinguish from C-to-T transitions at epigenetic sites.[8] The expansion of the genetic alphabet with a base that pairs with labeled 5fC-bases could allow for the direct readout of unmodified and modified cytosines simultaneously.[8] This approach eliminates the need for comparative analysis and lowers the risk of information loss.[8] An unnatural base pair with a hydrogen-bonding pattern that differs from the standard Watson-Crick pairs could support this approach.[8]

Recently, 3,7-Dideazaadenine (D) has been shown to form a base pair with 5fC-M and to allow direct 5fC detection by Sanger sequencing.[8] The 5fC-M:D base pair is depicted in Figure 64 (middle) and shows that the second hydrogen bond is formed by the protonated N1 in D, which is the predominant form below pH 8.6.[8, 144]

In this work, we modified D by adding an amino group to the 2-position to create 2,6-diamino-3,7-dideazapurine (2,6-Diamino-DDP). It should also be protonated at physiological pH and form an even more stable base pair with 5fC-M, fitting with their three hydrogen bond acceptors[75], as depicted in Figure 64 (right).

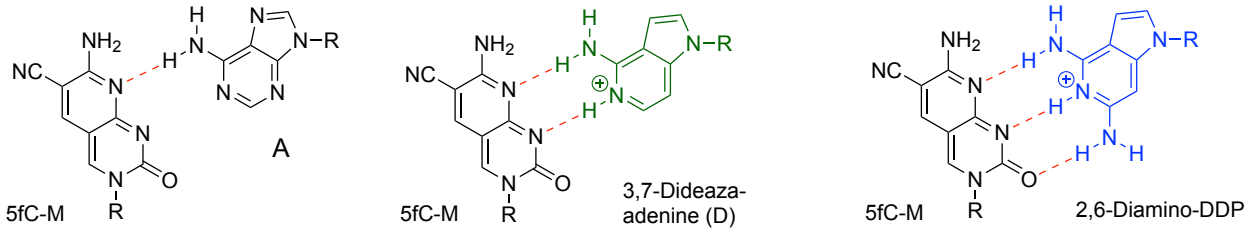



*Figure 64: Model of 5fC-M with potential base pair partners, forming one hydrogen bond with adenine, two with 3,7-Dideazaadenine (D, green)[8] and three with 2,6-Diamino-DDP (right, blue).*

The same concept could also be applied to acetylacetone-labeled 5fC-A, as discussed in chapter 5 of this thesis. The novel base pair of 5fC-A and 2,6-Diamino-DDP (Figure 65, right side) could have a high stability while 5fC-A additionally avoids the formation of a wobble base pair with G.

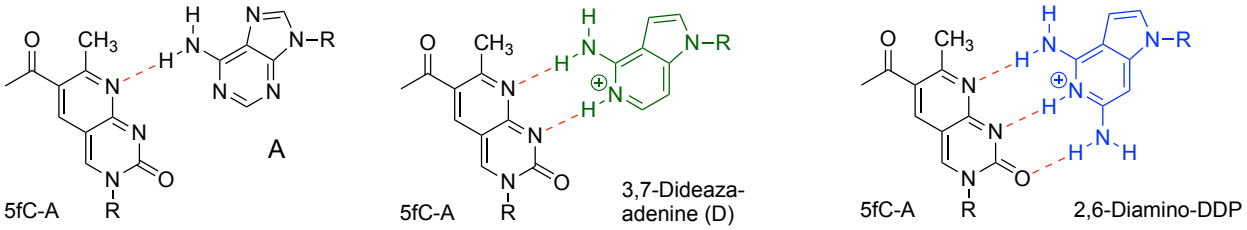



*Figure 65: Model of 5fC-A with potential base pair partners, forming one hydrogen bond with adenine, two with 3,7-Dideazaadenine (D, green) and three with 2,6-Diamino-DDP (right, blue).*

Overall, the combination of the novel base 2,6-Diamino-DDP with labeled 5fC-bases 5fC-M or 5fC-A could allow for an advanced 5fC-sequencing approach. Furthermore, it could enable the detection of 5mC and 5hmC, as there are reported processes that convert 5mC into 5fC (photo-catalytic oxidation by $Na_4W_{10}O_{32}$)[145] and 5hmC into 5fC (oxidation with $KRuO_4$)[73].

In the future, this unnatural base pair could be employed to read a multitude of epigenetic information. Additionally, DNA data storage[90] or aptamer generation[86] are possible applications for this additional base pair.

## 7.1 2,6-Diamino-DDP phosphoramidite

The first aim of this project was the synthesis of 2,6-Diamino-DDP as a phosphoramidite. Utilizing an optimized synthesis route that is shown in Figure 66, the production of over 100 milligrams of the benzoyl-protected 2,6-Diamino-DDP phosphoramidite was achieved in five steps.

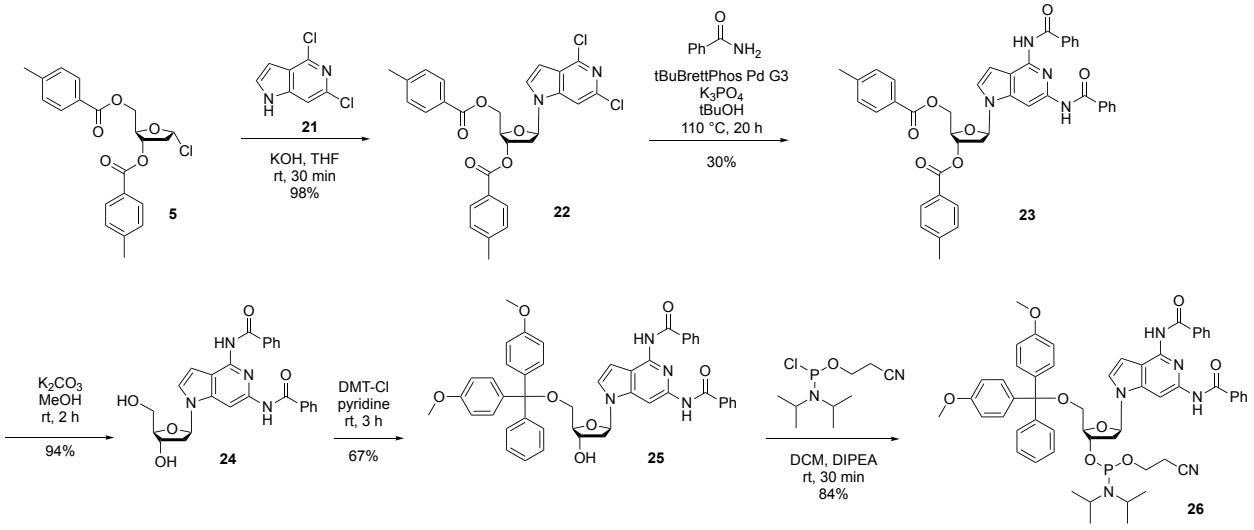



*Figure 66: Synthesis of 2,6-Diamino-DDP as DMT-protected phosphoramidite.*

In the first step, 3,7-Dideaza-2,6-dichloropurine **21** was glycosylated with 3,5-Di-*O*-benzoyl-2-deoxy-α-D-ribosyl chloride **5** (*Hoffer's chlorosugar*). In the following Buchwald-Hartwig-Amidation, the product was coupled with benzamide to convert the chlorine residue to amide **23**. When using the published procedure's reaction conditions[128], the product could be detected via LC-MS, but it could not be isolated in sufficient amounts. Therefore, the reaction procedure was optimized with a variety of conditions including different solvents (*tert*-butanol, toluene or 1,4-dioxane), Pd-sources (acetate or dba-complex), ligands (Xantphos or tBu Brettphos Pd G3) and bases ($Cs_2CO_3$ or $K_3PO_4$), as shown in Table 7.

*Table 7: Tested conditions for the Buchwald-Hartwig-coupling of 22 to 23. Benzamide (6.00 eq.), base (5.00 eq.), catalyst (0.02 eq.) and 22 (1.00 eq.) and the solvent were heated in a dry pressure tube.*

| Base | Catalyst+Ligand | Solvent | Temp. | Duration | Yield (UV) |
|---|---|---|---|---|---|
| $K_3PO_4$ | $Pd(OAc)_2$ + Xantphos | tBuOH | 110 °C | 24 h | 0% |
| $Cs_2CO_3$ | $Pd(OAc)_2$ + Xantphos | Dioxane / water 1000:1 | 100 °C | 16 h | 0% |
| $Cs_2CO_3$ | $Pd_2(dba)_3$ + Xantphos | dioxane | 100 °C | 3 d | traces |
| $Cs_2CO_3$ | $Pd_2(dba)_3$ + Xantphos | toluene / dioxane 1:1 | 100 °C | 3 d | 0% |
| $K_3PO_4$ | $Pd_2(dba)_3$ + Xantphos | toluene / dioxane 1:9 | 100 °C | 3 d | 0% |
| $Cs_2CO_3$ | $Pd_2(dba)_3$ + tBuBrettPhos | dioxane | 100 °C | 16 h | 0% |

| | | | | | |
|---|---|---|---|---|---|
| $Cs_2CO_3$ | tBuBrettPhos Pd G3 | dioxane | 100 °C | 16 h | traces |
| $Cs_2CO_3$ | $Pd_2(dba)_3$ + tBuBrettPhos | tBuOH | 80 °C | 16 h | 0% |
| $Cs_2CO_3$ | tBuBrettPhos Pd G3 | tBuOH | 80 °C | 19 h | traces |
| $Cs_2CO_3$ | $Pd_2(dba)_3$ + Xantphos | dry toluene | 100 °C | 10 h | 0% |
| $Cs_2CO_3$ | tBuBrettPhos Pd G3 | dry toluene | 100 °C | 10 h | 0% |
| $Cs_2CO_3$ | tBuBrettPhos Pd G3 | dry tBuOH | 80 °C | 29 h | 11% |
| $Cs_2CO_3$ | Xantphos Pd G3 | dioxane | 80 °C | 29 h | traces |
| $K_3PO_4$ | tBuBrettPhos Pd G3 | dry tBuOH | 100 °C | 16 h | 18% |
| $Cs_2CO_3$ | Xantphos Pd G3 | dioxane | 130 °C | 5 h | 0% |
| $Cs_2CO_3$ | tBuBrettPhos Pd G3 | dry tBuOH | 80 °C | 3 d | traces |
| $K_3PO_4$ | tBuBrettPhos Pd G3 | dry tBuOH | 110 °C | 20 h | 35% |

The reaction with tBu Brettphos Pd G3 and $K_3PO_4$ in *tert*-butanol at 110 °C achieved the highest yield of **23** with 35% UV yield and 30% isolated yield. After the following deprotection of the sugar unit to **24**, the 5'-OH-group of the deoxyribose was protected using a DMT (dimethoxytrityl) protecting group to give **25**. In the next step, the 3'-OH-group of the nucleoside reacted with 2-Cyanethyl-*N*,*N*-diisopropylchlorphosphoramidite to give phosphoramidite **26**.

The purified phosphoramidite **26** was sent to the company *ATDbio* to be incorporated into the 10mer-oligonucleotide GAAGCXTGC using solid phase synthesis. The incorporation was successful and the 10mer was used for melting point measurements to determine the thermodynamic stability. For that purpose, a 10mer-oligonucleotide with the matching sequence containing 5fC was labeled with malononitrile, according to the CLEVER-Seq protocol[9], to get TGCA-**5fC-M**-GCTTC. Together with canonical 10mers, the melting points of seven different DNA duplexes were measured at pH 7.3. The results are shown in Figure 67.

| Oligo 1 | Oligo 2 | mp/ °C | ± |
|---|---|---|---|
| TGCA**M**GCTTC | GAAGC**X**TGCA | 37,73 | 0,43 |
| TGCA**M**GCTTC | GAAGC**A**TGCA | 46,24 | 0,25 |
| TGCA**M**GCTTC | GAAGC**G**TGCA | 45,76 | 0,24 |
| TGCA**C**GCTTC | GAAGC**X**TGCA | 36,38 | 0,25 |
| TGCA**T**GCTTC | GAAGC**X**TGCA | 33,37 | 0,50 |
| TGCA**C**GCTTC | GAAGC**G**TGCA | 50,72 | - |
| TGCA**T**GCTTC | GAAGC**A**TGCA | 46,26 | - |

*Figure 67: DNA melting points of different 10mers, partially containing 2,6-Diamino-DDP (X) and 5fC-M (M).*

While the melting point of the duplex with 2,6-Diamino-DDP opposite 5fC-M was at 37.7 °C, the melting points of the duplexes with 2,6-Diamino-DDP opposite C and T were at 36.4 °C and 33.4 °C, respectively. The DNA duplexes that did not contain 2,6-Diamino-DDP (5fC-M opposite C and G, as well as C-G and A-T) had a melting temperature between 45-51 °C. This substantial difference of over 10 °C was found to be inconsistent with our expectations, therefore the oligonucleotide provided by ATDbio was subjected to additional analysis using high-resolution MS/MS. The detected mass revealed that the amide protecting groups on 2,6-Diamino-DDP were still attached, while all other bases were deprotected.

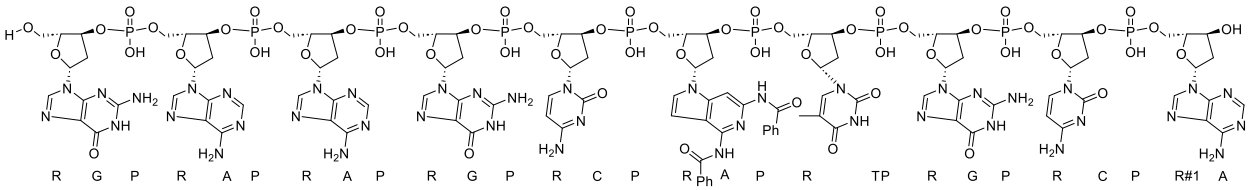



*Figure 68: 10mer DNA-strand with both protective groups still attached to 2,6-Diamino-DDP.*

The remarkable stability of the amide group underscores the substantial impact of the nitrogen count within the aromatic system. The higher electron density of the system in comparison to 2-Amino-DDP is predicted to result in an elevated p$K_a$ value, along with an enhanced stability of the amides. A series of conditions were tested in order to deprotect the DNA strand. The effects of concentrated ammonia or methylamine solutions and NaOH were examined at varying concentrations and temperatures. In the presence of 7 M ammonia in methanol at 80°C or 0.1 M NaOH at 60°C, mono-deprotection was achieved after two to three days. However, no conditions were identified under which both protective groups were cleaved without complete degradation of the DNA. Therefore, an alternative protective group strategy was required.

The strategy involved the direct removal of the benzamide groups following the Buchwald-Hartwig-Amidation, with the subsequent isolation of 2,6-Diamino-DDP as nucleoside, allowing for the attachment of a protecting group of choice to the exocyclic amines. A difficulty that arises from the described strategy is the strong electrophilic nature of most protecting group reagents, such as acid anhydrides or

chloroformates. They react not only with the amines but also with the free 5'- and 3'-hydroxyl groups of the deoxyribose unit. In order to selectively protect the amines with such groups, it is first necessary to protect the hydroxyl groups with a reagent that is selective for alcohols. For this purpose, silyl groups can be used, that must be deprotected again afterwards. Given that each additional step in a multi-step synthesis results in a decreased overall yield, the employment of a protecting group that is selective for amines is advantageous. One example known from the literature[146] is the commercially available *N,N*-Dimethylformamide dimethylacetal (dmf-dma), which reacts with amines to form amidines, while hydroxyl groups remain unprotected. This acetal-based strategy is shown in Figure 69.

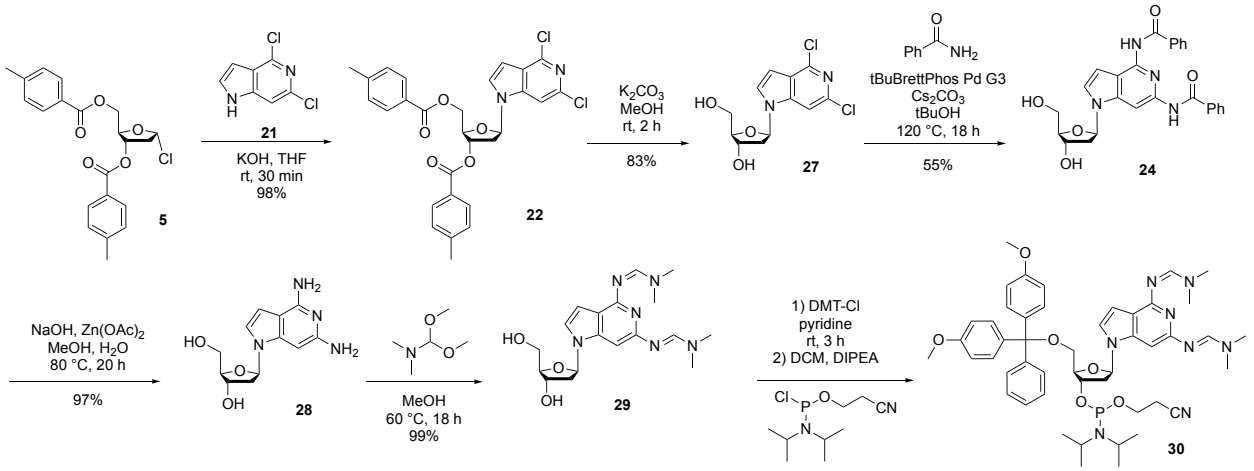



*Figure 69: Acetal-protection-based synthesis plan for 2,6-Diamino-DDP as DMT-protected phosphoramidite.*

Another improvement in this strategy compared to the previous one is the deprotection of the glycosylated chloro-base from **22** to **27** prior to the Buchwald coupling. This approach has been shown to enhance solubility and lead to an increase in the yield of the Buchwald coupling to 55%. After the following deprotection under strongly basic conditions, the nucleoside **28** could be isolated in 97% yield. The addition of zinc acetate in catalytic amounts accelerated the reaction, potentially forming a complex that is a mimic of metallo-exopeptidase activity[147].

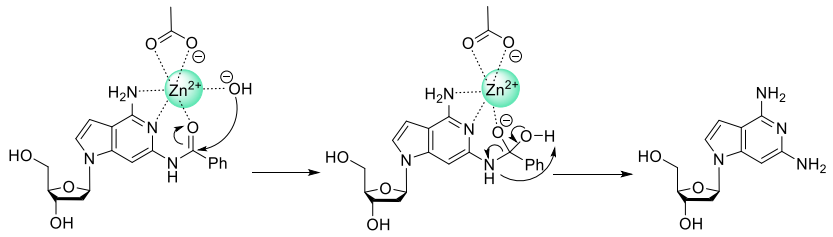


*Figure 70: Potential mechanism explaining the accelerated deprotection using $Zn^{2+}$ and high temperatures.*

The reaction of the deprotected 2,6-Diamino-DDP nucleoside with dmf-dma to get **29** worked with 99% yield. However, the amidine is very labile, especially under basic conditions, and got partially deprotected during the following DMT-protection.

In order to enhance the stability of amidines, the use of acetals with greater steric demand is advantageous. Therefore, the dibutyl- (dbf) and diisopropyl- (dif) analogues of dmf-dma were prepared and purified via distillation under high vacuum and used in reactions with the attempt to protect 2,6-Diamino-DDP.

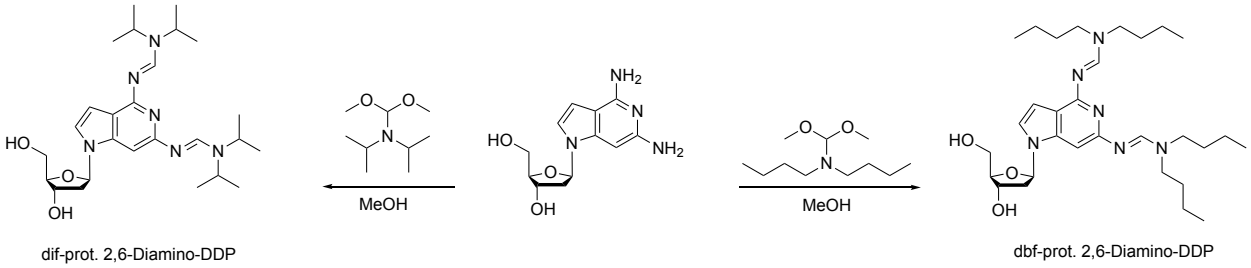



*Figure 71: Protection of 2,6-Diamino-DDP nucleoside with sterically demanding dif-dma (left) and dbf-dma (right).*

However, when using these larger acetals, a new challenge emerged. After one of the two exocyclic amines of 2,6-Diamino-DDP was protected, the protection of the other amine became significantly slower. Prolonged reaction times of up to seven days and elevated temperatures of 100°C led to the protection of both amines, but also resulted in the formation of numerous byproducts. The purification process proved challenging due to the comparable polarities of the compounds. Dif-protected 2,6-Diamino-DDP was ultimately obtained in a yield of 18%. However, a greater number of byproducts were observed in the subsequent reactions, leading to the decision to investigate alternative protective groups.

Amides remain a viable option due to their minimal reactivity with trichloroacetic acid (TCA) or oxidizers during DNA synthesis, and with $POCl_3$ during monophosphate synthesis. However, all aliphatic or aromatic amides that were previously utilized resulted in the formation of overly stable bonds. Therefore, the stronger electrophilic formamide was used to introduce *N*-formyl-protecting groups. As anticipated, the initial Buchwald-Hartwig-Amidations in *tert*-butanol exhibited low yields, even with 6.0 to 10.0 eq. of formamide. Several reaction conditions were investigated and it was observed that the most significant improvement was made by omitting the tert-butanol. The implementation of formamide as both, solvent and reagent, has been demonstrated to accelerate the reaction and increase the yields. The employment of triethylamine in conjunction with adjustments to reaction temperature and purification methods has led to a substantial enhancement in yield, reaching 85%.

In contrast to benzamide or acetamide, the smaller and less stable formamide protecting group can be cleaved at room temperature with ammonia, which renders it more suitable for this synthetic route.

Formamide-protected 2,6-Diamino-DDP **31** was observed to be stable for several minutes under acidic and oxidizing conditions, thereby indicating its potential for use in DNA synthesis. One issue that has been identified is the acetylation of the amide with the DNA capping solution (a mixture of acetic anhydride and N-methylimidazole). For the synthesis of short oligonucleotides, such as 10mers, the capping steps can be excluded. However, the reactivity of the nitrogen in the formamide also gives rise to other side reactions, including those with DMT-chloride. In order to circumvent undesirable side reactions, a conversion of to *N,N,N',N'*-tetraformyl-protected 2,6-Diamino-DDP **32** was performed through a formylation reaction by heating **31** in formic acid, as shown in Figure 72.

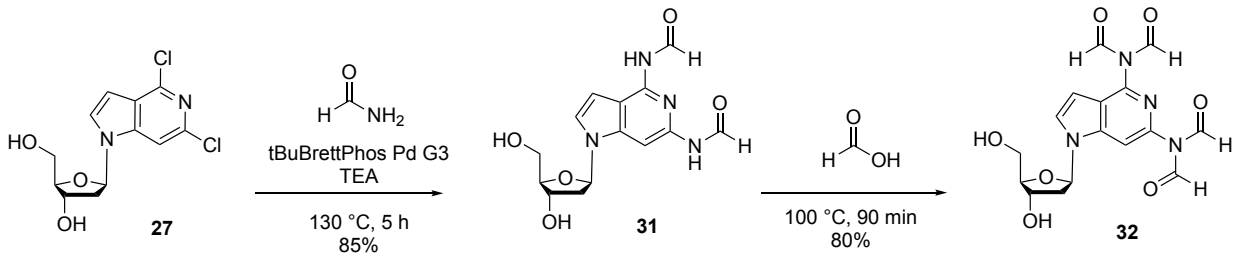



*Figure 72: Buchwald-coupling of 2,6-Dichloro-DDP-nucleoside **27** with formamide and following conversion to N,N,N',N'-tetraformyl-protected 2,6-Diamino-DDP with formic acid.*

However, the imides have been observed to react with nucleophiles, resulting in partial deprotection, even in aqueous solutions with neutral pH. Consequently, a protecting group strategy based on carbamates was tested.

In particular, carbamates that are deprotected by β-elimination appeared to be a favorable option, given that the electron-rich aromatic system present in 2,6-Diamino-DDP is not expected to interfere with the deprotonation at the C-H-acidic position. A cyanoethyl-carbamate compound was chosen as a suitable substrate for DNA synthesis. During its base-catalyzed deprotection, acrylamide and $CO_2$ are released, as shown in the following mechanism.

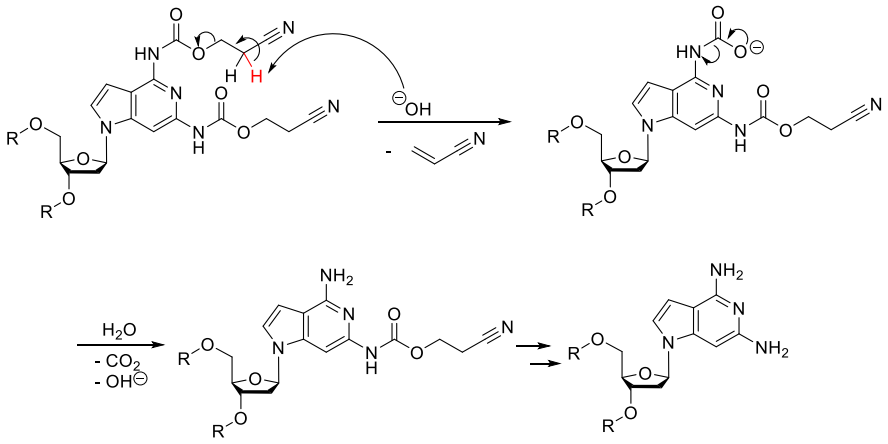


*Figure 73: Proposed cyanoethyl carbamate deprotection by a β-elimination mechanism.*

Consequently, the protecting agent 2-cyanoethyl chloroformate was prepared from *β*-cyanoethanol and triphosgene. Thereafter, 2,6-Diamino-DDP was directly protected using the prepared agent.

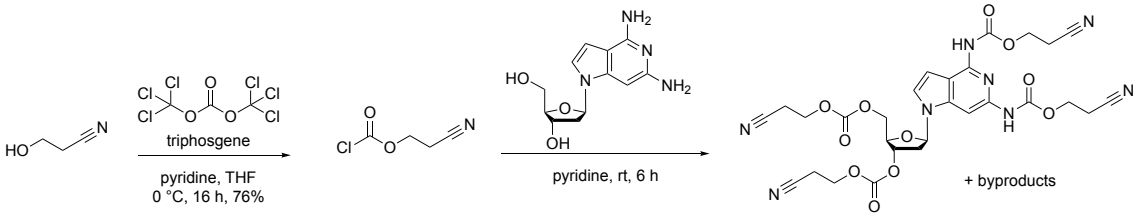



*Figure 74: Carbamate-protection of 2,6-Diamino-DDP.*

The resultant mixture contained 2,6-Diamino-DDP with one to five protecting groups. The exocyclic amines were converted to carbamates, while the hydroxyl groups were converted to carbonates. The attempt to selectively deprotect the carbonates while leaving the carbamates intact was unsuccessful because the reactivity was too similar. When using an aqueous $NaHCO_3$-solution, a $K_2CO_3$ solution in methanol, or an aqueous 25% ammonia solution at room temperature, no deprotection occurred. Heating in concentrated ammonia led to a full deprotection. Consequently, the revised strategy included the protection of the hydroxyl groups with a selective silyl group.

Direct silyl-protection of 2,6-Diamino-DDP with different reagents always led to low yields and unidentified side-products. Therefore, the amines of 2,6-Diamino-DDP were protected first with dmf-dma to **29.** Different silyl groups were attached afterwards, which proved to be a more effective approach. Figure 75 presents a model plan that utilizes a diisopropyl-based "silyl-clamp" to get **33**.

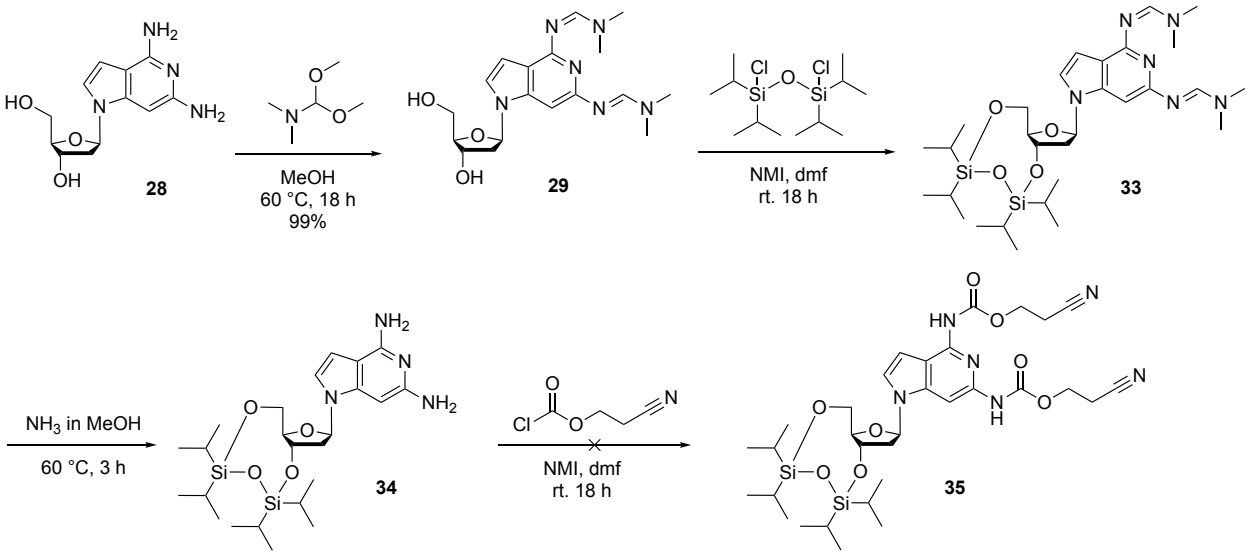



*Figure 75: Plan for the 2,6-Diamino-DDP phosphoramidite synthesis using silyl-protection for the deoxyribose and cyanoethyl carbamate for the amines.*

The amidine protecting groups could be removed with ammonia while preserving the silyl groups in **34**. However, the subsequent reaction with the chloroformate to **35** was not successful, despite the fact that all reactants had previously worked well in test reactions with other, commercially available reagents. Potential byproducts of this process could result in the quenching of the protection reaction. Despite several reactions with different conditions, synthesizing **35** was not successful.

In general, the search for one suitable protection group to protect both exocyclic amines proved to be a highly challenging task, also because the 2- and 6-Amino group have different reactivities. Consequently, a potential future direction for 2,6-Diamino-DDP would be to implement two different protection groups. A plan for this task is depicted below.

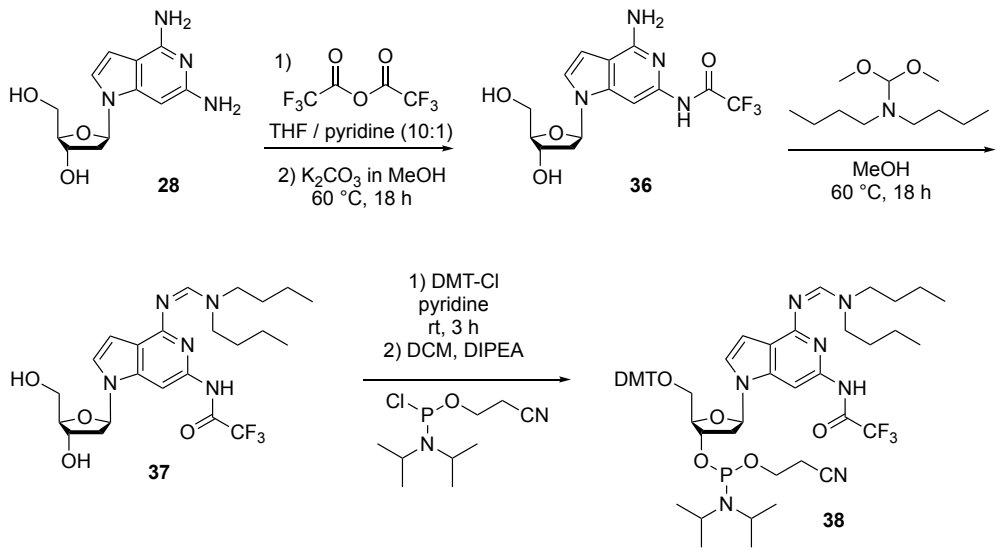



*Figure 76: Plan for the synthesis of 2,6-Diamino-DDP phosphoramidite using TFA as protecting group for the 2-Amino and dbf as protecting group for the 6-Amino group.*

Trifluoroacetic anhydride (TFA) could be used to fully protect nucleoside **28.** Because the 2-Amino position is expected and observed to be more stable, it should be possible to selectively cleave the other groups to get **36**. The 6-Amino group could then be protected as amidine, for example with dbf-dma, to get **37**. In the last two steps, **37** should be converted to the DMT-protected phosphoramidite **38**. The completion of this synthesis was not feasible within the scope of this thesis. However, it would be desirable to finalize this phosphoramidite to enable the solid phase DNA synthesis in future studies.

## 7.2 2,6-Diamino-DDP triphosphate

In order to synthesize the triphosphate of 2,6-Diamino-DDP, which is essential for its enzymatic incorporation into oligonucleotides, different synthetic strategies were conducted. To reduce the number of synthetic steps required to produce the triphosphate, some one-pot reaction methods were tested. One approach is the multi-step Ludwig-Eckstein-method, which was described in 1989.[148] The reaction of the 5'-hydroxyl group with salicylchlorophosphit is followed by two nucleophilic substitution reactions with pyrophosphate, which is added as tributylammonium salt.[149] Afterwards, the phosphorous (III) species is oxidized by iodine. The major advantage of the Ludwig-Eckstein-method is the one-flask-protocol, because all steps should work without purification in between.[148-149] Moreover, the method was also used for the synthesis of unnatural triphosphates.[82, 149-150] The major disadvantage is the low yield around 8%.[150] For this work, the Ludwig-Eckstein-method was tested using dmf-dma-protected nucleoside **29**, as shown in Figure 77.

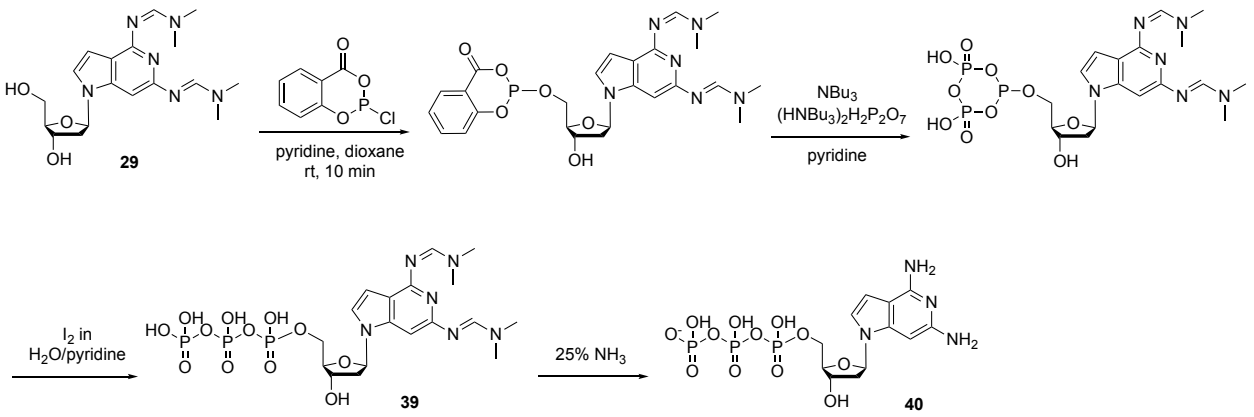



*Figure 77: Ludwig-Eckstein-method using dmf-dma-protected nucleoside to give 2,6-Diamino DDP triphosphate.*

The Ludwig-Eckstein-method proved to be ineffective for 2,6-Diamino-DDP, yielding multiple side products, while just traces of the 2,6-Diamino-DDP-triphosphate **40** were observed in LC/MS. Consequently, alternative methods were investigated. In the one-pot triphosphate synthesis shown in Figure 78, $POCl_3$ is used to phosphorylate the 5'-OH, directly followed by pyrophosphate addition.

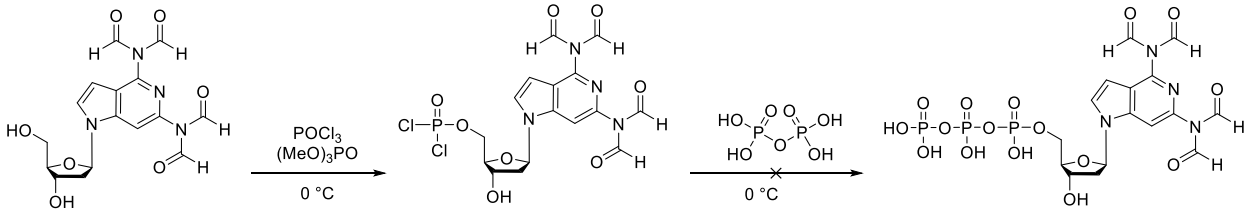



*Figure 78: Tested one-pot synthesis of fully formyl-protected 2,6-Diamino DDP triphosphate.*

However, this method was ineffective as well, yielding only byproducts. Several other conditions were tested trying to synthesize the triphosphate in a one-pot reaction directly from the nucleoside, as described in literature[148, 151]. One example is the conversion of unprotected 2,6-Diamino DDP with $POCl_3$ directly followed by pyrophosphate addition (Figure 79).

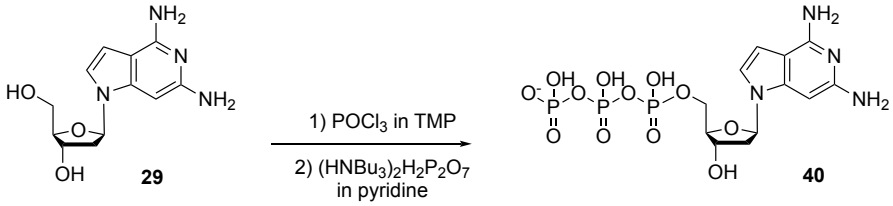



*Figure 79: Conversion of unprotected 2,6-Diamino-DDP with $POCl_3$ followed by pyrophosphate addition.*

Although the triphosphate **40** could be detected in the mass spectrum, the isolated yields were around 1%. Multiple byproducts occurred, including mono- and diphosphates, as shown in the LC/MS-spectrum in Figure 80.

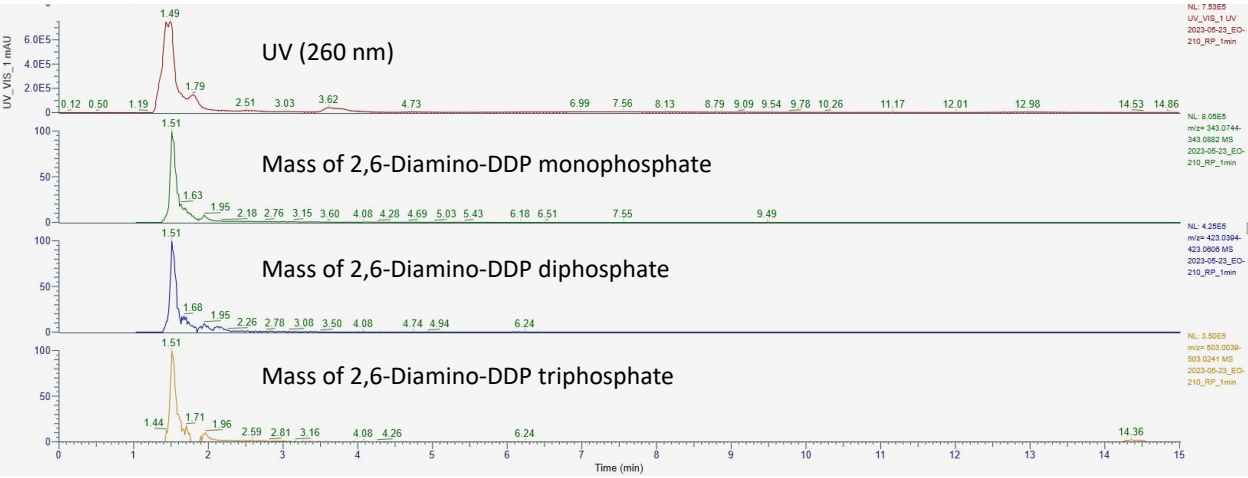



*Figure 80: LC/MS-spectrum of a one-pot triphosphate synthesis using conditions described in the literature.[148]*

Consequently, the decision was made to proceed with a step-by-step approach, employing HPLC after each step. This should ensure the isolation of each product in a pure state to minimize the occurrence of side reactions.

The first step for the synthesis of nucleoside triphosphates is the phosphorylation of the nucleoside to the 5'-monophosphate. The relatively selective phosphorylation of nucleosides in 5'-position can be achieved by using $POCl_3$ in trimethyl phosphate and was described by Yoshikawa in 1967[152].

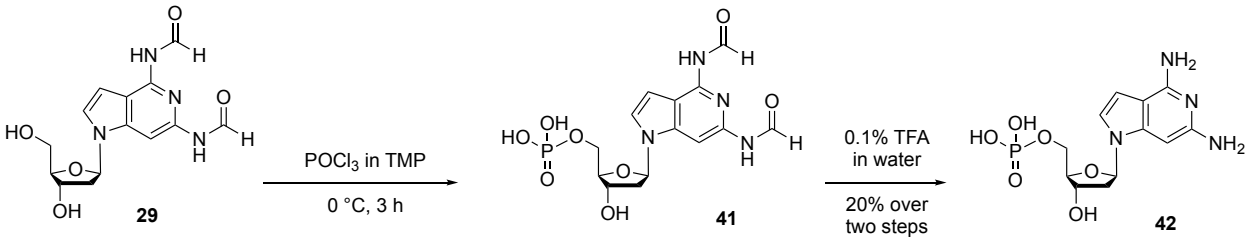



*Figure 81: Synthesis of 2,6-Diamino DDP monophosphate **42**.*

The formyl-protected 2,6-Diamino-DDP **29** was converted to the monophosphate **41**. Purification via HPLC using water and acetonitrile with 0.1% TFA each led to the deprotection of the formyl-groups and therefore to the product, as seen in the HRMS-spectrum in Figure 82. 2,6-Diamino-DDP monophosphate was isolated with 20% yield. Small amounts of 3',5'-Diphosphate species formed (m/z = 402.9764), but could be separated after the next step.

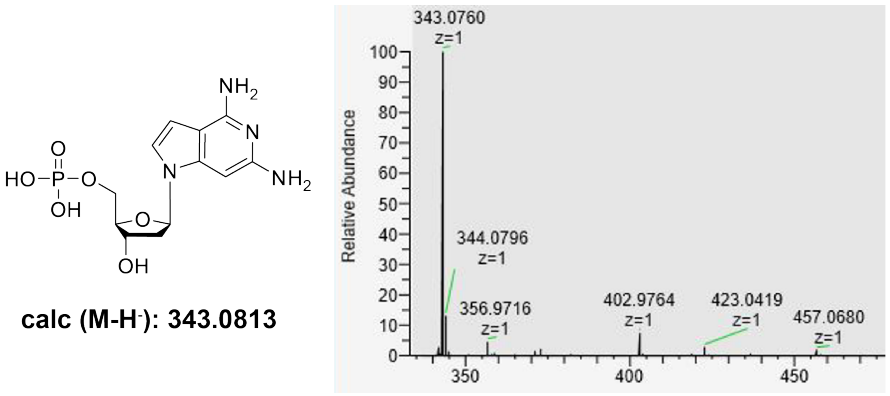



*Figure 82: Negative HRMS-spectrum of monophosphate **42**.*

The monophosphate synthesis achieved greater efficacy for benzoyl-protected 2,6-Diamino-DDP. However, further development was not pursued due to the high stability of the benzoyl group. Higher yields might be achieved by the protection of the 3'-hydroxy group to decrease the amount of undesired phosphorylation products.[149] But in order to selectively protect the 3'-hydroxyl group of a nucleoside, the 5'-OH group needs to be protected first by a bulky group, such as DMT, before the 3'-OH group can be acetylated. In the next step, the DMT protecting group is removed again.[149] Given that this would have required three additional steps, I proceeded with the monophosphate.

In the next step, the monophosphate must be activated. For this purpose, many reactions have been published.[131, 148-151, 153-154] One of the earliest ones was the morpholine-based activation, described as a viable group in 1961[155], where diphosphates were synthesized. In 1965, Reiss et al. demonstrated the synthesis of nucleoside triphosphates through condensation of the phosphoromorpholidate with inorganic pyrophosphate.[132] Based on this work, the triphosphate of the 2,6-Diamino-DDP nucleoside was conducted. The examination of a series of reactions, with and without various protecting groups (formyl, dma, and dif), did not result in the formation of the triphosphate product.

One potential explanation for this phenomenon is the enhanced reactivity of the N1 position, which has been observed to trigger side reactions with electrophilic reagents. Multiple byproducts were observed in the mass spectra of the tested reactions. Consequently, an alternative approach was implemented, which does not necessitate the use of reactive chemicals: enzymatic triphosphate synthesis.

### 7.3 Enzymatic Triphosphate Synthesis

Polyphosphate kinases (PPKs) are enzymes that use different phosphate sources to add additional phosphate groups to nucleoside mono- and diphosphates. Two of them (PPK04 and PPK05, both produced by *BioNukleo*) were tested in a variety of different conditions to convert 2,6-Diamino-DDP monophosphate

to its corresponding triphosphate. The best results were obtained when using PPK05 with sodium hexametaphosphate as phosphate source and 50 °C as reaction temperature. The maximum triphosphate concentration was reached after four days, before the triphosphate hydrolyzes back to di- and monophosphate when continuing the reaction. This hydrolysis process might be accelerated by the $Mg^{2+}$ ions present in the solution, but as $Mg^{2+}$ is a necessary cofactor for the kinase, it cannot be omitted. In an optimized workflow, fresh enzyme was added every 24 h for 4 days, before the reaction was stopped. At that point, about half of the 2,6-Diamino-DDP-monophosphate was converted into diphosphate and about 20% into triphosphate, as shown in Figure 83.

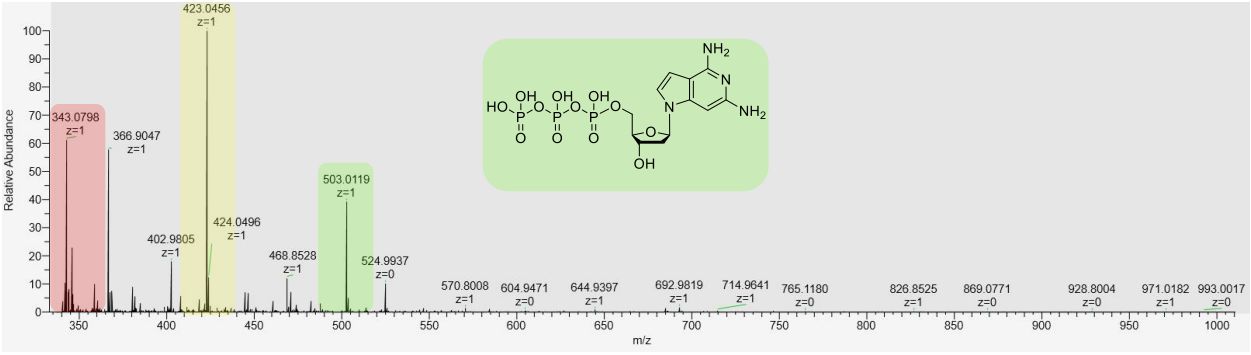



*Figure 83: Mass spectrum of the enzymatic 2,6-Diamino-DDP triphosphate synthesis after a reaction time of 4 days at 50 °C. The triphosphate is shown in green, the diphosphate in yellow and the monophosphate in red.*

The RP-HPLC-purification was challenging, because even with long run times and optimized gradients, mono-, di-, and triphosphate eluted at almost the same time. An ion exchange column could lead to better purification results, as shown for 2-Amino-DDP. However, 1% of pure triphosphate could be isolated and used in single nucleotide incorporation experiments which are described in chapter 7.4.

### 7.3.1 Enzymatic Triphosphate Synthesis using further monophosphates

As shown in the previous chapter 7.3, some polyphosphate kinases can convert unnatural dideaza-derivatives of nucleoside monophosphates into triphosphates (dNTPs). It is also possible to convert thio[156]- and chloro[156]-modified purines as well as 2,6-Diaminopurine[157] to triphosphates by using modified polyphosphate kinases. In order to evaluate the potential of deaza- and dideaza-derivatives of purines as substrates for polyphosphate kinases, a library of nucleosides and monophosphates (dNMPs) was synthesized (Figure 84).

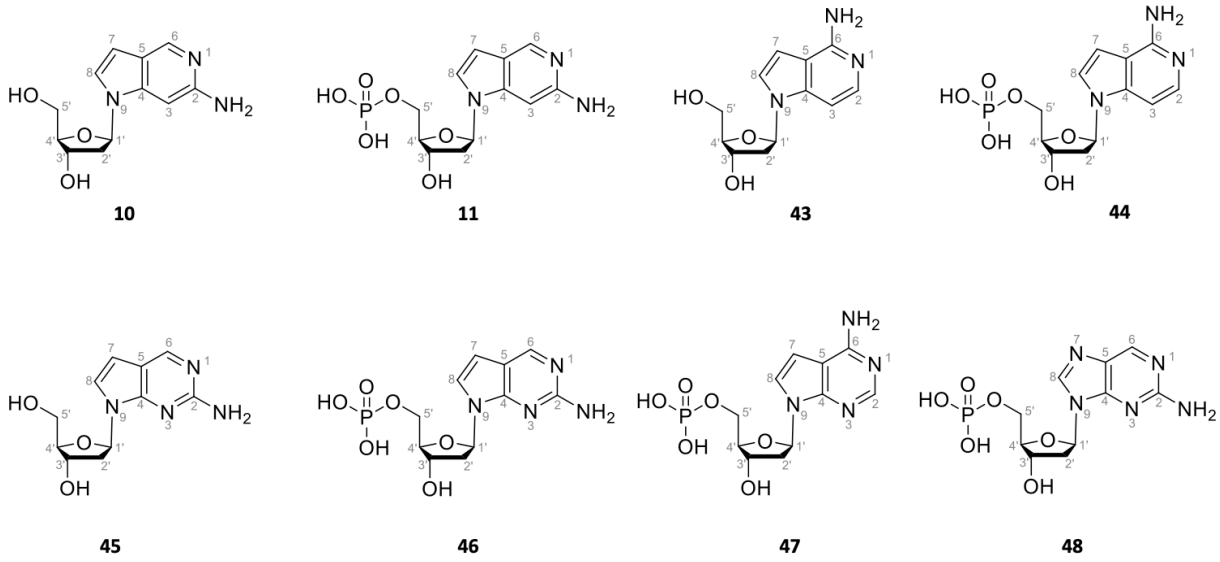



*Figure 84: Library of synthesized nucleosides and monophosphates (dNMPs).*

Our collaboration partner, Dr. Nicolas Cornelissen, is conducting experiments to assess the efficacy of the compounds as substrates for polyphosphate kinases.

## 7.4 Single nucleotide incorporation experiments

With great support from Susanne Nguyen, a master student under my supervision, we tested different polymerases to incorporate 2,6-Diamino-DDP into DNA by extending a biotin-labeled 22mer-primer. It was annealed to a malononitrile-labeled 5fCM-25mer-template.

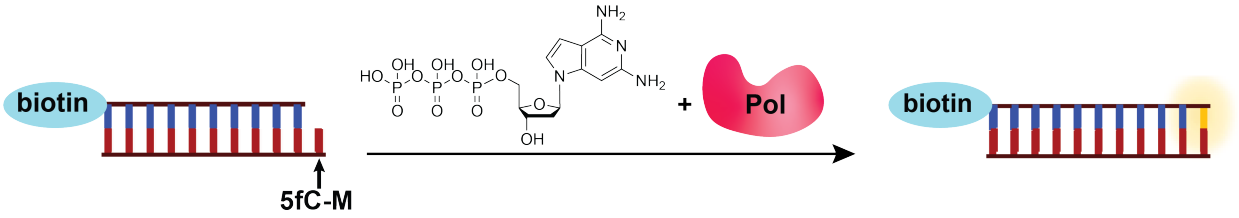



*Figure 85: Schematic illustration of the single nucleotide incorporation of 2,6-Diamino-DDP with a DNA polymerase.*

The sequence of the 25mer is AA5TGAAGACACGGCTATACATACT (5 = 5-formyl cytosine) and the 22mer is (Btn)AGTATGTATAGCCGTGTCTTCA. Five polymerases that have 3'-5'-exonuclease activity (proofreading) were tested: Q5U® Hot Start High-Fidelity DNA Polymerase, Phusion® High-Fidelity DNA Polymerase, T7 DNA Polymerase, DNA Polymerase I (E. coli), DNA Polymerase I Large Fragment'(Klenow). Because all polymerases with 3'-5'-exonuclease activity had issues with the incorporation or even degraded the primer, we decided to focus on polymerases without proofreading

activity. The following DNA polymerases were tested: Taq polymerase, Bst full length, Bst large fragment, Bst 3.0, Bsu large fragment, Klenow Fragment (exo-) and Deep Vent® (exo-).

While the other tested polymerases had been unsuccessful in fully extending the primer, Klenow Fragment (exo-), could fully incorporate 2,6-Diamino-DDP opposite 5fC-M in the absence of other dNTPs.

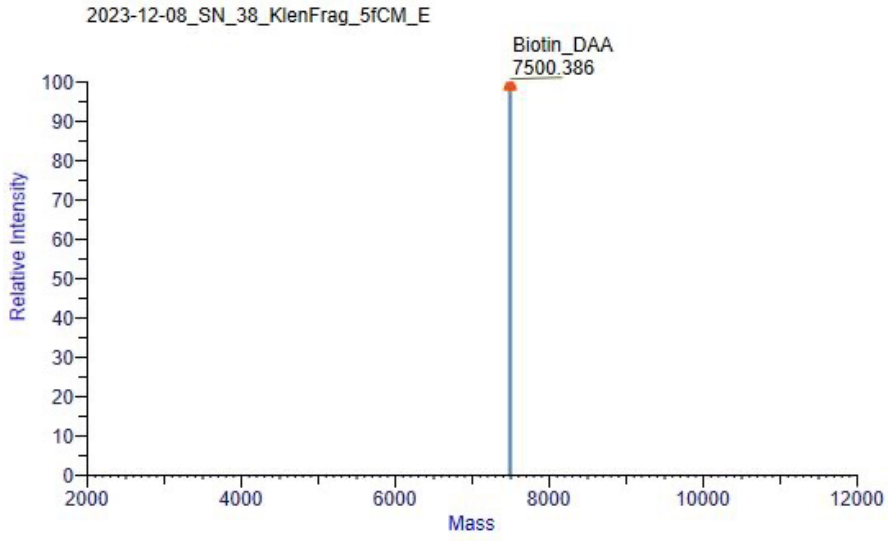



*Figure 86: Deconvoluted mass spectrum of the single nucleotide incorporation of 2,6-Diamino-DDP with Klenow Fragment (exo-). Deconvoluted mass: 7500.386. Calculated mass: 7500.365.*

The calculated mass of the 22mer primer is 7174.2872 and the calculated mass of the 23mer primer including 2,6-Diamino-DDP is 7500.3652. As the mass of 7500.386 was found (Figure 86), the successful incorporation of 2,6-Diamino-DDP was confirmed.

In competition with the natural triphosphates dATP and dGTP, no 2,6-Diamino-DDP was incorporated (Figure 87). Around 20% dGTP and approximately 80% dATP were incorporated opposite 5fC-M.

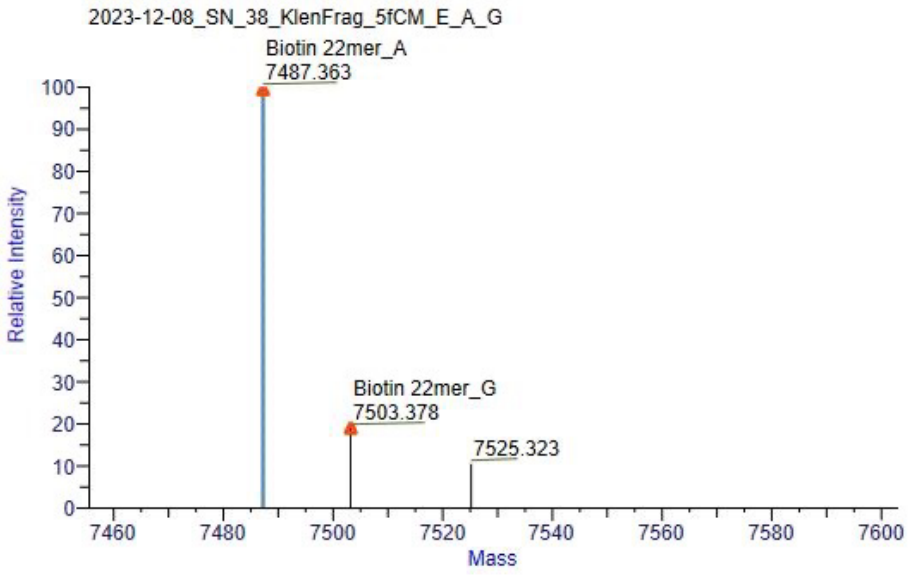



*Figure 87: Deconvoluted mass spectrum of the single nucleotide incorporation using 2,6-Diamino-DDP, dATP and dGTP with Klenow Fragment (exo-).*

The observed lack of competition from 2,6-Diamino-DDP can be explained by the polymerase bias for the naturally occurring building blocks. In the future, the incorporation of 2,6-Diamino-DDP could be made more efficient, and the selectivity of the process could be improved by using polymerases that have been specially engineered for this purpose.

## 7.5 2,6-Diamino-DDP: Conclusion and future perspectives

This chapter explored the design and synthesis of the unnatural nucleobase 2,6-Diamino-DDP, which is intended to have a high p$K_a$ and become protonated easily. In its protonated form, it is believed to form a novel base pair with three hydrogen bonds with labeled 5fC-bases, such as 5fC-M and 5fC-A, thereby expanding the genetic alphabet. This approach could be used to improve 5fC-sequencing, but also has the potential to support applications such as DNA data storage or aptamer development. However, the synthesis process was notably challenging. The high electron density within the aromatic ring, in addition to the altered reactivity of the two exocyclic amino groups, further complicated the synthesis. By employing an optimized Buchwald-Hartwig amidation the nucleoside was produced. Despite extensive efforts, no suitable protecting group could be identified for the phosphoramidite synthesis. All amides commonly used for other nucleosides, as well as other tested amides, were found to be too stable and could not be cleaved under DNA-compatible conditions. In the case of amidines, it was observed that the protection of a single amine was possible with sufficient yield, but the attachment of two amidines was not successful.

The 2,6-Diamino-DDP triphosphate could be produced enzymatically using a modified polyphosphate kinase. During primer extension experiments, the commercially available polymerases exhibited a low affinity for 2,6-Diamino-DDP. This was evidenced by experiments that revealed a clear preference for natural bases during primer extension. But in the absence of other dNTPs, Klenow Fragment (exo-) could fully incorporate 2,6-Diamino-DDP opposite 5fC-M, therefore forming a novel base pair.

In the future, the employment of a mixed protective group strategy would be advantageous. One electron-withdrawing group, such as the trifluoroacetyl group, could be attached to one exocyclic amide. The other one could be protected with an amidine, such as dbf-dma. This method will allow for the phosphoramidite synthesis, thus enabling the synthesis of oligonucleotides that contain 2,6 Diamino-DDP. In order to perform the enzymatic incorporation, an engineered polymerase could be designed and optimized to prefer the incorporation of 2,6-Diamino-DDP opposite 5fC-M or 5fC-A.

# 8. Summary and Outlook

In this dissertation, multiple modified DNA building blocks were investigated, synthesized and incorporated into nucleic acids. The aim of this work was to pave the way for new tools to study epigenetic mechanisms and allow the development of innovative diagnostic tools, as well as molecular machines.

The naturally occurring DNA modification 5-formylcytosine (5fC) plays a crucial role in gene expression, chromatin organization or stem cell differentiation. Conventional methods to detect 5fC, such as malononitrile-based labeling, are constrained by incomplete conversion and DNA degradation. Furthermore, the undesired incorporation of guanine and polymerase stalling are additional disadvantages of this method. This motivated us to explore more efficient 5fC-labeling techniques to convert 5fC into a base without exocyclic amine, 5fC-A. We introduced AceF-Seq, a novel technique that uses acetylacetone to achieve highly efficient and selective C-to-T conversion of 5fC without causing DNA damage or polymerase stalling. The superior performance of AceF-Seq was confirmed through Sanger sequencing using multiple polymerases. The sequencing results demonstrated the higher reliability and precision of AceF-Seq compared to existing methods. A graphical summary of the AceF-Seq technique and of the biological role of 5fC is shown in Figure 88. This advancement paves the way for high-resolution, genome-wide studies of 5fC. Moreover, novel diagnostic tools could detect changes in 5fC levels and monitor the effectiveness of treatments.

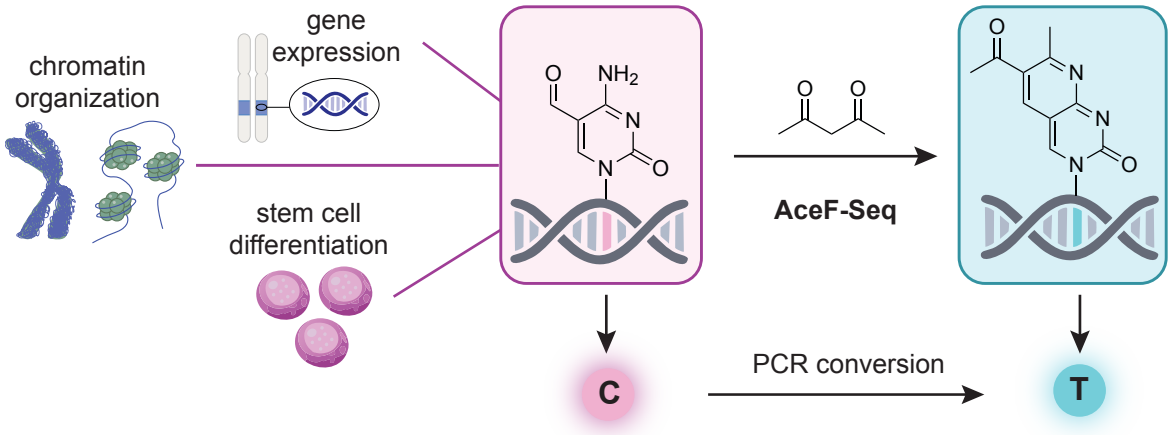



*Figure 88: Graphical summary of the biological role of 5fC and the AceF-Seq technique causing a C-to-T conversion.*

The objective of the second research project was to develop a programmable, pH-responsive nucleoside for DNA nanotechnology. The modified nucleobase 2-Amino-dideaza purine (2-Amino-DDP) was synthesized as triphosphate and as phosphoramidite. The synthesis of oligonucleotides containing 2-Amino-DDP enabled melting point measurements and revealed a clear pH-dependence and a strong destabilization of the 2-Amino-DDP:T pair at a pH below 7. Consequently, this could enable DNA origami structures to respond dynamically to environmental changes. Furthermore, the pH-dependent incorporation

of 2-Amino-DDP triphosphate opposite C and T templates was shown by multiple DNA polymerases. Overall, 2-Amino-DDP has the potential to act as a new, pH-responsive building block for diverse applications in DNA-based nanotechnology or nucleic acid enzymes.

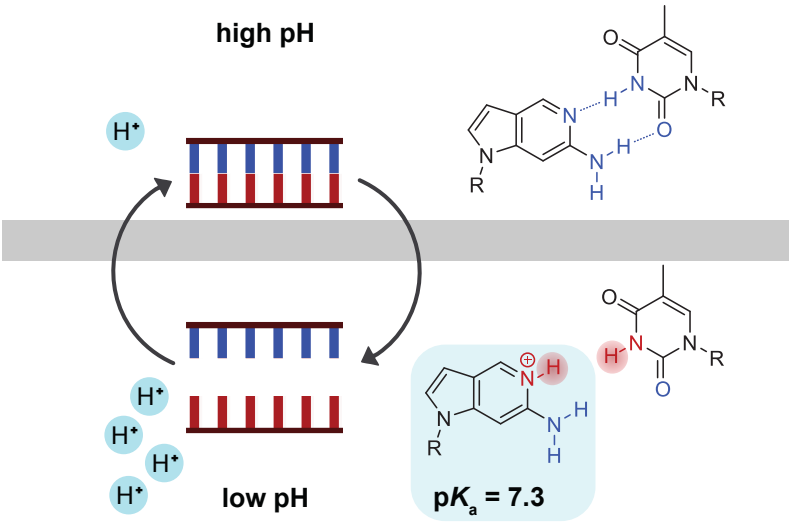



*Figure 89: Concept of 2-Amino-DDP forming a pH-responsive base pair with T.*

The third research project focused on the investigation of a new base pair to expand the genetic alphabet. For this approach, 2,6-Diamino-dideaza purine (2,6-Diamino-DDP) was synthesized. In its protonated form, it is hypothesized to form a novel base pair with three hydrogen bonds with labeled 5fC-bases, such as 5fC-M and 5fC-A, thereby expanding the genetic alphabet and enabling advanced 5fC-sequencing. Initially, when using an amide-protected 2,6-Diamino-DDP phosphoramidite, the solid phase DNA synthesis appeared to be effective, but the protecting groups were too stable to be removed from the oligonucleotide. Therefore, multiple different approaches were tested, including different amides, carbamates and amidines. However, the unsuitability of all of the tested protecting group strategies resulted in further solid-phase synthesis attempts being unsuccessful. In the future, a strategy that involves different protecting groups for the two exocyclic amines would be advantageous. The triphosphate of 2,6-Diamino-DDP was synthesized and successfully incorporated opposite to 5fC-M by Klenow Fragment (exo-), thereby forming a novel base pair. Ultimately, the novel base pair has the potential to support applications such as DNA data storage, sequencing, aptamer development, and PCR diagnostics.

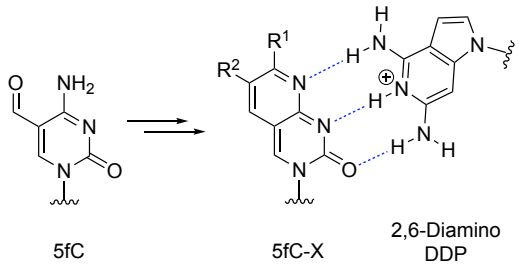



*Figure 90: 2,6-Diamino-DDP forming a novel base pair with labeled 5fC.*

# 9. Experimental part

## 9.1 General information

Unless otherwise specified, all commercially available reagents and solvents were utilized without additional purification. Solvents used for flash column chromatography were of laboratory-grade quality. Dry solvents were procured from Fisher Scientific, Acros, or VWR, and were used without additional processing. The commercially available oligonucleotides were purchased from Sigma-Aldrich / Merck, in HPLC-purified quality. They were ordered in a dry state, and diluted with Milli-Q-water to produce 100 µM solutions. Purification of crude compounds was achieved through silica gel flash column chromatography (Merck 60, particle size 0.040-0.063 mm) using indicated solvents.

Analytical thin-layer chromatography (TLC) was conducted on silica-coated aluminum plates (Merck 60 F254) and visualization of synthetic products was achieved through UV irradiation (254 nm and/or 356 nm) and potassium permanganate staining (1.5 g $KMnO_4$, 10 g $K_2CO_3$ in 1.25 mL of 10% aq. NaOH and 200 mL water). Analytical ultra-high-performance liquid chromatography-mass spectrometry (uHPLC-MS) and liquid chromatography-mass spectrometry (LC-MS) were performed and evaluated using an Agilent 1290 Infinity system equipped with a mass detector (column: Zorbax Eclipse C18 Rapid Resolution 2.1x50 mm 1.8µm, flow rate: 0.5 mL/min) or an Agilent Infinity HPLC system with a 3x50 mm, 1.8 µm Macherey-Nagel Nucleodur C18 Gravity column (flow rate: 0.5 mL/min). For LC-MS analysis, a gradient was applied, starting with 10% acetonitrile (incl. 0.1% formic acid) in water (incl. 0.1% formic acid) and progressing to 100% acetonitrile (incl. 0.1% formic acid).

Nuclear magnetic resonance (NMR) spectra were recorded using one of the following spectrometers: Bruker AV 400 Avance III HD (NanoBay), Bruker AV 500 Avance III HD (Prodigy), Bruker Avance NEO - 500 MHz, Bruker AV 600 Avance III HD (CryoProbe), or a Bruker AV 700 Avance III HD (CryoProbe) spectrometer. Chemical shifts are reported in parts per million (ppm) with reference to the deuterated solvent. Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, dd = double doublet, and m = multiplet. Coupling constant values are expressed in Hertz. Signals were assigned to their corresponding hydrogen atoms or carbon atoms based on 2D NMR correlations ($^1H/^1H$ COSY, $^1H/^{13}C$ HSQC, $^1H/^{13}C$ HMBC). For phosphorus-containing compounds, $^{31}P$ NMR analysis was conducted as well.

High-resolution mass spectrometry (HRMS) in positive mode was performed using an LTQ Orbitrap mass spectrometer with electron spray ionization coupled to an Accela HPLC-system (HPLC column: Hypersil GOLD, 50 mm x 1 mm, particle size 1.9 µm).

HRMS in negative mode was measured for phosphate-containing molecules and oligonucleotides using an Orbitrap Exploris 120 ESI-MS (Thermo Fisher Scientific). The HPLC was used with a flow rate of 0.20 mL/min in a solvent system of solvent A (50 mM HFIP and 15 mM TEA in water at pH 9.00) and solvent B (methanol) at 80 °C on a DNAPac RP 4 µm column (2.1 × 100 mm, Thermo Fisher Scientific).

For the analysis of PAGE gels, a ChemiDoc MP Imaging System from Bio-Rad Laboratories Inc. (Hercules, CA, USA) was used.

### 9.2 Solid phase DNA synthesis

Oligonucleotides were synthesized on CPG (Controlled Pore Glass) resin with 500 Angstrom Pore Size and 25-35 µmol/g loading from Applied Biosystems™, following modified DNA standard methods. Detritylation was achieved with dichloroacetic acid in DCM (3% v/v) for 45 seconds. For coupling, 100 µL of 100 mM phosphoramidite solution in acetonitrile was activated with 250 µL of 0.25 M ethyl-thiotetrazole (ETT) in acetonitrile. The coupling time was 30 seconds for canonical nucleosides and 10 min for 2-Amino-DDP. Capping was achieved with a 1:1 (v/v) mixture of acetic anhydride in acetonitrile and *N*-Methylimidazole in acetonitrile for 45 seconds. Oxidation was performed with 20 mM $I_2$ in pyridine/water (9:1 v/v) for 60 seconds. All sequences were synthesized with final trityl-deprotection (DMT-off). Cleavage from the solid support and deprotection was achieved after treatment with 25% ammonia solution for 6 h at 60 °C. HPLC purification was conducted using a Thermo Scientific™ DNAPac™ RP-HPLC-C18 column with 100 mM aqueous TEAA buffer (A) and a mixture of 80% acetonitrile and 20% 500 mM aqueous TEAA buffer (B). The oligonucleotides were analyzed by LC/MS using an Orbitrap Exploris 120 Mass Spectrometer.

For oligonucleotide TCAGXGTAAG (with X=2-Amino-DDP), 128 µg (41.6 nmol) were isolated.

### 9.3 General procedure for triphosphate purification

The initial purification process, carried out using high-performance liquid chromatography (HPLC), involved the use of an ion exchange column with a Tris-HCl buffer and an aqueous, 1.25-M NaCl solution. The process enabled the removal of byproducts, such as the nucleoside monophosphate, while the nucleoside diphosphate remained unseparated from the triphosphate. Following this initial purification, an additional HPLC purification was conducted, using a Thermo Scientific™ DNAPac™ RP-HPLC-C18 column with 100 mM aqueous TEAA buffer (A) and a mixture of 80% acetonitrile and 20% 500 mM aqueous TEAA buffer (B). The triphosphate was isolated, dried under high vacuum, and stored at -70 °C.

## 9.4 UV/Vis and melting point measurements

The UV-Vis-absorption for melting point measurements and for the p$K_a$-determination was measured on an Agilent Cary 3500 Multicell Peltier UV-Vis spectrophotometer with Xenon Flashlamp Source. For the UV-Vis absorption, nucleoside **5** (1 mM) was dissolved in 50 mM sodium acetate buffer (for pH 5.0) or in 50 mM sodium phosphate buffer (for pH 6.5 to 9.5). The absorption at a wavelength of 200 to 600 nm was measured with a spectral bandwidth of 2.0 nm.

The 10mer DNA used for melting point measurements had the sequence TCAGXGTAAG (X = 2-Amino-DDP). It was mixed with the corresponding template CTTACCCTGA or CTTACTCTGA at 3 µM oligonucleotide concentration in 10 mM phosphate buffer with pH from 6.0 to 9.5 and a total of 250 mM NaCl. Initially, the DNA was denatured at 80 °C and the strands were annealed by cooling to 10 °C for 10 min. The temperature was then raised to 60 °C at 2 °C/min, and the absorption was measured in 0.5 °C steps. After cooling down again to 10 °C, the measurements were repeated 3 times in total. The absorption at 260 nm and 420 nm was measured at each temperature and the difference was used to determine the melting points. The averaging time was set to 1.0 s and the spectral bandwidth to 2.0 nm.

## 9.5 Conditions for biological experiments

### 9.5.1 Oligonucleotide sequences

10mer DNA for melting point measurements: TCAGXGTAAG (X = 2-Amino-DDP), annealed to CTTACCCTGA or CTTACTCTGA

**Primer P1:** (Btn)-AGTATGTATAGCCGTGTCTTCA

**Primer P2**: (FAM)-GTTTTGGCTACCTGTTACTAAGCA

**Primer P3**: (forward):
AGCTCGTTTAGTGAACCGTGAATTATACCCTCGCGTTACTCCAAGAGATTGATCATAC

**Primer P4**: (reverse):
TGTAAAACGACGGCCAGACATAAACAGATATTCTCCACTGTGACCTAAAGTCCGCTAC

**Primer P5**: (Btn)-CGGGCGGACCAGAACCCTTGAGCACAGAAA

**Template T1**: (25mer): AA**5**TGAAGACACGGCTATACATACT (5 = 5-formylcytosine)

**Template T2**: (70mer):
CTCCAAGAGATTGATCATACATATCGGCACAGAAGT**5**ACACGACGCCGATGGGTAGCGGACTTTAGG
TCA (5 = 5-formylcytosine)

**Template T3** (25mer): **C**TGCTTAGTAACAGGTAGCCAAAAC

**Template T4** (25mer): **T**TGCTTAGTAACAGGTAGCCAAAAC

**Template T5** (148mer):
AGCTCGTTTAGTGAACCGTGAATTATACCCTCGCGTTACTCCAAGAGATTGATCATACATATCGGCAC
AGAAGT**5**ACACGACGCCGATGGGTAGCGGACTTTAGGTCACAGTGGAGAATATCTGTTTATGTCTGG
CCGTCGTTTTACA (5 = 5-formylcytosine)

**Template T6** (49mer): CTTCGGCTAACTCGACGTTTTTCTGTGCTCAAGGGTTCTGGTCCGCCCG

**Mass table of the used oligonucleotides**

*Table 8: Monoisotopic masses and molecular weights oligonucleotides with and without label and primer extension products.*

| Oligonucleotide length | Name | Monoisotopic Mass [Da] | Molecular Weight [g/mol] |
|---|---|---|---|
| 25mer | Template T1 | 7667.34 | 7671.46 |
| | 5fC-N | 7735.22 | 7739.09 |
| | 5fC-M | 7717.27 | 7719.10 |
| | 5fC-A | 7731.34 | 7735.14 |
| [biotin]-22mer | Primer P1 | 7174.27 | 7176.86 |
| | P-A | 7487.32 | 7491.08 |
| | P-AT | 7791.39 | 7795.27 |
| | P-ATT | 8095.44 | 8099.47 |
| | P-G | 7503.32 | 7506.07 |
| | P-GT | 7806.38 | 7810.27 |
| | P-GTT | 8110.42 | 8114.46 |
| 70mer | Primer P2 | 21633.58 | 21644.32 |
| | 5fC-A | 21697.63 | 21708.02 |
| | 5fC-M | 21681.60 | 21691.71 |

### 9.5.2 Ace-Labeling conditions

**2-Aminobenzaldehyde** (150 mM) was incubated with acetylacetone (1.5 M) and $FeCl_3 \cdot 6\,H_2O$ (0 to 3.00 eq.) in water for 16 h at room temperature and 550 rpm. The labeled product (ABA-A) was filtered and subjected to LC-MS (VelosPRO) without further purification.

**5fC-nucleoside** (10 mM) was incubated with acetylacetone (1.0 M) in NaOAc buffer (500 mM, pH 5.5) for 16 h at 60 °C and 550 rpm. The labeled nucleoside (5fC-Nu-A) was filtered and subjected to LC-MS (VelosPRO) without further purification.

5fC-containing template **T1** or **T2** (10 µM) was incubated with acetylacetone (1.0 M) in NaOAc buffer (500 mM, pH 5.5) for 16 h at 60 °C and 550 rpm. The labeled oligonucleotide was purified with a DNA

purification kit (Oligo Clean & Concentrator™ of Zymo Research) and analyzed by LC-MS (Orbitrap Exploris 120).

### 9.5.3 Malononitrile-Labeling conditions

5fC-containing template **T1** or **T2** (10 µM) was incubated with malononitrile (150 mM) in $NH_4OAc$ buffer (200 mM, pH 7.0) for 72 h at 20 °C and 550 rpm. The labeled oligonucleotide was purified with a DNA purification kit (Oligo Clean & Concentrator™ of Zymo Research) and analyzed by LC-MS (Orbitrap Exploris 120).

### 9.5.4 Primer Extension Experiments for 5fC-oligonucleotides

***Bst* 3.0 DNA polymerase** (NEB): The template (labeled **T1**) (0.5 µM) was assembled with Isothermal Amplification Buffer II Pack (1X, 5 µL), dATP and dGTP (each 200 µM), biotin-labeled primer **P1** (0.25 µM) and *Bst* 3.0 DNA Polymerase (1 U/50 µL). The reaction volume was 50 µL. The mixture was incubated at 65 °C for 30 min for primer extension and inactivated at 80 °C for 5 minutes.

***Bst* DNA polymerase, Large Fragment** (NEB): The template (labeled **T1**) (0.5 µM) was assembled with ThermoPol® Reaction Buffer (1X, 5 µL), dATP and dGTP (each 200 µM), biotin-labeled primer **P1** (0.25 µM) and *Bst* DNA Polymerase, Large Fragment (1 U/50 µL). The reaction volume was 50 µL. The mixture was incubated at 65 °C for 30 min for primer extension and inactivated at 80 °C for 20 minutes.

***Bsu* DNA polymerase, Large Fragment** (NEB): The template (labeled **T1**) (0.5 µM) was assembled with NEBuffer™ 2 (1X, 5 µL), either dATP and dGTP or a mixture of dNTPs (each 200 µM), biotin-labeled primer **P1** (0.25 µM) and *Bsu* DNA Polymerase, Large Fragment (1 U/50 µL). The reaction volume was 50 µL. The mixture was incubated at 37 °C for 30 min for primer extension and inactivated at 75 °C for 20 minutes.

**Deep Vent® (exo-) DNA polymerase** (NEB): The template (labeled **T1**) (0.5 µM) was assembled with ThermoPol® Reaction Buffer (1X, 5 µL), either dATP and dGTP or a mixture of dNTPs (each 200 µM), biotin-labeled primer **P1** (0.25 µM) and Deep Vent® (exo-) DNA Polymerase (1 U/50 µL). The reaction volume was 50 µL. The mixture was incubated at 75 °C for 30 min for primer extension and inactivated by direct purification.

**Klenow Fragment (3´→5´ exo-)** (NEB): The template (labeled **T1**) (0.5 µM) was assembled with NEBuffer™ 2 (1X, 5 µL), either dATP and dGTP or a mixture of dNTPs (each 200 µM), biotin-labeled primer **P1** (0.25 µM) and Klenow Fragment (3´→5´ exo-) (1 U/50 µL). The reaction volume was 50 µL. The mixture was incubated at 37 °C for 30 min for primer extension and inactivated at 75 °C for 20 minutes.

***Sulfolobus* DNA polymerase IV** (NEB): The template (labeled **T1**) (0.5 µM) was assembled with ThermoPol® Reaction Buffer (1X, 5 µL), dATP and dGTP (each 200 µM), biotin-labeled primer **P1** (0.25 µM) and *Sulfolobus* DNA Polymerase IV (1 U/50 µL). The reaction volume was 50 µL. The mixture was incubated at 55 °C for 30 min for primer extension and inactivated by direct purification.

***Taq* DNA polymerase** (NEB): The template (labeled **T1**) (0.5 µM) was assembled with ThermoPol® Reaction Buffer (1X, 5 µL), dATP and dGTP (each 200 µM), biotin-labeled primer **P1** (0.25 µM) and *Taq* DNA Polymerase (1 U/50 µL). The reaction volume was 50 µL. The mixture was incubated at 75 °C for 30 min for primer extension and inactivated by direct purification.

**Therminator™ DNA polymerase** (NEB): The template (labeled **T1**) (0.5 µM) was assembled with ThermoPol® Reaction Buffer (1X, 5 µL), dATP and dGTP (each 200 µM), biotin-labeled primer **P1** (0.25 µM) and Therminator™ DNA Polymerase (1 U/50 µL). The reaction volume was 50 µL. The mixture was incubated at 75 °C for 30 min for primer extension and inactivated by direct purification.

All primer extension reactions were conducted in a ProFlex™ thermocycler (Applied Biosystems™). The products were purified with the Oligo Clean & Concentrator™ kit (Zymo Research) and analyzed by LC-MS (Orbitrap Exploris 120).

### 9.5.5 Polymerase Chain Reactions (PCR)

**AmpliTaq Gold™ 360 DNA-Polymerase** (Thermo Fisher Scientific): The template (**T2**, 0.3 nM) labeled with acetylacetone or malononitrile was assembled with AmpliTaq Gold™ 360 Buffer (1X, 5 µL), $MgCl_2$ (2 mM, 4 µL), dNTPs (200 µM, 1 µL), forward primer **P3** and reverse Primer **P4** (each 0.5 µM) and AmpliTaq Gold™ 360 DNA polymerase (1.25 U/50 µL, 0.25 µL). The reaction volume was 50 µL.

*Table 9: PCR conditions for AmpliTaq Gold™ 360 DNA polymerase.*

| Step | Temperature | Time |
|---|---|---|
| Initial Denaturation | 95 °C | 10 min |
| 20 Cycles | 95 °C<br>49 °C<br>72 °C | 30 s<br>30 s<br>60 s |
| Final Extension | 72 °C | 7 min |

**One*Taq*® DNA Polymerase (NEB):** The template (**T2**) labeled with acetylacetone or malononitrile (0.3 nM) was assembled with 1X One*Taq*® Standard Reaction Buffer (10 µL), dNTPs (200 µM, 1 µL), forward primer **P3** and reverse Primer **P4** (each 0.5 µM) and One*Taq*® DNA polymerase (1.25 U/50 µL, 0.25 µL). The reaction volume was 50 µL.

*Table 10: PCR conditions for OneTaq® DNA polymerase.*

| Step | Temperature | Time |
|---|---|---|
| Initial Denaturation | 94 °C | 30 s |
| 20 Cycles | 94 °C<br>43 °C<br>68 °C | 30 s<br>60 s<br>60 s |
| Final Extension | 68 °C | 10 min |

**Deep Vent® and Deep Vent ® (exo-) DNA Polymerase (NEB):** The template (**T2**) labeled with acetylacetone or malononitrile (0.3 nM) was assembled with 1X ThermoPol Reaction Buffer (5 µL), dNTPs (200 µM, 1 µL), forward primer **P3** and reverse Primer **P4** (each 0.5 µM) and Deep Vent® DNA polymerase (1 U/50 µL, 0.5 µL) or Deep Vent® (exo-) DNA polymerase (1 U/50 µL, 0.5 µL). The reaction volume was 50 µL.

***Table 11:*** *PCR conditions for Deep Vent® and Deep Vent® (exo-) DNA polymerase.*

| Step | Temperature | Time |
|---|---|---|
| Initial Denaturation | 95 °C | 5 min |
| 20 Cycles | 95 °C<br>45 °C<br>72 °C | 30 s<br>30 s<br>60 s |
| Final Extension | 72 °C | 5 min |

**KOD Hot Start DNA Polymerase (Sigma-Aldrich):** The template (**T2**) labeled with acetylacetone or malononitrile (325.6 pg, 0.3 nM) was assembled with forward primer **P3** and reverse Primer **P4** (each 0.5 µM) and KOD Hot Start Master Mix DNA Polymerase (2 U/50 µL, 25 µL). The reaction volume was 50 µL.

***Table 12:*** *PCR conditions for KOD Hot Start DNA polymerase.*

| Step | Temperature | Time |
|---|---|---|

| | | |
|---|---|---|
| Initial Denaturation | 95 °C | 2 min |
| 20 Cycles | 95 °C<br>45 °C<br>70 °C | 20 s<br>10 s<br>30 s |
| Final Extension | 70 °C | 2 min |

**Q5® Hot Start High-Fidelity DNA Polymerase (NEB):** The template (**T2**) labeled with acetylacetone or malononitrile (325.6 pg, 0.3 nM) was assembled with 1X Q5® Reaction Buffer (10 µL), dNTPs (200 µM, 1 µL), forward primer **P3** and reverse Primer **P4** (each 0.5 µM) and Q5® Hot Start High-Fidelity DNA Polymerase (1 U/50 µL, 0.5 µL). The reaction volume was 50 µL.

***Table 13:*** *PCR conditions for Q5® Hot Start DNA polymerase.*

| Step | Temperature | Time |
|---|---|---|
| Initial Denaturation | 98 °C | 30 s |
| 20 Cycles | 98 °C<br>58 °C<br>72 °C | 10 s<br>30 s<br>30 s |
| Final Extension | 72 °C | 2 min |

All PCR reactions were conducted in a ProFlex™ thermocycler (Applied Biosystems™). The PCR products were purified with the QIAquick PCR Purification Kit (QIAGEN) and analyzed by a High Sensitivity D1000 ScreenTape Assay in a TapeStation (Agilent).

The concentrations of the PCR products were determined using a NanoDrop 2000c (Thermo Scientific), diluted to 2-4 ng/µL and sent to *Microsynth* AG for Sanger-Sequencing.

### 9.5.6 Primer extension experiments for 2-Amino-DDP incorporation

To test the incorporation of a single 2-Amino-DDP, the 24-mer primer **P2** with the fluorophore 6-FAM attached at the 5'-end was annealed to a 25-mer template with a C overhang (**T3**) or T overhang (**T4**) at its 5'-end. The resulting oligonucleotide strands were separated using Urea-PAGE and detected by the fluorescence signal caused by 6-FAM. If not stated otherwise, primer extension reactions were performed in a 20 µL reaction volume containing 0.25 µM of the 6-FAM primer, 0.5 µM template, 25 to 50 µM dNTP,

0.1 units/µL polymerase, 20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl and 5 mM $MgSO_4$. The mixtures were incubated at 37 °C (for Bsu and Kle) or at 60 °C (all other tested polymerases) for the stated times.

For the extension of the 49-mer template **T6** with a 5'-biotin-modified 30-mer primer **P5**, in the first step, 2-Amino-DDP triphosphate was added and successfully incorporated two times. The conditions are: 100 µL total reaction volume, 5.0 µM template, 2.5 µM primer, 20 µM dNTP, 0.05 units/µL Therminator DNA polymerase, 20 min. Temp: 75 °C. In the next step, all other dNTPs were added and the residual 17 units were added. The conditions are: 100 µL total reaction volume, 5.0 µM template, 2.5 µM primer, 20 µM dNTP, 100 µM each of dATP, dGTP, dCTP, dTTP, 0.05 units/µL Therminator DNA polymerase, 40 min. Temp: 75 °C.

### 9.5.7 Urea-PAGE preparation

To prepare the Urea-PAGE gels, ROTIPHORESE® Sequenziergel concentrate (240 mL), ROTIPHORESE® Sequenziergel diluent (30 mL) and ROTIPHORESE® Sequenziergel buffer concentrate (30 mL) were mixed. The Concentrate contains 237.5 g/L of acrylamide, 12.5 g/L of *N*,*N*'-Methylenebisacrylamide, and 7.5 M urea in aqueous solution. The Diluent contains 7.5 M urea in aqueous solution. The Buffer contains 0.89 M tris-borate 20 mM EDTA buffer at pH 8.3 (10X TBE) and urea. This leads to a final monomer concentration of 20% which was suitable to separate DNA fragments of around 20-30 nucleotides length. APS (2.4 mL of a freshly made 10% aqueous solution) and TEMED (120 µL) were added and the gels casted, the comb was inserted and it was left for polymerization for at least 60 min. The electrophoresis unit was filled with TBE-buffer, and the gel was heated by a pre-run for 30 min, before the samples were loaded and separated at 800 W for 3:00 h. Readout of the fluorescent bands was performed with an imager (FX, Bio-Rad).

### 9.5.8 Primer extension experiments for 2,6-Diamino-DDP incorporation

For the primer extension experiments using 2,6-Diamino-DDP, the 5fC-containing 25mer template **T1** was used. The 22mer-primer **P1** was annealed. The sequences are listed in chapter 9.5.1. The 5'biotin-modified primer was used to achieve a better separation on the LC/MS.

5fC-containing template **T1** (10 µM) was incubated with malononitrile (150 mM) in $NH_4OAc$ buffer (200 mM, pH 7.0) for 72 h at 20 °C and 550 rpm. The labeled oligonucleotide was purified with a DNA purification kit (Oligo Clean & Concentrator™ of Zymo Research) and analyzed by LC-MS (Orbitrap Exploris 120). After purification, different polymerases were used for the incorporation of 2,6-Diamino-DDP triphosphate.

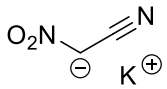
**Bsu DNA Polymerase**, Large Fragment: Into an Eppendorf tube, 2 µL of NEBuffer (10x) along with 1 µL each of template **T1** (10 µM) and primer **P1** (10 µM) were added. To this mixture, chosen dNTPs were added to achieve a final concentration of 200 µM, followed by the addition of Polymerase to reach a final concentration of 1 U/mL. The mixture was then incubated at 37 °C for 5 min and inactivated at 75 °C for 20 minutes. Subsequently, purification was performed using a Zymo kit.

**Deep Vent (exo-):** Into an Eppendorf tube, 2 µL of Thermopol Reaction Buffer (10x) along with 1 µL each of template **T1** (10 µM) and primer **P1** (10 µM) were added. To this mixture, chosen dNTPs were added to achieve a final concentration of 200 µM, followed by the addition of Polymerase to reach a final concentration of 1 U/mL. The mixture was then incubated at 65 °C for 60 s for primer annealing and at 72 °C for primer extension. Subsequently, purification was performed using the Zymo kit.

**Klenow Fragment (exo-)**: Into an Eppendorf tube, 2 µL of NEBuffer 2 (10x) along with 1 µL each of a template **T1** (10 µM) and primer **P1** (10 µM) were added. Subsequently, selected dNTPs were added to achieve a final concentration of 100 µM, followed by the addition of the Polymerase to attain a final concentration of 1 U/mL. The mixture was incubated at 37°C for 30 minutes and subsequently inactivated at 75°C for 20 minutes. Post-incubation, purification was performed using the Zymo kit.

## 9.6 Experimental procedures and product characterization

**Potassium nitroacetonitrile (NAN)**



Chemical Formula: $KC_2HN_2O_2$
Molecular Weight: 85,0425

Step 1: Ethyl cyanoglyoxylate-2-oxime (*Oxyma Pure*) (2.00 g, 14.1 mmol, 1.00 eq.) was mixed with a solution of KOH (261 mg, 4.64 mmol, 0.33 eq.) in water (30 ml). A solution of $KMnO_4$ (3.36 g, 21.1 mmol, 1.50 eq.) in water (50 mL) was added dropwise over a course of 30 min at 40 °C. The mixture was stirred at 40 °C for additional 30 min. The precipitate was removed by filtration and the filtrate was concentrated in vacuo.

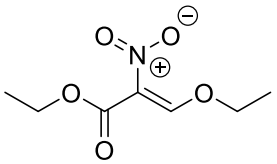
Step 2: The resulting salt was added to a solution of KOH (790 mg, 14.1 mmol, 1.00 eq.) in water (10 mL) and the mixture was stirred overnight at room temperature. After 16 h, the solvent was removed in vacuo and the dry residue was washed with cold ethanol and dried. This material was extracted with methanol (3×50 mL) and a solution of 10% $H_2O$ in methanol (50 mL). The extract was concentrated in vacuo, and the residue was washed with ethanol and filtered to give pale-yellow crystals.

$^1$H NMR (500 MHz, DMSO-$d_6$) δ/ppm = 5.58 (s, 1H, CH).

$^{13}$C NMR (126 MHz, DMSO-$d_6$) δ/ppm = 120.1 (CN), 79.9 ($CNO_2$).

The product is not stable under LC-MS conditions; therefore, no exact mass could be determined.

**Compound 3**



Chemical Formula: $C_7H_{11}NO_5$
Molecular Weight: 189,1670

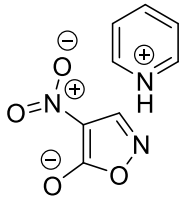
Ethyl nitroacetate (3.33 mL, 30.0 mmol, 1.00 eq), Triethyl orthoformate (7.49 mL, 45.0 mmol, 1.50 eq) and Acetic anhydride (7.02 mL, 75.0 mmol, 2.50 eq) were stirred for 24 h at 100°C. Residual starting material and impurities were removed via distillation in vacuo to give the product (2.30 g, 12.2 mmol, 41%) as an orange oil.

$^{1}$H NMR (600 MHz, $CDCl_3$) δ/ppm = 8.21 (s, 1H, C=C<u>H</u>-O (*E* or *Z*)), 7.52 (s, 1H, C=C<u>H</u> (*E* or *Z*), 4.37 – 4.24 (m, 4H, OC<u>$H_2$</u>$CH_3$), 1.43 (t, *J* = 7.1 Hz, 3H, C<u>$H_3$</u> (*E* or *Z*)), 1.39 (t, *J* = 7.1 Hz, 3H, C<u>$H_3$</u> (*E* or *Z*)), 1.31 (t, *J* = 7.1 Hz, 3H, C<u>$H_3$</u> (*E* or *Z*)), 1.28 (t, *J* = 7.1 Hz, 3H, C<u>$H_3$</u> (*E* or *Z*)).

$^{13}$C NMR (151 MHz, $CDCl_3$) δ/ppm = 163.1 (C=C<u>H</u>-O (*E* or *Z*)), 160.0 (<u>C</u>=O), 156.2 (C=C<u>H</u>-O (*E* or *Z*)), 109.6 (<u>C</u>-$NO_2$), 74.8 (O<u>C</u>$H_2CH_3$ (*E* or *Z*)), 74.1 (O<u>C</u>$H_2CH_3$ (*E* or *Z*)), 62.1 (O<u>C</u>$H_2CH_3$ (*E* or *Z*)), 62.0 (O<u>C</u>$H_2CH_3$ (*E* or *Z*)), 15.2 (<u>C</u>$H_3$), 14.2 (<u>C</u>$H_3$).

The product is not stable under LC-MS conditions; therefore, no exact mass could be determined.

**Compound 4**



Chemical Formula: $C_3HN_2O_4^-$
Molecular Weight: 129,0515

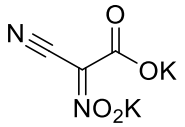
A mixture of **3** (2.20 g, 11.6 mmol, 1.00 eq.), hydroxylamine hydrochloride (1.05 g, 15.2 mmol, 1.30 eq.), and pyridine (2.34 mL, 29.1 mmol, 2.50 eq.) was stirred in ethanol (40 mL) at 60 °C for 2 h. The mixture was cooled in an ice bath and the precipitate was collected by filtration. Recrystallization from ethanol gave the pure pyridinium salt **4** (1.06 g, 5.09 mmol, 44%) as yellow needles.

$^{1}$H NMR (400 MHz, DMSO-$d_6$) δ/ppm = 8.95 – 8.92 (m, 2H, C$^{meta}$-<u>H</u>, Pyridinium), 8.62 (tt, *J* = 7.8, 1.5 Hz, 1H, C$^{para}$-<u>H</u>, Pyridinium), 8.47 (s, 1H, C<u>H</u>, Nitroisoxazolone), 8.11 – 8.06 (m, 2H, C$^{ortho}$-<u>H</u>, Pyridinium).

$^{13}$C NMR (101 MHz, DMSO-$d_6$) δ/ppm = 166.7 (CO, Nitroisoxazolone) 148.8 (CH, Nitroisoxazolone), 146.5 ($C^{para}$-H, Pyridinium), 142.2 ($C^{meta}$-H, Pyridinium), 127.3 ($C^{ortho}$-H, Pyridinium), 107.4 (C-$NO_2$, Nitroisoxazolone).

The product is not stable under LC-MS conditions; therefore, no exact mass could be determined.

**Carboxy-nitroacetonitrile (C-NAN)**



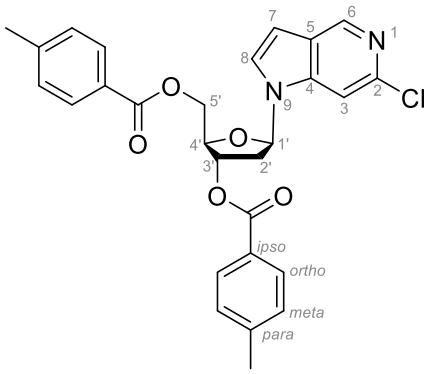

Chemical Formula: $C_3K_2N_2O_4$
Molecular Weight: 206,2396

Compound **4** (50.0 mg, 0.239 mmol) was stirred in a 0.5 M solution of KOH in methanol (2.0 mL) for 15 min at rt. The precipitate that formed during the reaction was collected by filtration and air-dried to give the dipotassium salt C-NAN (49.0 mg, 0.238 mmol, 99%) as a white solid.

$^{13}$C NMR (101 MHz, $D_2O$) δ/ppm = 167.3 (C=O), 119.3 (CN).

The product is not stable under LC-MS conditions; therefore, no exact mass could be determined.

**Compound 6b**




Chemical Formula: $C_{28}H_{25}ClN_2O_5$
Exact Mass: 504,1452
Molecular Weight: 504,9670

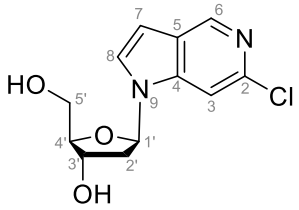
Freshly powdered KOH (3.67 g, 65.5 mmol, 5.00 eq.) was quickly mixed with dry THF (50 mL) under argon atmosphere. 6-Chloro-1H-pyrrolo[3,2-c]pyridine (2.00 g, 13.1 mmol, 1.00 eq.) and 1-Chloro-2-deoxy-3,5-di-*O*-toluoyl-ribose (5.61 g, 14.4 mmol, 1.10 eq.) were added and the resulting mixture was stirred for 30 min at room temperature under argon atmosphere. The brown mixture was filtered through Celite, the solid residue was washed with ethyl acetate and the filtrate was evaporated in vacuo. The crude product was purified via flash chromatography (20 to 50% EA in PE) to give **6b** (6.51 g, 12.9 mmol, 98%) as a yellow solid.

$^1$H NMR (500 MHz, $CDCl_3$) δ/ppm = 8.58 – 8.55 (m, 1H, H-6), 7.92 – 7.89 (m, 2H, $C^{5'OBz\text{-}ortho}$-H), 7.85 – 7.81 (m, 2H, $C^{3'OBz\text{-}ortho}$-H), 7.40 – 7.36 (m, 1H, H-3), 7.24 (d, *J* = 3.5 Hz, 1H, H-8), 7.23 – 7.15 (m, 4H, $C^{5'OBz\text{-}meta}$-H & $C^{3'OBz\text{-}meta}$-H), 6.53 (dd, *J* = 3.5, 0.9 Hz, 1H, H-7), 6.30 (dd, *J* = 8.4, 5.6 Hz, 1H, H-1'), 5.63 (dt, *J* = 6.3, 2.5 Hz, 1H, H-3'), 4.59 – 4.51 (m, 3H, H-5'a & H-5'b & H-4'), 2.75 – 2.61 (m, 2H, H-2'a & H-2'b), 2.37 (s, 3H, $C^{5'OBz}$-C$H_3$), 2.34 (s, 3H, $C^{3'OBz}$-C$H_3$).

$^{13}$C NMR (126 MHz, $CDCl_3$) δ/ppm 166.3 (COOC-5'), 166.1 (COOC-3'), 144.8 ($C^{5'OBz\text{-}ipso}$), 144.4 ($C^{3'OBz\text{-}ipso}$), 143.4 (C-2), 142.9 (C-6), 142.9 (C-4), 129.9 ($C^{5'OBz\text{-}ortho}$), 129.8 ($C^{3'OBz\text{-}ortho}$), 129.5 ($C^{5'OBz\text{-}meta}$), 129.4 ($C^{3'OBz\text{-}meta}$), 126.8 ($C^{5'OBz\text{-}para}$), 126.5 ($C^{3'OBz\text{-}para}$), 126.3 (C-8), 125.4 (C-5), 105.2 (C-3), 103.3 (C-7), 85.6 (C-1'), 82.4 (C-4'), 74.9 (C-3'), 64.2 (C-5'), 38.4 (C-2'), 21.9 (Ar-C$H_3$), 21.8 (Ar-C$H_3$).

HRMS: Calculated for $[M+H]^+$ 505.1525, found: 505.1542.

**Compound 7**




Chemical Formula: $C_{12}H_{13}ClN_2O_3$
Exact Mass: 268,0615
Molecular Weight: 268,6970

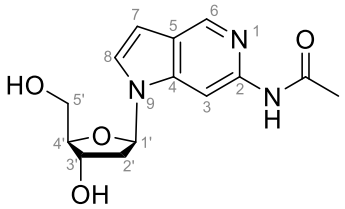
$Cs_2CO_3$ (620 mg, 1.90 mmol, 3.00 eq.) was added to a solution of **6b** (320 mg, 0.635 mmol, 1.00 eq.) in MeOH (50 mL) at room temperature and the mixture was stirred for 2 h. After removing the solvent in vacuo, water (50 mL) was added and it was extracted with EA (3 x 100 mL). The solvent was removed under reduced pressure and the residue was purified by flash column chromatography (1 to 10% MeOH in DCM) to give the nucleoside (173 mg, 0.571 mmol, 90%) as a white solid.

$^1$H NMR (700 MHz, Methanol-$d_4$) δ/ppm = 8.56 (s, 1H, H-6), 7.70 (s, 1H, H-3), 7.66 (d, *J* = 3.4 Hz, 1H, H-8), 6.70 (d, *J* = 3.4 Hz, 1H, H-7), 6.40 (dd, *J* = 6.8 Hz, 1H, H-1'), 4.51 (dt, *J* = 6.2, 3.6 Hz, 1H, H-3'), 4.00 – 3.95 (m, 1H, H-4'), 3.75 (dd, *J* = 12.1, 3.8 Hz, 1H, H-5'a), 3.70 (dd, *J* = 12.0, 4.3 Hz, 1H, H-5'b), 2.60 – 2.54 (m, 1H, H-2'a), 2.39 – 2.36 (m, 1H, H-2'b).

$^{13}$C NMR (176 MHz, Methanol-$d_4$) δ/ppm = 142.0 (C-2), 141.9 (C-4), 141.8 (C-6), 127.9 (C-8), 125.6 (C-5), 105.3 (C-3), 102.0 (C-7), 87.3 (C-4'), 85.2 (C-1'), 71.1 (C-3'), 61.8 (C-5'), 39.9 (C-2').

HRMS: Calculated for $[M+H]^+$ 269.0687, found: 269.0677.

**Compound 8**

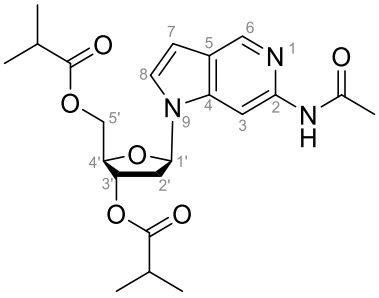


Chemical Formula: $C_{14}H_{17}N_3O_4$
Exact Mass: 291,1219
Molecular Weight: 291,3070

Nucleoside **7** (1.00 g, 3.72 mmol, 1.00 eq.) and acetamide (4.40 g, 74.4 mmol, 20.0 eq.) were mixed and heated to 130 °C. KOtBu (1.25 g, 11.2 mmol, 3.00 eq.) and tBu-BrettPhos Pd G3 (63.6 mg, 74.4 µmol, 0.02 eq.) were added under argon. The mixture was stirred at 130 °C for 16 h. After cooling down to room temperature, the volatiles were evaporated in vacuo. The crude product was purified via flash chromatography on silica gel (MeOH 2 to 20% in DCM) to give the acetyl-protected nucleoside mixed with acetamide as orange crystals. It was directly used for the hydroxyl-group protection.

HRMS: Calculated for $[M+H]^+$ 292.1292, found: 292.1293.

**Compound 9**



Chemical Formula: $C_{22}H_{29}N_3O_6$
Exact Mass: 431,2056
Molecular Weight: 431,4890

In a dry Schlenk tube, **8** (600 mg, 2.06 mmol, 1.00 eq.) was dissolved in anhydrous pyridine (30.0 mL). Isobutyric anhydride (3.43 mL, 20.6 mmol, 10.0 eq.) and *N*-methylimidazole (51 µL, 0.21 mmol, 0.10 eq.) were added and the resulting solution was stirred at room temperature under argon atmosphere for 30 min. The reaction mixture was quenched with methanol (30.0 mL), stirred for 10 minutes and evaporated in vacuo. The crude product was mixed with water (50 mL) and extracted with DCM (3x100 mL). The combined organic layers were washed with water (2x50 mL), dried over $MgSO_4$ and evaporated in vacuo.

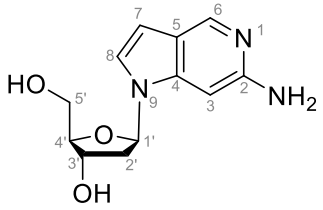
The product was purified via flash column chromatography (2 to 50% acetone in DCM) to give 732 mg (1.70 mmol, 82% over two steps) of the protected nucleoside as a pale yellow solid.

$^{1}$H NMR (500 MHz, $CDCl_3$) δ/ppm = 9.65 (s, 1H, NH), 8.49 (s, 1H, H-6), 8.41 (s, 1H, H-3), 7.34 (d, *J* = 3.5 Hz, 1H, H-8), 6.64 (d, *J* = 3.5 Hz, 1H, H-7), 6.38 (dd, *J* = 8.2, 5.9 Hz, 1H, H-1'), 5.34 – 5.30 (m, 1H, H-3'), 4.36 – 4.29 (m, 2H, H-5'a, H-5'b), 4.28 – 4.26 (m, 1H, H-4'), 2.70 – 2.63 (m, 1H, $C^{iBu}$-H), 2.62 – 2.58 (m, 1H, $C^{iBu}$-H), 2.57 – 2.49 (m, 2H, H-2'a, H-2'b), 2.25 (s, 3H, $C^{Acetyl}$-$CH_3$), 1.22 (dd, *J* = 6.9, 2.5 Hz, 6H, $C^{iBu}$-$CH_3$), 1.18 (dd, *J* = 6.9, 1.7 Hz, 6H, $C^{iBu}$-$CH_3$).

$^{13}$C NMR (126 MHz, $CDCl_3$) δ/ppm = 176.8 ($C^{iBu}$=O), 176.7 ($C^{iBu}$=O), 169.4 ($C^{Acetyl}$=O), 145.2 (C-2), 142.4 (C-4), 139.3 (C-6), 125.9 (C-8), 123.1 (C-5), 103.6 (C-7), 94.8 (C-3), 84.9 (C-1'), 82.2 (C-4'), 74.2 (C-3'), 64.0 (C-5'), 38.2 (C-2'), 34.1 ($C^{iBu}$H), 33.9 ($C^{iBu}$H), 24.8 ($C^{Acetyl}\underline{C}H_3$), 19.1 ($C^{iBu}\underline{C}H_3$), 19.1 ($C^{iBu}\underline{C}H_3$), 19.0 ($C^{iBu}\underline{C}H_3$), 19.0 ($C^{iBu}\underline{C}H_3$).

HRMS: Calculated for $[M+H]^+$ 432.2129, found: 432.2125.

**2-Amino-DDP nucleoside (Compound 10)**




Chemical Formula: $C_{12}H_{15}N_3O_3$
Exact Mass: 249,1113
Molecular Weight: 249,2700

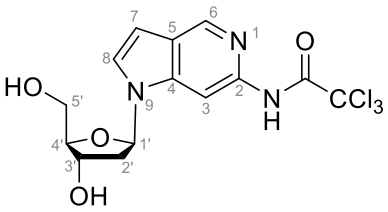
Compound **9** (3.00 g, 6.95 mmol, 1.00 eq.) was stirred in a solution of 2 M NaOMe in MeOH at 80 °C for 18 h. The reaction mixture was evaporated in vacuo and fully dried via lyophilization, the crude product was purified via HPLC (5 to 20% ACN in water with 0.1% TFA) to give nucleoside **10** (1.13 g, 4.52 mmol, 65%) as white solid.

$^{1}$H NMR (600 MHz, Methanol-*d*$_4$) δ/ppm = 8.28 (s, 1H, H-6), 7.67 (d, *J* = 3.7 Hz, 1H, H-8), 6.93 (s, 1H, H-3), 6.67 (d, *J* = 3.7 Hz, 1H, H-7), 6.29 (dd, *J* = 6.8 Hz, 1H, H-1'), 4.50 (dt, *J* = 6.2, 3.1 Hz, 1H, H-3'), 4.00 (q, *J* = 3.8 Hz, 1H, H-4'), 3.78 – 3.64 (m, 2H, H-5'a, H-5'b), 2.55 (ddd, *J* = 13.7, 6.9 Hz, 1H, H-2'a), 2.38 (ddd, *J* = 13.6, 6.0, 3.2 Hz, 1H, H-2'b).

$^{13}$C NMR (151 MHz, Methanol-*d*$_4$) δ/ppm = 151.0 (C-2), 147.7 (C-4), 131.7 (C-8), 130.9 (C-6), 121.5 (C-5), 104.3 (C-7), 90.1 (C-3), 88.9 (C-4'), 86.5 (C-1'), 72.5 (C-3'), 63.2 (C-5'), 41.0 (C-2').

HRMS: Calculated for $[M+H]^+$ 250.1186, found: 250.1185.

**Compound 10c**



Chemical Formula: $C_{14}H_{14}Cl_3N_3O_4$
Exact Mass: 393,0050
Molecular Weight: 394,6330

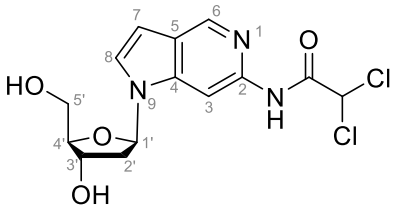
**10** (100 mg, 0.265 mmol, 1.00 eq.) was stirred in dry THF (4.0 mL) at rt under argon atmosphere. Trichloroacetic anhydride (425 mg, 1.38 mmol, 5.00 eq.) was added dropwise to the reaction mixture at 0 °C. After removal of the ice bath and stirring at rt for 30 min, the reaction was quenched with the addition of MeOH (5.0 mL) and the volatiles were removed *in vacuo*. Flash column chromatography (5 to 20% methanol in DCM) gave the product (45 mg, 0.11 mmol, 41%) as yellow solid.

$^1$H NMR (700 MHz, Methanol-$d_4$) δ/ppm = δ 8.95 (s, 1H, H-6), 8.16 (s, 1H, H-3), 8.08 (d, *J* = 3.6 Hz, 1H, H-8), 7.05 (d, *J* = 3.6 Hz, 1H, H-7), 6.52 (dd, *J* = 7.3, 6.1 Hz, 1H, H-1'), 4.56 (dt, *J* = 6.4, 3.3 Hz, 1H, H-3'), 4.07 (q, *J* = 3.7 Hz, 1H, H-4'), 3.79 (dd, *J* = 12.0, 3.6 Hz, 1H, H-5'a), 3.77 – 3.70 (m, 1H, H-5'b), 2.63 (ddd, *J* = 13.6, 7.3, 6.2 Hz, 1H, H-2'a), 2.56 – 2.49 (m, 1H, H-2'b).

$^{13}$C NMR (176 MHz, Methanol-$d_4$) δ/ppm = 162.8 (C$\underline{C}$OCCl$_3$), 144.5 (C-2), 138.8 (C-4), 135.8 (C-6), 133.8 (C-8), 125.4 (C-5), 105.9 (C-7), 101.6 (C-3), 90.1 (CO$\underline{C}$Cl$_3$), 89.3 (C-4'), 87.5 (C-1'), 72.4 (C-3'), 63.0 (C-5'), 41.6 (C-2').

HRMS: The product is not stable under LC-MS conditions, but the exact mass of the hydrolyzed product could be determined. Calculated for $[M+H]^+$ 250.1186, found: 250.1166.

**Compound 10d**



Chemical Formula: $C_{14}H_{15}Cl_2N_3O_4$
Exact Mass: 359,0440
Molecular Weight: 360,1910

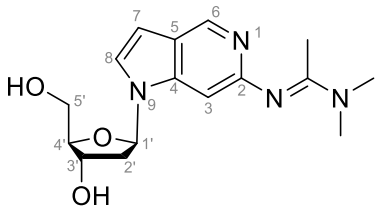
**10** (120 mg, 0.33 mmol, 1.00 eq.) was stirred in dry THF (5.0 mL) at rt under argon atmosphere. Dichloroacetyl chloride (0.159 mL, 1.65 mmol, 5.00 eq.) was added dropwise to the reaction mixture at 0 °C. After stirring at rt for 16 h, the reaction was quenched with the addition of saturated aqueous $NaHCO_3$-solution. The crude product was purified via flash column chromatography (10 to 20% methanol in DCM) to give the product (26.0 mg, 0.072 mmol, 22%) as white solid.

$^1$H NMR (500 MHz, Methanol-$d_4$) δ/ppm = 8.57 (s, 1H, H-6), 8.26 (s, 1H, H-3), 7.60 (d, *J* = 3.4 Hz, 1H, H-8), 6.66 (d, *J* = 3.5 Hz, 1H, H-7), 6.54 (s, 1H, C<u>H</u>$Cl_2$), 6.42 (dd, *J* = 7.8, 6.0 Hz, 1H, H-1'), 4.52 – 4.47 (m, 1H, H-3'), 4.02 – 3.97 (m, 1H, H-4'), 3.78 – 3.70 (m, 2H, H-5'), 2.63 – 2.55 (m, 1H, H-2'a), 2.44 – 2.36 (m, 1H, H-2'b).

$^{13}$C NMR (126 MHz, Methanol-$d_4$) δ/ppm = 164.1 (<u>C</u>O$CHCl_2$), 145.2 (C-2), 142.6 (C-4), 142.6 (C-6), 128.6 (C-8), 125.8 (C-5), 103.3 (C-7), 97.3 (C-3), 88.5 (C-4'), 86.4 (C-1'), 72.7 (C-3'), 68.1 (<u>C</u>$HCl_2$), 63.5 (C-5'), 41.1 (C-2').

HRMS: The product is not stable under LC-MS conditions, but the exact mass of the hydrolyzed product could be determined. Calculated for $[M+H]^+$ 250.1186, found: 250.1166.

**Compound 10e**



Chemical Formula: $C_{16}H_{22}N_4O_3$
Exact Mass: 318,1692
Molecular Weight: 318,3770

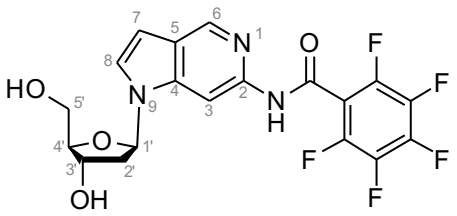
**10** (100 mg, 0.401 mmol, 1.00 eq.) and dimethylacetamide dimethylacetal (984 mg, 4.01 mmol, 10.00 eq.) were mixed and stirred at 60 °C for 4 days. The volatiles were evaporated in vacuo and the mixture was purified by HPLC (2 to 60% ACN in water with 0.1% TFA) to give 80 mg (0.251 mmol, 63%) of the product as an orange oil.

$^1$H NMR (500 MHz, Methanol-$d_4$) δ/ppm = 8.40 – 8.33 (m, 1H, H-6), 7.77 (d, $J$ = 3.5 Hz, 1H, H-8), 7.44 (s, 1H, H-3), 6.79 – 6.75 (m, 1H, H-7), 6.40 (dd, $J$ = 7.7, 5.9 Hz, 1H, H-1'), 4.50 (ddd, $J$ = 12.7, 6.5, 3.5 Hz, 1H, H-3'), 3.98 (p, $J$ = 3.8 Hz, 1H, H-4'), 3.77 – 3.66 (m, 2H, H-5'a and H-5'b), 2.67 (s, 6H, N(C$\underline{H}_3$)$_2$), 2.61 – 2.49 (m, 1H, H-2'a), 2.45 – 2.33 (m, 1H, H-2'b), 2.22 (s, 3H, N=CC$\underline{H}_3$).

$^{13}$C NMR (126 MHz, Methanol-$d_4$) δ/ppm = 165.2 (N=$\underline{C}$CH$_3$), 153.2 (C-2), 144.9 (C-4), 144.9 (C-6), 131.2 (C-8), 125.1 (C-5), 104.2 (C-7), 101.9 (C-3), 88.9 (C-4'), 86.9 (C-1'), 72.5 (C-3'), 63.2 (C-5'), 41.3 (C-2'), 35.3 (N(C$\underline{H}_3$)$_2$), 16.5 (N=C$\underline{C}$H$_3$).

HRMS: Calculated for $[M+H]^+$ 319.1765, found: 319.1760.

**Compound 10f**



Chemical Formula: $C_{19}H_{14}F_5N_3O_4$
Exact Mass: 443,0904
Molecular Weight: 443,3300

Compound **10** (400 mg, 1.60 mmol, 1.00 eq.) was stirred in dry pyridine (5.0 mL) at rt under argon atmosphere. Pentafluorobenzoyl chloride (1.16 mL, 8.02 mmol, 5.00 eq.) was added dropwise to the reaction mixture at rt. After stirring at rt for 30 min, methanol (10 mL) and ammonia in methanol (5.0 mL) were added and the mixture was stirred for 3 h at rt, before it the volatiles were evaporated in vacuo and the crude product was purified via flash column chromatography (1 to 10% MeOH in DCM) to give 390 mg (0.88 mmol, 55%) of the product as yellow solid.

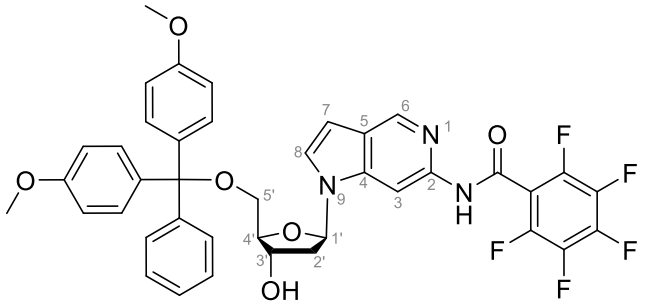
$^1$H NMR (700 MHz, Methanol-$d_4$) δ/ppm = 8.56 (s, 1H, H-6), 8.36 (s, 1H, H-3), 7.81 (d, *J* = 3.5 Hz, 1H, H-8), 7.60 (d, *J* = 3.5 Hz, 1H, H-7), 6.45 (dd, *J* = 7.7, 4.2 Hz, 1H, H-1'), 4.52 – 4.49 (m, 1H, H-3'), 4.16 (q, *J* = 4.1 Hz, 1H, H-4'), 3.77 – 3.72 (m, 2H, H-5'a, H-5'b), 2.61 (ddd, *J* = 13.9, 7.8, 6.3 Hz, 1H, H-2'a), 2.43 – 2.31 (m, 1H, H-2'b).

$^{13}$C NMR (176 MHz, Methanol-$d_4$) δ/ppm = 157.4 (CONHR), 150.0 (C-2), 145.5 ($CF^{Ar}$), 145.4 ($CF^{Ar}$), 142.5 (C-6), 138.4 ($CF^{Ar}$), 135.7 (C-4), 129.5 (C-8), 128.5 (C-7), 125.7 (C-5), 113.7 (CCONHR), 97.3 (C-3), 89.8 (C-4'), 86.5 (C-1'), 72.7 (C-3'), 63.5 (C-5'), 41.4 (C-2').

$^{19}$F NMR (470 MHz, Methanol-$d_4$) δ/ppm = -143.64 – -143.80 (m, 2F, $CF^{ortho}$), -155.07 (ddt, *J* = 19.8, 17.2, 2.9 Hz, 1F, $CF^{para}$), -163.83 (tdt, *J* = 20.0, 7.6, 4.0 Hz, 2F, $CF^{meta}$).

MS: Calculated for $[M+H]^+$ 444.1, found: 444.4.

**Compound 10g**



Chemical Formula: $C_{40}H_{32}F_5N_3O_6$
Exact Mass: 745,2211
Molecular Weight: 745,7030

In a dry Schlenk tube, compound **10f** (200 mg, 0.45 mmol, 1.00 eq.) and 4,4´-dimethoxytrityl chloride (459 mg, 1.35 mmol, 3.00 eq.) were dried under high vacuum, before anhydrous pyridine (3.0 mL) was added at rt. The resulting yellow mixture was stirred at room temperature for 30 min, before MeOH (5 mL) was added and stirring was continued for another 5 min. The reaction mixture was evaporated to dryness under reduced pressure. The crude product was purified via flash column chromatography (25 to 100% EA in CH with 0.1% TEA) to give the product (235 mg, 0.31 mmol, 70%) as yellow solid.

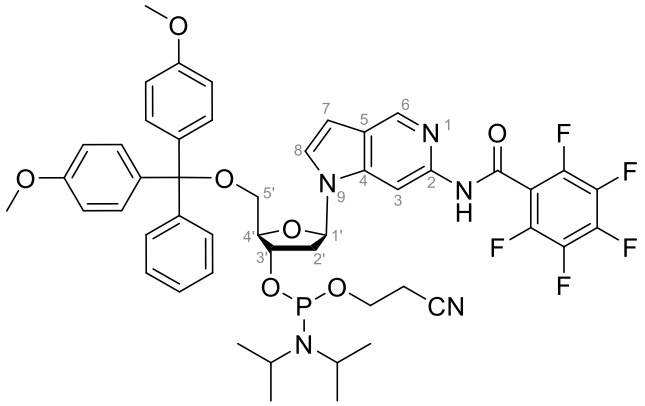
$^1$H NMR (700 MHz, $CDCl_3$) δ/ppm = 8.61 (s, 1H, H-6), 8.40 (s, 1H, H-3), 7.70 (d, *J* = 4.1 Hz, 1H, H-8), 7.68 (dt, *J* = 7.6, 1.8 Hz, 1H, $H^{Ar}$), 7.45 – 7.41 (m, 1H, $H^{Ar}$), 7.35 – 7.26 (m, 5H, $H^{Ar}$), 7.24 – 7.16 (m, 1H, $H^{Ar}$), 6.88 – 6.79 (m, 3H, $H^{Ar}$), 6.58 (d, *J* = 3.5 Hz, 1H, $H^{Ar}$), 6.52 (d, *J* = 3.5 Hz, 1H, $H^{Ar}$), 6.46 – 6.43 (m, 1H, H-1'), 4.61 (dt, *J* = 6.3, 3.7 Hz, 1H, H-3'), 4.09 (q, *J* = 4.2 Hz, 1H, H-4'), 3.40 (dd, *J* = 10.2, 4.2 Hz, 1H, H-5'a), 3.80 (s, 3H, $OCH_3$), 3.78 (s, 3H, $OCH_3$), 3.37 (dd, *J* = 10.2, 4.8 Hz, 1H, H-5'b), 2.49 (ddd, *J* = 9.7, 6.1, 3.0 Hz, 1H, H-2'a), 2.40 (dt, *J* = 14.4, 3.3 Hz, 1H, H-2'b).

$^{13}$C NMR (176 MHz, $CDCl_3$) δ/ppm = 158.7 (CONHR), 149.9 (C-2), 144.7 (C-6), 139.6 ($C^{Ar}$), 136.2 ($C^{Ar}$), 135.8 (C-4), 130.2 (C-8), 129.3 ($C^{Ar}$), 128.3 ($C^{Ar}$), 128.1 ($C^{Ar}$), 127.9 ($C^{Ar}$), 127.1 ($C^{Ar}$), 123.9 ($C^{Ar}$), 113.4 (CCONHR), 102.7 ($C^{Ar}$), 102.3 ($C^{Ar}$), 95.8 (C-3), 86.8 (C-4'), 85.5 (C-1'), 72.8 (C-3'), 64.0 (C-5'), 55.4 ($OCH_3$), 55.3($OCH_3$), 40.8 (C-2').

$^{19}$F NMR (470 MHz, $CDCl_3$) δ/ppm = -140.07 (dt, *J* = 23.3, 9.4 Hz, 2F, $CF^{ortho}$), -150.21 (q, *J* = 21.3 Hz, 1F, $CF^{para}$), -159.84 (dtt, *J* = 28.3, 15.3, 6.3 Hz, 2F, $CF^{meta}$).

MS: Calculated for $[M+H]^+$ 746.2, found: 746.8.

**Compound 10h**



Chemical Formula: $C_{49}H_{49}F_5N_5O_7P$
Exact Mass: 945,3290
Molecular Weight: 945,9248

After drying in high vacuum, compound **10g** (150 mg, 0.20 mmol, 1.00 eq.) was dissolved in dry DCM (3.0 mL) containing DIPEA (105 µL, 0.60 mmol, 3.00 eq.). 2-Cyanoethyl-*N*,*N*-diisopropylphosphoramidochloridite (90 µL, 0.40 mmol, 2.00 eq.) was added to the solution under argon and the reaction mixture was stirred for 30 min at room temperature. DCM (10 mL) was added to the mixture, and the organic layer was washed with aqueous saturated $NaHCO_3$ (4 x 30 mL), dried over $Na_2SO_4$, and concentrated in vacuo. The residue was purified via flash column chromatography (10 to 50% EA in CH) to give the phosphoramidite (136 mg, 0.144 mmol, 71%) as yellow solid.

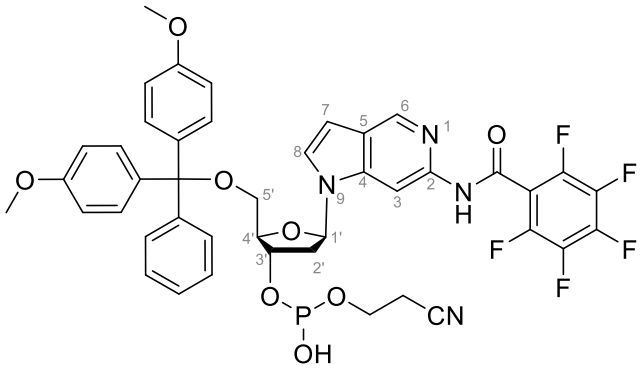
$^1$H NMR (700 MHz, $CDCl_3$) δ/ppm = 8.50 (s, 1H, H-6), 8.44 (s, 1H, H-3), 7.38 (t, *J* = 3.5 Hz, 1H, H-8), 7.37 – 7.29 (m, 4H, $H^{Ar}$), 7.25 – 7.18 (m, 1H, $H^{Ar}$), 6.87 (ddt, *J* = 9.0, 6.7, 1.8 Hz, 2H, $H^{Ar}$), 6.80 – 6.76 (m, 3H, $H^{Ar}$), 6.54 (d, *J* = 3.3 Hz, 1H, $H^{Ar}$), 6.45 (dd, *J* = 6.8 Hz, 1H, H-1'), 4.66 (q, *J* = 8.4, 7.9 Hz, 1H, H-3'), 4.12 (q, *J* = 7.2 Hz, 1H, H-4'), 3.79 (s, 3H, $OC\underline{H}_3$), 3.78 (s, 3H, $OC\underline{H}_3$), 3.68 – 3.55 (m, 2H, $POC\underline{H}_2CH_2CN$), 3.36 (ddd, *J* = 10.2, 7.4, 4.0 Hz, 1H, H-5'a), 3.30 – 3.27 (m, 1H, H-5'b), 3.20 (td, *J* = 10.1, 3.8 Hz, 2H, $N^{iPr}(C\underline{H}(CH_3)_2)_2$), 2.65 (t, *J* = 6.3 Hz, 1H, H-2'a), 2.53 (t, *J* = 6.3 Hz, 1H, H-2'b), 2.62 – 2.58 (m, 2H, $POCH_2C\underline{H}_2CN$), 1.21 – 1.16 (m, 12H, $N^{iPr}(CH(C\underline{H}_3)_2)_2$.

$^{13}$C NMR (176 MHz, $CDCl_3$) δ/ppm = 158.7 ($\underline{C}ONHR$), 144.8 (C-2), 141.6 (C-6), 141.5 ($C^{Ar}$), 141.3 ($C^{Ar}$), 141.2 ($C^{Ar}$), 135.9 ($C^{Ar}$), 130.3 (C-8), 128.4 ($C^{Ar}$), 128.1 ($C^{Ar}$), 127.0 ($C^{Ar}$), 126.5 ($C^{Ar}$), 124.5 ($C^{Ar}$), 124.4, ($C^{Ar}$), 117.7 ($POCH_2CH_2\underline{C}N$), 113.4 ($C^{Ar}$), 102.7 ($C^{Ar}$), 102.2 ($C^{Ar}$), 95.7 (C-3), 86.8 (C-4'), 85.2 (C-1'), 74.1 (C-3'), 63.7 (C-5'), 58.6 ($PO\underline{C}H_2CH_2CN$), 55.4 ($O\underline{C}H_3$), 53.6 ($N^{iPr}(\underline{C}H(CH_3)_2)_2$), 40.2 (C-2'), 24.8 (4 C, $N^{iPr}(CH(\underline{C}H_3)_2)_2$), 21.2 ($POCH_2\underline{C}H_2CN$).

$^{19}$F NMR (470 MHz, $CDCl_3$) δ/ppm = -139.99 (p, *J* = 10.9, 10.2 Hz, 2F, $CF^{ortho}$), -149.76 – -150.05 (m, 1F, $CF^{para}$), -159.66 (qd, *J* = 21.4, 7.6 Hz, 2F, $CF^{meta}$).

$^{31}$P NMR (283 MHz, $CDCl_3$) δ/ppm = 148.75.

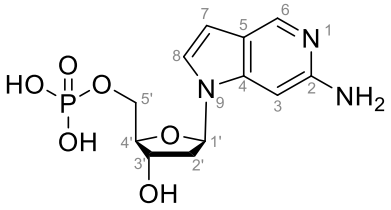
The product hydrolyzes under LC-MS conditions to the structure depicted below.



Chemical Formula: $C_{43}H_{36}F_5N_4O_8P$
Exact Mass: 862,2191
Molecular Weight: 862,7468

MS: Calculated for $[M+H]^+$ 864.2, found: 864.2.

**2-Amino-DDP monophosphate (Compound 11)**



Chemical Formula: $C_{12}H_{16}N_3O_6P$
Exact Mass: 329,0777
Molecular Weight: 329,2488

In a dry Schlenk flask, nucleoside **10** (50 mg, 0.138 mmol, 1.00 eq.) was dried under high vacuum, before trimethyl phosphate (3.0 mL) was added under an argon atmosphere. The mixture was cooled in an ice bath and $POCl_3$ (0.126 mL, 1.38 mmol, 10.0 eq.) was added dropwise. The reaction mixture was stirred at 0 °C for 20 min and at rt for 7 h, before a solution of 10% TEA in water (3 mL) was added and the mixture was stirred for another 15 min at rt. The mixture was concentrated under reduced pressure and was purified via RP-HPLC, with 5 to 20% ACN in water containing 0.1% TFA to give the nucleoside monophosphate (25.0 mg, 56.4 µmol, 41%) as yellow oil.

$^1$H NMR (700 MHz, $D_2O$) δ/ppm = δ 8.16 (s, 1H, H-6), 7.53 (d, *J* = 3.7 Hz, 1H, H-8), 7.06 (s, 1H, H-3), 6.64 (d, *J* = 3.6 Hz, 1H, H-7), 6.31 (dd, *J* = 8.5, 6.0 Hz, 1H, H-1'), 4.69 (dt, *J* = 5.9, 2.7 Hz, 1H, H-3'), 4.23 (p, *J*

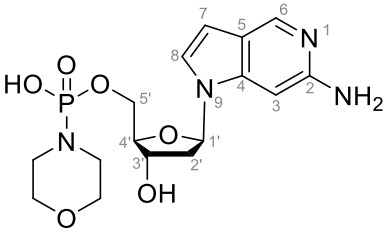
= 2.8 Hz, 1H, H-4'), 4.10 (dd, *J* = 4.8, 3.2 Hz, 2H, H-5'a, H-5'b), 2.66 (ddd, *J* = 14.5, 8.6, 6.3 Hz, 1H, H-2'a), 2.39 (ddd, *J* = 14.1, 6.1, 2.6 Hz, 1H, H-2'b).

$^{13}$C NMR (176 MHz, $D_2O$) δ/ppm = 148.5 (C-4), 145.6 (C-2), 131.1 (C-8), 129.8 (C-6), 120.2 (C-5), 103.4 (C-7), 90.3 (C-3), 86.1 (C-1'), 85.3 (C-4'), 71.3 (C-3'), 64.9 (C-5'), 38.4 (C-2').

$^{31}$P NMR (283 MHz, $D_2O$) δ/ppm = 0.04.

HRMS: Calculated for $[M-H]^-$ 328.0704, found: 328.0706.

**2-Amino-DDP morpholidate (Compound 12)**



Chemical Formula: $C_{16}H_{23}N_4O_6P$
Exact Mass: 398,1355
Molecular Weight: 398,3558

In a dry Schlenk flask, compound **11** (20.0 mg, 0.045 mmol, 1.00 eq.) was dissolved in dry DMSO (1.0 mL). Morpholine (39.3 mg, 0.451 mmol, 10.0 eq.) was added and the reaction mixture was stirred at room temperature for 5 min. Dipyridyl disulfide (29.8 mg, 0.135 mmol, 3.00 eq.) and triphenylphosphine (35.5 mg, 0.135 mmol, 3.00 eq.) were added and the reaction mixture was stirred for 2 h at room temperature. Water (10 mL) was added and the mixture was extracted with DCM (3 x 20 mL) to remove lipophilic components. The aqueous phase was dried by lyophilization and the crude product was purified via HPLC (5 to 20 % ACN in water) to give 11.0 mg (0.028 mmol, 61%) of the morpholidate as yellow solid.

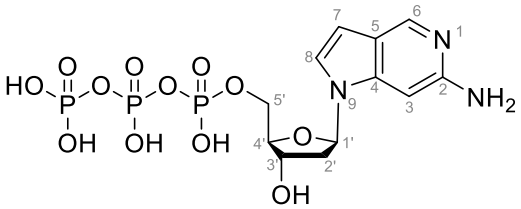
$^1$H NMR (600 MHz, $D_2O$) δ/ppm = 8.27 (s, 1H, H-6), 7.57 (d, *J* = 3.7 Hz, 1H, H-8), 7.14 (s, 1H, H-3), 6.72 (d, *J* = 3.4 Hz, 1H, H-7), 6.36 (dd, *J* = 7.7, 6.2 Hz, 1H, H-1'), 4.70 (dt, *J* = 6.5, 3.3 Hz, 1H, H-3'), 4.22 – 4.20 (m, 1H, H-4'), 3.97 – 3.95 (m, 2H, H-5'a, H-5'b), 3.62 – 3.54 (m, 4H, $O^{morpholidate}CH_2$), 2.98 – 2.93 (m, 4H, $N^{morpholidate}CH_2$), 2.76 (ddd, *J* = 14.1, 7.7, 6.4 Hz, 1H, H-2'a), 2.46 (ddd, *J* = 14.1, 6.2, 3.4 Hz, 1H, H-2'b).

$^{13}$C NMR (151 MHz, $D_2O$) δ/ppm = 148.7 (C-2), 146.0 (C-4), 130.9 (C-8), 130.1 (C-6), 120.2 (C-5), 103.5 (C-7), 90.4 (C-3), 85.9 (C-1'), 85.4 (C-4'), 71.1 (C-3'), 66.9 ($O^{morpholidate}\underline{C}H_2$), 64.4 (C-5'), 44.69 ($N^{morpholidate}\underline{C}H_2$), 38.2 (C-2').

$^{31}$P NMR (243 MHz, $D_2O$) δ/ppm = 7.37.

HRMS: Calculated for $[M-H]^-$ 397.1282, found: 397.1282.

**2-Amino-DDP triphosphate (Compound 13)**



Chemical Formula: $C_{12}H_{18}N_3O_{12}P_3$
Exact Mass: 489,0103
Molecular Weight: 489,2063

Compound **12** (8.0 mg, 20.1 µmol, 1.00 eq.) was dried in a Schlenk flask. Anhydrous DMF (500 µL) was added and the solution was stirred at room temperature. Tributyl ammonium pyrophosphate (36.5 mg, 100.4 µmol, 5.00 eq.) was added and the reaction mixture was stirred at rt for 24 h, before it was lyophilized. The crude product was purified using an HPLC with ion exchange column with Tris-buffer and a NaCl-solution with raising concentration (0 to 1.25 M NaCl). The resulting product fractions were desalted and purified using an HPLC with C18-column using a system of 100 mM TEAA-buffer (A) and a mixture of 20% 500 mM TEAA-buffer and 80% acetonitrile (B). The resulting triphosphate was lyophilized and obtained as TEA-salt (2.84 mg, 3.58 µmol, 18%) and stored at -70 °C.

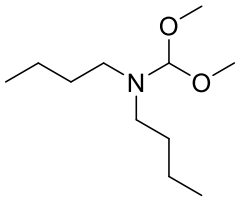
$^1$H NMR (600 MHz, $D_2O$) δ/ppm = 8.30 (d, *J* = 4.6 Hz, 1H, H-6), 7.55 (d, *J* = 3.7 Hz, 1H, H-8), 7.17 (s, 1H, H-3), 6.70 (d, *J* = 3.6 Hz, 1H, H-7), 6.36 (dd, *J* = 8.3, 6.2 Hz, 1H, H-1'), 4.85 – 4.83 (m, 1H, H-3'), 4.44 – 4.40 (m, 1H, H-4'), 4.31 – 4.27 (m, 1H, H-5'a), 4.23 – 4.19 (m, 1H, H-5'b), 2.75 (dt, *J* = 14.5, 7.5 Hz, 1H, H-2'a), 2.41 (dt, *J* = 14.1, 4.0 Hz, 2H, H-2'b).

$^{13}$C NMR (151 MHz, $D_2O$) δ/ppm = 149.6 (C-4), 145.3 (C-2), 132.2 (C-6), 130.5 (C-8), 120.7 (C-5), 103.3 (C-7), 90.7 (C-3), 86.2 (C-1'), 85.6 (C-4'), 70.8 (C-3'), 65.0 (C-5'), 38.4 (C-2').

$^{31}$P NMR (243 MHz, $D_2O$) δ/ppm = -7.38 – -7.56 (m, 1P, γ-P), -11.07 (dd, *J* = 66.0, 19.9 Hz, 1P, α-P), -22.15 – -22.42 (m, 1P, β-P).

HRMS: Calculated for $[M-H]^-$ 488.0031, found: 488.0019.

***N,N*-dibutyl formamide dimethyl acetal (17)**



Chemical Formula: $C_{11}H_{25}NO_2$
Molecular Weight: 203,3260

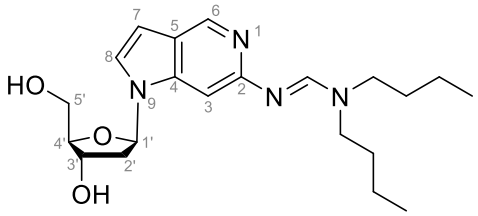
*N*,*N*-dibutyl formamide (28.9 mL, 159 mmol, 1.00 eq.) and dimethyl sulfate (15.1 mL, 159 mmol, 1.00 eq.) were mixed and stirred at 90 °C for 3 d. The resulting brown reaction mixture was then cooled to room temperature, before diethyl ether (20 mL) was slowly added over 10 min while stirring. The upper, ethereal layer was removed with a cannula, and diethyl ether (20 mL) was added to the residue. The upper, ethereal layer was removed again. The remaining brown mixture was slowly poured into a rapidly stirring, cold (0 °C) solution of 5.4 M sodium methoxide in methanol (88.3 mL, 477 mmol, 3.00 eq.) and was stirred for 16 h at room temperature, before it was extracted with cyclohexane (5 × 100 mL). The combined organic extracts were dried over sodium sulfate and concentrated in a low vacuum to a yellow liquid. The mixture was distilled under high vacuum with a slowly raising bath temperature (final temperature: 130 °C) to give 15.9 g (78.0 mmol, 49%) of *N*,*N*-dibutyl formamide dimethyl acetal as a clear, colorless liquid (b.p. 62 °C at 1.5 mbar). The collected fraction contained 11% *N*,*N*-dibutyl formamide (determined by NMR).

$^1$H NMR (500 MHz, $CDCl_3$) δ/ppm = 4.52 (s, 1H, $NC\underline{H}(OCH_3)_2$), 3.31 (s, 6H, $OCH_3$), 2.63 – 2.57 (m, 4H, $N(C\underline{H}_2CH_2CH_2CH_3)_2$), 1.45 – 1.38 (m, 4H, $N(CH_2C\underline{H}_2CH_2CH_3)_2$), 1.34 – 1.26 (m, 4H, $N(CH_2CH_2C\underline{H}_2CH_3)_2$), 0.90 (t, *J* = 7.3 Hz, 6H, $N(CH_2CH_2CH_2C\underline{H}_3)_2$).

$^{13}$C NMR (126 MHz, $CDCl_3$) δ/ppm = 112.8 ($N\underline{C}H(OCH_3)_2$), 54.1 ($OCH_3$), 47.3 ($N(\underline{C}H_2CH_2CH_2CH_3)_2$), 31.2 ($N(CH_2\underline{C}H_2CH_2CH_3)_2$), 20.7($N(CH_2CH_2\underline{C}H_2CH_3)_2$), 14.2 ($N(CH_2CH_2CH_2\underline{C}H_3)_2$).

The product is not stable under LC-MS conditions; therefore, no exact mass could be determined.

**Compound 18**



Chemical Formula: $C_{21}H_{32}N_4O_3$
Exact Mass: 388,2474
Molecular Weight: 388,5120

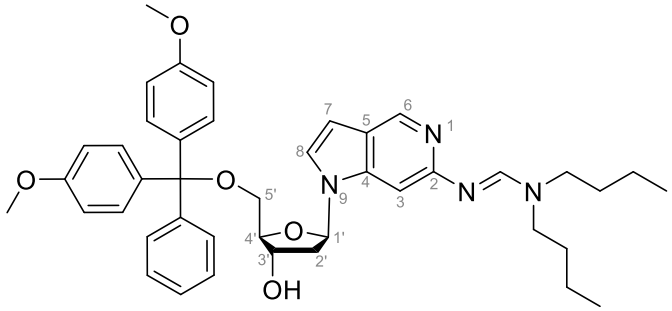
Nucleoside **10** (160 mg, 0.64 mmol, 1.00 eq.) was dissolved in THF (5 mL). *N*,*N*-Dibutyl formamide dimethyl acetal (0.35 mL, 1.93 mmol, 3.00 eq.) was added and the mixture was stirred at rt for 3 days. The reaction mixture was concentrated under reduced pressure. The crude product was purified via HPLC to give 240 mg (0.62 mmol, 95%) of **18**.

$^1$H NMR (700 MHz, Methanol-$d_4$) δ/ppm = 8.44 (s, 1H, H-6), 8.44 (s, 1H, N=C<u>H</u>N), 7.66 (d, *J* = 3.5 Hz, 1H, H-8), 7.51 (s, 1H, H-3), 6.67 (d, *J* = 3.5 Hz, 1H, H-7), 6.36 (dd, *J* = 7.5, 6.2 Hz, 1H, H-1'), 4.46 (dt, *J* = 6.2, 3.1 Hz, 1H, H-3'), 3.95 (q, *J* = 3.8 Hz, 1H, H-4'), 3.73 – 3.65 (m, 2H, H-5'), 3.52 (t, *J* = 7.8 Hz, 2H, (N(C<u>H</u>$_2$CH$_2$CH$_2$CH$_3$)$_2$)), 3.44 (t, *J* = 7.4 Hz, 2H, (N(C<u>H</u>$_2$CH$_2$CH$_2$CH$_3$)$_2$)), 2.50 (ddd, *J* = 13.8, 7.5, 6.3 Hz, 1H, H-2'a), 2.34 (ddd, *J* = 13.6, 6.0, 3.2 Hz, 1H, H-2'b), 1.64 – 1.60 (m, 4H, (N(CH$_2$C<u>H</u>$_2$CH$_2$CH$_3$)$_2$)), 1.38 – 1.30 (m, 4H, (N(CH$_2$CH$_2$C<u>H</u>$_2$CH$_3$)$_2$)), 0.95 – 0.88 (m, 6H, (N(CH$_2$CH$_2$CH$_2$C<u>H</u>$_3$)$_2$)).

$^{13}$C NMR (176 MHz, Methanol-$d_4$) δ/ppm = 156.7 (N<u>C</u>H=N), 152.1 (C-2), 145.8 (C-4), 135.9 (C-6), 131.1 (C-8), 124.1 (C-5), 104.3 (C-7), 94.7 (C-3), 88.9 (C-4'), 86.8 (C-1'), 72.5 (C-3'), 63.2 (C-5'), 53.7 (N(<u>C</u>H$_2$CH$_2$CH$_2$CH$_3$)$_2$), 47.1 (N(<u>C</u>H$_2$CH$_2$CH$_2$CH$_3$)$_2$), 41.3 (C-2'), 29.2 (N(CH$_2$<u>C</u>H$_2$CH$_2$CH$_3$)$_2$), 20.7 (N(CH$_2$CH$_2$<u>C</u>H$_2$CH$_3$)$_2$), 14.0 (N(CH$_2$CH$_2$CH$_2$<u>C</u>H$_3$)$_2$).

HRMS: Calculated for $[M+H]^+$ 389.2547, found: 389.2537.

**Compound 19**



Chemical Formula: $C_{42}H_{50}N_4O_5$
Exact Mass: 690,3781
Molecular Weight: 690,8850

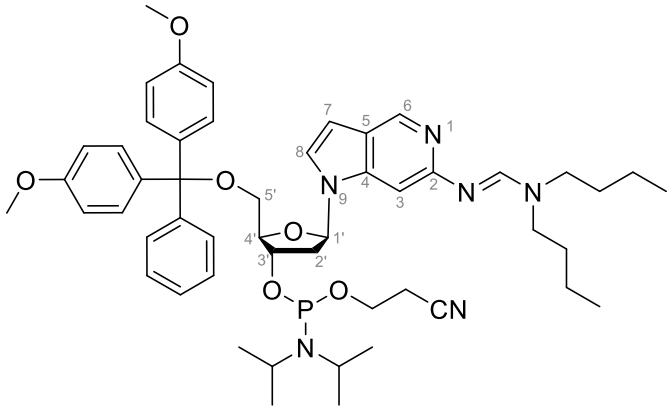
In a dry Schlenk tube, **18** (200 mg, 0.515 mmol, 1.00 eq.) and 4,4´-dimethoxytrityl chloride (523 mg, 1.54 mmol, 3.00 eq.) were dried under high vacuum, before anhydrous THF (15.0 mL) and DIPEA (0.44 mL, 2.57 mmol, 5.00 eq.) were added at rt. The resulting yellow mixture was stirred at rt for 16 h, before MeOH (5 mL) was added and stirring was continued for another 5 min. The reaction mixture was evaporated to dryness under reduced pressure. The crude product was purified via flash column chromatography (30% to 70% acetone in DCM with 0.1% TEA) to give 92.0 mg (0.133 mmol, 26%) of the DMT-protected product as yellow solid.

$^1$H NMR (600 MHz, Methanol-$d_4$) δ/ppm = 8.45 (d, $J$ = 0.9 Hz, 1H, H-6), 8.20 (s, 1H, N=CHN), 7.37 – 7.36 (m, 1H, $C^{DMT}$-$H^{para}$), 7.36 – 7.35 (m, 1H, $C^{Ar}$H), 7.33 (d, $J$ = 3.5 Hz, 1H, H-8), 7.24 – 7.21 (m, 4H, $C^{DMT}$-$H^{meta}$), 7.18 – 7.09 (m, 4 H, $C^{Phenyl}$H), 6.73 – 6.70 (m, 4H, $C^{DMT}$-$H^{ortho}$), 6.50 (d, $J$ = 3.5 Hz, 1H, H-7), 6.37 (dd, $J$ = 6.4 Hz, 1H, H-1'), 4.53 (dt, $J$ = 6.4, 4.1 Hz, 1H, H-3'), 4.06 (dt, $J$ = 4.9, 3.6 Hz, 1H, H-4'), 3.70 (s, 6H, $OCH_3$), 3.32 – 3.25 (m, 4H, $N(C\underline{H}_2CH_2CH_2CH_3)_2$), 2.65 (ddd, $J$ = 13.3, 6.5 Hz, 1H, H-2'a), 2.40 (ddd, $J$ = 13.5, 6.2 Hz, 1H, H-2'b), 1.68 – 1.56 (m, 4H, $N(CH_2C\underline{H}_2CH_2CH_3)_2$), 1.34 – 1.26 (m, 4H, $N(CH_2CH_2C\underline{H}_2CH_3)_2$), 0.96 (tt, $J$ = 14.9, 7.4 Hz, 6H, $N(CH_2CH_2CH_2C\underline{H}_3)_2$).

$^{13}$C NMR (151 MHz, Methanol-$d_4$) δ/ppm = 160.0 ($\underline{C}^{DMT}$-O-$CH_3$), 156.9 (C-2), 156.8 (N=$\underline{C}$HN), 146.3 (C-4), 141.1 (C-6), 137.1 ($C^{DMT}$-$H^{para}$), 131.3 ($C^{DMT}$-$H^{meta}$), 129.3 ($C^{DMT}$-$H^{para}$), 128.7 ($C^{DMT}$-$H^{meta}$), 127.4 (C-8), 124.4 (C-5), 114.0 ($C^{DMT}$-$H^{meta}$), 103.3 (C-7), 97.5 (C-3), 87.5 ($C^{DMT}$-O-5'), 87.1 (C-4'), 86.1 (C-1'), 72.6 (C-3'), 65.2 (C-5'), 55.7 (O-$CH_3$), 52.9 ($N(\underline{C}H_2CH_2CH_2CH_3)_2$), 40.8 (C-2'), 32.3 ($N(CH_2\underline{C}H_2CH_2CH_3)_2$), 30.3 $N(CH_2\underline{C}H_2CH_2CH_3)_2$), 21.2 ($N(CH_2CH_2\underline{C}H_2CH_3)_2$), 20.8 ($N(CH_2CH_2\underline{C}H_2CH_3)_2$), 14.3 ($N(CH_2CH_2CH_2\underline{C}H_3)_2$), 14.1 ($N(CH_2CH_2CH_2\underline{C}H_3)_2$).

HRMS: Calculated for $[M+H]^+$ 691.3854, found: 691.3827.

**Compound 20**



Chemical Formula: $C_{51}H_{67}N_6O_6P$
Exact Mass: 890,4860
Molecular Weight: 891,1068

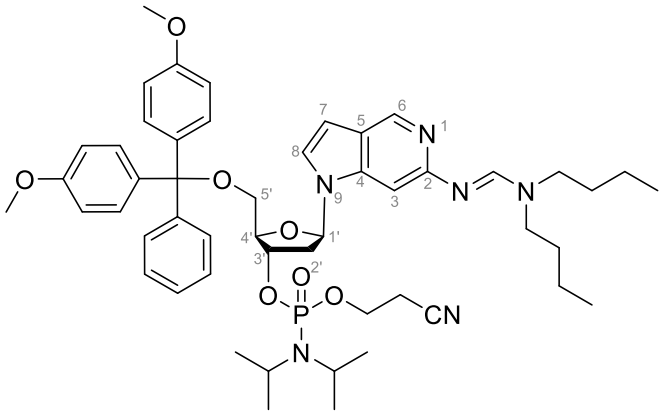
After drying in high vacuum, **19** (92.0 mg, 0.133 mmol, 1.00 eq.) was dissolved in dry DCM (1.0 mL) containing DIPEA (69.8 µL, 0.399 mmol, 3.00 eq.). 2-Cyanoethyl-*N*,*N*-diisopropylphosphoramidochloridite (Compound 12, 59.5 µL, 0.266 mmol, 2.00 eq.) was added to the solution under argon and the reaction mixture was stirred for 30 min at rt. The mixture was concentrated in vacuo and purified via flash column chromatography (20 to 70% EA in CH) to give the phosphoramidite (104 mg, 0.117 mmol, 88%).

$^{1}$H NMR (600 MHz, Acetonitrile-$d_3$) δ/ppm = 8.54 (s, 1H, N=CHN), 8.46 (d, *J* = 0.9 Hz, 1H, H-6), 7.39 – 7.36 (m, 2H, $H^{Ar,DMT}$), 7.27 – 7.21 (m, 7H, H-8, $H^{Ar,DMT}$), 7.21 – 7.17 (m, 1H, $H^{Ar,DMT}$), 6.90 – 6.89 (m, 1H, H-3), 6.78 – 6.75 (m, 4H, $H^{ortho,DMT}$), 6.50 (dd, *J* = 3.4, 0.9 Hz, 1H, H-7), 6.33 (t, *J* = 6.3 Hz, 1H, H-1'), 4.72 (ddt, *J* = 11.0, 6.6, 4.5 Hz, 1H, H-3'), 4.13 (tdd, *J* = 4.6, 3.3, 0.9 Hz, 1H, H-4'), 3.74 (s, 3H, $OCH_3$), 3.73 (s, 3H, $OCH_3$), 3.71 – 3.57 (m, 3H, $POCH_2CH_2CN$, $N^{iPr}(CH(CH_3)_2)_2$), 3.48 (t, *J* = 7.6 Hz, 2H, $N(CH_2CH_2CH_2CH_3)_2$), 3.33 – 3.29 (m, 2H, $N(CH_2CH_2CH_2CH_3)_2$), 3.28 (dd, *J* = 10.5, 3.4 Hz, 1H, H-5'a), 3.13 (dd, *J* = 10.5, 4.7 Hz, 1H, H-5'b), 2.72 – 2.65 (m, 1H, H-2'a), 2.55 – 2.48 (m, 3H, H-2'b, $POCH_2CH_2CN$), 1.67 – 1.55 (m, 4H, $N(CH_2CH_2CH_2CH_3)_2$), 1.39 – 1.30 (m, 4H, $N(CH_2CH_2CH_2CH_3)_2$), 1.20 – 1.15 (m, 12H, $N^{iPr}(CH(CH_3)_2)_2$), 0.95 (q, *J* = 8.9, 8.5 Hz, 6H, $N(CH_2CH_2CH_2CH_3)_2$).

$^{13}$C NMR (151 MHz, Acetonitrile-$d_3$) δ/ppm = 159.6 ($C^{DMT}$-O-$CH_3$), 157.7 (C-2), 155.2 (N=CHN), 146.1 ($C^{Ar,DMT}$), 144.0 (C-4), 142.2 (C-6), 136.8 ($C^{Ar,DMT,meta}$), 131.0 ($C^{Ar,DMT,meta}$), 129.0 ($C^{Ar,DMT,meta}$), 128.8 (C-8), 127.7 ($C^{Ar,DMT}$), 125.6 ($C^{Ar,DMT}$), 123.6 (C-5), 119.4 ($POCH_2CH_2CN$), 114.0 ($C^{Ar,DMT,ortho}$), 102.9 (C-7), 98.0 (C-3), 86.9 ($C^{DMT}$-O-5'), 85.6 (C-1'), 85.5 (C-4'), 73.8 (C-3'), 64.3 (C-5'), 59.5 ($POCH_2CH_2CN$), 55.9 (2 C, $OCH_3$), 51.9 ($N^{iPr}(CH(CH_3)_2)_2$), 51.9 ($N(CH_2CH_2CH_2CH_3)_2$), 45.5 ($N(CH_2CH_2CH_2CH_3)_2$), 39.4 (C-2'), 32.1 ($N(CH_2CH_2CH_2CH_3)_2$), 30.1 ($N(CH_2CH_2CH_2CH_3)_2$), 24.9 (4 C, $N^{iPr}(CH(CH_3)_2)_2$), 21.0 ($POCH_2CH_2CN$), 21.0 ($N(CH_2CH_2CH_2CH_3)_2$), 20.9 ($N(CH_2CH_2CH_2CH_3)_2$), 14.3 ($N(CH_2CH_2CH_2CH_3)_2$), 14.1 ($N(CH_2CH_2CH_2CH_3)_2$).

$^{31}$P NMR (243 MHz, Acetonitrile-$d_3$) δ/ppm = 147.96.

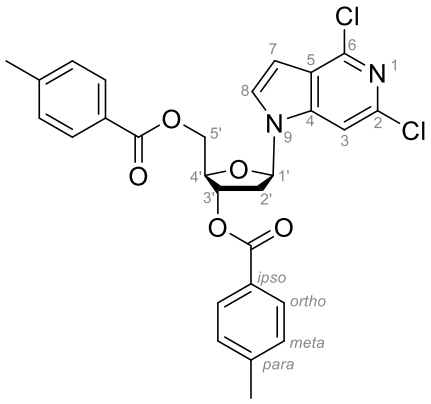
The product oxidizes under LC-MS conditions to the structure depicted below.



Chemical Formula: $C_{51}H_{67}N_6O_7P$
Exact Mass: 906,4809
Molecular Weight: 907,1058

HRMS: Calculated for $[M+H]^+$ 907.4882, found: 907.4886.

**Compound 22**



Chemical Formula: $C_{28}H_{24}Cl_2N_2O_5$
Molecular Weight: 539,4090

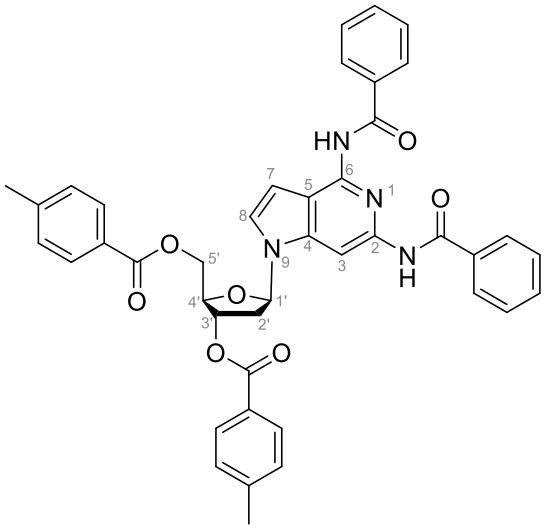
Freshly powdered KOH (3.75 g, 66.8 mmol, 5.00 eq.) was dissolved in dry THF (100 mL) in a dried Schlenk tube. 4,6-Dichloro-1H-pyrrolo[3,2-c]pyridine (2.50 g, 13.4 mmol, 1.00 eq.) and 1-Chloro-2-deoxy-3,5-di-O-toluoyl-ribose (5.72 g, 14.7 mmol, 1.10 eq.) were added and the resulting mixture was stirred for 30 min at room temperature. The mixture was filtered through Celite, washed with ethyl acetate and the solvent was removed *in vacuo*. The crude product was purified *via* column chromatography (0 to 50% EA in PE) to give the product (6.50 g, 12.1 mmol, 90%) as a yellow solid.

$^{1}$H NMR (500 MHz, $CDCl_3$) δ/ppm = 7.98 – 7.96 (m, 2H, $C^{5'OBz\text{-}ortho}$-H), 7.87 – 7.86 (m, 2H, $C^{3'OBz\text{-}ortho}$-H), 7.39 (s, 1H, H-3), 7.33 (d, *J* = 3.5 Hz, 1H, H-8), 7.30 – 7.28 (m, 2H, $C^{5'OBz\text{-}meta}$-H), 7.24 – 7.22 (m, 2H, $C^{3'OBz\text{-}meta}$-H), 6.61 (d, *J* = 3.5, 1H, H-7), 6.33 (dd, *J* = 8.1, 5.7 Hz, 1H, H-1'), 5.70 (dt, *J* = 5.7, 2.5 Hz, 1H, H-3'), 4.71 – 4.66 (m, 1H, H-5'a), 4.64 – 4.59 (m, 2H, H-4'& H-5'b), 2.82 – 2.70 (m, 2H H-2'a & H-2'b), 2.44 (s, 3H Ar-C$H_3$), 2.42 (s, 3H Ar-C$H_3$).

$^{13}$C NMR (126 MHz, $CDCl_3$) δ/ppm = 166.3 (COOC-3'), 166.1 (COOC-5'), 144.8 ($C^{5'OBz\text{-}ipso}$), 144.5 ($C^{3'OBz\text{-}ipso}$), 142.5 ($C^{Ar}$-H), 142.3 ($C^{Ar}$-H), 141.7 ($C^{Ar}$-H), 129.9 ($C^{5'OBz\text{-}ortho}$), 129.7 ($C^{3'OBz\text{-}ortho}$), 129.5 ($C^{5'OBz\text{-}meta}$), 129.4 $C^{3'OBz\text{-}meta}$), 126.6 ($C^{3'OBz\text{-}para}$), 126.5 ($C^{Ar}$-H), 126.4 ($C^{5'OBz\text{-}para}$), 123.8 ($C^{Ar}$-H), 104.9 (C-3), 103.2 (C-7), 85.9 (C-1'), 82.6 (C-4'), 74.8 (C-3'), 64.0 (C-5'), 38.4 (C-2'), 21.9 (Ar-C$H_3$), 21.8 (Ar-C$H_3$).

HRMS: Calculated for $[M+H]^+$ 539.1135, found: 539.1134.

**Compound 23**



Chemical Formula: $C_{42}H_{36}N_4O_7$
Molecular Weight: 708,7710

A dried pressure tube was charged with 4 A molecular sieves, 1.35 g (11.1 mmol, 6.00 eq.) of benzamide, 1.97 g (9.27 mmol, 5.00 eq.) of $K_3PO_4$, and 1.00 g (1.85 mmol, 1.00 eq.) of **22**. 63.4 mg (0.074 mmol, 0.04 eq.) of tBu-BrettPhos Pd G3 was added. Dry *tert*-butanol (30 mL) was added under argon and the mixture was stirred at 110 °C for 20 h and, after being cooled to rt, diluted with ethyl acetate (30 mL), and washed with water (5 x 50 mL). The organic layer was dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The

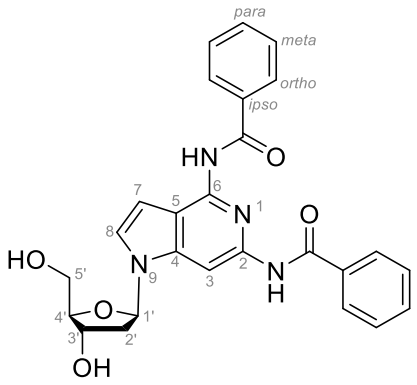
crude product was purified *via* column chromatography (0 to 100% EA in PE with 0.1% $NEt_3$) to give the product (393 mg, 0.554 mmol, 30%) as a yellow solid.

$^1$H NMR (600 MHz, $CDCl_3$) δ/ppm = 8.73 (s, 1H, $N^{Amide}$-H), 8.58 (s, 1H, $N^{Amide}$-H), 8.41 (s, 1H, H-3), 8.01 (m, 2H, $C^{Ar}$-H), 7.95 – 7.91 (m, 4H, $C^{Ar}$-H), 7.90 – 7.87 (m, 2H, $C^{Ar}$-H), 7.52 (m, 2H, $C^{Ar}$-H), 7.48 – 7.42 (m, 5H, $C^{Ar}$-H), 7.29 (m, 2H, $C^{Ar}$-H), 7.25 – 7.23 (m, 1H, $C^{Ar}$-H), 7.22 (m, 2H, $C^{Ar}$-H), 6.77 (d, *J* = 3.6 Hz, 1H, H-7), 6.51 (dd, *J* = 8.5, 5.7 Hz, 1H, H-1'), 5.70 (dt, *J* = 6.6, 2.4 Hz, 1H, H-3'), 4.67 – 4.62 (m, 2H, H-5'a & H-5'b), 4.58 – 4.52 (m, 1H, H-4'), 2.81 – 2.75 (m, 1H, H-2'b), 2.71 (ddd, *J* = 14.3, 5.8, 2.2 Hz, 1H, H-2'a), 2.44 (s, 3H, Ar-C$\underline{H}_3$), 2.38 (s, 3H, Ar-C$\underline{H}_3$).

$^{13}$C NMR (151 MHz, $CDCl_3$) δ/ppm = 166.4 (C=O), 166.3 (C=O), 165.5 (C=O), 165.3 (C=O), 144.5 ($C^{Ar}$), 144.5 ($C^{Ar}$), 144.2 ($C^{Ar}$), 143.9 ($C^{Ar}$), 142.3 ($C^{Ar}$), 134.5 ($C^{Ar}$), 134.2 ($C^{Ar}$), 132.4 ($C^{Ar}$), 132.1 ($C^{Ar}$), 130.0 ($C^{Ar}$), 129.7 ($C^{Ar}$), 129.4 ($C^{Ar}$), 128.9 ($C^{Ar}$), 128.8 ($C^{Ar}$), 127.6 ($C^{Ar}$), 127.2 ($C^{Ar}$), 126.9 ($C^{Ar}$), 126.5 ($C^{Ar}$), 123.9 ($C^{Ar}$), 114.9 (C-5), 105.3 (C-7), 92.9 (C-3), 85.1 (C-1'), 82.3 (C-4'), 75.1 (C-3'), 64.5 (C-5'), 38.4 (C-2'), 21.9 (Ar-$\underline{C}H_3$), 21.8 (Ar-$\underline{C}H_3$).

HRMS: Calculated for $[M+H]^+$ 709.2657, found: 709.2655.

**Compound 24**




Chemical Formula: $C_{26}H_{24}N_4O_5$
Molecular Weight: 472,5010

$K_2CO_3$ (175 mg, 1.27 mmol, 3.00 eq.) was added to a solution of **23** (300 mg, 0.423 mmol, 1.00 eq.) in MeOH (15 mL) at room temperature and the mixture was stirred for 2 h. Water (30 mL) was added and it was extracted with EA (3 x 50 mL). The organic layer was dried over $MgSO_4$ solvent was removed under

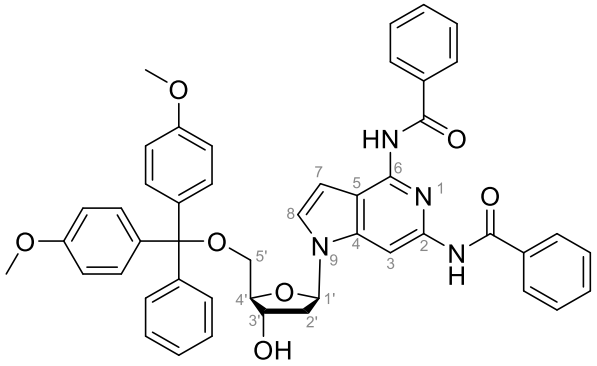
reduced pressure and the residue was purified *via* column chromatography (0 to 10% MeOH in DCM) to give the product (188 mg, 0.398 mmol, 94%) as a yellow solid.

$^{1}$H NMR (600 MHz, DMSO-$d_6$) δ/ppm = 10.64 (s, 1H, $N^{Amide}$-H), 10.49 (s, 1H, $N^{Amide}$-H), 8.26 (s, 1H, H-3), 8.05 – 7.97 (m, 4H, $C^{ortho}$-H), 7.58 – 7.41 (m, 7H, $C^{meta}$-H & $C^{para}$-H & H-8), 6.43 (s, 1H, H-7), 6.33 (d, *J* = 7.2 Hz, 1H, H-1'), 5.34 – 5.30 (m, 1H, 3'-OH), 4.86 (s, 1H, 5'-OH), 4.32 (s, 1H, H-3'), 3.84 (s, 1H, H-4'), 3.54 – 3.50 (m, 1H, H-5'a), 3.50 – 3.47 (m, 1H, H-5'b), 2.45 – 2.38 (m, 1H, H-2'a), 2.27 – 2.18 (m, 1H, H-2'b).

$^{13}$C NMR (151 MHz, DMSO-$d_6$) δ/ppm = 165.5 (C=O), 165.4 (C=O), 143.0 (C-6), 142.7 (C-2), 134.4 ($C^{ipso}$), 134.0 ($C^{ipso}$), 131.9 ($C^{para}$), 131.7 ($C^{para}$), 128.4 (4 C, $C^{meta}$), 128.4 (C-8), 127.9 (2 C, $C^{ortho}$), 127.8 (2 C, $C^{ortho}$), 125.9 (C-4), 117.4 (C-5), 103.1 (C-7), 94.5 (C-3), 87.3 (C-4'), 84.5 (C-1'), 71.0 (C-3'), 62.0 (C-5'), 39.4 (C-2').

HRMS: Calculated for $[M+H]^+$ 473.1819, found: 473.1817.

**Compound 25**




Chemical Formula: $C_{47}H_{42}N_4O_7$
Molecular Weight: 774,8740

In a dry Schlenk tube, 4,4-Dimethoxytrityl chloride (323 mg, 0.952 mmol, 3.00 eq.), **24** (150 mg, 0.317 mmol, 1.00 eq.) and DMAP (3.9 mg, 0.032 mmol, 0.10 eq.) were solved in anhydrous pyridine (5 mL). The resulting orange solution was stirred at room temperature for 6 h. MeOH (1 mL) was added and stirring was continued for another 10 min. The reaction mixture was poured into a saturated $NaHCO_3$ solution (10 mL) and the aqueous phase was extracted with DCM (3x20 mL), and the combined organic phase was dried over $Na_2SO_4$, filtered and evaporated to dryness under reduced pressure. The crude product was purified

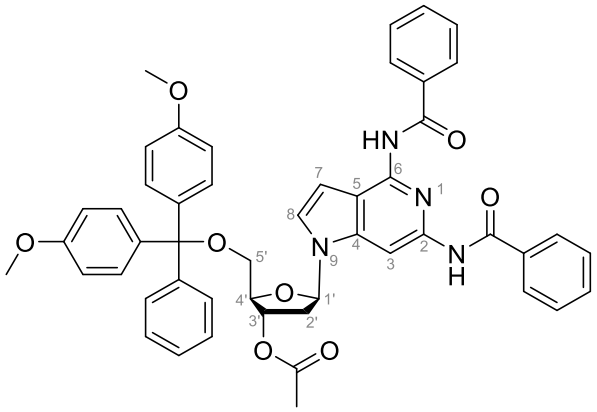
*via* column chromatography (0 to 10% MeOH in DCM) to yield the product (147 mg, 0.19 mmol, 60%) as a yellow solid.

$^{1}$H NMR (500 MHz, $CDCl_3$) δ/ppm = 8.67 (bs, 1H, NH), 8.46 (bs, 1H, NH), 8.37 (s, 1H, H-3), 7.99 – 7.95 (m, 3H, $C^{Ar}$-H), 7.93 – 7.90 (m, 3H, $C^{Ar}$-H), 7.60 – 7.56 (m, 2H, $C^{Ar}$-H), 7.55 – 7.48 (m, 6H, $C^{Ar}$-H), 7.43 – 7.40 (m, 3H, $C^{Ar}$-H), 7.33 – 7.28 (m, 5H, $C^{Ar}$-H), 7.26 (s, 2H, $C^{Ar}$-H), 6.80 – 6.79 (m, 2H, $C^{Ar}$-H), 6.78 – 6.77 (m, 4H, $C^{Ar}$-H), 6.43 (dd, *J* = 6.6 Hz, 1H, H-1'), 4.59 (dt, *J* = 7.0, 3.8 Hz, 1H, H-3'), 4.10 – 4.08 (m, 1H, H-4'), 3.76 (s, 3H, $OCH_3$), 3.75 (s, 3H, $OCH_3$), 3.39 (dd, *J* = 10.2, 4.4 Hz, 1H, H-5'a), 3.34 (dd, *J* = 10.1, 4.8 Hz, 1H, H-5'b), 2.55 (dt, *J* = 13.5, 6.7 Hz, 1H, H-2'a), 2.44 (ddd, *J* = 13.6, 6.1, 3.8 Hz, 1H, H-2'b).

$^{13}$C NMR (126 MHz, $CDCl_3$) δ/ppm = 165.4 (C=O), 165.4 (C=O), 158.6 ($\underline{C}^{DMT}OCH_3$), 144.7 ($C^{Ar}$), 144.2 ($C^{Ar}$), 135.9 ($C^{Ar}$), 132.5 ($C^{Ar}$), 132.2 ($C^{Ar}$), 130.2 ($C^{Ar}$), 129.0 ($C^{Ar}$), 128.9 ($C^{Ar}$), 128.3 ($C^{Ar}$), 128.0 ($C^{Ar}$), 127.7 ($C^{Ar}$), 127.2 ($C^{Ar}$), 127.0 ($C^{Ar}$), 124.8 ($C^{Ar}$), 113.3 ($C^{Ar}$), 104.4 ($C^{Ar}$), 93.1 (C-3), 86.6 ($O\underline{C}^{DMT}(CAr)_3$), 85.4 (C-4'), 84.9 (C-1'), 72.8 (C-3'), 64.1 (C-5'), 55.3 ($O\underline{C}H_3$), 40.7 (C-2').

HRMS: Calculated for $[M+H]^+$ 775.3126, found: 775.3127.

**Compound 25b**



Chemical Formula: $C_{49}H_{44}N_4O_8$
Exact Mass: 816,3159
Molecular Weight: 816,9110

In a dry Schlenk tube, **25** (140 mg, 0.181 mmol, 1.00 eq.) and DMAP (2.2 mg, 0.018 mmol, 0.10 eq.) were dissolved in anhydrous pyridine (2.0 mL). Acetic anhydride (51 µL, 0.54 mmol, 3.00 eq.) was added and the reaction mixture was stirred for 15 min at room temperature. MeOH (3 mL) was added and stirring was

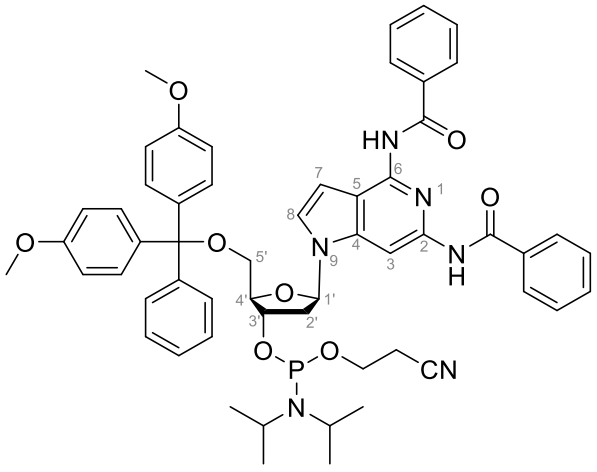
continued for another 10 min before the solvent was removed under reduced pressure. The residue was dissolved in EA (50 mL) and washed with saturated $NaHCO_3$ solution (3x25 mL) and water (25 mL). The separated organic phase was dried over $Na_2SO_4$ and concentrated *in vacuo*. The residue was purified via column chromatography (0 to 100% EA in PE) to yield 48.0 mg (0.059 mmol, 33%) of the product as a yellow solid.

$^1$H NMR (500 MHz, $CDCl_3$) δ/ppm = 8.16 – 8.11 (m, 4H, $C^{Ar}$-H), 7.74 (s, 1H, $C^{Ar}$-H), 7.64 – 7.59 (m, 2H, $C^{Ar}$-H), 7.55 – 7.48 (m, 4H, $C^{Ar}$-H), 7.39 – 7.37 (m, 1H, $C^{Ar}$-H), 7.29 – 7.22 (m, 5H, $C^{Ar}$-H), 7.16 – 7.11 (m, 3H, $C^{Ar}$-H), 6.81 – 6.77 (m, 4H, $C^{Ar}$-H), 6.12 (dd, *J* = 9.1, 5.4 Hz, 1H, H-1'), 5.27 – 5.25 (m, 1H, H-3'), 3.77 (s, 6H, $OCH_3$), 3.73 – 3.69 (m, 1H, H-5'a), 3.62 – 3.58 (m, 1H, H-5'b), 3.46 (s, 1H, H-4'), 2.54 – 2.45 (m, 1H, H-2'a), 2.33 – 2.28 (m, 1H, H-2'b), 2.03 (s, 3H, $C^{Ac}CH_3$).

$^{13}$C NMR (126 MHz, $CDCl_3$) δ/ppm = 170.4 (C=O), 162.5 (C=O), 162.3 (C=O), 158.7 ($\underline{C}^{DMT}OCH_3$), 147.4 ($C^{Ar}$), 142.6 ($C^{Ar}$), 139.6 ($C^{Ar}$), 139.1 ($C^{Ar}$), 138.2 ($C^{Ar}$), 133.9 ($C^{Ar}$), 133.7 ($C^{Ar}$), 131.5 ($C^{Ar}$), 129.2 ($C^{Ar}$), 129.1 ($C^{Ar}$), 128.8 ($C^{Ar}$), 128.0 ($C^{Ar}$), 127.9 ($C^{Ar}$), 127.2 ($C^{Ar}$), 113.3 ($C^{Ar}$), 105.1 ($C^{Ar}$), 86.1 (C-1'), 86.1 (C-4'), 81.5 ($O\underline{C}^{DMT}(CAr)_3$), 74.7 (C-3'), 61.4 (C-5'), 55.4($O\underline{C}H_3$), 38.4 (C-2'), 21.0 ($C^{Ac}\underline{C}H_3$).

MS: Calculated for $[M+H]^+$ 817.3, found: 817.0.

**Compound 26**



Chemical Formula: $C_{56}H_{59}N_6O_8P$
Molecular Weight: 975,0958

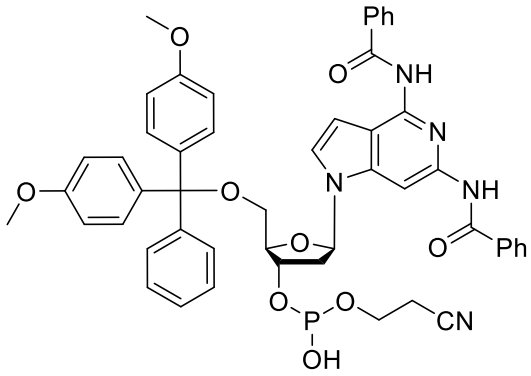
**25** (110 mg, 0.142 mmol, 1.00 eq.) was dissolved in dry DCM (2.0 mL) containing DIPEA (74 µL, 0.426 mmol, 3.00 eq.) under argon atmosphere. 2-Cyanoethyl-*N,N*-diisopropylphosphoramidochloridite (48 µL, 0.213 mmol, 1.50 eq.) was added to the solution and the reaction mixture was stirred for 30 min at room temperature. DCM (10 mL) was added to the mixture, and the organic layer was washed with aqueous saturated $NaHCO_3$ (4 x 30 mL), dried over $Na_2SO_4$, and concentrated *in vacuo*. The residue was purified *via* column chromatography (0 to 100% EA in PE) to yield the product (116 mg, 0.119 mmol, 84%) as a yellow oil.

$^1$H NMR (500 MHz, $CDCl_3$) δ/ppm = 8.50 – 8.48 (m, 1H, NH), 8.42 – 8.40 (m, 1H, NH), 8.35 (s, 1H, H-3), 7.98 – 7.96 (m, 2H, $C^{Ar}$-H), 7.93 – 7.90 (m, 2H, $C^{Ar}$-H), 7.61 – 7.55 (m, 2H, $C^{Ar}$-H), 7.54 – 7.49 (m, 5H, $C^{Ar}$-H), 7.45 – 7.41 (m, 2H, $C^{Ar}$-H), 7.36 – 7.30 (m, 5H, $C^{Ar}$-H), 7.25 – 7.17 (m, 1H, $C^{Ar}$-H), 6.82 – 6.77 (m, 5H, $C^{Ar}$-H), 6.45 (dd, *J* = 6.7 Hz, 1H, H-1'), 4.71 – 4.67 (m, 1H, H-3'), 4.24 (dd, *J* = 7.2, 3.6 Hz, 1H, H-4'), 3.77 – 3.76 (m, 7H, $O^{DMT}CH_3$ & $POCH_2$), 3.74 – 3.67 (m, 2H, $POCH_2$), 3.66 – 3.57 (m, 2H, NCH), 3.43 – 3.37 (m, 1H, H-5'b), 3.36 – 3.29 (m, 1H, H-5'a), 2.68 (dd, *J* = 6.4 Hz, 1H, $CH_2CN$), 2.61 – 2.57 (m, 2H, H-2'a & H-2'b), 2.49 (dd, *J* = 6.5 Hz, 1H, $CH_2CN$), 1.21 – 1.17 (m, 12H, $N(CH(C\underline{H}_3)_2)_2$).

$^{13}$C NMR (126 MHz, $CDCl_3$) δ/ppm = 165.3 (C=O), 165.2 (C=O), 158.6 ($\underline{C}^{DMT}OCH_3$), 144.7 ($C^{Ar}$), 144.5 ($C^{Ar}$), 143.7 ($C^{Ar}$), 142.2 ($C^{Ar}$), 135.9 ($C^{Ar}$), 134.7 ($C^{Ar}$), 134.3 ($C^{Ar}$), 132.4 ($C^{Ar}$), 132.1 ($C^{Ar}$), 130.3 ($C^{Ar}$), 130.2 ($C^{Ar}$), 129.0 ($C^{Ar}$), 128.4 ($C^{Ar}$), 128.3 ($C^{Ar}$), 128.0 ($C^{Ar}$), 127.6 ($C^{Ar}$), 127.1 ($C^{Ar}$), 127.0 ($C^{Ar}$), 124.7 ($C^{Ar}$), 117.8 ($C^{Ar}$), 117.6 ($C^{Ar}$), 114.9 ($C^{Ar}$), 113.2 ($C^{Ar}$), 104.8 ($C^{Ar}$), 93.0 (C-3), 86.5 ($O\underline{C}^{DMT}(CAr)_3$), 85.1, 77.4, 77.2, 76.9, 63.7, 63.5, 58.7, 58.6, 55.4, 53.6, 43.4, 43.4, 43.3, 43.3, 40.1 (C-2'), 24.8, 24.8, 24.8, 24.7, 24.7, 24.7, 24.6, 20.5, 20.5, 20.4, 20.3.

$^{31}$P NMR (243 MHz, $CDCl_3$) δ/ppm = 148.81, 148.68.

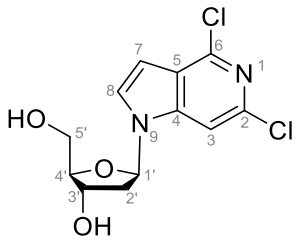
Under LC-MS conditions, the structure hydrolyzed to the structure depicted below.



Chemical Formula: $C_{50}H_{46}N_5O_9P$
Exact Mass: 891,3033
Molecular Weight: 891,9178

HRMS: Calculated for $[M+H]^+$ 892.3106, found: 892.3109.

**Compound 27**



Chemical Formula: $C_{12}H_{12}Cl_2N_2O_3$
Molecular Weight: 303,14

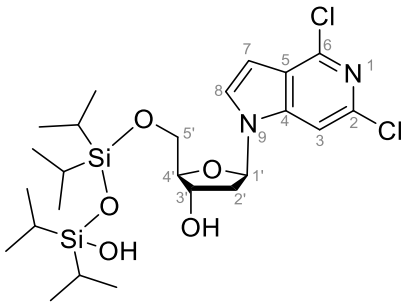
$K_2CO_3$ (846 mg, 6.12 mmol, 3.00 eq.) was added to a solution of **22** (1.10 g, 2.04 mmol, 1.00 eq.) in MeOH (100 mL) at room temperature and the mixture was stirred for 2 h. After removing the solvent in vacuo, water (50 mL) was added and it was extracted with EA (3 x 100 mL). The solvent was removed under reduced pressure and the residue was purified via column chromatography (1 to 10% MeOH in DCM) to give the nucleoside (601 mg, 1.98 mmol, 97%) as a yellow solid.

$^{1}$H NMR (500 MHz, DMSO-$d_6$) δ/ppm = 7.94 (s, 1H, H-3), 7.87 (d, *J* = 3.5 Hz, 1H, H-8), 6.67 (d, *J* = 3.4 Hz, 1H, H-7), 6.40 (dd, *J* = 6.7 Hz, 1H, H-1'), 5.38 – 5.33 (m, 1H, 3'-O<u>H</u>), 5.06 – 4.97 (m, 1H, 5'-O<u>H</u>), 4.41 – 4.34 (m, 1H, H-3'), 3.87 – 3.82 (m, 1H, H-4'), 3.64 – 3.52 (m, 2H, H-5'a & H-5'b), 2.46 – 2.40 (m, 1H, H-2'a), 2.30 – 2.25 (m, 1H, H-2'b).

$^{13}$C NMR (126 MHz, DMSO-$d_6$) δ/ppm = 142.3 (C-4), 140.8 (C-6), 140.1 (C-2), 130.0 (C-8), 123.4 (C-5), 106.4 (C-3), 101.7 (C-7), 88.0 (C-4'), 85.9 (C-1'), 70.9 (C-3'), 61.9 (C-5'), 40.1 (C-2').

HRMS: Calculated for $[M+H]^+$ 539.1135, found: 539.1134.

**Compound 27b**




Chemical Formula: $C_{24}H_{40}Cl_2N_2O_5Si_2$
Exact Mass: 562,1853
Molecular Weight: 563,6630

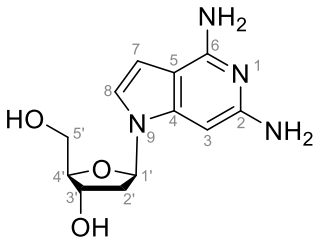
**27** (400 mg, 1.32 mmol, 1.00 eq.) was co-evaporated with pyridine and dissolved in dry pyridine (10.0 mL) under argon atmosphere. Imidazole (270 mg, 3.96 mmol, 3.00 eq.) and 1,3-Dichloro-1,1,3,3-tetraisopropyl disiloxane (0.45 mL, 1.45 mmol, 1.10 eq.) were added and the mixture was stirred at rt for 16 h. After evaporating the solvent, the residue was purified via column chromatography (5 to 10% acetone in PE) to give the product (560 mg, 1.03 mmol, 78%) as pale yellow solid.

$^{1}$H NMR (600 MHz, $CDCl_3$) δ/ppm = 7.41 (d, *J* = 3.5 Hz, 1H, H-8), 7.28 (s, 1H, H-3), 6.63 (d, *J* = 3.4 Hz, 1H, H-7), 6.14 (dd, *J* = 7.1, 3.4 Hz, 1H, H-1'), 4.69 (q, *J* = 7.6 Hz, 1H, H-3'), 4.05 (dd, *J* = 12.5, 3.3 Hz, 1H, H-5'a), 3.95 (dd, *J* = 12.5, 4.8 Hz, 1H, H-5'b), 3.88 (ddd, *J* = 7.0, 4.8, 3.3 Hz, 1H, H-4'), 2.58 (ddd, *J* = 13.3, 8.2, 7.1 Hz, 1H, H-2'a), 2.44 (ddd, *J* = 13.3, 7.5, 3.4 Hz, 1H, H-2'b), 1.12 – 1.02 (m, 28H, $C^{isopropyl}$-H).

$^{13}$C NMR (151 MHz, $CDCl_3$) δ/ppm = 142.5 (C-6), 141.6 (C-4), 141.5 (C-2), 126.9 (C-8), 124.0 (C-5), 104.7 (C-3), 102.5 (C-7), 85.3 (C-4'), 84.2 (C-1'), 69.9 (C-3'), 61.8 (C-5'), 40.3 (C-2'), 17.6 ($C^{isopropyl}$), 17.5 ($C^{isopropyl}$).

HRMS: Calculated for $[M+H]^+$ 563.1926, found: 563.1918.

**2,6-Diamino-DDP nucleoside (Compound 28)**




Chemical Formula: $C_{12}H_{16}N_4O_3$
Exact Mass: 264,1222
Molecular Weight: 264,2850

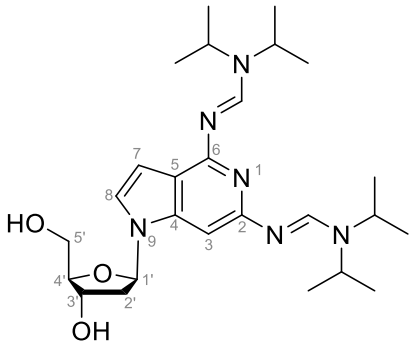
**24** (350 mg, 0.741 mmol, 1.00 eq.) was added to a solution of $Zn(OAc)_2$ dihydrate (163 mg, 0.741 mmol, 1.00 eq.) in methanol (20 mL) and 5 M NaOH (20 mL) and stirred at 80 °C for 20 h. The solvent was removed in vacuo and the crude residue was dissolved in a mixture of water (3 mL) and acetonitrile (3 mL) and purified via HPLC (2 to 50% ACN in $H_2O$ with 0.1% TFA) to give the nucleoside as TFA-salt (273 mg, 0.722 mmol, 97%) as a yellow solid.

$^1$H NMR (500 MHz, $D_2O$) δ/ppm = 7.15 (d, J = 3.6, 1H, H-8), 6.53 (d, J = 3.5 Hz, 1H, H-7), 6.24 (dd, J = 6.9 Hz, 1H, H-1'), 6.22 (s, 1H, H-3), 4.52 (dt, J = 7.2, 3.8 Hz, 1H, H-3'), 4.01 – 3.96 (m, 1H, H-4'), 3.75 (ddd, J = 12.3, 4.2, 1.1 Hz, 1H, H-5'a), 3.66 (ddd, J = 12.2, 5.7, 1.1 Hz, 1H, H-5'b), 2.66 – 2.59 (m, 1H, H-2'a), 2.41 – 2.34 (m, 1H, H-2'b).

$^{13}$C NMR (126 MHz, $D_2O$) δ/ppm = 152.4 (C-6), 151.5 (C-2), 144.2 (C-4), 121.9 (C-8), 105.7 (C-5), 101.2 (C-7), 86.1 (C-4'), 84.4 (C-1'), 80.7 (C-3), 71.1 (C-3'), 61.7 (C-5'), 38.2 (C-2').

HRMS: Calculated for $[M+H]^+$ 265.1295, found: 265.1298.

**Compound 28c**



Chemical Formula: $C_{26}H_{42}N_6O_3$
Exact Mass: 486,3318
Molecular Weight: 486,6610

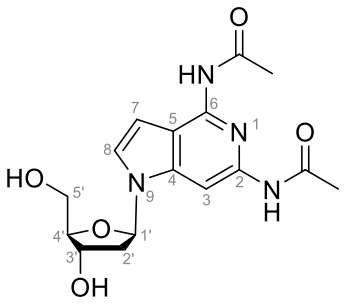
A dried pressure tube was charged with **28** (13 mg, 0.034 mmol, 1.00 eq.) and dry pyridine (1.0 mL) was added. **Dif-dma** (0.067 mL, 0.34 mmol, 10.00 eq.) was added to the suspension and the mixture was stirred at 120 °C for 2 d. The reaction mixture was concentrated under reduced pressure. The crude product was purified via HPLC (5 to 60% ACN in $H_2O$ with 0.1% TFA) to give 8.0 mg (0.017 mmol, 48%) of the product as brown solid.

$^1$H NMR (600 MHz, $D_2O$) δ/ppm = 8.97 (s, 1H, N=C<u>H</u>-N), 8.91 (s, 1H, N=C<u>H</u>-N), 7.63 (d, *J* = 3.7 Hz, 1H, H-8), 7.31 (s, 1H, H-3), 6.93 (d, *J* = 3.6 Hz, 1H, H-7), 6.52 (t, *J* = 6.9 Hz, 1H, H-1'), 4.66 (hept, *J* = 6.7 Hz, 1H, C<u>H</u>$^{iPr}$), 4.61 (dt, *J* = 6.9, 3.6 Hz, 1H, H-3') 4.47 (hept, *J* = 6.0 Hz, 1H, C<u>H</u>$^{iPr}$), 4.27 (hept, *J* = 6.7 Hz, 1H, C<u>H</u>$^{iPr}$), 4.19 (dt, *J* = 13.4, 6.7 Hz, 1H), 4.10 (q, *J* = 4.0 Hz, 1H, H-4'), 3.82 – 3.77 (m, 1H, H-5'a), 3.72 (dd, *J* = 12.4, 5.3 Hz, 1H, H-5'b), 2.76 (m, 1H, H-2'a), 2.52 (m, 1H, H-2'b), 1.47 (m, 24H, C<u>H</u>$_3^{iPr}$).

$^{13}$C NMR (151 MHz, $D_2O$) δ/ppm = 150.6 (N=<u>C</u>H-N), 149.3 (N=<u>C</u>H-N), 144.3 (C-4), 141.7 (C-6), 140.7 (C-2), 127.5 (C-8), 114.1 (C-5), 101.3 (C-7), 93.0 (C-3), 86.4 (C-4'), 84.9 (C-1'), 71.0 (C-3'), 61.5 (C-5'), 52.0 (<u>C</u>H$^{iPr}$), 51.4 (<u>C</u>H$^{iPr}$), 50.3 (<u>C</u>H$^{iPr}$), 50.0 (<u>C</u>H$^{iPr}$), 38.4 (C-2'), 23.2 (<u>C</u>H$_3^{iPr}$), 18.5 (<u>C</u>H$_3^{iPr}$).

HRMS: Calculated for $[M+H]^+$ 487.3391, found: 487.3434.

**Compound 28f**



Chemical Formula: $C_{16}H_{20}N_4O_5$
Exact Mass: 348,1434
Molecular Weight: 348,3590

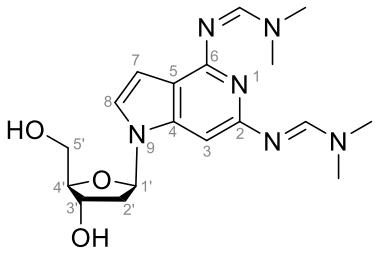
A dried pressure tube was charged with **28** (850 mg, 2.80 mmol, 1.00 eq.), acetamide (16.6 g, 280 mmol, 100 eq.) and tBu-BrettPhos Pd G3 (47.9 mg, 0.056 mmol, 0.02 eq.), before triethylamine (850 mg, 8.41 mmol, 3.00 eq.) was added under argon. The mixture was stirred at 140 °C for 16 h. After cooling down to room temperature, the volatiles were evaporated in vacuo. The crude product was purified via flash chromatography on silica gel (MeOH 2 to 10% in DCM) to give the product (650 mg, 1.87 mmol, 67%) as white solid.

$^1$H NMR (700 MHz, $D_2O$) δ/ppm = 7.45 (s, 1H, H-8), 6.86 (s, 1H, H-3), 6.61 (s, 1H, H-7), 6.13 (dd, *J* = 6.9 Hz, 1H, H-1'), 4.53 (dt, *J* = 6.7, 3.5 Hz, 1H, H-3'), 4.08 (dt, *J* = 5.7, 3.7 Hz, 1H, H-4'), 3.82 (dd, *J* = 12.4, 3.8 Hz, 1H, H-5'a), 3.72 (dd, *J* = 12.4, 5.8 Hz, 1H, H-5'b), 2.55 (ddd, *J* = 14.0, 6.9 Hz, 1H, H-2'a), 2.45 (ddd, *J* = 14.1, 6.3, 3.5 Hz, 1H, H-2'b), 2.25 (s, 3H, $CH_3$), 2.14 (s, 3H, $CH_3$).

$^{13}$C NMR (176 MHz, $D_2O$) δ/ppm = 173.6 (C=O), 172.8 (C=O), 142.7 (C-6), 142.7 (C-4), 139.8 (C-2), 126.7 (C-8), 103.5 (C-3), 103.5 (C-7), 102.1 (C-5), 86.5 (C-4'), 84.7 (C-1'), 70.9 (C-3'), 61.7 (C-5'), 38.7 (C-2'), 23.5 ($CH_3$), 23.3($CH_3$).

HRMS: Calculated for $[M+H]^+$ 349.1506, found: 349.1500.

**Compound 29**



Chemical Formula: $C_{18}H_{26}N_6O_3$
Molecular Weight: 374,4450

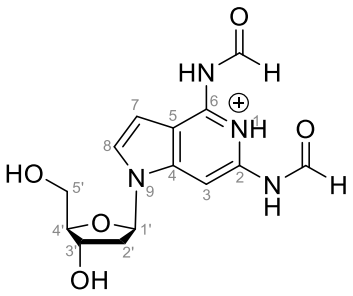
A dried pressure tube was charged with **28** (100 mg, 0.264 mmol, 1.00 eq.) and argon-flushed HPLC-grade methanol (10.0 mL) was added. Dimethylformamide dimethyl acetal (2.64 mmol, 10.0 eq.) was added to the suspension and the mixture was heated at 80 °C for 20 h. The reaction was cooled to room temperature and concentrated under reduced pressure. The crude product was purified via HPLC to give 128 mg (0.262 mmol, 99%) of the product as TFA-salt.

$^1$H NMR (600 MHz, $D_2O$) δ/ppm = 9.12 (s, 1H, N=C<u>H</u>-N), 9.04 (s, 1H, N=C<u>H</u>-N), 7.61 (dd, *J* = 3.8, 1.4 Hz, 1H, H-8), 7.28 – 7.26 (m, 1H, H-3), 6.93 (d, *J* = 3.6 Hz, 1H, H-7), 6.49 (dd, *J* = 7.0, 2.5 Hz, 1H, H-1'), 4.61 – 4.59 (m, 1H, H-3'), 4.11 – 4.08 (m, 1H, H-4'), 3.78 (dd, *J* = 12.4, 3.9 Hz, 1H, H-5'a), 3.71 (dd, *J* = 12.4, 5.3 Hz, 1H, H-5'b), 3.55 (s, 3H, NC<u>H</u>$_3$), 3.50 (s, 3H, NC<u>H</u>$_3$), 3.43 (s, 3H, NC<u>H</u>$_3$), 3.33 (s, 3H, NC<u>H</u>$_3$), 2.77 – 2.72 (m, 1H, H-2'a), 2.54 – 2.49 (m, 1H, H-2'b).

$^{13}$C NMR (151 MHz, $D_2O$) δ/ppm = 153.2 (N=<u>C</u>H-N), 152.3 (N=<u>C</u>H-N), 144.2 (C-4), 141.4 (C-6), 140.3 (C-2), 127.3 (C-8), 113.6 (C-5), 101.3 (C-7), 92.6 (C-3), 86.4 (C-4'), 84.9 (C-1'), 71.0 (C-3'), 61.5 (C-5'), 44.5 (N<u>C</u>H$_3$), 44.0 (N<u>C</u>H$_3$), 38.5 (C-2'), 37.6 (N<u>C</u>H$_3$), 37.2 (N<u>C</u>H$_3$).

MS: Calculated for $[M+H]^+$ 375.2, found: 374.8.

**Compound 31**



Chemical Formula: $C_{14}H_{17}N_4O_5^+$
Exact Mass: 321,1193
Molecular Weight: 321,3125

A dried pressure tube was charged with **27** (500 mg, 1.65 mmol, 1.00 eq.), formamide (13.1 mL, 330 mmol, 200 eq.) and tBu-BrettPhos Pd G3 (28.2 mg, 0.033 mmol, 0.02 eq.), before triethylamine (0.68 mL, 4.95 mmol, 3.00 eq.) was added under argon. The mixture was stirred at 140 °C for 3 h. The crude product was purified via flash chromatography on silica gel (2 to 10% MeOH in DCM) to give 423 mg (1.32 mmol, 80%) of the product as yellow solid.

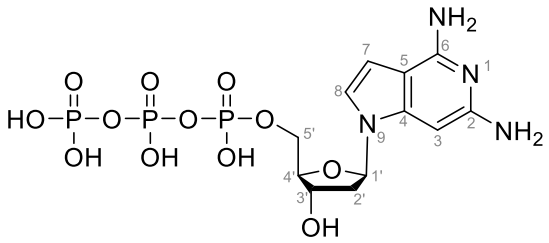
$^1$H NMR (500 MHz, methanol-$d_4$) δ/ppm = 8.43 (s, 1H, $C^{Formyl}$-H), 7.61 (d, *J* = 3.6 Hz, 1H, H-8), 6.91 (d, *J* = 3.5 Hz, 1H, H-7), 6.85 (s, 1H, H-3), 6.30 (dd, *J* = 7.6, 6.0 Hz, 1H, H-1'), 4.51 (dt, *J* = 6.0, 3.1 Hz, 1H, H-3'), 4.01 (q, *J* = 3.9 Hz, 1H, H-4'), 3.79 – 3.68 (m, 2H, H-5'), 2.56 (ddd, *J* = 13.5, 7.5, 6.1 Hz, 1H, H-2'a), 2.41 (ddd, *J* = 13.5, 6.0, 3.2 Hz, 1H, H-2'b).

$^{13}$C NMR (126 MHz, methanol-$d_4$) δ/ppm = 163.3 ($C^{Formyl}$), 142.2 (C-6), 142.2 (C-4), 137.5 (C-2), 127.3 (C-8), 108.4 (C-5), 104.5 (C-7), 89.1 (C-4'), 87.1 (C-3), 87.0 (C-1'), 72.5 (C-3'), 63.2 (C-5'), 41.5 (C-2').

HRMS: The amides are cleaved under LC-MS conditions.

Calculated deprotected mass $[M+H]^+$ 265.1295, found: 265.1308.

**2,6-Diamino-DDP triphosphate (Compound 40)**

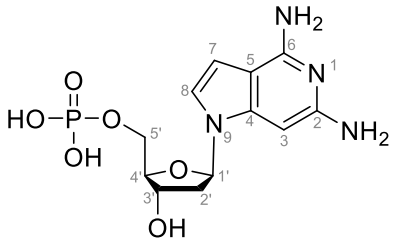


Chemical Formula: $C_{12}H_{19}N_4O_{12}P_3$
Exact Mass: 504,0212
Molecular Weight: 504,2213

A solution of **42** (5.0 µmol), $MgCl_2$ (10 mM), Tris-HCl (pH 7.5, 70 mM) and sodium hexametaphosphate (10 mM) in a total reaction volume of 1000 µL was prepared. PPK05 (0.1 nmol) was added and the mixture was heated to 50 °C for 4 days, before it was lyophilized. The crude mixture was purified via HPLC with TEAA buffer (pH 7.0, 100 mM) to give the triphosphate (ca. 1%). Due to the very small quantity of the substance, no NMR could be obtained and the concentration was estimated based on the UV absorption of the solution to give an isolated yield of 1%.

HRMS: Calculated for $[M-H]^-$: 503.0140, found: 503.0070.

**Compound 42**



Chemical Formula: $C_{12}H_{17}N_4O_6P$
Exact Mass: 344,0886
Molecular Weight: 344,2638

In a dry Schlenk flask, compound 29 (20.0 mg, 0.042 mmol, 1.00 eq.) was dissolved in trimethylphosphate (1.0 mL) under an argon atmosphere. $POCl_3$ (7.7 µL, 0.084 mmol, 2.00 eq.) was added dropwise. The reaction mixture was stirred at 0 °C for 15 min, before a solution of 10% TEA in water (2 mL) was added

and it was stirred for 15 min at rt. The mixture was concentrated under reduced pressure and purified via HPLC to isolate the product (4.0 mg, 0.013 mmol, 20%) as white solid.

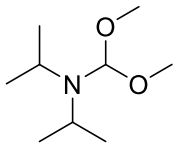
$^{1}$H NMR (400 MHz, $D_2O$) δ/ppm = 7.33 (d, J = 3.3 Hz, 1H, H-8), 6.98 (d, J = 7.6 Hz, 1H, H-3), 6.75 (d, J = 3.4 Hz, 1H, H-7), 6.24 (dd, J = 7.1, 2.1 Hz, 1H, H-1'), 4.73 – 4.70 (m, 1H, H-4'), 4.60 (dt, J = 6.1, 1.8 Hz, 1H, H-3'), 4.13 – 4.02 (m, 2H, H-5'a, H-5'b), 2.94 (dt, J = 14.6, 6.6 Hz, 1H, H-2'a), 2.27 (dt, J = 14.8, 2.1 Hz, 1H, H-2'b).

$^{13}$C NMR (101 MHz, $D_2O$) δ/ppm = 147.6 (C-6), 138.7 (C-4), 126.5 (C-8), 125.5 (C-2), 110.2 (C-5), 102.0 (C-7), 101.0 (C-3), 90.2 (C-1'), 88.5 (C-4'), 71.2 (C-3'), 65.5 (C-5'), 40.1 (C-2').

$^{31}$P NMR (162 MHz, $D_2O$) δ/ppm = 0.07 (s, 1P).

HRMS: Calculated for $[M-H]^-$ 343.0813, found: 343.0760.

**Diisopropylformamide dimethylacetal (dif-dma)**




Chemical Formula: $C_9H_{21}NO_2$
Molecular Weight: 175,2720

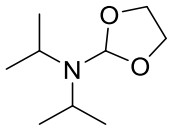
*N*,*N*-Diisopropylformamide (14.0 mL, 96.7 mmol, 1.00 eq.) and dimethyl sulfate (9.18 mL, 96.7 mmol, 1.00 eq.) were stirred at 90 °C for 3 d. The resulting brown reaction mixture was then cooled to room temperature. Diethyl ether (10 mL) was added to the cooled solution while stirring for 10 min. The upper, ethereal layer was removed with a cannula, and an additional portion of diethyl ether (10 mL) was added to the residue. The upper, ethereal layer was removed again. The remaining brown mixture was poured into a rapidly stirring, cold 5.4 M solution of sodium methoxide in methanol (53.8 mL, 290 mmol, 3.00 eq.) and was stirred for 16 h at room temperature, before it was extracted with cyclohexane (5 × 50 mL). The combined extracts were dried over sodium sulfate and concentrated to a yellow liquid. The mixture was distilled (bath 110 °C) to give 9.32 g (53.2 mmol, 55%) of dif-dma as a clear, colorless liquid (bp 54 °C, 5.7 mbar), containing 5% *N*,*N*-diisopropylformamide.

$^{1}$H NMR (500 MHz, $CDCl_3$) δ/ppm = 4.61 (s, 1H, NC<u>H</u>$(OCH_3)_2$), 3.34 (hept, *J* = 6.8 Hz, 2H, N(C<u>H</u>$(CH_3)_2)_2$), 3.25 (s, 6H, $OCH_3$), 1.06 (d, *J* = 6.8 Hz, 12H, N(CH(C<u>H</u>$_3)_2)_2$).

$^{13}$C NMR (126 MHz, $CDCl_3$) δ/ppm = 109.2 (N<u>C</u>H$(OCH_3)_2$), 53.6 ($OCH_3$), 42.8 (N(<u>C</u>H$(CH_3)_2)_2$), 23.1 (N(CH(<u>C</u>H$_3)_2)_2$).

The product is not stable under LC-MS conditions; therefore, no exact mass could be determined.

**Diisopropylformamide ethylene glycol acetal**




Chemical Formula: $C_9H_{19}NO_2$
Molecular Weight: 173,2560

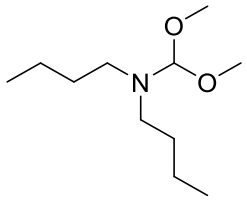
**Dif-dma** (1.00 mL, 5.08 mmol, 1.00 eq.) and ethylene glycol (1.42 mL, 25.4 mmol, 5.00 eq.) were stirred at 50 °C and 50 mbar pressure for 2 hours. The resulting yellow mixture was extracted with petroleum ether (3x10 mL). The combined extracts were concentrated in vacuo to yield the product as a colorless liquid (853 mg, 4.93 mmol, 97%).

$^{1}$H NMR (400 MHz, $CDCl_3$) δ/ppm = 5.67 (s, 1H, NC<u>H</u>$O_2$), 3.88 (dt, *J* = 58.4, 7.0 Hz, 4H, OC<u>H</u>$_2$C<u>H</u>$_2$O), 3.34 (hept, *J* = 6.6 Hz, 2H, N(C<u>H</u>$(CH_3)_2)_2$), 1.09 (d, *J* = 6.7 Hz, 12H, N(CH(C<u>H</u>$_3)_2)_2$).

$^{13}$C NMR (101 MHz, $CDCl_3$) δ/ppm = 109.3 (N<u>C</u>H$O_2$), 63.3 (O<u>C</u>H$_2$<u>C</u>H$_2$O), 43.2 (N(<u>C</u>H$(CH_3)_2)_2$), 23.2 (N(CH(<u>C</u>H$_3)_2)_2$).

The product is not stable under LC-MS conditions; therefore, no exact mass could be determined.

**Dibutylformamide dimethylacetal (dbf-dma)**



Chemical Formula: $C_{11}H_{25}NO_2$
Molecular Weight: 203,3260

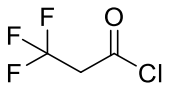
*N*,*N*-Dibutyl formamide (28.9 mL, 159 mmol, 1.00 eq.) and dimethyl sulfate (15.1 mL, 159 mmol, 1.00 eq.) were stirred at 90 °C for 3 d. The resulting brown reaction mixture was then cooled to room temperature. Diethyl ether (20 mL) was added to the cooled solution while stirring for 10 min. The upper, ethereal layer was removed with a cannula, and an additional portion of diethyl ether (20 mL) was added to the residue. The upper, ethereal layer was removed again. The remaining brown mixture was poured into a rapidly stirring, cold 5.4 M solution of sodium methoxide in methanol (88.3 mL, 477 mmol, 3.00 eq.) and was stirred for 16 h at room temperature, before it was extracted with cyclohexane (5 × 100 mL). The combined extracts were dried over sodium sulfate and concentrated to a yellow liquid. The mixture was distilled (bath 130 °C) to give 15.9 g (78.0 mmol, 49%) of dbf-dma as a clear, colorless liquid (bp 62 °C, 1.5 mbar), contaminated with 11% *N*,*N*-Dibutyl formamide.

$^1$H NMR (500 MHz, $CDCl_3$) δ/ppm = 4.52 (s, 1H, $NC\underline{H}(OCH_3)_2$), 3.31 (s, 6H, $OCH_3$), 2.63 – 2.57 (m, 4H, $N(C\underline{H}_2CH_2CH_2CH_3)_2$), 1.45 – 1.38 (m, 4H, $N(CH_2C\underline{H}_2CH_2CH_3)_2$), 1.34 – 1.26 (m, 4H, $N(CH_2CH_2C\underline{H}_2CH_3)_2$), 0.90 (t, *J* = 7.3 Hz, 6H, $N(CH_2CH_2CH_2C\underline{H}_3)_2$).

$^{13}$C NMR (126 MHz, $CDCl_3$) δ/ppm = 112.8 ($N\underline{C}H(OCH_3)_2$), 54.1 ($OCH_3$), 47.3 ($N(\underline{C}H_2CH_2CH_2CH_3)_2$), 31.2 ($N(CH_2\underline{C}H_2CH_2CH_3)_2$), 20.7($N(CH_2CH_2\underline{C}H_2CH_3)_2$), 14.2 ($N(CH_2CH_2CH_2\underline{C}H_3)_2$).

The product is not stable under LC-MS conditions; therefore, no exact mass could be determined.

**3,3,3-Trifluoropropionyl chloride**




Chemical Formula: $C_3H_2ClF_3O$
Molecular Weight: 146,4932

Trifluoro propionic acid (25.0 g, 195 mmol, 1.00 eq.) was slowly added to phosphorus pentachloride (61.0 g, 293 mmol, 1.50 eq.) under cooling with an ice bath. After stirring at 0 °C for 20 min, the mixture was stirred for 20 min at room temperature. Then the mixture was slowly heated to 60 °C and stirred for 2 h. The product was distilled under ambient pressure (b.p. 69-70 °C) to give the product (28.2 g, 193 mmol, 99%) as a colorless liquid.

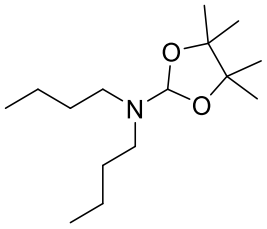
$^{1}$H NMR (600 MHz, $CDCl_3$) δ/ppm = δ 3.73 (q, *J* = 9.2 Hz, 2H).

$^{13}$C NMR (151 MHz, $CDCl_3$) δ/ppm = 164.55 (q, *J* = 4.2 Hz, COCl), 121.82 (q, *J* = 277.4 Hz, $CF_3$), 49.82 (q, *J* = 31.6 Hz, $CH_2$).

$^{19}$F NMR (565 MHz, $CDCl_3$) δ/ppm = -64.47 (t, *J* = 9.3 Hz, 3F).

The product is not stable under LC-MS conditions; therefore, no exact mass could be determined.

**Dibutylformamide pinacol acetal**




Chemical Formula: $C_{15}H_{31}NO_2$
Molecular Weight: 257,4180

**Dbf-dma** (1.00 mL, 4.38 mmol, 1.00 eq.) and pinacol (2.59 g, 21.9 mmol, 5.00 eq.) were stirred at 60 °C and 70 mbar pressure for 2 hours. The resulting mixture was mixed with petroleum ether (20 mL) and

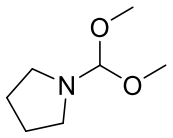
filtered. The filtrate was concentrated in vacuo to yield the product as a colorless liquid (1.08 g, 4.20 mmol, 96%).

$^{1}$H NMR (700 MHz, $CDCl_3$) δ/ppm = δ 5.51 (d, *J* = 1.1 Hz, 1H, NC<u>H</u>$O_2$R), 3.21 (t, *J* = 7.7 Hz, 1H), 2.62 (t, *J* = 8.4, 7.7 Hz, 4H, N(C<u>H</u>$_2$CH$_2$CH$_2$CH$_3$)$_2$), 1.38 (tt, *J* = 7.9, 6.3 Hz, 4H, N(CH$_2$C<u>H</u>$_2$CH$_2$CH$_3$)$_2$), 1.28 – 1.19 (m, 4H, N(CH$_2$CH$_2$C<u>H</u>$_2$CH$_3$)$_2$), 1.15 (s, 6H, (OC(C<u>H</u>$_3$)$_2$)$_2$), 1.13 (s, 6H, (OC(C<u>H</u>$_3$)$_2$)$_2$), 0.85 – 0.81 (m, 6H, N(CH$_2$CH$_2$CH$_2$C<u>H</u>$_3$)$_2$).

$^{13}$C NMR (176 MHz, $CDCl_3$) δ/ppm = δ 109.6 (N<u>C</u>$HO_2$R), 47.0 (N(<u>C</u>H$_2$CH$_2$CH$_2$CH$_3$)$_2$), 31.1 (N(CH$_2$<u>C</u>H$_2$CH$_2$CH$_3$)$_2$), 24.5 (OC(<u>C</u>H$_3$)$_2$)$_2$), 22.9 (OC(<u>C</u>H$_3$)$_2$)$_2$), 20.5 (N(CH$_2$CH$_2$<u>C</u>H$_2$CH$_3$)$_2$), 14.0 (N(CH$_2$CH$_2$CH$_2$<u>C</u>H$_3$)$_2$).

The product is not stable under LC-MS conditions; therefore, no exact mass could be determined.

**_N_-Formyl-pyrrolidine-dimethyl acetal**




Chemical Formula: $C_7H_{15}NO_2$
Molecular Weight: 145,2020

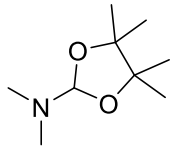
*N*-Formyl-pyrrolidine (12.0 mL, 126 mmol, 1.00 eq.) and dimethyl sulfate (14.4 mL, 151 mmol, 1.20 eq.) were stirred at 90 °C for 3 d. The resulting brown reaction mixture was then cooled to room temperature and slowly dropped into a rapidly stirring, cold 5.4 M solution of sodium methoxide in methanol (70.1 mL, 378 mmol, 3.00 eq.) and was stirred for 16 h at room temperature, before it was extracted with cyclohexane (5 × 100 mL). The combined extracts were dried over sodium sulfate and concentrated to a yellow liquid. The mixture was distilled under reduced pressure to give the acetal (4.50 g, 63 mmol, 50%) as a clear, colorless liquid (bp = 32 °C at 8 mbar).

$^{1}$H NMR (700 MHz, $CDCl_3$) δ/ppm = 4.36 (s, 1H, NC<u>H</u>(OCH$_3$)$_2$), 3.24 (s, 6H, NCH(OC<u>H</u>$_3$)$_2$), 2.60 (td, *J* = 6.9, 2.5 Hz, 4H, NC<u>H</u>$_2$CH$_2$), 1.68 (td, *J* = 6.1, 5.1, 2.7 Hz, 4H, NCH$_2$C<u>H</u>$_2$).

$^{13}$C NMR (176 MHz, $CDCl_3$) δ/ppm = 111.8 (N<u>C</u>H(OCH$_3$)$_2$), 51.5 (NCH(O<u>C</u>H$_3$)$_2$), 47.8 (N<u>C</u>H$_2$CH$_2$), 23.9 (NCH$_2$<u>C</u>H$_2$).

The product is not stable under LC-MS conditions; therefore, no exact mass could be determined.

**Dimethylformamide pinacol acetal**



Chemical Formula: $C_9H_{19}NO_2$
Exact Mass: 173,1416

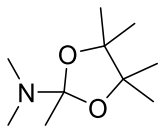
*N*,*N*-Dimethylformamide dimethyl acetal (5.30 g, 44.5 mmol, 2.00 eq.) and pinacol (2.63 g, 22.2 mmol, 1.00 eq.) were stirred at rt for 16 h. Afterwards, the mixture was stirred at 60 °C and 200 mbar pressure for 2 hours. The resulting mixture was concentrated in vacuo to yield the product (3.79 g, 21.9 mmol, 98%) as a colorless liquid.

$^1$H NMR (500 MHz, DMSO-$d_6$) δ/ppm = 5.29 (s, 1H NC<u>H</u>$(OR)_2$), 2.19 (s, 6H, N($C\underline{H}_3)_2$), 1.12 (d, *J* = 2.8 Hz, 12H, OC($C\underline{H}_3)_2)_2$).

$^{13}$C NMR (126 MHz, DMSO) δ/ppm = 109.0 (N<u>C</u>H$(OR)_2$), 80.4 <u>C</u>$(CH_3)_2$), 37.0 (N($\underline{C}H_3)_2$), 24.3(C($\underline{C}H_3)_2$), 22.8(C($\underline{C}H_3)_2$).

The product is not stable under LC-MS conditions; therefore, no exact mass could be determined.

**Dimethylacetamide pinacol acetal**



Chemical Formula: $C_{10}H_{21}NO_2$
Exact Mass: 187,1572

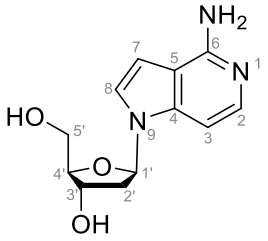
*N*,*N*-Dimethylacetamide dimethyl acetal (890 mg, 6.68 mmol, 1.00 eq.) and pinacol (870 mg, 7.35 mmol, 1.10 eq.) were stirred at 60 °C and 200 mbar pressure for 2 hours. 1.19 g (6.35 mmol, 95%) of the product were obtained.

$^{1}$H NMR (500 MHz, $CDCl_3$) δ/ppm = 2.19 (s, 6H, N($C\underline{H}_3$)$_2$), 1.26 (s, 6H, C($C\underline{H}_3$)$_2$), 1.24 (s, 3H, NC$C\underline{H}_3$), 1.15 (s, 6H, C($C\underline{H}_3$)$_2$).

$^{13}$C NMR (126 MHz, $CDCl_3$) δ/ppm = 114.1 (N$\underline{C}CH_3$(OR)$_2$), 83.4 $\underline{C}$(CH$_3$)$_2$, 38.6 (N($\underline{C}H_3$)$_2$), 25.2 (C($\underline{C}H_3$)$_2$), 24.6 (C($\underline{C}H_3$)$_2$), 17.3 (NC$\underline{C}H_3$(OR)$_2$).

The product is not stable under LC-MS conditions. Therefore, no exact mass could be determined.

**6-Amino-3,7-dideazapurine nucleoside (43)**




Chemical Formula: $C_{12}H_{15}N_3O_3$
Molecular Weight: 249,2700

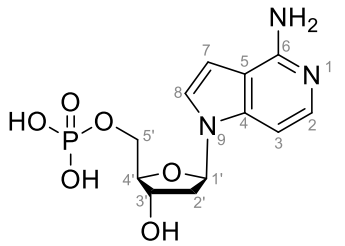
DIPEA (2.11 mL, 12.4 mmol, 5.00 eq.) was stirred in dry THF (50 mL). 1H-Pyrrolo[3,2-c]pyridine-4-amine (330 mg, 2.48 mmol, 1.00 eq.) was added to the stirring solution, before 1-Chloro-2-deoxy-3,5-di-*O*-toluoyl-ribose (1.06 g, 2.73 mmol, 1.10 eq.) was added 2 minutes later. The resulting brown mixture was stirred for 90 min at room temperature, before a 0.1 M solution of sodium methoxide in methanol (50 mL) was added and the mixture was stirred for 16 h at rt. The mixture was filtered, washed with methanol, and the filtrate was concentrated in vacuo. The crude product was purified via HPLC (5 to 25% ACN in water + 0.1% TFA) to give 530 mg (2.12 mmol, 85%) of the product as yellow oil.

$^{1}$H NMR (500 MHz, $D_2O$) δ/ppm = δ 7.34 (d, *J* = 7.5 Hz, 1H, H-2), 6.91 (d, *J* = 3.4 Hz, 1H, H-8), 6.50 (d, *J* = 7.6 Hz, 1H, H-3), 6.26 (d, *J* = 3.3 Hz, 1H, H-7), 5.84 (dd, *J* = 7.0, 2.4 Hz, 1H, H-1'), 4.39 (td, *J* = 4.7, 3.8, 2.4 Hz, 1H, H-4'), 4.35 (dt, *J* = 6.2, 2.3 Hz, 1H, H-3'), 3.66 (dd, *J* = 12.3, 3.8 Hz, 1H, H-5'a), 3.57 (dd, *J* = 12.3, 5.3 Hz, 1H, H-5'b), 2.69 (dt, *J* = 14.8, 6.6 Hz, 1H, H-2'a), 2.01 (dt, *J* = 14.7, 2.4 Hz, 1H, H-2'b).

$^{13}$C NMR (126 MHz, $D_2O$) δ/ppm = 146.7 (C-6), 137.8 (C-4), 126.0 (C-8), 124.8 (C-2), 109.5 (C-5), 101.5 (C-7), 100.4 (C-3), 89.7 (C-4'), 89.6 (C-1'), 70.6 (C-3'), 61.3 (C-5'), 39.7 (C-2').

HRMS: Calculated for $[M+H]^+$ 250.1186, found: 250.1181.

**6-Amino-3,7-dideazapurine nucleoside monophosphate (44)**




Chemical Formula: $C_{12}H_{16}N_3O_6P$
Exact Mass: 329,0777
Molecular Weight: 329,2488

In a dry Schlenk flask, compound 43 (10 mg, 0.040 mmol, 1.00 eq.) was dried under high vacuum, before trimethyl phosphate (2.0 mL) was added under an argon atmosphere. The mixture was cooled in an ice bath and $POCl_3$ (36.6 µL, 0.401 mmol, 10.0 eq.) was added dropwise. The reaction mixture was stirred at 0 °C for 60 min, before a solution of 10% TEA in water (3 mL) was added dropwise and the mixture was stirred for another 15 min at rt. The mixture was concentrated under reduced pressure and was purified via RP-HPLC, with 5 to 20% ACN in water containing 0.1% TFA to give the nucleoside monophosphate (9.0 mg, 27.3 µmol, 68%) as yellow oil.

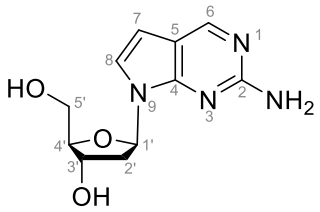
$^1$H NMR (400 MHz, $D_2O$) δ/ppm = 7.76 (d, J = 7.3 Hz, 1H, H-2), 7.33 (d, J = 3.3 Hz, 1H, H-8), 6.98 (d, J = 7.6 Hz, 1H, H-3), 6.75 (d, J = 3.4 Hz, 1H, H-7), 6.24 (dd, J = 7.1, 2.1 Hz, 1H, H-1'), 4.73 – 4.70 (m, 1H, H-4'), 4.60 (dt, J = 6.1, 1.8 Hz, 1H, H-3'), 4.13 – 4.02 (m, 2H, H-5'a, H-5'b), 2.94 (dt, J = 14.6, 6.6 Hz, 1H, H-2'a), 2.27 (dt, J = 14.8, 2.1 Hz, 1H, H-2'b).

$^{13}$C NMR (101 MHz, $D_2O$) δ/ppm = 147.6 (C-6), 138.7 (C-4), 126.5 (C-8), 125.5 (C-2), 110.2 (C-5), 102.0 (C-7), 101.0 (C-3), 90.2 (C-1'), 88.5 (C-4'), 71.2 (C-3'), 65.5 (C-5'), 40.1 (C-2').

$^{31}$P NMR (162 MHz, $D_2O$) δ/ppm = 0.07 (s, 1P).

HRMS: Calculated for $[M-H]^-$ 328.0704, found: 328.0649.

**2-Amino-7-deazapurine nucleoside (45)**



Chemical Formula: $C_{11}H_{14}N_4O_3$
Exact Mass: 250,1066
Molecular Weight: 250,2580

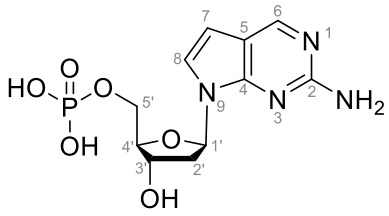
DIPEA (1.90 mL, 11.2 mmol, 5.00 eq.) was stirred in dry THF (50 mL). 7H-Pyrrolo[2,3-d]pyrimidine-2-amine (300 mg, 2.24 mmol, 1.00 eq.) was added, before 1-Chloro-2-deoxy-3,5-di-O-toluoyl-ribose (957 mg, 2.46 mmol, 1.10 eq.) was added 2 minutes later. The resulting brown mixture was stirred for 2 at room temperature, before sodium methoxide (0.1 M solution in methanol, 100 mL) was added and the mixture was stirred for 20 h at rt. The mixture was filtered, washed with methanol, and the filtrate was concentrated in vacuo. The crude product was purified via HPLC (5 to 25% acetonitrile in water with 0.1% TFA) to give the product (254 mg, 1.01 mmol, 45%) as white solid.

$^1$H NMR (700 MHz, Methanol-$d_4$) δ/ppm = 8.52 (s, 1H, H-6), 7.78 (d, *J* = 4.0 Hz, 1H, H-8), 6.67 (d, *J* = 4.0 Hz, 1H, H-7), 6.55 (dd, *J* = 7.8, 3.6 Hz, 1H, H-1'), 4.47 (dt, *J* = 6.9, 3.4 Hz, 1H, H-3'), 4.28 (dt, *J* = 4.8, 3.6 Hz, 1H, H-4'), 3.69 (dd, *J* = 12.2, 3.8 Hz, 1H, H-5'a), 3.63 (dd, *J* = 12.2, 4.8 Hz, 1H, H-5'b), 2.85 (ddd, *J* = 14.6, 7.8, 6.9 Hz, 1H, H-2'a), 2.34 (dt, *J* = 14.6, 3.5 Hz, 1H, H-2'b).

$^{13}$C NMR (176 MHz, Methanol-$d_4$) δ/ppm = 155.9 (C-4), 153.5 (C-2), 139.4 (C-6), 131.6 (C-8), 113.9 (C-5), 103.7 (C-7), 89.1 (C-4'), 85.0 (C-1'), 72.1 (C-3'), 62.7 (C-5'), 40.5 (C-2').

HRMS: Calculated for $[M+H]^+$ 251.1139, found: 251.1140.

**2-Amino-7-deazapurine nucleoside monophosphate (46)**



Chemical Formula: $C_{11}H_{15}N_4O_6P$
Exact Mass: 330,0729
Molecular Weight: 330,2368

In a dry Schlenk flask, compound 45 (50.0 mg, 200 µmol, 1.00 eq.) was dried under high vacuum, before trimethyl phosphate (1.0 mL) was added under an argon atmosphere. The mixture was cooled in an ice bath and $POCl_3$ (0.018 mL, 0.200 mmol, 1.00 eq.) was added dropwise. The reaction mixture was stirred at 0 °C for 30 min, before it was added to a solution of 10% TEA in water (10 mL) and stirred for 15 min at rt. The mixture was concentrated under reduced pressure and was purified via RP-HPLC, with 5 to 20% ACN in water containing 0.1% TFA to give the nucleoside monophosphate (20.0 mg, 60.6 µmol, 30%) as white solid.

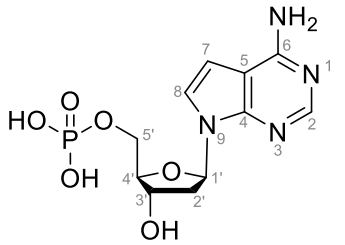
$^{1}$H NMR (600 MHz, $D_2O$) δ/ppm = 8.48 (s, 1H, H-6), 7.62 (d, *J* = 4.1 Hz, 1H, H-8), 6.69 (d, *J* = 4.0 Hz, 1H, H-7), 6.61 (dd, *J* = 7.1 Hz, 1H, H-1'), 4.69 (dt, *J* = 6.0, 2.9 Hz, 1H, H-3'), 4.23 (p, *J* = 3.4 Hz, 1H, H-4'), 4.09 – 4.07 (m, 2H, H-5'a, H-5'b), 2.70 (ddd, *J* = 14.0, 7.6, 6.3 Hz, 1H, H-2'a), 2.45 (ddd, *J* = 14.1, 6.3, 3.1 Hz, 1H, H-2'b).

$^{13}$C NMR (151 MHz, $D_2O$) δ/ppm = 155.1 (C-4), 152.2 (C-2), 138.6 (C-6), 129.3 (C-8), 112.8 (C-5), 103.5 (C-7), 85.4 (C-4'), 83.3 (C-1'), 71.5 (C-3'), 65.1 (C-5'), 38.4 (C-2').

$^{31}$P NMR (283 MHz, $D_2O$) δ/ppm = 0.07 (s).

HRMS: Calculated for $[M-H]^-$ 329.0656, found: 329.0601.

**6-Amino-7-deazapurine nucleoside monophosphate (47)**



Chemical Formula: $C_{11}H_{15}N_4O_6P$
Exact Mass: 330,0729
Molecular Weight: 330,2368

In a dry Schlenk flask, 7-Deaza 6-Aminopurine nucleoside (100 mg, 0.400 mmol, 1.00 eq.) was dried under high vacuum, before trimethyl phosphate (3.0 mL) was added under an argon atmosphere. The mixture was cooled in an ice bath and $POCl_3$ (0.182 mL, 2.00 mmol, 5.00 eq.) was added dropwise. The reaction mixture was stirred at 0 °C for 60 min, before a solution of 10% TEA in water (3 mL) was added dropwise and the mixture was stirred for another 15 min at rt. The mixture was concentrated under reduced pressure and was purified via RP-HPLC, with 5 to 20% ACN in water containing 0.1% TFA to give the nucleoside monophosphate (68.0 mg, 0.206 mmol, 52%) as white solid.

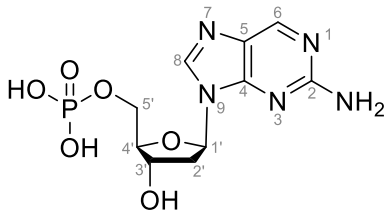
$^1$H NMR (700 MHz, $D_2O$) δ/ppm = 8.22 (s, 1H, H-2), 7.67 (d, *J* = 3.9 Hz, 1H, H-8), 6.82 (d, *J* = 3.8 Hz, 1H, H-7), 6.68 (dd, *J* = 7.6, 6.4 Hz, 1H, H-1'), 4.71 (dt, *J* = 6.1, 3.1 Hz, 1H, H-3'), 4.26 (qd, *J* = 3.3, 1.8 Hz, 1H, H-4'), 4.09 (ddd, *J* = 5.4, 3.5, 1.8 Hz, 2H, H-5'a, H-5'b), 2.70 (ddd, *J* = 13.9, 7.7, 6.1 Hz, 1H, H-2'a), 2.51 (ddd, *J* = 14.1, 6.4, 3.3 Hz, 1H, H-2'b).

$^{13}$C NMR (176 MHz, $D_2O$) δ/ppm = 150.6 (C-6), 146.7 (C-4), 142.0 (C-2), 124.8 (C-8), 102.5 (C-7), 102.2 (C-5), 85.6 (C-4'), 83.6 (C-1'), 71.4 (C-3'), 64.9 (C-5'), 39.2 (C-2').

$^{31}$P NMR (283 MHz, $D_2O$) δ/ppm = 0.17 (s, 1P).

HRMS: Calculated for $[M-H]^-$ 329.0656, found: 329.0599.

**2-Aminopurine nucleoside monophosphate (48)**



Chemical Formula: $C_{10}H_{14}N_5O_6P$
Exact Mass: 331,0682
Molecular Weight: 331,2248

In a dry Schlenk flask, 2-Aminopurine nucleoside (10.0 mg, 40.0 µmol, 1.20 eq.) was dried under high vacuum, before trimethyl phosphate (1.0 mL) was added under an argon atmosphere. The mixture was cooled in an ice bath and $POCl_3$ (3.0 µL, 33 µmol, 1.00 eq.) was added. The reaction mixture was stirred at 0 °C for 30 min, before it was added to a solution of 10% TEA in water (10 mL) and stirred for 15 min at rt. The mixture was concentrated under reduced pressure and was purified via RP-HPLC, with 5 to 20% ACN in water, and again by ion-pairing-HPLC with 100 mM TEAA-buffer and 3 to 20% ACN to give the nucleoside monophosphate (0.5 mg, 1.5 µmol, 5%) as colorless solid.

$^1$H NMR (700 MHz, $D_2O$) δ/ppm = 8.66 (s, 1H, H-6), 8.40 (s, 1H, H-8), 6.48 (dd, *J* = 6.9 Hz, 1H, H-1'), 4.74 (dt, *J* = 6.1, 3.2 Hz, 1H, H-3'), 4.27 (d, *J* = 3.8 Hz, 1H, H-4'), 4.08 – 4.04 (m, 2H, H-5'a, H-5'b), 2.87 (dt, *J* = 13.9, 6.7 Hz, 1H, H-2'a), 2.55 (ddd, *J* = 14.0, 6.3, 3.4 Hz, 1H, H-2'b).

$^{13}$C NMR (176 MHz, $D_2O$) δ/ppm = 159.8 (C-2), 152.8 (C-4), 149.0 (C-6), 142.5 (C-8), 127.2 (C-5), 85.9 (C-4'), 83.4 (C-1'), 71.4 (C-3'), 64.7 (C-5'), 38.6 (C-2').

$^{31}$P NMR (283 MHz, $D_2O$) δ/ppm = 0.55.

HRMS: Calculated for $[M-H]^-$, 330.0609$^-$, found: 330.0550$^-$.

## 9.7 NMR- and MS-spectra of key compounds

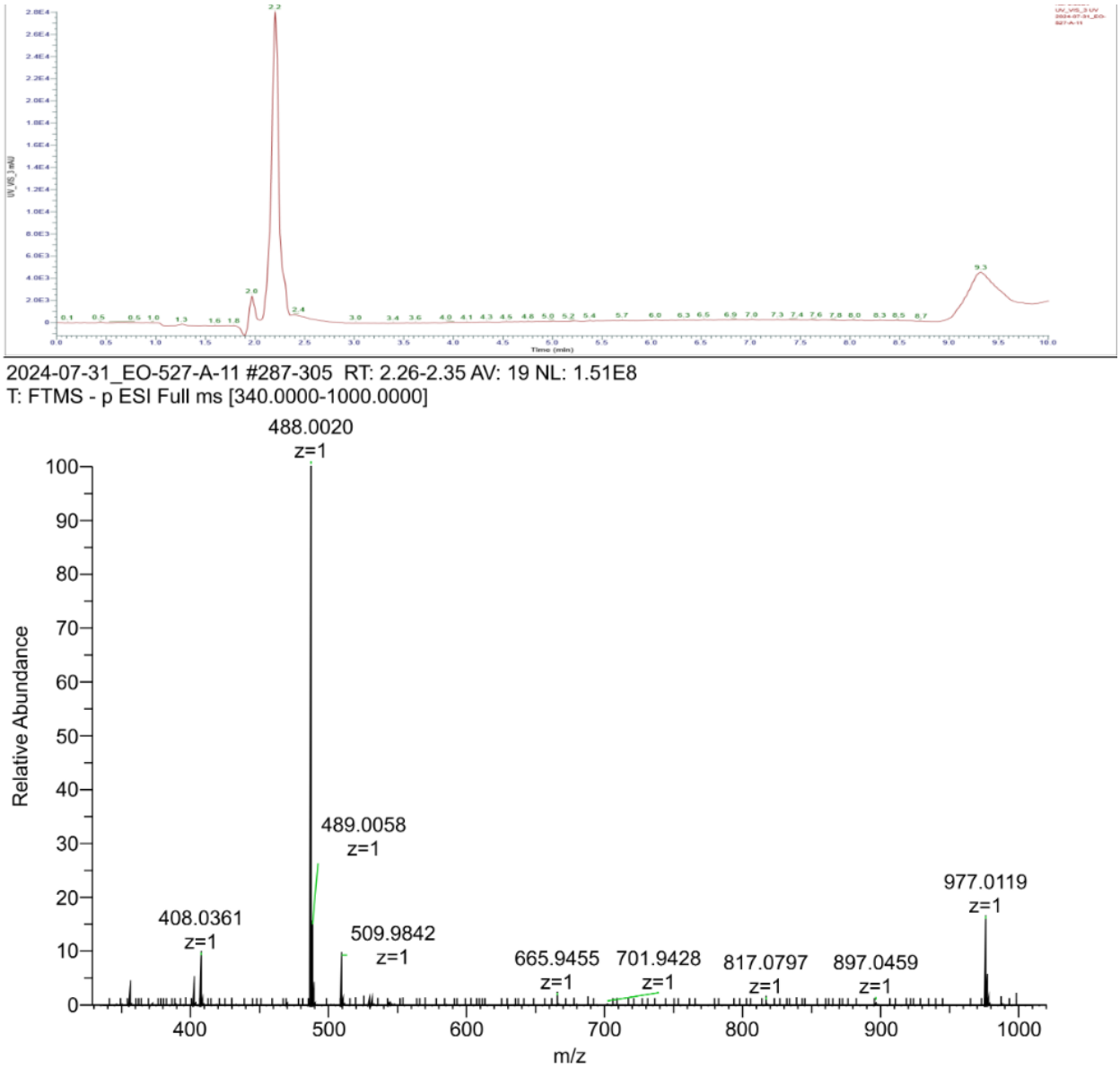



*Figure 91: HRMS-Spectrum of triphosphate 13.*

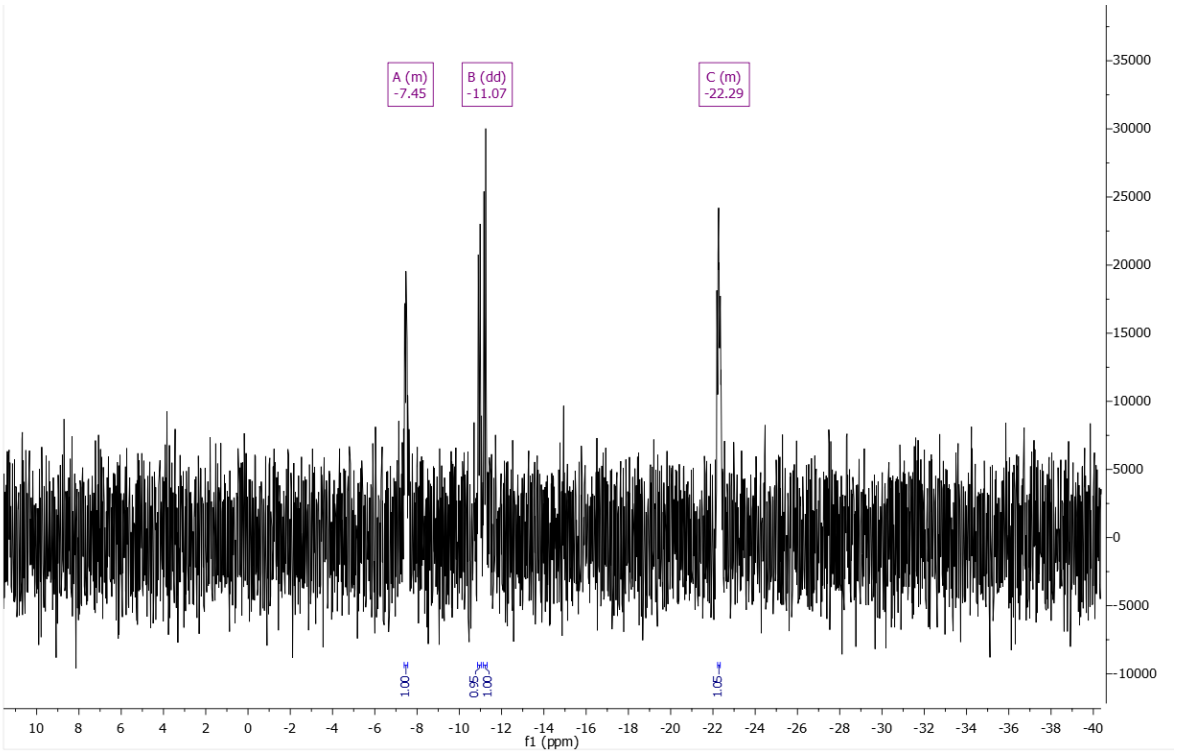



*Figure 92: $^{31}$P-NMR-Spectrum of compound 13.*

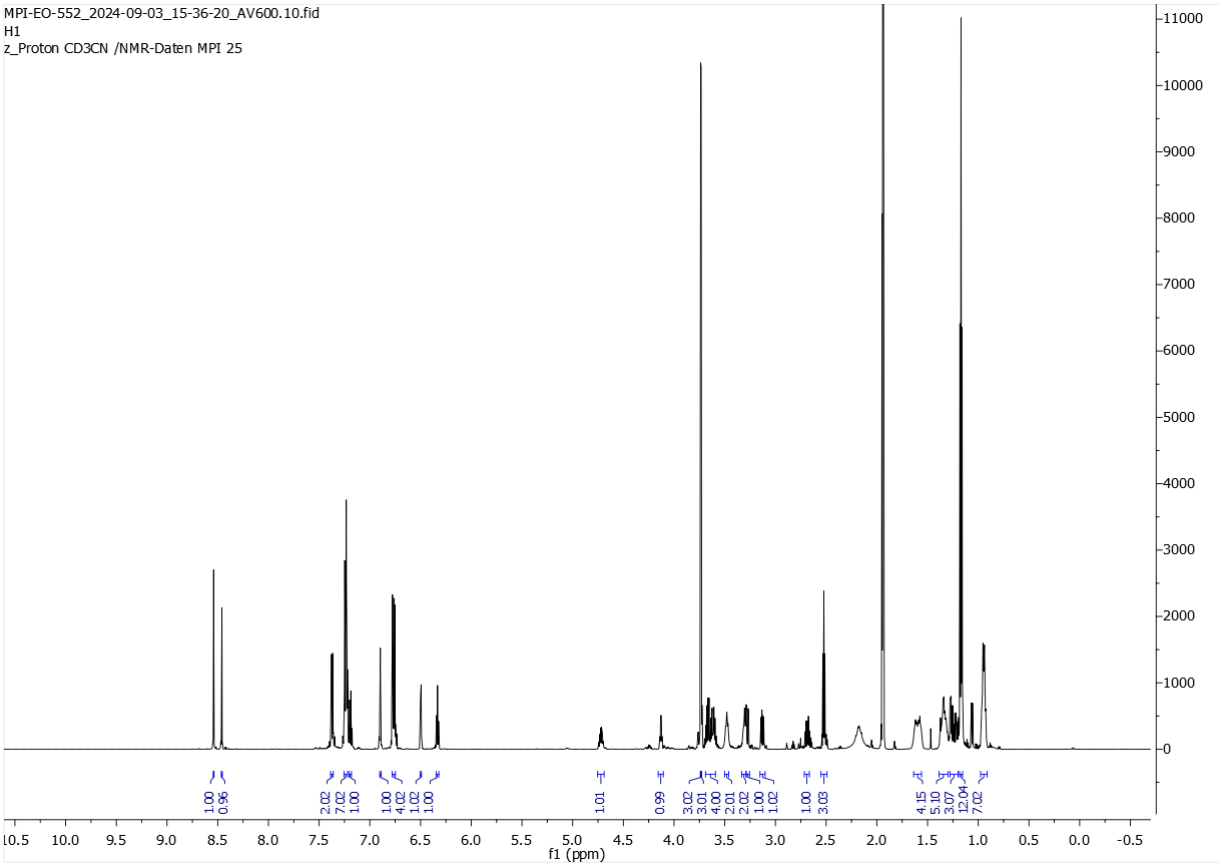



*Figure 93: $^{1}$H NMR-Spectrum of compound 2-Amino-DDP-Phosphoramidite 20.*

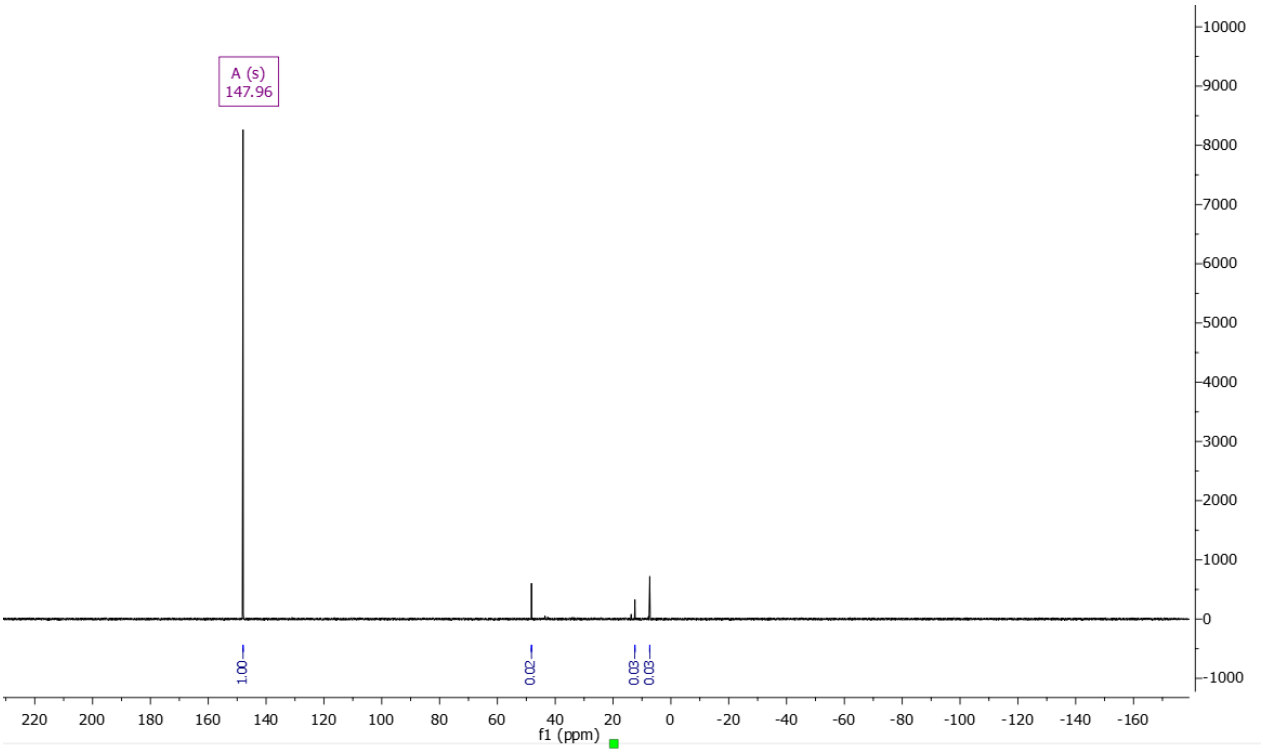



*Figure 94: $^{31}$P NMR-Spectrum of 2-Amino-DDP-Phosphoramidite 20.*

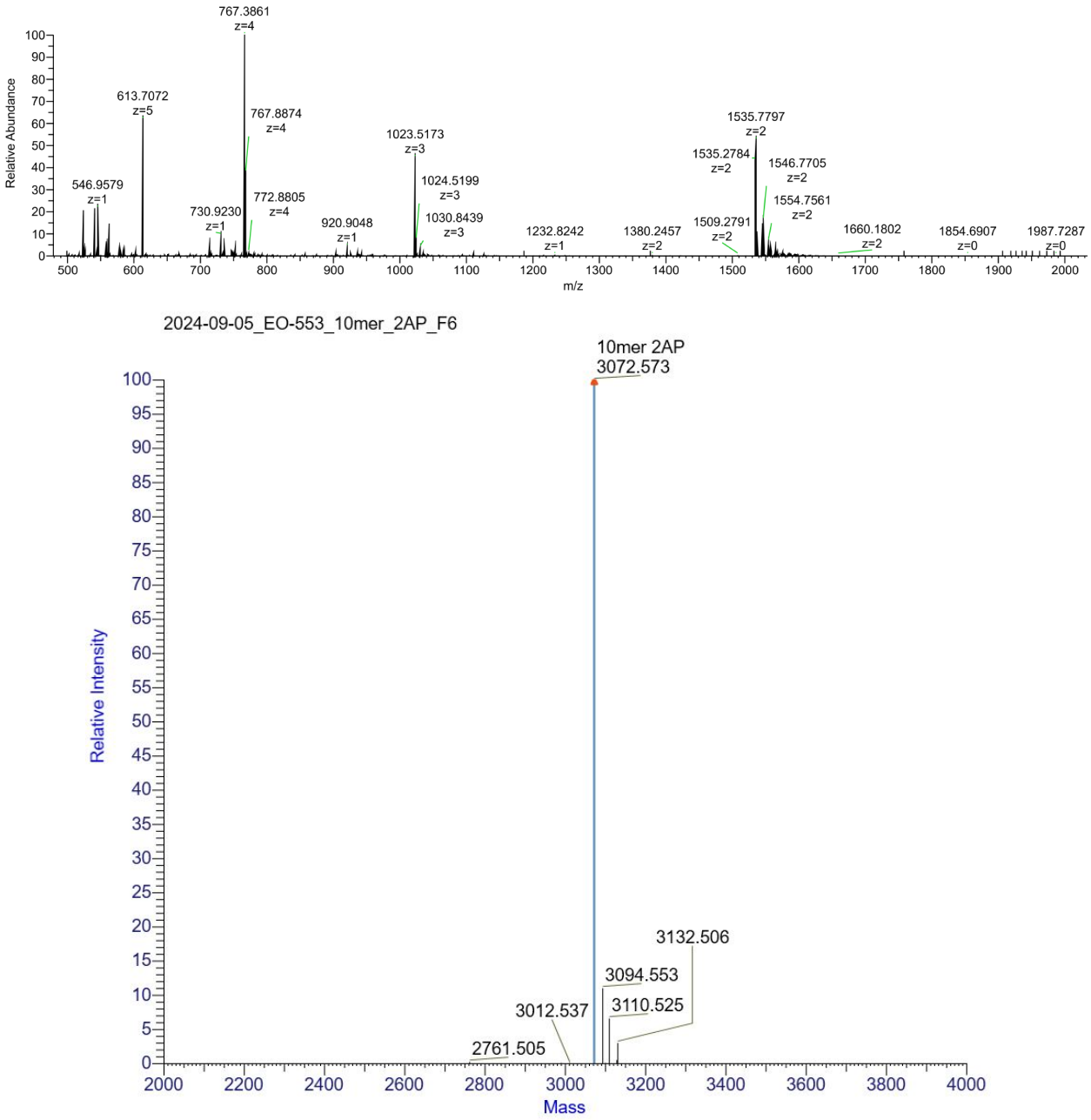



*Figure 95: HRMS-Spectrum (top) and deconvoluted mass (bottom) of DNA-strand TCAGXGTAAG (with X = 2-Amino-DDP). The calculated mass is 3072.580.*

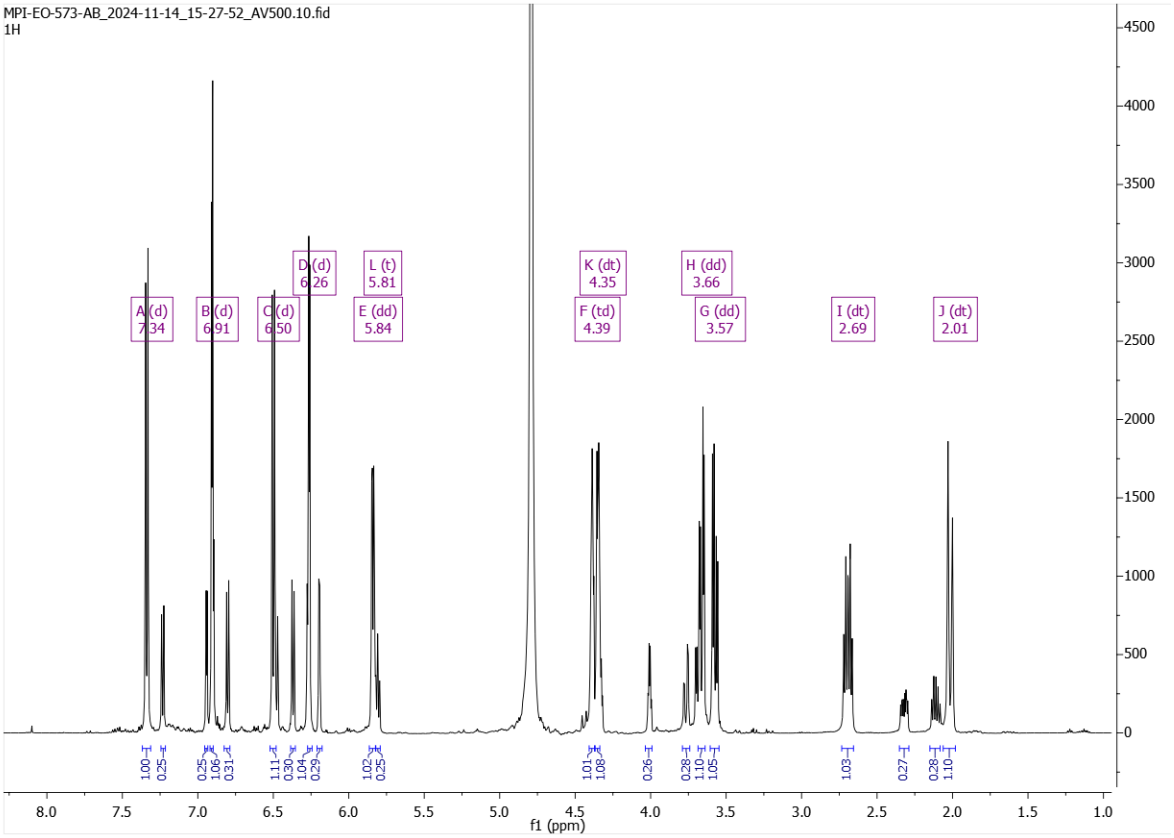



*Figure 96: [1]H NMR spectrum of compound 43.*

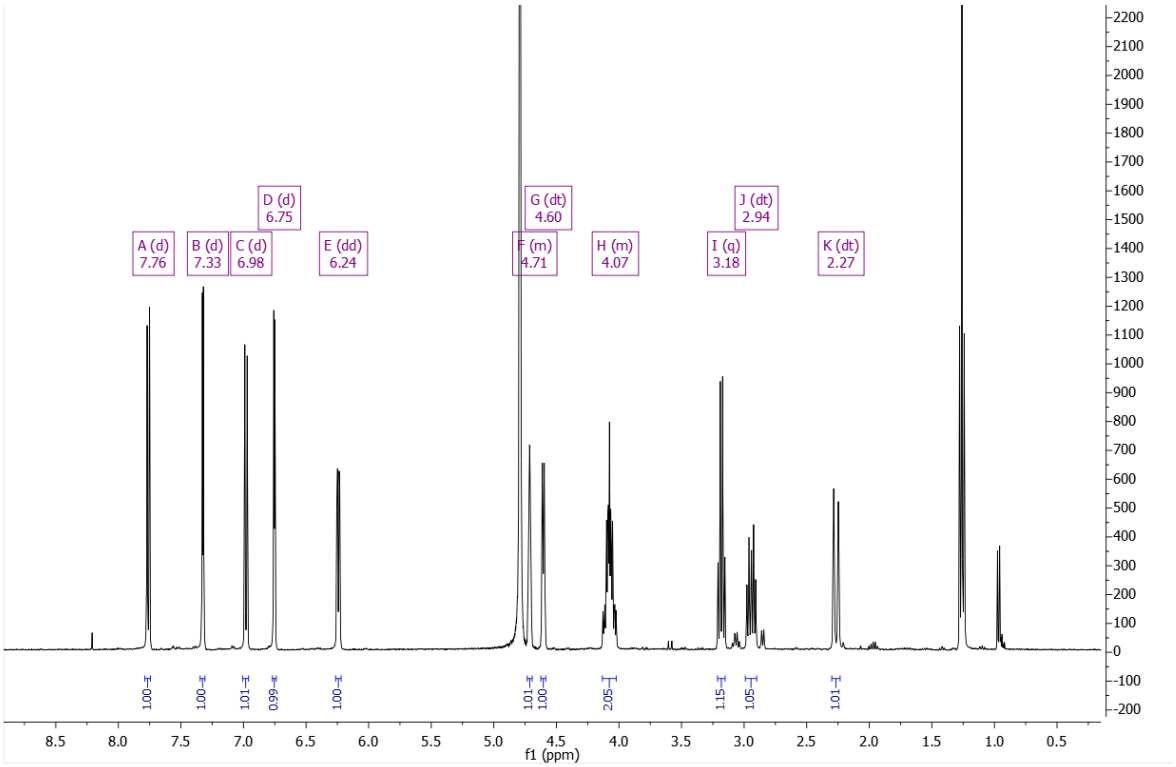



*Figure 97: [1]H NMR spectrum of compound 44.*

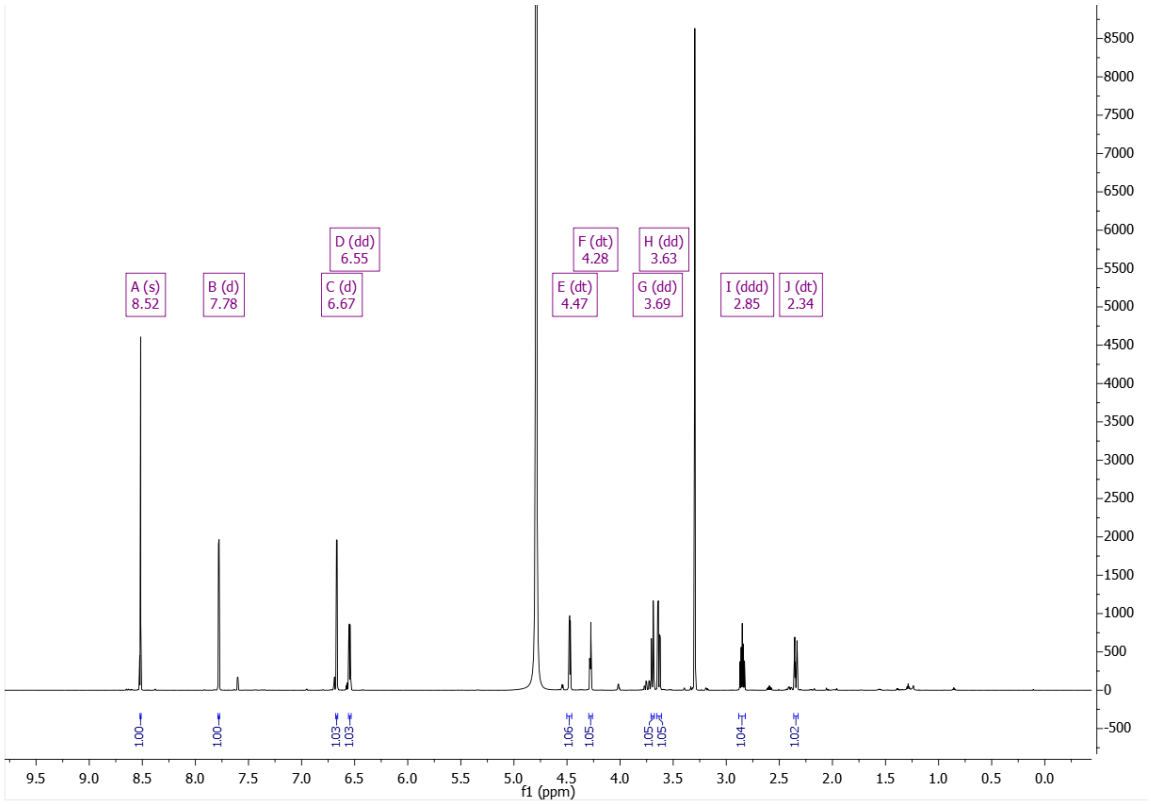



*Figure 98: [1]H NMR spectrum of compound 45.*

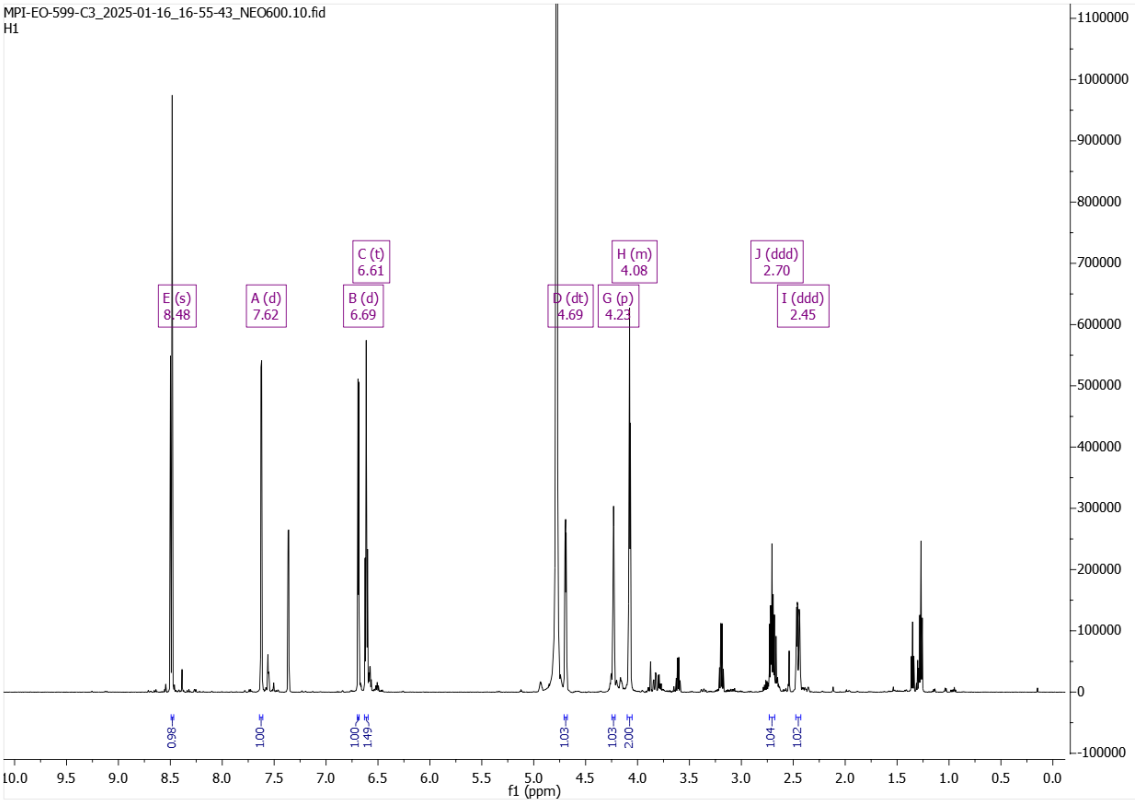



*Figure 99: [1]H NMR spectrum of compound 46.*

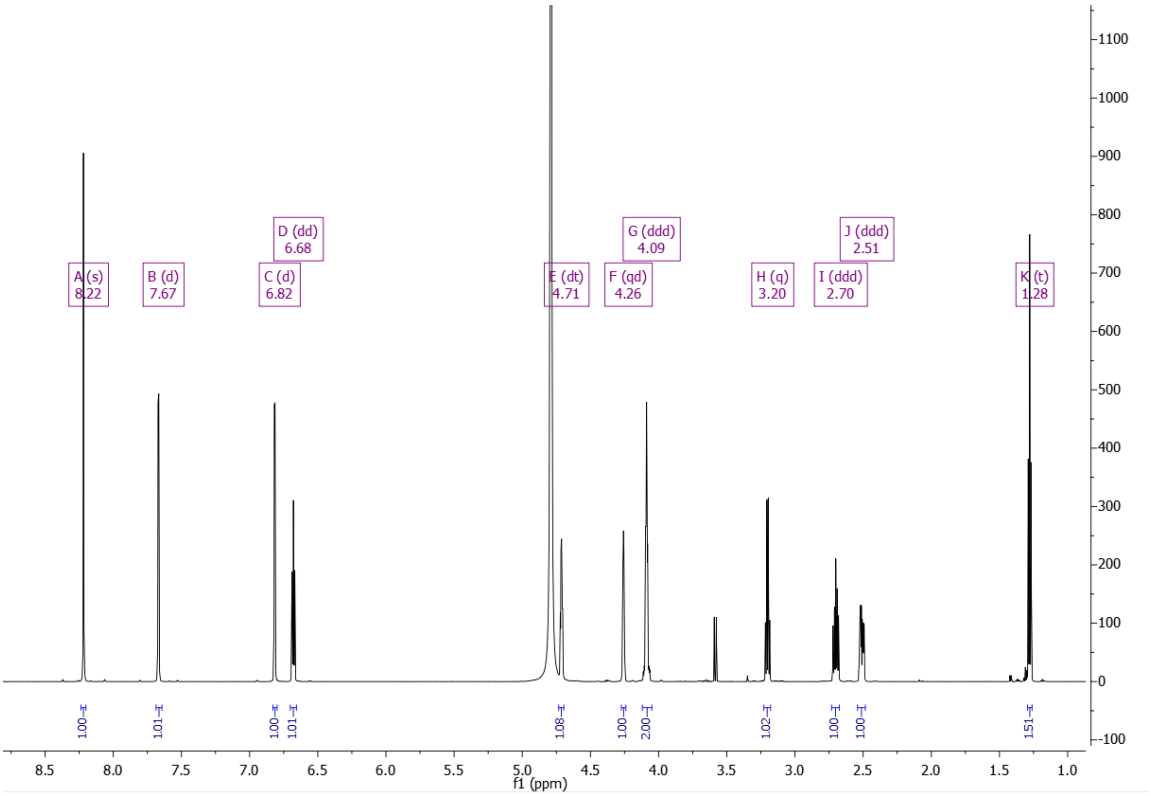



*Figure 100: $^{1}$H NMR spectrum of compound 47.*

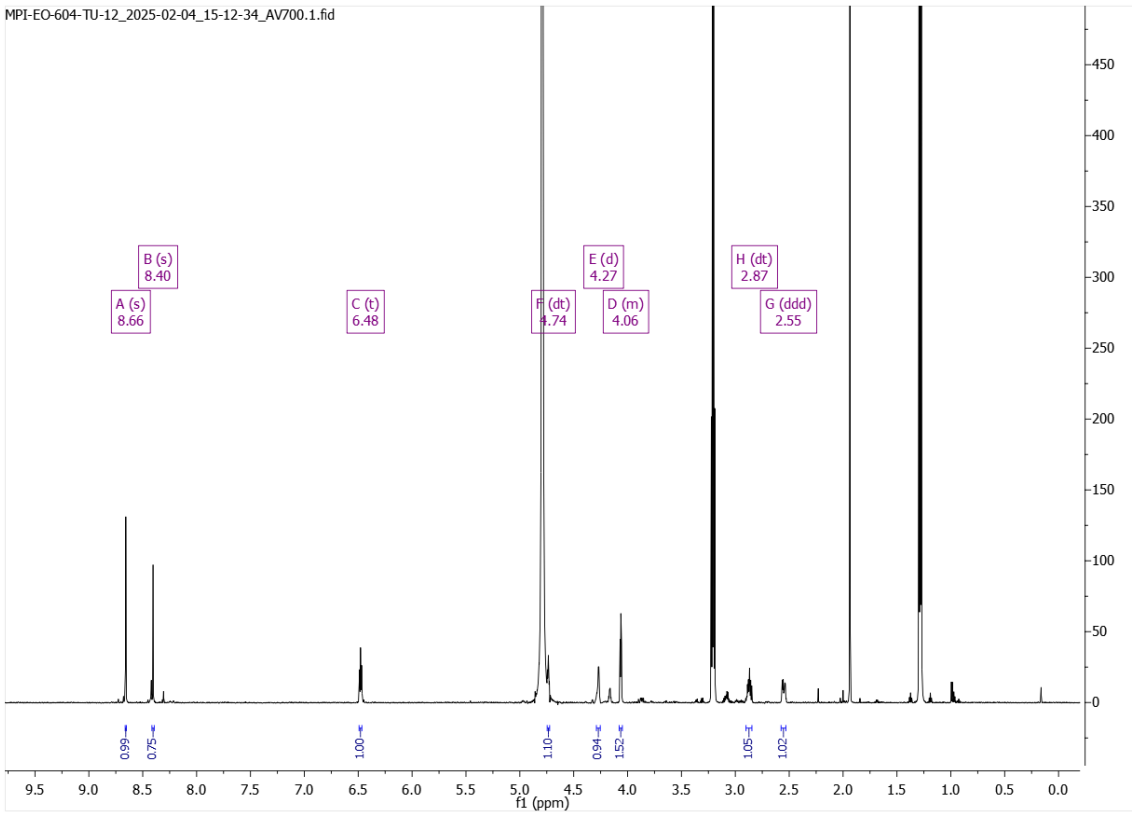



*Figure 101: $^{1}$H NMR spectrum of compound 48.*

## 9.8 Unedited gel images

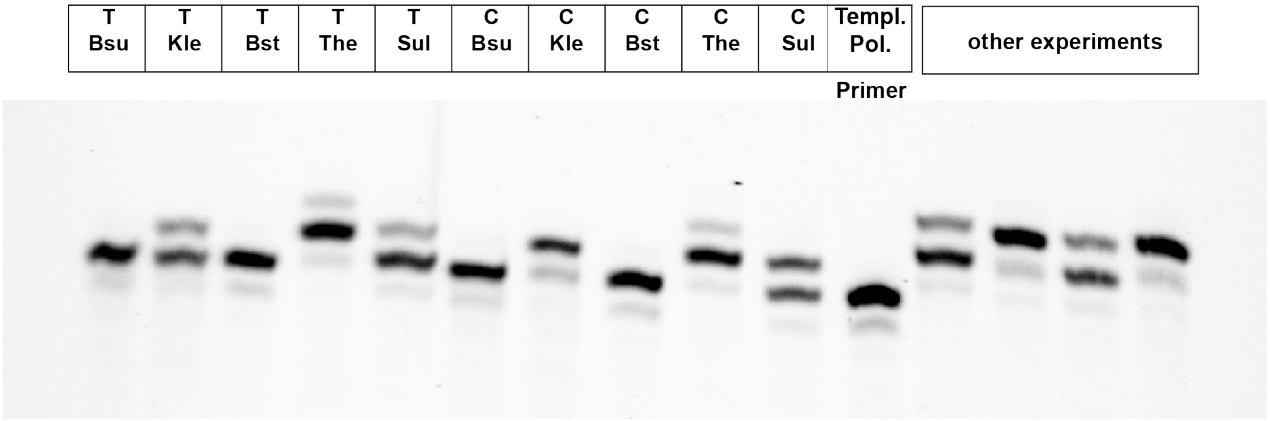



*Figure 102: Unedited image of Figure 45. Urea-PAGE of the enzymatic incorporation of 2-Amino-DDP with different polymerases. Conditions: 0.5 µM template, 0.25 µM primer, 100 µM dNTP, 0.05 units/µL polymerase, 10 min. Temp: 60 °C for Bst, Kle and Sul, 37 °C for Bsu and Kle. Bsu (Bsu DNA Polymerase, Large Fragment), Kle (Klenow-Fragment), Bst (Bst DNA Polymerase, Large Fragment), The (Therminator™ DNA Polymerase), Sul (Sulfolobus DNA Polymerase IV).*

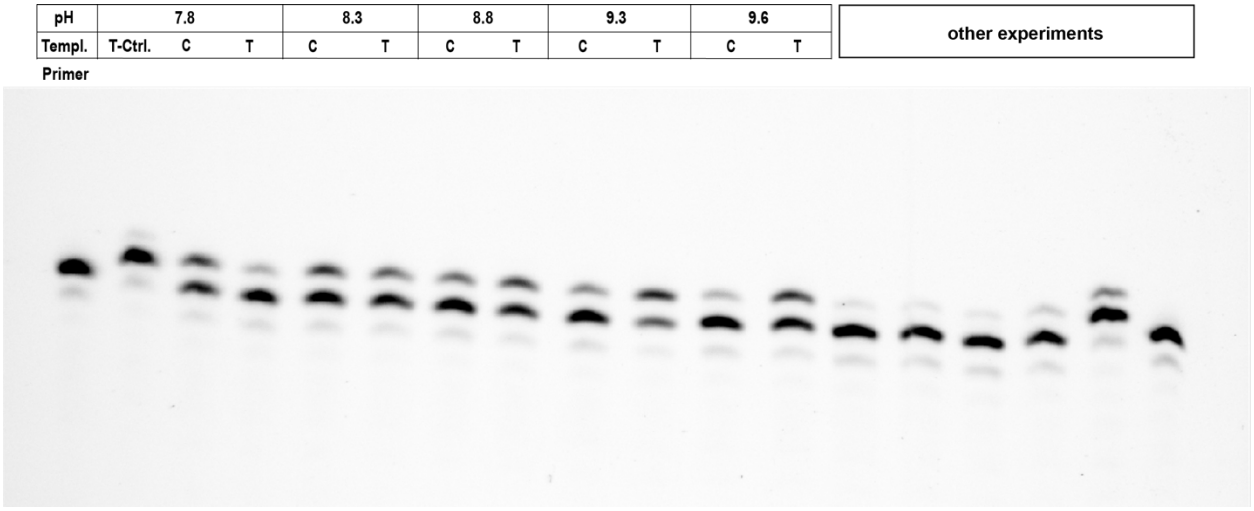



*Figure 103: Unedited image of Figure 46. Urea-PAGE of the enzymatic and pH-dependent incorporation of 2-Amino-DDP with Thermo Sequenase™ DNA Polymerase (Cytiva), conditions: 0.5 µM template, 0.25 µM primer, 25 µM dNTP, 0.05 units/µL polymerase, 30 mM Tris-HCl, 7.5 mM $MgSO_4$, 60 °C, 10 min.*

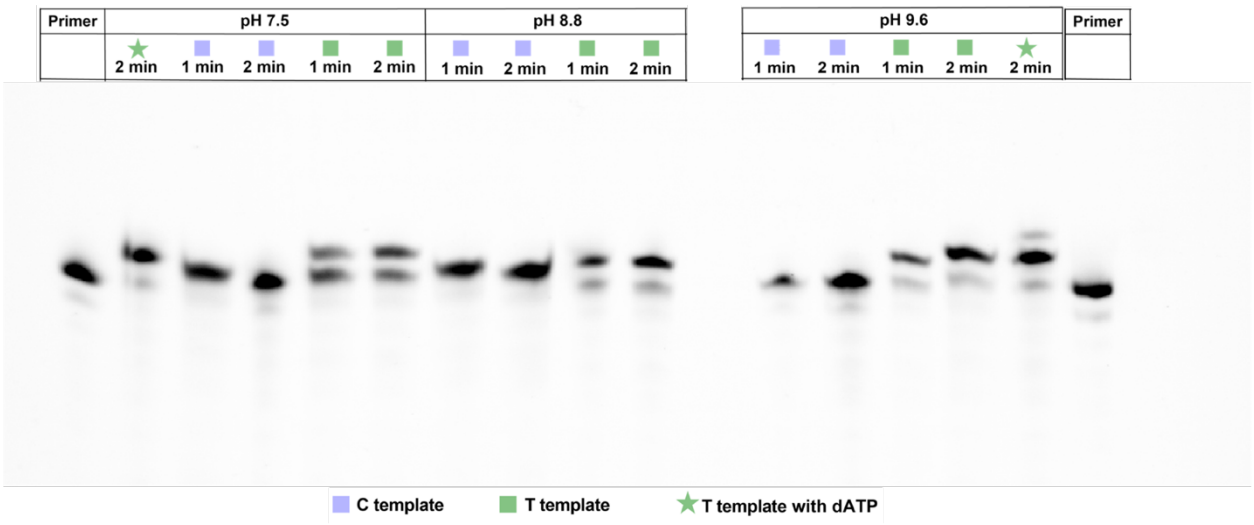



*Figure 104: Unedited image of Figure 52. Urea-PAGE of the incorporation of 2-Aminopurine with KlenTaq. Conditions: 0.5 µM template, 0.25 µM primer, 25 µM dNTP, 0.05 units/µL polymerase, 25 mM Tris-HCl, 40 mM KCl, 5 mM $MgSO_4$, 60 °C. The experiments where dATP was incorporated opposite a T template were conducted as positive control.*

## 9.9 Software

*Table 14: Software used within this thesis.*

| Software | Supplier | Description |
|---|---|---|
| Adobe Illustrator 2025 | Adobe Inc. | Illustration of figures, graphics and schemes |
| DeepL Write 2025<br>https://www.deepl.com/en/write | DeepL SE, subscription plan of the Max Planck Society | Checking grammar and spelling and correcting self-written sentences |
| ChemDraw Professional 23.1.2 | Revvity Signals Inc. | Creating and modifying chemical structures and reaction mechanisms |
| GraphPad Prism 10 | GraphPad Software, Inc. | Statistical analysis and quantification of measured data |
| Image Lab 6.0.1 | Bio-Rad Laboratories Inc. (Hercules, CA, USA) | Visualizing and quantifying bands in PAGE gels |
| Mnova 15.0.1 | Mestrelab Research | Analyzing NMR data |
| EndNote 20 | Clarivate Analytics | Reference management |

# 10. References


[1] S. Nguyen, Master's thesis, "Chemical labelling of 5-formylcytosine in DNA and its potential use for DNA sequencing", TU Dortmund **2024**.

[2] A. Schöne, Master's thesis, "Development of novel chemical labeling strategies for the detection of 5-formylcytosine in DNA", TU Dortmund **2025**.

[3] E. Ogel, S. Becker, "Synthesis and Incorporation of a pH-Responsive Nucleoside Into DNA Sequences", *Chembiochem* **2025**, e70086.

[4] M. J. Booth, T. W. B. Ost, D. Beraldi, N. M. Bell, M. R. Branco, W. Reik, S. Balasubramanian, "Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine", *Nat Protoc* **2013**, *8*, 1841-1851.

[5] T. M. Storebjerg, S. H. Strand, S. Hoyer, A. S. Lynnerup, M. Borre, T. F. Orntoft, K. D. Sorensen, "Dysregulation and prognostic potential of 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) levels in prostate cancer", *Clin Epigenetics* **2018**, *10*.

[6] C. Loy, L. Ahmann, I. De Vlaminck, W. Gu, "Liquid Biopsy Based on Cell-Free DNA and RNA", *Annu Rev Biomed Eng* **2024**, *26*, 169-195.

[7] B. A. Flusberg, D. R. Webster, J. H. Lee, K. J. Travers, E. C. Olivares, T. A. Clark, J. Korlach, S. W. Turner, "Direct detection of DNA methylation during single-molecule, real-time sequencing", *Nature Methods* **2010**, *7*, 461-472.

[8] D. Schmidl, S. M. Becker, J. M. Edgerton, S. Balasubramanian, "An unnatural base pair for the detection of epigenetic cytosine modifications in DNA", *Nat Chem* **2025**.

[9] C. Zhu, Y. Gao, H. Guo, B. Xia, J. Song, X. Wu, H. Zeng, K. Kee, F. Tang, C. Yi, "Single-Cell 5-Formylcytosine Landscapes of Mammalian Early Embryos and ESCs at Single-Base Resolution", *Cell Stem Cell* **2017**, *20*, 720-731.

[10] J. L. T. Jeremy M. Berg, Gregory J. Gatto jr., Lubert Stryer, *Stryer Biochemie*, 8 ed., Springer Spektrum Berlin, Heidelberg, **2017**.

[11] J. D. Watson, F. H. Crick, "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid", *Nature* **1953**, *171*, 737-738.

[12] R. E. Franklin, R. G. Gosling, "Molecular Configuration in Sodium Thymonucleate", *Nature* **1953**, *171*, 740-741.

[13] X. M. Cai, P. J. Gray, D. D. Von Hoff, "DNA minor groove binders: Back in the groove", *Cancer Treat Rev* **2009**, *35*, 437-450.

[14] P. L. Hamilton, D. P. Arya, "Natural product DNA major groove binders", *Nat Prod Rep* **2012**, *29*, 134-143.

[15] M. K. Bilyard, S. Becker, S. Balasubramanian, "Natural, modified DNA bases", *Curr Opin Chem Biol* **2020**, *57*, 1-7.

[16] S. Kriaucionis, N. Heintz, "The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain", *Science* **2009**, *324*, 929-930.

[17] E. R. Gibney, C. M. Nolan, "Epigenetics and gene expression", *Heredity (Edinb)* **2010**, *105*, 4-13.

[18] Z. D. Smith, A. Meissner, "DNA methylation: roles in mammalian development", *Nature Reviews Genetics* **2013**, *14*, 204-220.

[19] C. X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S. Q. Mao, L. Lin, C. Street, Y. J. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G. L. Xu, P. Jin, C. He, "Genome-wide Profiling of 5-Formylcytosine Reveals Its Roles in Epigenetic Priming", *Cell* **2013**, *153*, 678-691.

[20] A. L. Mattei, N. Bailly, A. Meissner, "DNA methylation: a historical perspective", *Trends Genet* **2022**, *38*, 676-707.

[21] J. Fullgrabe, W. S. Gosal, P. Creed, S. Liu, C. K. Lumby, D. J. Morley, T. W. B. Ost, A. J. Vilella, S. Yu, H. Bignell, P. Burns, T. Charlesworth, B. Fu, H. Fordham, N. J. Harding, O. Gandelman, P. Golder, C. Hodson, M. Li, M. Lila, Y. Liu, J. Mason, J. Mellad, J. M. Monahan, O. Nentwich, A. Palmer, M. Steward, M. Taipale, A. Vandomme, R. S. San-Bento, A. Singhal, J. Vivian, N. Wojtowicz, N. Williams, N. J. Walker, N. C. H. Wong, G. N. Yalloway, J. D. Holbrook, S. Balasubramanian, "Simultaneous sequencing of genetic and epigenetic bases in DNA", *Nat Biotechnol* **2023**, *41*, 1457–1464.

[22] A. Jambhekar, A. Dhall, Y. Shi, “Roles and regulation of histone methylation in animal development”, *Nat Rev Mol Cell Bio* **2019**, *20*, 625-641.
[23] C. R. Clapier, B. R. Cairns, “The biology of chromatin remodeling complexes”, *Annu Rev Biochem* **2009**, *78*, 273-304.
[24] A. M. Fleming, Y. Ding, C. J. Burrows, “Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair”, *P Natl Acad Sci USA* **2017**, *114*, 2604-2609.
[25] B. Searle, M. Müuller, T. Carell, A. Kellett, “Third-Generation Sequencing of Epigenetic DNA”, *Angew Chem Int Edit* **2023**, *135*.
[26] M. R. Branco, G. Ficz, W. Reik, “Uncovering the role of 5-hydroxymethylcytosine in the epigenome”, *Nat Rev Genet* **2011**, *13*, 7-13.
[27] M. Frommer, L. E. McDonald, D. S. Millar, C. M. Collis, F. Watt, G. W. Grigg, P. L. Molloy, C. L. Paul, “A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands”, *Proc Natl Acad Sci U S A* **1992**, *89*, 1827-1831.
[28] T. B. Johnson, R. D. Coghill, “Researches on pyrimidines. CIII. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus”, *Journal of the American Chemical Society* **1925**, *47*, 2838-2844.
[29] R. D. Hotchkiss, “The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography”, *J Biol Chem* **1948**, *175*, 315-332.
[30] M. Gold, J. Hurwitz, M. Anders, “The Enzymatic Methylation of Rna and DNA, Ii. On the Species Specificity of the Methylation Enzymes”, *Proc Natl Acad Sci U S A* **1963**, *50*, 164-169.
[31] C. A. Gillespie, A. Chowdhury, K. A. Quinn, M. W. Jenkins, A. M. Rollins, M. Watanabe, S. M. Ford, “Fundamentals of DNA methylation in development”, *Pediatr Res* **2024**.
[32] X. J. Wu, Y. Zhang, “TET-mediated active DNA demethylation: mechanism, function and beyond”, *Nature Reviews Genetics* **2017**, *18*, 517-534.
[33] M. Okano, D. W. Bell, D. A. Haber, E. Li, “DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development”, *Cell* **1999**, *99*, 247-257.
[34] A. Hermann, R. Goyal, A. Jeltsch, “The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites”, *Journal of Biological Chemistry* **2004**, *279*, 48350-48359.
[35] M. Bostick, J. K. Kim, P. O. Estève, A. Clark, S. Pradhan, S. E. Jacobsen, “UHRF1 plays a role in maintaining DNA methylation in mammalian cells”, *Science* **2007**, *317*, 1760-1764.
[36] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, “Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine”, *Science* **2011**, *333*, 1300-1303.
[37] P. Melamed, Y. Yosefzon, C. David, A. Tsukerman, L. Pnueli, “Tet Enzymes, Variants, and Differential Effects on Function”, *Front Cell Dev Biol* **2018**, *6*.
[38] S. G. Jin, Z. M. Zhang, T. L. Dunwell, M. R. Harter, X. Wu, J. Johnson, Z. Li, J. Liu, P. E. Szabó, Q. Lu, G. L. Xu, J. Song, G. P. Pfeifer, “Tet3 Reads 5-Carboxylcytosine through Its CXXC Domain and Is a Potential Guardian against Neurodegeneration”, *Cell Rep* **2016**, *14*, 493-505.
[39] M. Yu, G. C. Hon, K. E. Szulwach, C. X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park, J. H. Min, P. Jin, B. Ren, C. He, “Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome”, *Cell* **2012**, *149*, 1368-1380.
[40] E. K. Schutsky, J. E. DeNizio, P. Hu, M. Y. Liu, C. S. Nabel, E. B. Fabyanic, Y. Hwang, F. D. Bushman, H. Wu, R. M. Kohli, “Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase”, *Nat Biotechnol* **2018**, *36*, 1083-1090.
[41] H. Xu, J. Chen, J. Cheng, L. Kong, X. Chen, M. Inoue, Y. Liu, S. Kriaucionis, M. Zhao, C. X. Song, “Modular Oxidation of Cytosine Modifications and Their Application in Direct and Quantitative Sequencing of 5-Hydroxymethylcytosine”, *J Am Chem Soc* **2023**, *145*, 7095-7100.
[42] B. Xia, D. Han, X. Lu, Z. Sun, A. Zhou, Q. Yin, H. Zeng, M. Liu, X. Jiang, W. Xie, C. He, C. Yi, “Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale”, *Nat Methods* **2015**, *12*, 1047-1050.
[43] Y. Gao, L. Li, P. Yuan, F. Zhai, Y. X. Ren, L. Y. Yan, R. Li, Y. Lian, X. H. Zhu, X. L. Wu, K. Kee, L. Wen, J. Qiao, F. C. Tang, “5-Formylcytosine landscapes of human preimplantation embryos at single-cell resolution”, *Plos Biol* **2020**, *18*.
[44] A. Li, X. Sun, A. E. Arguello, R. E. Kleiner, “Chemical Method to Sequence 5-Formylcytosine on RNA”, *ACS Chem Biol* **2022**, *17*, 503-508.

[45] T. Pfaffeneder, B. Hackner, M. Truss, M. Munzel, M. Muller, C. A. Deiml, C. Hagemeier, T. Carell, "The discovery of 5-formylcytosine in embryonic stem cell DNA", *Angew Chem Int Ed Engl* **2011**, *50*, 7008-7012.
[46] E. Parasyraki, M. Mallick, V. Hatch, V. Vastolo, M. U. Musheev, E. Karaulanov, A. Gopanenko, S. Moxon, M. Méndez-Lago, D. D. Han, L. Schomacher, D. Mukherjee, C. Niehrs, "5-Formylcytosine is an activating epigenetic mark for RNA Pol III during zygotic reprogramming", *Cell* **2024**, *187*.
[47] S. F. Ji, H. Z. Shao, Q. Y. Han, C. L. Seiler, N. Y. Tretyakova, "Reversible DNA-Protein Cross-Linking at Epigenetic DNA Marks", *Angew Chem Int Edit* **2017**, *56*, 14130-14134.
[48] Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, "Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA", *Science* **2011**, *333*, 1303-1307.
[49] L. F. Wang, Y. Zhou, L. Xu, R. Xiao, X. Y. Lu, L. Chen, J. Chong, H. R. Li, C. He, X. D. Fu, D. Wang, "Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex", *Nature* **2015**, *523*, 621.
[50] B. Kosel, K. Bigler, B. C. Buchmuller, S. R. Acharyya, R. Linser, D. Summerer, "Evolved Readers of 5-Carboxylcytosine CpG Dyads Reveal a High Versatility of the Methyl-CpG-Binding Domain for Recognition of Noncanonical Epigenetic Marks", *Angew Chem Int Edit* **2024**, *63*.
[51] D. R. Bentley, S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, J. M. Boutell, J. Bryant, R. J. Carter, R. Keira Cheetham, A. J. Cox, D. J. Ellis, M. R. Flatbush, N. A. Gormley, S. J. Humphray, L. J. Irving, M. S. Karbelashvili, S. M. Kirk, H. Li, X. Liu, K. S. Maisinger, L. J. Murray, B. Obradovic, T. Ost, M. L. Parkinson, M. R. Pratt, I. M. Rasolonjatovo, M. T. Reed, R. Rigatti, C. Rodighiero, M. T. Ross, A. Sabot, S. V. Sankar, A. Scally, G. P. Schroth, M. E. Smith, V. P. Smith, A. Spiridou, P. E. Torrance, S. S. Tzonev, E. H. Vermaas, K. Walter, X. Wu, L. Zhang, M. D. Alam, C. Anastasi, I. C. Aniebo, D. M. Bailey, I. R. Bancarz, S. Banerjee, S. G. Barbour, P. A. Baybayan, V. A. Benoit, K. F. Benson, C. Bevis, P. J. Black, A. Boodhun, J. S. Brennan, J. A. Bridgham, R. C. Brown, A. A. Brown, D. H. Buermann, A. A. Bundu, J. C. Burrows, N. P. Carter, N. Castillo, E. C. M. Chiara, S. Chang, R. Neil Cooley, N. R. Crake, O. O. Dada, K. D. Diakoumakos, B. Dominguez-Fernandez, D. J. Earnshaw, U. C. Egbujor, D. W. Elmore, S. S. Etchin, M. R. Ewan, M. Fedurco, L. J. Fraser, K. V. Fuentes Fajardo, W. Scott Furey, D. George, K. J. Gietzen, C. P. Goddard, G. S. Golda, P. A. Granieri, D. E. Green, D. L. Gustafson, N. F. Hansen, K. Harnish, C. D. Haudenschild, N. I. Heyer, M. M. Hims, J. T. Ho, A. M. Horgan, et al., "Accurate whole human genome sequencing using reversible terminator chemistry", *Nature* **2008**, *456*, 53-59.
[52] S. D. Mooney, "Progress towards the integration of pharmacogenomics in practice", *Hum Genet* **2015**, *134*, 459-465.
[53] M. Ignatiadis, G. W. Sledge, S. S. Jeffrey, "Liquid biopsy enters the clinic - implementation issues and future challenges", *Nat Rev Clin Oncol* **2021**, *18*, 297-312.
[54] M. Zalis, G. G. V. Veloso, P. J. r. Aguiar, N. Gimenes, M. X. Reis, S. Matsas, C. G. Ferreira, "Next-generation sequencing impact on cancer care: applications, challenges, and future directions", *Front Genet* **2024**, *15*.
[55] F. Sanger, A. R. Coulson, "Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA-Polymerase", *J Mol Biol* **1975**, *94*, 441.
[56] F. Sanger, S. Nicklen, A. R. Coulson, "DNA sequencing with chain-terminating inhibitors", *Proc Natl Acad Sci U S A* **1977**, *74*, 5463-5467.
[57] E. E. Schadt, S. Turner, A. Kasarskis, "A window into third-generation sequencing", *Hum Mol Genet* **2010**, *19*, R227-240.
[58] J. Shendure, S. Balasubramanian, G. M. Church, W. Gilbert, J. Rogers, J. A. Schloss, R. H. Waterston, "DNA sequencing at 40: past, present and future", *Nature* **2017**, *550*, 345-353.
[59] P. Nyren, B. Pettersson, M. Uhlen, "Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay", *Anal Biochem* **1993**, *208*, 171-175.
[60] J. M. Rothberg, W. Hinz, T. M. Rearick, J. Schultz, W. Mileski, M. Davey, J. H. Leamon, K. Johnson, M. J. Milgrew, M. Edwards, J. Hoon, J. F. Simons, D. Marran, J. W. Myers, J. F. Davidson, A. Branting, J. R. Nobile, B. P. Puc, D. Light, T. A. Clark, M. Huber, J. T. Branciforte, I. B. Stoner, S. E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J. A. Fidanza, E.

Namsaraev, K. J. McKernan, A. Williams, G. T. Roth, J. Bustillo, “An integrated semiconductor device enabling non-optical genome sequencing”, *Nature* **2011**, *475*, 348-352.

[61] B. Merriman, J. M. Rothberg, I. T. R. D. Team, “Progress in Ion Torrent semiconductor chip based sequencing”, *Electrophoresis* **2012**, *33*, 3397-3417.

[62] M. J. Levene, J. Korlach, S. W. Turner, M. Foquet, H. G. Craighead, W. W. Webb, “Zero-mode waveguides for single-molecule analysis at high concentrations”, *Science* **2003**, *299*, 682-686.

[63] J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. Dewinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach, S. Turner, “Real-Time DNA Sequencing from Single Polymerase Molecules”, *Science* **2009**, *323*, 133-138.

[64] A. H. Laszlo, I. M. Derrington, B. C. Ross, H. Brinkerhoff, A. Adey, I. C. Nova, J. M. Craig, K. W. Langford, J. M. Samson, R. Daza, K. Doering, J. Shendure, J. H. Gundlach, “Decoding long nanopore sequencing reads of natural DNA”, *Nature Biotechnology* **2014**, *32*, 829-833.

[65] Y. Liu, P. Siejka-Zielinska, G. Velikova, Y. Bi, F. Yuan, M. Tomkova, C. Bai, L. Chen, B. Schuster-Bockler, C. X. Song, “Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution”, *Nat Biotechnol* **2019**, *37*, 424-429.

[66] T. Wang, J. M. Fowler, L. Liu, C. E. Loo, M. Luo, E. K. Schutsky, K. N. Berrios, J. E. DeNizio, A. Dvorak, N. Downey, S. Montermoso, B. Y. Pingul, M. Nasrallah, W. S. Gosal, H. Wu, R. M. Kohli, “Direct enzymatic sequencing of 5-methylcytosine at single-base resolution”, *Nat Chem Biol* **2023**.

[67] A. Hecht, M. Grunstein, “Mapping DNA interaction sites of chromosomal proteins using immunoprecipitation and polymerase chain reaction”, *Method Enzymol* **1999**, *304*, 399-414.

[68] A. Lentini, C. Lagerwall, S. Vikingsson, H. K. Mjoseng, K. Douvlataniotis, H. Vogt, H. Green, R. R. Meehan, M. Benson, C. E. Nestor, “A reassessment of DNA-immunoprecipitation-based genomic profiling”, *Nature Methods* **2018**, *15*, 499-504.

[69] Y. Huang, W. A. Pastor, Y. Shen, M. Tahiliani, D. R. Liu, A. Rao, “The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing”, *PLoS One* **2010**, *5*, e8888.

[70] K. Tanaka, A. Okamoto, “Degradation of DNA by bisulfite treatment”, *Bioorg Med Chem Lett* **2007**, *17*, 1912-1915.

[71] Y. Liu, J. Cheng, P. Siejka-Zielinska, C. Weldon, H. Roberts, M. Lopopolo, A. Magri, V. D'Arienzo, J. M. Harris, J. A. McKeating, C. X. Song, “Accurate targeted long-read DNA methylation and hydroxymethylation sequencing with TAPS”, *Genome Biol* **2020**, *21*, 54.

[72] M. Yu, G. C. Hon, K. E. Szulwach, C. X. Song, P. Jin, B. Ren, C. He, “Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine”, *Nat Protoc* **2012**, *7*, 2159-2170.

[73] M. J. Booth, G. Marsico, M. Bachman, D. Beraldi, S. Balasubramanian, “Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution”, *Nature Chemistry* **2014**, *6*, 435-440.

[74] X. Y. Lu, C. X. Song, K. Szulwach, Z. P. Wang, P. Weidenbacher, P. Jin, C. He, “Chemical Modification-Assisted Bisulfite Sequencing (CAB-Seq) for 5-Carboxylcytosine Detection in DNA”, *Journal of the American Chemical Society* **2013**, *135*, 9315-9317.

[75] H. Zeng, M. Mondal, R. Song, J. Zhang, B. Xia, M. Liu, C. Zhu, B. He, Y. Q. Gao, C. Yi, “Unnatural Cytosine Bases Recognized as Thymines by DNA Polymerases by the Formation of the Watson-Crick Geometry”, *Angew Chem Int Ed Engl* **2019**, *58*, 130-133.

[76] K. Futami, M. Kimoto, Y. W. S. Lim, I. Hirao, “Genetic Alphabet Expansion Provides Versatile Specificities and Activities of Unnatural-Base DNA Aptamers Targeting Cancer Cells”, *Mol Ther Nucleic Acids* **2019**, *14*, 158-170.

[77] S. Hoshika, N. A. Leal, M. J. Kim, M. S. Kim, N. B. Karalkar, H. J. Kim, A. M. Bates, N. E. Watkins, Jr., H. A. SantaLucia, A. J. Meyer, S. DasGupta, J. A. Piccirilli, A. D. Ellington, J. SantaLucia, Jr., M. M. Georgiadis, S. A. Benner, “Hachimoji DNA and RNA: A genetic system with eight building blocks”, *Science* **2019**, *363*, 884-887.

[78] M. Kimoto, I. Hirao, “Genetic alphabet expansion technology by creating unnatural base pairs”, *Chem Soc Rev* **2020**, *49*, 7602-7626.

[79] C. Switzer, S. E. Moroney, S. A. Benner, “Enzymatic incorporation of a new base pair into DNA and RNA”, *Journal of the American Chemical Society* **1989**, *111*, 8322-8323.

[80] J. A. Piccirilli, T. Krauch, S. E. Moroney, S. A. Benner, “Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet”, *Nature* **1990**, *343*, 33-37.
[81] J. D. Bain, C. Switzer, A. R. Chamberlin, S. A. Benner, “Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code”, *Nature* **1992**, *356*, 537-539.
[82] Z. Yang, A. M. Sismour, P. Sheng, N. L. Puskar, S. A. Benner, “Enzymatic incorporation of a third nucleobase pair”, *Nucleic Acids Res* **2007**, *35*, 4238-4249.
[83] D. A. Malyshev, K. Dhami, T. Lavergne, T. Chen, N. Dai, J. M. Foster, I. R. Correa, Jr., F. E. Romesberg, “A semi-synthetic organism with an expanded genetic alphabet”, *Nature* **2014**, *509*, 385-388.
[84] Y. Zhang, J. L. Ptacin, E. C. Fischer, H. R. Aerni, C. E. Caffaro, K. San Jose, A. W. Feldman, C. R. Turner, F. E. Romesberg, “A semi-synthetic organism that stores and retrieves increased genetic information”, *Nature* **2017**, *551*, 644-647.
[85] I. Hirao, T. Ohtsuki, T. Fujiwara, T. Mitsui, T. Yokogawa, T. Okuni, H. Nakayama, K. Takio, T. Yabuki, T. Kigawa, K. Kodama, T. Yokogawa, K. Nishikawa, S. Yokoyama, “An unnatural base pair for incorporating amino acid analogs into proteins”, *Nat Biotechnol* **2002**, *20*, 177-182.
[86] M. Kimoto, R. Yamashige, K. Matsunaga, S. Yokoyama, I. Hirao, “Generation of high-affinity DNA aptamers using an expanded genetic alphabet”, *Nat Biotechnol* **2013**, *31*, 453-457.
[87] K. I. Matsunaga, M. Kimoto, V. W. Lim, H. P. Tan, Y. Q. Wong, W. Sun, S. Vasoo, Y. S. Leo, I. Hirao, “High-affinity five/six-letter DNA aptamers with superior specificity enabling the detection of dengue NS1 protein variants beyond the serotype identification”, *Nucleic Acids Res* **2021**, *49*, 11407-11424.
[88] A. D. Ellington, J. W. Szostak, “Invitro Selection of Rna Molecules That Bind Specific Ligands”, *Nature* **1990**, *346*, 818-822.
[89] I. Hirao, M. Kimoto, K. H. Lee, “DNA aptamer generation by ExSELEX using genetic alphabet expansion with a mini-hairpin DNA stabilization method”, *Biochimie* **2018**, *145*, 15-21.
[90] L. Ceze, J. Nivala, K. Strauss, “Molecular digital data storage using DNA”, *Nat Rev Genet* **2019**, *20*, 456-466.
[91] M. Kimoto, S. H. G. Soh, I. Hirao, “Sanger Gap Sequencing for Genetic Alphabet Expansion of DNA”, *Chembiochem* **2020**, *21*, 2287-2296.
[92] Q. Q. Hu, H. Li, L. H. Wang, H. Z. Gu, C. H. Fan, “DNA Nanotechnology-Enabled Drug Delivery Systems”, *Chemical Reviews* **2019**, *119*, 6459-6506.
[93] X. Mao, M. Liu, Q. Li, C. Fan, X. Zuo, “DNA-Based Molecular Machines”, *JACS Au* **2022**, *2*, 2381-2399.
[94] P. L. Anelli, N. Spencer, J. F. Stoddart, “A molecular shuttle”, *J Am Chem Soc* **1991**, *113*, 5131-5133.
[95] R. Van Noorden, D. Castelvecchi, “World's tiniest machines win chemistry Nobel”, *Nature* **2016**, *538*, 152-153.
[96] H. Ramezani, H. Dietz, “Building machines with DNA molecules”, *Nat Rev Genet* **2020**, *21*, 5-26.
[97] P. W. Rothemund, “Folding DNA to create nanoscale shapes and patterns”, *Nature* **2006**, *440*, 297-302.
[98] J. Ji, D. Karna, H. Mao, “DNA origami nano-mechanics”, *Chem Soc Rev* **2021**, *50*, 11966-11978.
[99] J. Li, T. P. Chiu, R. Rohs, “Predicting DNA structure using a deep learning method”, *Nat Commun* **2024**, *15*, 1243.
[100] R. D. Astumian, “Thermodynamics and kinetics of a Brownian motor”, *Science* **1997**, *276*, 917-922.
[101] A. K. Pumm, W. Engelen, E. Kopperger, J. Isensee, M. Vogt, V. Kozina, M. Kube, M. N. Honemann, E. Bertosin, M. Langecker, R. Golestanian, F. C. Simmel, H. Dietz, “A DNA origami rotary ratchet motor”, *Nature* **2022**, *607*, 492-498.
[102] P. D. Boyer, “The ATP synthase - A splendid molecular machine”, *Annual Review of Biochemistry* **1997**, *66*, 717-749.
[103] J. C. Gonzalez-Olvera, M. Durec, R. Marek, R. Fiala, M. Morales-Garcia, E. Gonzalez-Jasso, R. C. Pless, “Protonation of Nucleobases in Single- and Double-Stranded DNA”, *Chembiochem* **2018**, *19*, 2088-2098.
[104] T. I. Smol'janinova, V. A. Zhidkov, G. V. Sokolov, “Analysis of difference spectra of protonated DNA: determination of degree of protonation of nitrogen bases and the fractions of disordered nucleotide pairs”, *Nucleic Acids Res* **1982**, *10*, 2121-2134.

[105] R. K. Sigel, E. Freisinger, B. Lippert, “Effects of N7-methylation, N7-platination, and C8-hydroxylation of guanine on H-bond formation with cytosine: platinum coordination strengthens the Watson-Crick pair”, *J Biol Inorg Chem* **2000**, *5*, 287-299.
[106] H. Mei, S. A. Ingale, F. Seela, “Imidazolo-dC Metal-Mediated Base Pairs: Purine Nucleosides Capture Two Ag Ions and Form a Duplex with the Stability of a Covalent DNA Cross-Link”, *Chem-Eur J* **2014**, *20*, 16248-16257.
[107] R. Krishnamurthy, “Role of pK(a) of nucleobases in the origins of chemical evolution”, *Acc Chem Res* **2012**, *45*, 2035-2044.
[108] K. Seio, T. Sasami, R. Tawarada, M. Sekine, “Synthesis of 2'-O-methyl-RNAs incorporating a 3-deazaguanine, and UV melting and computational studies on its hybridization properties”, *Nucleic Acids Res* **2006**, *34*, 4324-4334.
[109] R. Guimil Garcia, A. S. Brank, J. K. Christman, V. E. Marquez, R. Eritja, “Synthesis of oligonucleotide inhibitors of DNA (Cytosine-C5) methyltransferase containing 5-azacytosine residues at specific sites”, *Antisense Nucleic Acid Drug Dev* **2001**, *11*, 369-378.
[110] X. Peng, H. Li, F. Seela, “pH-Dependent mismatch discrimination of oligonucleotide duplexes containing 2'-deoxytubercidin and 2- or 7-substituted derivatives: protonated base pairs formed between 7-deazapurines and cytosine”, *Nucleic Acids Res* **2006**, *34*, 5987-6000.
[111] M. D. Dore, M. G. Rafique, T. P. Yang, M. Zorman, C. M. Platnich, P. F. Xu, T. Trinh, F. J. Rizzuto, G. Cosa, J. N. Li, A. Guarne, H. F. Sleiman, “Heat-activated growth of metastable and length-defined DNA fibers expands traditional polymer assembly”, *Nature Communications* **2024**, *15*.
[112] B. Chen, Y. T. Wang, W. J. Ma, H. Cheng, H. H. Sun, H. Z. Wang, J. Huang, X. X. He, K. M. Wang, “A Mimosa-Inspired Cell-Surface-Anchored Ratiometric DNA Nanosensor for High-Resolution and Sensitive Response of Target Tumor Extracellular pH”, *Anal Chem* **2020**, *92*, 15104-15111.
[113] S. Modi, M. G. Swetha, D. Goswami, G. D. Gupta, S. Mayor, Y. Krishnan, “A DNA nanomachine that maps spatial and temporal pH changes inside living cells”, *Nat Nanotechnol* **2009**, *4*, 325-330.
[114] Y. W. Li, S. Z. Yue, J. Y. Cao, C. Z. Zhu, Y. X. Wang, X. Hai, E. L. Song, S. Bi, “pH-responsive DNA nanomicelles for chemo-gene Synergetic Therapy of Anaplastic Large Cell Lymphoma”, *Theranostics* **2020**, *10*, 8250-8263.
[115] E. K. Voinkov, E. N. Ulomskiy, V. L. Rusinov, K. V. Savateev, V. V. Fedotov, E. B. Gorbunov, M. L. Isenov, O. S. Eltsov, “New stable form of nitroacetonitrile”, *Mendeleev Communications* **2016**, *26*, 172-173.
[116] N. Nishiwaki, Y. Takada, Y. Inoue, Y. Tohda, M. Ariga, “Ring-Opening Reaction of the Pyridinium Salt of 4-Nitro-3-Isoxazolin-5-One - a Preparation of Trifunctionalized Methane Derivatives”, *J Heterocyclic Chem* **1995**, *32*, 473-475.
[117] M. J. Kamlet, “Methyl 2-Nitro-3-Ethoxyacrylate and Related Compounds”, *Journal of Organic Chemistry* **1959**, *24*, 714-715.
[118] K. Iwai, N. Nishiwaki, “Nitroacetonitrile and Its Synthetic Equivalents”, *J Org Chem* **2021**, *86*, 13177-13185.
[119] X. Y. Jin, Z. R. Huang, L. J. Xie, L. Liu, D. L. Han, L. Cheng, “Photo-Facilitated Detection and Sequencing of 5-Formylcytidine RNA”, *Angew Chem Int Edit* **2022**, *61*.
[120] T. Nevesely, M. Wienhold, J. J. Molloy, R. Gilmour, “Advances in the E to Z Isomerization of Alkenes Using Small Molecule Photocatalysts”, *Chemical Reviews* **2022**, *122*, 2650-2694.
[121] C. Kielbassa, L. Roza, B. Epe, “Wavelength dependence of oxidative DNA damage induced by UV and visible light”, *Carcinogenesis* **1997**, *18*, 811-816.
[122] S. Tasqeeruddin, Y. I. Asiri, S. Shaheen, “An environmentally benign, simple and proficient synthesis of quinoline derivatives catalyzed by FeCl as a green and readily available catalyst”, *Green Chem Lett Rev* **2021**, *14*, 118-126.
[123] J. Marco-Contelles, E. Pérez-Mayoral, A. Samadi, M. D. Carreiras, E. Soriano, “Recent Advances in the Friedlander Reaction”, *Chemical Reviews* **2009**, *109*, 2652-2671.
[124] M. N. Mattath, D. Ghosh, S. Pratihar, S. Shi, T. Govindaraju, “Nucleic Acid Architectonics for pH-Responsive DNA Systems and Devices”, *Acs Omega* **2022**, *7*, 3167-3176.
[125] C. Kielar, S. Ramakrishnan, S. Fricke, G. Grundmeier, A. Keller, “Dynamics of DNA Origami Lattice Formation at Solid-Liquid Interfaces”, *Acs Appl Mater Inter* **2018**, *10*, 44844-44853.
[126] Y. Xin, X. Y. Ji, G. Grundmeier, A. Keller, “Dynamics of lattice defects in mixed DNA origami monolayers”, *Nanoscale* **2020**, *12*, 9733-9743.

[127] F. Seela, H. Rosemeyer, W. Bourgeois, “Synthesis of 5-aza-7-deazapurine and 3,7-dideazapurine 2'-deoxyribofuranosides by solid-liquid phase-transfer glycosylation”, *Nucleic Acids Symp Ser* **1987**, 49-52.
[128] B. P. Fors, K. Dooleweerdt, Q. Zeng, S. L. Buchwald, “An Efficient System For the Pd-Catalyzed Cross-Coupling of Amides and Aryl Chlorides”, *Tetrahedron* **2009**, *65*, 6576-6583.
[129] S. L. Beaucage, R. P. Iyer, “Advances in the Synthesis of Oligonucleotides by the Phosphoramidite Approach”, *Tetrahedron* **1992**, *48*, 2223-2311.
[130] M. P. Reddy, N. B. Hanna, F. Farooqui, “Ultrafast cleavage and deprotection of oligonucleotides synthesis and use of C-Ac derivatives”, *Nucleos Nucleot* **1997**, *16*, 1589-1598.
[131] J. Y. Liao, S. Bala, A. K. Ngor, E. J. Yik, J. C. Chaput, “P(V) Reagents for the Scalable Synthesis of Natural and Modified Nucleoside Triphosphates”, *J Am Chem Soc* **2019**, *141*, 13286-13289.
[132] J. Reiss, J. Moffatt, “Dismutation Reactions of Nucleoside Polyphosphates. III. The Synthesis of α, ı-Dinucleoside 5'-Polyphosphates1”, *The Journal of Organic Chemistry* **1965**, *30*, 3381-3387.
[133] M. A. Reeve, C. W. Fuller, “A Novel Thermostable Polymerase for DNA-Sequencing”, *Nature* **1995**, *376*, 796-797.
[134] K. Betz, D. A. Malyshev, T. Lavergne, W. Welte, K. Diederichs, T. J. Dwyer, P. Ordoukhanian, F. E. Romesberg, A. Marx, “KlenTaq polymerase replicates unnatural base pairs by inducing a Watson-Crick geometry”, *Nature Chemical Biology* **2012**, *8*, 612-614.
[135] C. H. Martinez, C. Dardonville, “Rapid Determination of Ionization Constants (pK a) by UV Spectroscopy Using 96-Well Microtiter Plates”, *ACS Med Chem Lett* **2013**, *4*, 142-145.
[136] L. J. Mcbride, R. Kierzek, S. L. Beaucage, M. H. Caruthers, “Amidine Protecting Groups for Oligonucleotide Synthesis .16.”, *Journal of the American Chemical Society* **1986**, *108*, 2040-2048.
[137] Y. Y. Chai, D. Kondhare, A. G. Zhang, P. Leonard, F. Seela, “The 2-Amino Group of 8-Aza-7-deaza-7-bromopurine-2,6-diamine and Purine-2,6-diamine as Stabilizer for the Adenine-Thymine Base Pair in Heterochiral DNA with Strands in Anomeric Configuration”, *Chem-Eur J* **2021**, *27*, 2093-2103.
[138] H. Bredereck, F. Effenberger, G. Simchen, “Reaktionsfahige Saureamid-Dimethylsulfat-Komplexe”, *Angew Chem Int Edit* **1961**, *73*, 493-493.
[139] G. S. Ti, B. L. Gaffney, R. A. Jones, “Transient Protection - Efficient One-Flask Syntheses of Protected Deoxynucleosides”, *Journal of the American Chemical Society* **1982**, *104*, 1316-1319.
[140] N. D. Sinha, J. Biernat, H. Koster, “Beta-Cyanoethyl N,N-Dialkylamino/N-Morpholinomonochloro Phosphoamidites, New Phosphitylating Agents Facilitating Ease of Deprotection and Work-up of Synthesized Oligonucleotides”, *Tetrahedron Letters* **1983**, *24*, 5843-5846.
[141] J. Kypr, I. Kejnovska, D. Renciuk, M. Vorlickova, “Circular dichroism and conformational polymorphism of DNA”, *Nucleic Acids Res* **2009**, *37*, 1713-1725.
[142] A. Dallmann, L. Dehmel, T. Peters, C. Mügge, C. Griesinger, J. Tuma, N. P. Ernsting, “2-Aminopurine Incorporation Perturbs the Dynamics and Structure of DNA”, *Angew Chem Int Edit* **2010**, *49*, 5989-5992.
[143] T. Lindahl, “Instability and Decay of the Primary Structure of DNA”, *Nature* **1993**, *362*, 709-715.
[144] F. Seela, W. Bourgeois, “Synthesis of 3,7-Dideaza-2'-Deoxyadenosine and Related Pyrrolo[3,2-C]Pyridine 2'-Deoxyribonucleosides and 2',3'-Dideoxyribonucleosides”, *Synthesis-Stuttgart* **1988**, 938-943.
[145] T. Yan, Y. Q. Chen, B. Mortishire-Smith, A. Simeone, A. Hofer, S. Balasubramanian, “Selective Photocatalytic C-H Oxidation of 5-Methylcytosine in DNA”, *Angew Chem Int Edit* **2025**, *64*.
[146] H. Bredereck, G. Simchen, S. Rebsdat, W. Kantlehner, P. Horn, R. Wahl, H. Hoffmann, P. Grieshaber, “Säureamid-Reaktionen, L; Orthoamide, I Darstellung und Eigenschaften der Amidacetale und Aminalester”, *Chemische Berichte* **2006**, *101*, 41-50.
[147] C. C. D. Wybon, C. Mensch, C. Hollanders, C. Gadais, W. A. Herrebout, S. Ballet, B. U. W. Maes, “Zn-Catalyzed tert-Butyl Nicotinate-Directed Amide Cleavage as a Biomimic of Metallo-Exopeptidase Activity”, *ACS Catalysis* **2017**, *8*, 203-218.
[148] J. Ludwig, F. Eckstein, “Rapid and efficient synthesis of nucleoside 5'-0-(1-thiotriphosphates), 5'-triphosphates and 2',3'-cyclophosphorothioates using 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one”, *The Journal of Organic Chemistry* **1989**, *54*, 631-635.
[149] M. Hollenstein, “Nucleoside triphosphates--building blocks for the modification of nucleic acids”, *Molecules* **2012**, *17*, 13569-13591.

[150] M. Flamme, P. Rothlisberger, F. Levi-Acobas, M. Chawla, R. Oliva, L. Cavallo, G. Gasser, P. Marliere, P. Herdewijn, M. Hollenstein, “Enzymatic Formation of an Artificial Base Pair Using a Modified Purine Nucleoside Triphosphate”, *ACS Chem Biol* **2020**, *15*, 2872-2884.
[151] R. S. Goody, M. Isakov, “Simple Synthesis and Separation of the Diastereomers of Alpha-Thio Analogs of Ribophosphates, and Deoxyribophosphates, Diphosphates, and Triphosphates”, *Tetrahedron Letters* **1986**, *27*, 3599-3602.
[152] M. Yoshikawa, T. Kato, T. Takenishi, “A novel method for phosphorylation of nucleosides to 5'-nucleotides”, *Tetrahedron Lett* **1967**, *50*, 5065-5068.
[153] Q. Sun, J. Sun, S.-S. Gong, C.-J. Wang, S.-Z. Pu, F.-D. Feng, “Efficient synthesis of 5-hydroxymethyl-, 5-formyl-, and 5-carboxyl-2′-deoxycytidine and their triphosphates”, *RSC Adv.* **2014**, *4*, 36036-36039.
[154] V. Borsenberger, M. Kukwikila, S. Howorka, “Synthesis and enzymatic incorporation of modified deoxyuridine triphosphates”, *Org Biomol Chem* **2009**, *7*, 3826-3835.
[155] S. Roseman, J. Distler, J. Moffatt, H. Khorana, “Nucleoside polyphosphates. XI. 1 An improved general method for the synthesis of nucleotide coenzymes. Syntheses of uridine-5', cytidine-5'and guanosine-5'diphosphate derivatives”, *Journal of the American Chemical Society* **1961**, *83*, 659-663.
[156] R. M. Mitton-Fry, R. Rasche, A. M. Lawrence-Dörner, J. Eschenbach, A. Tekath, A. Rentmeister, D. Kümmel, N. V. Cornelissen, “Structure-guided engineering of a polyphosphate kinase 2 class III from an bacterium to produce base-modified purine nucleotides”, *Rsc Chem Biol* **2025**, *6*, 1328-1335.
[157] R. M. Mitton-Fry, J. Eschenbach, H. Schepers, R. Rasche, M. Erguven, D. Kuemmel, A. Rentmeister, N. V. Cornelissen, “Chemo-enzymatic production of base-modified ATP analogues for polyadenylation of RNA”, *Chemical Science* **2024**, *15*, 13068-13073.

# 11. Appendix

## 11.1 Abbreviations

| | |
|---|---|
| 2-AP | 2-aminopurine |
| 5caC | 5-carboxycytosine |
| 5fC | 5-formylcytosine |
| 5fC-A | 5fC labeled with acetylacetone |
| 5fC-A* | non-cyclized form of 5fC-A |
| 5fC-M | 5fC labeled with malononitrile |
| 5fC-N | 5fC labeled with nitroacetonitrile |
| 5fC-N* | non-cyclized form of 5fC-N |
| 5hmC | 5-hydroxymethylcytosine |
| 5hmU | 5-hydroxymethyluracil |
| 5mC | 5-methylcytosine |
| A | adenine |
| ABA | 2-aminobenzaldehyde |
| ABA-A | ABA labeled with acetylacetone |
| Ac | acetyl |
| ACN | acetonitrile |
| APOBEC3A | apolipoprotein B mRNA-editing catalytic polypeptide-like 3A |
| APOBECA3A | apolipoprotein B mRNA-editing catalytic polypeptide-like A3A |
| AceF-seq | acetylacetone-induced 5-formylcytosine-to-thymine conversion sequencing |
| BER | base excision repair |
| bp | base pair |
| BHQ | black hole quencher |
| BS | bisulfite sequencing |
| C | cytosine |
| CLEVER-seq | chemical-labeling-enabled C-to-T conversion sequencing |
| COSY | correlated spectroscopy |
| ctDNA | circulating tumor-derived DNA |
| DCM | dichloromethane |
| DDP | 3,7-dideaza purine |
| ddNTPs | dideoxyribonucleoside triphosphates |
| DHU | dihydrouracil |
| DMSO | dimethylsulfoxide |
| DNA | deoxyribonucleic acid |

| | |
|---|---|
| DNMT | DNA methyltransferases |
| dNTPs | deoxyribonucleoside triphosphates |
| DTT | dithiothreitol |
| EDTA | ethylenediaminetetraacetic acid |
| eq. | equivalents |
| ESC | embryonic stem cells |
| EtOH | ethanol |
| FAM | 6-carboxyfluorescein |
| G | guanine |
| HFIP | 1,1,1,3,3,3-hexafluoro-2-propanol |
| HMBC | Heteronuclear multiple bond correlation |
| HPLC | high performance liquid chromatography |
| HRMS | high resolution mass spectrometry |
| HSQC | heteronuclear single quantum coherence |
| LC/MS | liquid chromatography/mass spectrometry |
| MeOH | methanol |
| mRNA | messenger RNA |
| NAN | nitroacetonitrile |
| NGS | next generation sequencing |
| NMR | nuclear magnetic resonance |
| PCR | polymerase chain reaction |
| PE | petroleum ether (b.p. = 40-60 °C) |
| Pfb | pentafluorobenzoyl |
| ppm | parts per million |
| RNA | ribonucleic acid |
| SAM | *S*-adenosylmethionine |
| T | thymine |
| TDG | thymine DNA glycosylase |
| TET | ten-eleven translocation oxygenase |
| TEA | triethylamine |
| TFA | trifluoroacetic acid |
| TLC | thin layer chromatography |
| TRIS | tris(hydroxymethyl)aminomethane |
| tRNA | transfer RNA |
| U | uracil |
| UV | ultraviolet |

# Eidesstattliche Versicherung (Affidavit)

Name, Vorname
(Surname, first name)

Matrikel-Nr.
(Enrolment number)

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden, § 63 Abs. 5 Hochschulgesetz NRW.

Die Abgabe einer falschen Versicherung an Eides statt ist strafbar.

Wer vorsätzlich eine falsche Versicherung an Eides statt abgibt, kann mit einer Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft werden, § 156 StGB. Die fahrlässige Abgabe einer falschen Versicherung an Eides statt kann mit einer Freiheitsstrafe bis zu einem Jahr oder Geldstrafe bestraft werden, § 161 StGB.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offence can be punished with a fine of up to EUR 50,000.00. The competent administrative authority for the pursuit and prosecution of offences of this type is the chancellor of the TU Dortmund University. In the case of multiple or other serious attempts at deception, the candidate can also be unenrolled, Section 63, paragraph 5 of the Universities Act of North Rhine-Westphalia.

The submission of a false affidavit is punishable.

Any person who intentionally submits a false affidavit can be punished with a prison sentence of up to three years or a fine, Section 156 of the Criminal Code. The negligent submission of a false affidavit can be punished with a prison sentence of up to one year or a fine, Section 161 of the Criminal Code.

I have taken note of the above official notification.

Ort, Datum
(Place, date)

Unterschrift
(Signature)

Titel der Dissertation:
(Title of the thesis):

Ich versichere hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel selbstständig und ohne unzulässige fremde Hilfe angefertigt habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht.
Die Arbeit hat in gegenwärtiger oder in einer anderen Fassung weder der TU Dortmund noch einer anderen Hochschule im Zusammenhang mit einer staatlichen oder akademischen Prüfung vorgelegen.

I hereby swear that I have completed the present dissertation independently and without inadmissible external support. I have not used any sources or tools other than those indicated and have identified literal and analogous quotations.

The thesis in its current version or another version has not been presented to the TU Dortmund University or another university in connection with a state or academic examination.*

*Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the PhD thesis is the official and legally binding version.

Ort, Datum
(Place, date)

Unterschrift
(Signature)