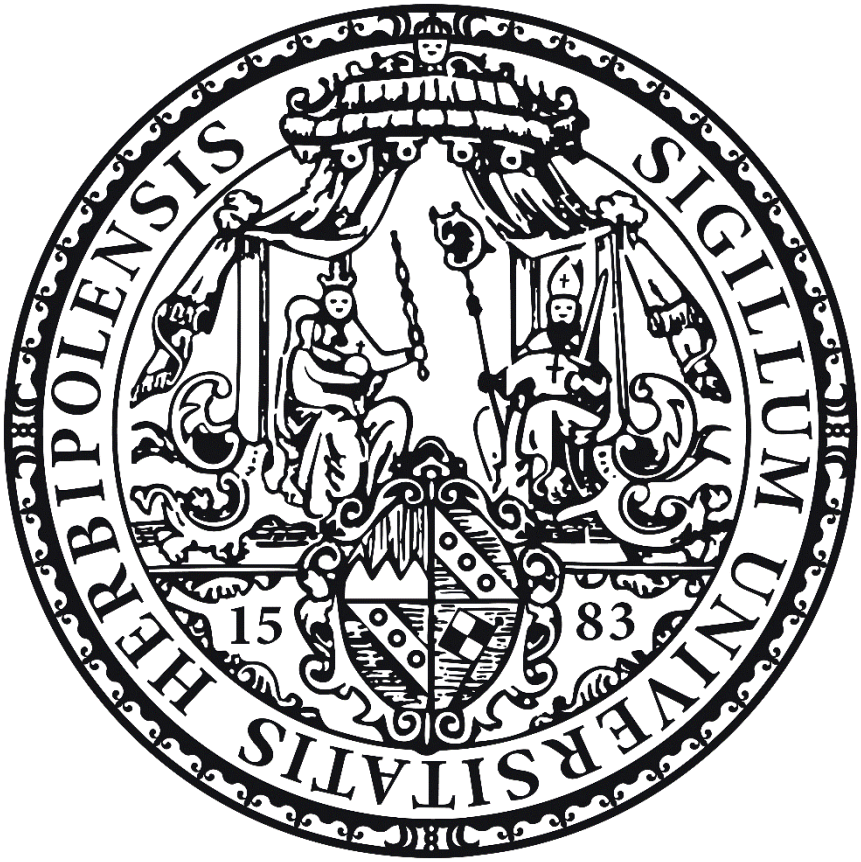
# Synthesis and applications of modified nucleosides and RNA nucleotides




Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrades der Julius-Maximilians-Universität Würzburg

vorgelegt von

Carina Stiller

aus Schweinfurt

Würzburg 2023

Eingereicht bei der Fakultät für Chemie und Pharmazie am

09.02.2023

Gutachter der schriftlichen Arbeit

1. Gutachter: Prof. Dr. Claudia Höbartner
2. Gutachter: Prof. Dr. Utz Fischer

Prüfer des öffentlichen Promotionskolloquiums

1.Prüfer: Prof. Dr. Claudia

2. Prüfer: Prof. Dr. Utz Fischer

3. Prüfer: Prof. Dr. Matthias Lehmann

Datum des öffentlichen Promotionskolloquiums

03.04.2023

Doktorurkunde ausgehändigt am

_____

*Für mein Patenkind Viviane*

*und*

*meinen Onkel Gerhard*

Mein besonderer Dank gebührt meiner Doktormutter Prof. Dr. Claudia Höbartner für die Möglichkeit meine Doktorarbeit in Ihrer Arbeitsgruppe (als erster Neuzugang in Würzburg) machen und dabei viele interessante Themen bearbeiten zu dürfen. Deine Tür stand jederzeit offen, um kleiner oder größere Probleme zu besprechen, und du hast mir mit vielen hilfreichen Tipps und Diskussionen oft sehr weitergeholfen. Ich habe sehr viel gelernt in deiner Arbeitsgruppe, sowohl wissenschaftlich als auch persönlich, und ich weiß deine Unterstützung sehr zu schätzen!

Weiterhin danke ich Herrn Prof. Dr. Utz Fischer für die Übernahme des Zweitgutachtens.

Des Weiteren bedanke ich mich bei meinen Kooperationspartnern Prof. Dr. Markus Bohnsack, Dr. Katherine Bohnsack, Prof. Dr. Patrick Cramer, Prof. Dr. Christian Häring, Nicole Kleiber, Fabian Kabinger und Julia Locherer für die spannenden Diskussionen und die Zusammenarbeit in den verschiedenen Projekten! Die gute Teamarbeit hat so manche Erkenntnisse hervorgebracht, die sich in den damit verbundenen Publikationen widerspiegeln, wofür ich sehr dankbar bin.

Ich möchte mich besonders bei Dr. Irene Bessi für die zahlreichen Messungen und Auswertungen der Oligo-NMR-Proben bedanken. Danke, dass du dir immer die Zeit genommen hast meine Proben zu messen und mir die Theorie hinter den Messungen zu erklären. Ohne dich hätte ich das nicht hinbekommen. Ich weiß deine Unterstützung sehr zu schätzen – du machst eine großartige Arbeit!

In diesem Zusammenhang möchte ich mich bei allen Mitarbeiten der NMR- und Massenabteilung bedanken, vor allem bei Sebastian Mayer und Michaela Schraut. Danke, dass ihr euch auch oft spontan die Zeit genommen habt, wenn es dringende Proben zum Messen gab. Ich weiß, Ihr hattet immer viel zu tun, daher weiß ich es sehr zu schätzen. Ich danke auch Dr. Matthias Grüne, Stefanie Schmitt, Patricia Altenberger und Dr. Juliane Adelmann für Ihre Unterstützung.

Außerdem möchte ich mich bei den Mitarbeitern der ‚Chemobar' bedanken, vor allem bei Joachim Wlaka. Danke für die lustigen Gespräche und die Unterstützung im Praktikum und die Kaffeelieferung während eines langen Praktikumstages – Schichtdienst sei Dank.

Mein Dank geht auch an Markus Braun, Bernd Brunner und Michael Ramold für die technische Unterstützung.

Für ihre Beteiligung an meinen Projekten bedanke ich mich auch bei meinen Bacheloranden und Praktikanten Lia Silbermann, Max Passargus (Danke für den Spargel), Joffrey Cadieu, Eva Schaller, Maria-Victoria Benicke Rojas, Till Schroeder und Timo Pfeufer, sowie meinen Auszubildenden Kirsten Schöpf, Hendrick Hebling und Celine Pfeuffer. Danke für eure engagierte Mitarbeit und die gute gemeinsame Zeit!

Ein großer Dank gilt auch allen ‚Leidensgenossen', die mir die Zeit während der Praktikumsbetreuung erträglicher gemacht haben. Vor allem Arthur Turkin, der ein sehr guter Freund geworden ist - danke für die kontinuierliche Nimm-2-Lieferung auch wenn du gar nicht im Praktikum eingeteilt warst. Daher sei es dir auch verziehen, dass du meinen Namen einfach nie richtig aussprichst. Außerdem Paul, Joshua, Hermann und Julia.

Mein Dank gilt auch meiner Freundin Sandra Gunesch für ihre moralische Unterstützung nicht nur während der Doktorarbeit, sondern auch außerhalb der Universität!

Ich bedanke mich bei allen Mitgliedern der Arbeitsgruppe Höbartner für die tolle gemeinsame Zeit, die konstruktiven Gespräche bei Problemen und die außeruniversitären Aktivitäten. Dabei möchte ich besonders meine beiden Bürokollegen Julia und Hermann hervorheben. Ich danke euch für die überaus amüsante Zeit, die wir zusammen im Büro verbracht haben. Danke auch an Florian Seitz (Flori-Pori) für die lustigen Gespräche und die fachlichen Diskussionen. Man konnte sich immer auf euch verlassen! Danke an meine Mittagscrew Julia, Manuela, Ann-Kathrin und Celine für die gemeinsamen Mittagspausen. Ich werde unsere lustigen Gespräche während des Mittagessens sehr vermissen. Danke auch Celine für deine tatkräftige Unterstützung im Labor. Danke Manuela, Ann-Kathrin und Doris für euren Einsatz im Laboralltag! Ihr haltet den Laden am Laufen! Thank you Anam for your support and the helpful discussions in the lab. Danke Sebastian für unsere gemeinsame Zeit im Labor. Ich vermisse unsere tollen Gespräche! Zudem Danke ich auch den Neuzugängen im Arbeitskreis Nicolas und Konstantin für die schöne, wenn auch kurze, Zeit und die verantwortungsvolle Übernahme der organisatorischen Aufgaben. Und nochmals Danke Julia für deine außerordentliche Hilfsbereitschaft und deine moralische Unterstützung! Du hattest immer ein offenes Ohr und hast dir Zeit genommen, was wirklich nicht selbstverständlich war! Wir waren ein gutes Team!

Danke Aaron, Julia, Hermann und vor allem Ashwin für das Korrekturlesen meiner Arbeit (oder sollte ich sagen, für das Löschen meiner Kommas?)

Mein größter Dank gilt meinen Eltern Christine und Richard, meinen Geschwistern Sandra und Alexander, meinen Nichten Viviane und Emilie, meinen Neffen Joshua und Benedikt und meiner Schwägerin Maria. Danke für eure Unterstützung, euer Verständnis und die ab und zu benötigte Ablenkung vom stressigen Arbeitsalltag. Danke, dass ihr immer hinter mir steht.

Thank you, Ashwin, for your constant support and love. You were always there for me and cheered me up when I was struggling. I know it has not always been easy! Thinking about how we met it is really *funny how sometimes you just find things*.

Parts of this thesis have been published and will be used and/or adapted with the allowance of the following journals (*shared first author):

- N. Kleiber*, N. Lemus-Diaz*, C. Stiller*, M. Heinrichs, M. Mong-Quyen Mai, P. Hackert, R. Richter-Dennerlein, C. Höbartner, K. E. Bohnsack, M. T. Bohnsack, The RNA methyltransferase METTL8 installs $m^3C_{32}$ in mitochondrial $tRNAs^{Thr/Ser(UCN)}$ to optimise tRNA structure and mitochondrial translation
*Nat. Commun.* **2022**, *13*, 209.
- F. Kabinger*, C. Stiller*, J. Schmitzová*, C. Dienemann, G. Kokic, H.S. Hillen, C. Höbartner, P. Cramer, Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis, *Nat. Struct. Mol. Biol.* **2021**, *28*, 740-746.
- A. Liaqat, M.V. Sednev, C. Stiller, C. Höbartner, RNA-cleaving deoxyribozymes differentiate methylated cytidine isomers in RNA, *Angew. Chem. Int. Ed.* **2021**, *60*, 19058–19062.
- A. Liaqat, C. Stiller, M. Michel, M.V. Sednev, C. Höbartner, $N^6$-isopentenyladenosine in RNA determines the cleavage site of endonuclease deoxyribozymes, *Angew. Chem. Int. Ed.* **2020**, *59*, 18627-18631.

*It always seems impossible until it's done.*

*-Nelson Mandela*

# Content

Abstract ..... I
Zusammenfassung ..... III
1. Introduction ..... 1
1.1 General Information about DNA and RNA ..... 1
1.2 RNA driving the epigenetic bus[15] ..... 2
1.3 Post-transcriptional modifications as a second layer of information ..... 5
1.3.1 Naturally occurring tRNA modifications ..... 5
1.3.2 Functions of post-transcriptional modifications ..... 7
1.3.3 'Writer' enzymes, interdependencies, and recognition elements ..... 8
1.3.4 'Eraser' enzymes (Dioxygenases) ..... 10
1.3.5 Detection of modifications ..... 13
1.4 Artificial modifications ..... 17
1.4.1 Fluorescent nucleoside analogs ..... 17
1.4.2 Crosslinkers ..... 23
1.4.3 Antiviral nucleoside analogs ..... 28
2. Research Objectives and thesis outline ..... 42
3. Results and discussion ..... 44
3.1 Naturally occurring tRNA modifications ..... 44
3.1.1 Synthesis of phosphoramidites of naturally occurring tRNA modifications ..... 44
3.1.2 Synthesis of oligonucleotides containing natural modified nucleosides ..... 58
3.1.3 Applications and investigations using synthetic modified RNAs ..... 62
3.1.4 Investigation of eraser enzymes: the dioxygenases ALKBH1 and ALKBH3 ..... 77
3.1.5 In vitro selection of RNA-cleaving DNAzymes for the detection of modifications ..... 91
3.2 The NHC antiviral nucleoside analog against SARS-CoV-2 ..... 98
3.2.1 Synthesis of the NHC phosphoramidite building block ..... 98
3.2.2 Incorporation of NHC into RNA oligonucleotides by solid phase synthesis ..... 99
3.2.3 RNA elongation and thermal denaturation experiments ..... 100
3.2.4 Structural basis of NHC-induced mutagenesis ..... 103
3.2.5 Thermal denaturation experiments on the Dickerson Drew Sequence ..... 106
3.2.6 Electrophoretic mobility shift assay ..... 111
3.2.7 NMR analysis of the duplex forming Dickerson Drew Sequences ..... 112
3.2.8 Imino-water exchange rates ..... 118
3.2.9 NMR analysis of the hairpin-forming Dickerson Drew Sequences ..... 119
3.2.10 UV-Vis and CD-spectroscopy of the Dickerson Drew Sequences ..... 120

3.2.11 $^{15}$N labeling of a mutated Dickerson Drew Sequence ..... 122
3.2.12 NMR analysis of the NHC-A containing Dickerson Drew sequences ..... 124
3.2.13 Crystallography ..... 127
3.2.14 NHC in another sequence context ..... 128
3.3 deaza Guanosine Crosslinker for DNA-Protein interactions ..... 134
3.3.1 Synthesis of the c7dG-crosslinker building blocks ..... 136
3.3.2 Incorporation of the c7dG-crosslinker triphosphate into DNA ..... 139
3.4 Fluorescent nucleosides analogues ..... 146
3.4.1 Synthesis of the v$^{8}$m$^{6}$A phosphoramidite and free nucleoside ..... 146
3.4.2 Incorporation of the m$^{6}$v$^{8}$A phosphoramidite into a palindromic sequence ..... 148
3.4.3 UV melting analysis ..... 150
3.4.4 Fluorescence spectroscopy ..... 152
4.Conclusion and outlook ..... 154
5. Experimental part ..... 159
5.1 Chemical Syntheses ..... 159
5.1.1 General Remarks ..... 159
5.1.2 Synthetic procedure ..... 160
5.2 Oligonucleotide Syntheses and characterization ..... 228
5.2.1 General Remarks ..... 228
5.2.2 Automated solid phase synthesis ..... 228
5.2.3 Deprotection and work-up ..... 228
5.2.4 PAGE purification ..... 229
5.2.5 HPLC ..... 230
5.3 Spectroscopy and Crystallography ..... 231
5.3.1 Mass spectrometry ..... 231
5.3.2 NMR spectroscopy ..... 231
5.3.3 UV melting curves ..... 232
5.3.4 CD spectroscopy ..... 234
5.3.5 Fluorescence ..... 234
5.3.6 Crystallography ..... 234
5.4 Enzymatic reactions ..... 235
5.4.1 General Remarks ..... 235
5.4.2 Copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) of 5' alkynes ..... 235
5.4.3 Labeling of non-functionalized 3'-ends with $NaIO_4$ ..... 235
5.4.4 Phosphorylation of 5'-OH ends with T4 PNK ..... 236
5.4.5 Dephosphorylation of 3'-phosphate ends with CIAP ..... 236

5.4.6 Primer Extension ..... 236
5.4.7 PCR ..... 237
5.4.8 Demethylation assay with ALKBH1/ALKBH3 ..... 237
5.4.9 ZipTip$_{C18}$ desalting ..... 237
5.4.10 Splint-Ligation ..... 238
5.4.11 Crosslinker functionalization ..... 238
6. Appendix ..... 239
6.1 Supplementary melting curves ..... 239
6.1.1 Dickerson Drew - duplex ..... 239
6.1.2 Dickerson Drew - hairpin ..... 241
6.1.3 NHC in another sequence context. ..... 241
6.2 Supplementary NMR data ..... 247
6.2.1 Dickerson Drew – duplex ..... 247
6.2.2 Dickerson Drew – hairpin ..... 248
6.2.3 Dickerson Drew – duplex ($^{15}$N) ..... 249
6.2.4 NHC in another sequence context ..... 249
6.2.5 2D-spectra of the Dickerson Drew Sequences ..... 250
6.2.6 Imino-water exchange rates ..... 253
6.2.7 Chemical shift perturbation (CSP) ..... 254
6.3 Crosslinker ..... 254
6.4 Abbreviations ..... 255
References ..... 261

# Abstract

As central components of life, DNA and RNA encode the genetic information. However, RNA performs several functions that exceed the competences stated in the 'central dogma of life'. RNAs undergo extensive post-transcriptional processing like chemical modifications. Among all classes of RNA, tRNAs are the most extensively modified. Their modifications are chemically diverse and vary from simple methylations (e.g. $m^3C$, $m^6A$) to more complex residues, like isopentenyl group (e.g. $i^6A$, hypermodifications: e.g. $ms^2i^6A$) or even amino acids (e.g. $t^6A$). Depending on their location within the overall structure, modifications can have an impact on tRNA stability and structure, as well as affinity for the ribosome and translation efficiency and fidelity. Given the importance of tRNA modifications new tools are needed for their detection and to study their recognition by proteins and enzymatic transformations.

The chemical synthesis of these naturally occurring tRNA modifications as phosphoramidite building blocks is a prerequisite to incorporate the desired modification *via* solid-phase synthesis into oligonucleotides. With the help of the $m^3C$, $(ms^2)i^6A$, and $t^6A$ oligonucleotides, the importance and impact of tRNA modifications was investigated in this thesis. To this end, the role of METTL8 as the methyltransferase responsible for the installation of the methyl group at C32 for mt-$tRNA^{Thr}$ and mt-$tRNA^{Ser(UCN)}$ was resolved. Thereby, the respective adenosine modification on position 37 is essential for the effectiveness of the enzyme. Besides, by means of NMR analysis, CD spectroscopy, thermal denaturation experiments, and native page separation, the impact of $m^3C_{32}$ on the structure of the tRNA ASLs was shown. The modification appeared to fine-tune the tRNA structure to optimize mitochondrial translation. To investigate the regulation of the dynamic modification pathway of $m^3C$, demethylation assays were performed with the modified tRNA-ASLs and the (α-KG)- and Fe(II)-dependent dioxygenase ALKBH1 and ALKHB3. A demethylation activity of ALKBH3 on the mt-tRNAs was observed, even though it has so far only been described as a cytoplasmic enzyme. Whether this is physiologically relevant and ALKBH3 present a mitochondrial localization needs further validation. In addition, ALKBH1 was confirmed to not be able to demethylate $m^3C$ on mt-tRNAs, but indications for a deprenylation and exonuclease activity were found. Furthermore, the aforementioned naturally occurring modifications were utilized to find analytical tools that can determine the modification levels by DNAzymes, which cleave RNA in the presence of a specific modification. Selective DNA enzymes for $i^6A$, as well as the three cytidine isomers $m^3C$, $m^4C$, and $m^5C$ have been identified and characterized.

Besides the naturally occurring tRNA modifications, the investigation on artificially modified nucleosides is also part of this thesis. Nucleosides with specific properties for desired applications can be created by modifying the scaffold of native nucleosides.

During the pandemic, the potential of antiviral nucleoside analogues was highlighted for the treatment of the SARS-CoV-2 infection. For examinations of the potential drug-candidate Molnupiravir, the $N^4$-hydroxycytidine phosphoramidite building block was synthesized and incorporated into several RNA oligonucleotides. A two-step model for the NHC-induced mutagenesis of SARS-CoV-2 was proposed based on RNA elongation, thermal denaturation, and cryo-EM experiments using the modified RNA strands with the recombinant SARS-CoV-2 RNA-dependent RNA polymerase. Two tautomeric forms of NHC enable base pairing with guanosine in the amino and with adenosine in the imino form, leading to error catastrophe after the incorporation into viral RNA. These findings were further corroborated by thermal melting curve analysis and NMR spectroscopy of the NHC-containing Dickerson Drew sequence. In conclusion, the *anti*-amino form in the NHC-G base pair was assigned by NMR analysis using a $^{15}$N-labeld NHC building block incorporated into the Dickerson Drew sequence.

This thesis also addressed the synthesis of a 7-deazaguanosine crosslinker with a masked aldehyde as a diol linker for investigations of DNA-protein interactions. The diol functional group can be unmasked to release the reactive aldehyde, which can specifically form a covalent bond with amino acids Lys or Arg within the protein complex condensin. The incorporation of the synthesized phosphoramidite and triphosphate building blocks were shown and the functionality of the PCR product containing the crosslinker was demonstrated by oxidation and the formation of a covalent bond with a fluorescein label.

The development of assays that detect changes in this methylation pattern of $m^6A$ could provide new insights into important biological processes. In the last project of this thesis, the influence of RNA methylation states on the structural properties of RNA was analyzed and a fluorescent nucleoside analog (8-vinyladenosine) as molecular tools for such assays was developed. Initial experiments with the fluorescent nucleoside analog $N^6$-methyl-8-vinyladenosine ($m^6v^8A$) were performed and revealed a strong fluorescence enhancement of the free $m^6v^8A$ nucleoside by the installation of the vinyl moiety at position 8.

Overall, this thesis contributes to various research topics regarding the application of naturally occurring and artificial nucleoside analogues. Starting with the chemical synthesis of RNA and DNA modifications, this thesis has unveiled several open questions regarding the dynamic (de-)methylation pathway of $m^3C$ and the mechanism of action of molnupiravir through in-depth analysis and provided the basis for further investigations of the protein complex condensin, and a new fluorescent nucleoside analog $m^6v^8A$.

# Zusammenfassung

Als zentrale Bestandteile des Lebens kodieren DNA und RNA die genetische Information. Die RNA erfüllt jedoch noch mehr Funktionen, die über die im 'zentralen Dogma des Lebens' genannten Kompetenzen hinausreichen. RNA Stränge werden posttranskriptionell verändert, wie zum Beispiel durch chemische Modifikationen. tRNAs sind unter allen RNA-Klassen am umfangreichsten und chemisch vielfältigsten modifiziert. Ihre Modifikationen reichen von einfachen Methylierungen (z. B. $m^3C$ oder $m^6A$) bis hin zu komplexeren Resten, wie einer Isopentenyl-Gruppe ($i^6A$, Hypermodifikation: z. B. $ms^2i^6A$) oder sogar Aminosäuren ($t^6A$). Abhängig von ihrer Position innerhalb der tRNA können Modifikationen Einfluss auf die tRNA Stabilität und Struktur, sowie die Affinität zu den Ribosomen und die Translationseffizienz und Genauigkeit, haben. Angesichts dieser Bedeutung von tRNA-Modifikationen werden zum einen Nachweisemethoden zur Detektion von Modifikationen, als auch Werkzeuge zur Untersuchung der Erkennungsmechanismen spezifischer Proteine und deren enzymatischer Funktionalisierung benötigt.

Dabei ist die chemische Synthese dieser natürlichen Modifikationen als Phosphoramidit-Bausteine die Voraussetzung, um die gewünschte Modifikation überhaupt erst über Festphasensynthese in Oligonukleotide einbauen zu können. Mit Hilfe der $m^3C$-, $(ms^2)i^6A$- und $t^6A$-modifizierten Oligonukleotide wurde die Bedeutung dieser tRNA-Modifikationen für die Struktur und Funktionalität des jeweiligen tRNA Anticodon-Loops (ACL) untersucht.

Ein Kapitel dieser Arbeit klärte die tatsächliche Rolle von METTL8 auf. Als Methyltransferase ist das Protein für den Einbau der Methylgruppe an C32 in den mitochondrialen $tRNA^{Thr}$ und $tRNA^{Ser(UCN)}$ verantwortlich, kann dies aber nur bewerkstelligen, wenn zuvor bereits eine entsprechende Adenosin-Modifikation an A37 installiert wurde. Außerdem wurde mittels NMR-Analyse, CD-Spektroskopie, Schmelzkurvenanalysen und Gelelektrophorese der Einfluss von $m^3C_{32}$ auf die Struktur der tRNA Anticodon-Stem-Loops (ASLs) gezeigt. Die Modifikation scheint die tRNA-Struktur anzupassen, um die mitochondriale Translation zu optimieren.

Um herauszufinden, wie die dynamischen Modifikationswege von $m^3C$ reguliert werden, wurden mit den modifizierten tRNA-ASLs und den (α-KG)- und Fe(II)-abhängigen Dioxygenasen ALKBH1 und ALKHB3 Demethylierungsassays durchgeführt. Obwohl ALKBH3 bisher nur als cytoplasmatisches Enzym bekannt war, konnte es mt-tRNAs demethylieren. Inwiefern diese Aktivität physiologisch relevant ist und ob ALKBH3 vielleicht zusätzlich auch eine mitochondriale Lokalisierung aufweist, muss noch weiter untersucht werden. Zudem wurde gezeigt, dass ALKBH1 $m^3C_{32}$-modifizierte mt-tRNAs nicht demethylieren kann, es jedoch Hinweise darauf gibt, dass ALKBH1 zusätzlich zu der bereits beschriebenen Aktivität der Oxidation von $m^5C$ zu

$f^5C$ in mitochondrialer $tRNA^{Met}$ eine noch näher zu untersuchende Deprenylierungs- und Exonuklease-Aktivität besitzt.

Außerdem wurden die zuvor erwähnten natürlichen Modifikationen verwendet, um DNA-Enzyme als analytische Werkzeuge zur Bestimmung des Modifikationsgrades zu finden. Die Enzyme katalysieren die Spaltung von RNA, falls eine spezielle Modifikation vorhanden ist. Es wurden selektive DNA-Enzyme für $i^6A$ sowie die drei Cytidin-Isomere $m^3C$, $m^4C$ und $m^5C$ identifiziert und charakterisiert.

Neben den posttranskriptionalen Modifikationen war auch die Untersuchung künstlich modifizierter Nukleoside ein Teil dieser Arbeit. Das Gerüst nativer Nukleoside kann so modifiziert werden, dass die Nukleoside spezifischen Eigenschaften für die gewünschte Anwendung erhalten.

Während der Pandemie wurde antiviralen Nukleosidanaloga zur Behandlung der SARS-CoV-2-Infektion eine große Bedeutung zugeschrieben. Um den potenziellen Arzneimittelkandidaten Molnupiravir zu untersuchen, wurde *N*$^4$-Hydroxycytidin als Phosphoramidit-Baustein synthetisiert und in mehrere RNA-Oligonukleotide mittels Festphasensynthese eingebaut. Basierend auf den Ergebnissen von RNA-Elongations-, thermischen Denaturierungs- und Cryo-EM-Experimenten, bei denen die modifizierten RNA-Stränge und die rekombinante SARS-CoV-2-RNA-abhängige RNA-Polymerase verwendet wurde, wurde ein zweistufiges Modell für die NHC-induzierte Mutagenese von SARS-CoV-2 postuliert. Dieser Mechanismus wird durch die zwei tautomeren Formen von NHC ermöglicht, wobei die Amino-Form ein Basenpaar mit Guanosin bildet und die Imino-Form mit Adenosin basenpaaren kann. Nach dem Einbau in die virale RNA kommt es zu Mutationen und zur sogenannten Fehlerkatastrophe. Diese Erkenntnisse wurden durch thermische Schmelzkurvenanalyse und NMR-Spektroskopie der NHC-haltigen Dickerson Drew Sequenz ergänzt. Mit Hilfe eines $^{15}N$-markierten NHC-Bausteins, der in die Dickerson Drew Sequenz eingebaut wurde, konnte schließlich die Anti-Amino Form in dem NHC-G Basenpaar durch NMR-Analyse eindeutig nachgewiesen werden.

Ein weiteres Forschungsprojekt dieser Arbeit befasste sich mit der Synthese eines 7-Deazaguanosin-Crosslinkers, welcher einen geschützten Aldehyd als Diol-Linker enthielt. Dieser Crosslinker sollte zur Untersuchung von DNA-Protein-Interaktionen dienen. Der Einbau der synthetisierten Phosphoramidit- und Triphosphat-Bausteine konnte erfolgreich durchgeführt werden und die Funktionalität des PCR-Produktes, welches den Crosslinker enthielt, wurde durch Oxidation und die Bildung einer kovalenten Bindung mit einem Fluorescein-Label demonstriert.

Der letzte Teil dieser Arbeit beschäftigte sich mit der der Entwicklung eines Assays, um Veränderungen im Methylierungslevel von $m^6A$ nachweisen zu können. Dies könnte neue Einblicke in wichtige biologische Prozesse liefern. Daher wurde im letzten Projekt der Einfluss von RNA-Methylierungszuständen auf die strukturellen Eigenschaften von RNA untersucht und dafür ein fluoreszierendes Nukleosidanalog (8-Vinyladenosin) als molekulares Werkzeug entwickelt. Die ersten Experimente mit dem Nukleosidanalog *N*$^6$-Methyl-8-Vinyladenosin ($m^6v^8A$) zeigten einen deutlichen Fluoreszenzanstieg durch den Einbau der Vinyleinheit an Position 8 im Vergleich zu dem nicht fluoreszierenden $m^6A$.

Insgesamt trägt diese Arbeit zu verschiedenen Forschungsthemen bezüglich der Anwendung von natürlich vorkommenden und künstlichen Nukleosidanaloga bei. Ausgehend von der chemischen Synthese von RNA- und DNA-Modifikationen hat diese Arbeit durch eingehende Analysen mehrere offene Fragestellungen zum dynamischen (De-)Methylierungsweg von $m^3C$ und zum Wirkmechanismus von Molnupiravir aufgedeckt und die Grundlage für weitere Untersuchungen des Proteinkomplexes Condensin und eines neuen fluoreszierenden Nukleosidanalogons $m^6v^8A$ geschaffen.

# 1. Introduction

Almost 70 years have passed since one of the most substantial progresses in scientific history was published – the discovery of the deoxyribonucleic acid (DNA) double helical structure, where the nucleobases form base pairs.[1, 2] With this fundamental concept of the genetic information, the way was paved for the development of advanced molecular biology.[3] Another key discovery was subsequently proposed in 1957 by Francis Crick, postulating the central dogma of life. It describes the correlation between the three biopolymers: DNA, ribonucleic acid (RNA), and proteins. The genetic information, which is encoded in the sequence, is getting transferred among nucleic acids and to polypeptides, Thus, the transfer of information is preserved from DNA to RNA to protein.[4, 5]

## 1.1 General Information about DNA and RNA

As the central components of life, DNA and RNA encode the genetic information. They consist of the four canonical nucleosides deoxyadenosine (dA), deoxyguanosine (dG), deoxycytosine (dC) and thymidine (T) for DNA, and the respective ribonucleosides A, C, G and uridine (U) for RNA (Figure 1).[6]

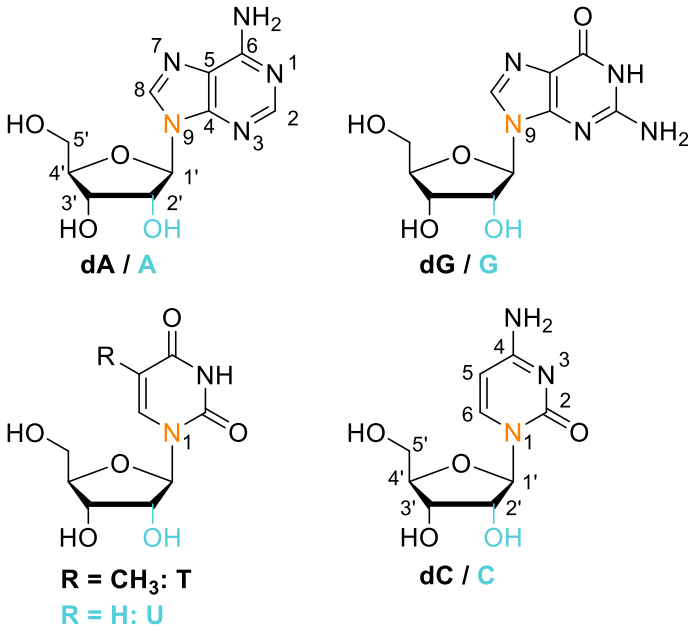



Figure 1. The four canonical DNA and RNA (light blue) nucleosides (d)A, (d)G, (d)T/U and (d)C.

The nucleobases are connected through the N9 of purines (A, G) and N1 of pyrimidines (T, U, C) to the 1'-carbon of a ribose sugar, which contains an additional hydroxy group at the 2'-position in RNA, in contrast to DNA (Figure 1, light blue).[6, 7] These nucleosides are linked to each other over a phosphodiester bond between the 3'-hydroxy and a 5'-phosphate group, forming a negatively charged phosphate backbone in oligonucleotides (Figure 2B).[8] Two of these single strands then form a double helical DNA structure (Figure 2A), as described in 1953 by James Watson, Francis

Crick[2] and Rosalind Franklin.[1] The strands are oriented antiparallelly to each other, with the nucleobases pointing into the inside of the helix, perpendicular to the axis (for B-form helices), and the phosphate backbone towards the outside.[2] The helical rise per nucleotide is 3.4 Å, resulting in a pitch of 34 Å consisting of 10.5 nucleotides for B-DNA.[9] The sugar in a commonly occurring right handed B-DNA (Figure 2A) is in a C2'-endo conformation, whereas RNA adopts a C3'-endo conformation (Figure 2C).[10] Conformational changes of the furanose ring might arise under different hydration conditions or with variations in metal ions.[10, 11] RNA-helices usually take A-form geometry, with 11 nucleotides in a pitch, with a wide and shallow minor groove, but a deep and narrow major groove.[12] Differences between the A-RNA and B-DNA helices can also be noticed from the top view (Figure 2A/C, small extract next to the helix structure). Besides double helical structures, RNA typically forms loops like hairpins, bulges and junctions.[13]

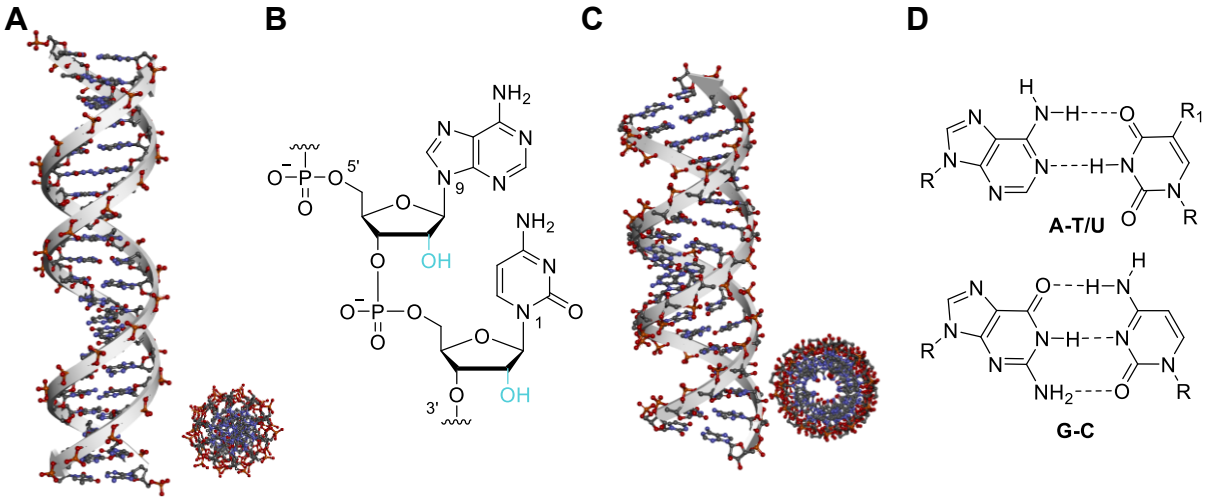



Figure 2. Double helical structures, with the negatively charged phosphate backbone (red/orange) pointing towards the outside, and the nucleobases (blue) into the inside. Between the two strands, Watson-Crick base pairs are formed. A: B-DNA duplex (drawn with Discovery Studio) with the top view on the right B: Excerpt of an RNA strand, showing two nucleotides connected through a 3'-5'-phosphodiester linkage C: A-RNA duplex (drawn with Discovery Studio) with the top view on the right D: A-T/U ($R^1$=$CH_3$/$R^1$=H) and G-C base-pairs with the hydrogen bonds shown as dashed lines.

In addition to stacking and hydrophobic interactions of the nucleobases, which are stabilizing the helical structure, specific hydrogen bonds form between A-T/U and G-C (Watson-Crick base pairs, Figure 2D) of the antiparallel strands.[2, 14] As non-covalent interactions, they provide the basis for genetic processes – the two strands of the DNA duplex can be separated into single strands during replication and translation.[6]

## 1.2 RNA driving the epigenetic bus[15]

DNA, the storage repository for genetic information, gets transcribed into RNA as an accurate copy. This copy is then further translated into proteins. This simplified schematic shown in the 'central dogma of life' does not reflect the complexity of the diverse functions and properties of RNA.[16] RNA performs a plethora of functions than just the transfer of information from DNA to proteins. Less than 1.5% of the genome encodes for proteins,[15, 17] while the majority is getting

transcribed into noncoding RNAs,[17] like transfer RNA (tRNA), ribosomal RNA (rRNA), small nuclear and nucleolar RNA (snRNA, snorRNA), and long noncoding RNAs (lnRNA). The diverse functions of non-coding RNAs range from catalytic activity[16, 18] to regulation of gene activity[19, 20]. Among these, particular attention will be paid onto tRNA in the following sections, which participates in the transfer of information to produce specific proteins based on a corresponding mRNA sequence.

As important as the discovery of the DNA double helical structure for the understanding of the genetic processes was the three-dimensional structure of tRNA, for a better understanding of how its structure is related to its function. Following the first successful attempts to crystalize tRNA in 1968,[21] high resolution X-Ray diffraction analysis revealed the structure of a naturally occurring RNA. Subsequently, the three-dimensional structure of the yeast phenylalanine tRNA ($tRNA^{Phe}$) was solved in 1974 with 3Å resolution.[22, 23] The clover leaf secondary structure is composed of an amino acid stem (Acceptor stem), a TΨC, D, and anticodon stem as well as an extra loop. It is stabilized by additional interactions among the different regions and forms an L-shaped tertiary structure. The TΨC stem is stacking on top of the acceptor-stem double helix and is connected through the D loop and the extra loop as the central part of the structure to the anticodon stem.[23] Thus, the amino acid stem as the amino acid acceptor and the anticodon loop (ACL) are the two ends of the L-shaped structure (Figure 3).[22] An improved tertiary structure of $tRNA^{Phe}$ with a higher resolution of 1.93 Å was published 26 years later, in 2000, with the same overall structure, but more precise details (Figure 3B,C).[24]

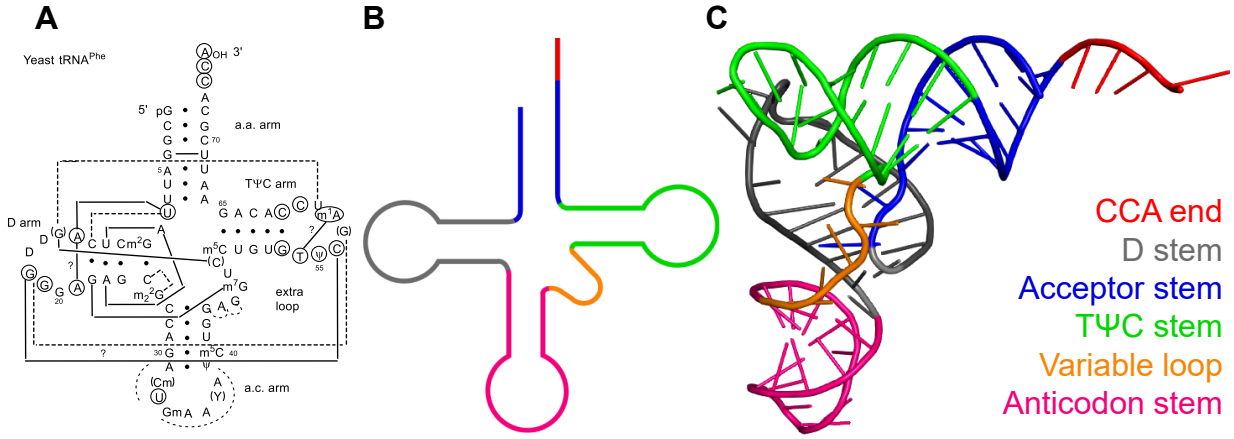



Figure 3. Structure of tRNAs. A: Predicted interactions in the clover leaf secondary structure, which stabilize the tertiary structure, from 1974. Figure reproduced from Ref.[23] B: Schematic clover leaf structure (color code as in C). C: L-shaped tertiary structure of $tRNA^{Phe}$ from yeast. 1EHZ-structure from 2000 (Pymol).[24]

Three nucleotides of the tRNA-anticodon base pair with the corresponding complementary nucleotides of the mRNA codon, which determines the amino acid to be incorporated. Therefore, at the 3'-end (CCA-End, Figure 3B/C, red), the amino acid corresponding to the respective anticodon is attached to the tRNA by the enzyme aminoacyl tRNA-synthetase (aaRS),[25] facilitating the transfer

of the amino acid to the growing peptide chain during the decoding of the mRNA sequence.[23] First, the amino acid is activated with ATP, forming aminoacyl-adenylate, followed by esterification of the 2'- or 3'-OH of the CCA-end. Each of the 20 available enzymes (aaRS) are specific for one amino acid and corresponding (iso-acceptor) tRNAs.[26] The enzymes are divided into two classes depending on their sequence motifs and their function to acylate either the 2'- or 3'- hydroxy group of the CAA-end. Class I consists of 'HIGH' and 'KMSKS' sequence motifs and a catalytic domain based on the Rossman fold (ATP binding domain). The enzymes attach the activated amino acid to the 2'-hydroxy group (subsequent transesterification to 3'-OH)[27].[28] Class II is made up of three sequence motifs (motif 1,2 and 3), with an antiparallel beta strand as active center and acylates the 3'-OH group (PheRS as exception).[28] Extensive interactions between the aaRS and the tRNAs, mainly at the two ends of the L-shaped structure (anticodon loop and acceptor stem),[29] ensure specificity for the correct acylation with the corresponding amino acid.[25, 29]

The aminoacyl-tRNA (aa-tRNA) can then participate in protein biosynthesis, which is divided into initiation (start codon of mRNA), elongation and termination (stop codon of mRNA). During this process, aa-tRNA gets associated with elongation factors (EF-Tu in prokaryotes, EF-1α in eukaryotes[25])[30] and forms the ternary aa-tRNA-EF-Tu-GTP complex in prokaryotes. One special characteristic of EF-Tu is that it has a higher substrate specificity, as it can bind all aa-tRNAs, in contrast to aaRS.[30] Additionally, the elongation factor prevents uncharged or mischarged RNAs from being incorporated.[31, 32] The crystal structure of aa-tRNA-EF-Tu-GTP complex containing yeast $tRNA^{Phe}$ was determined with 2.7Å resolution in 1995.[30] As the anticodon of the tRNA is exposed, it can bind to the codon of the current mRNA within the A-site of the ribosome. Upon formation of codon-anticodon interactions, GTP of the ternary complex gets hydrolyzed and releases EF-Tu-GDP and the amino acid can be transferred to the polypeptide chain.[33] An induced-fit-mechanism was proposed for the discrimination of aa-tRNA within the ribosome.[34] A conformational rearrangement of the decoding center (16S rRNA) has to take place to facilitate the GTPase activation and A-site accommodation of the aa-tRNA. This conformational change occurs faster for cognate than non/near-cognate codon:anticodon interactions, thus accelerating GTP hydrolysis and elongation of the peptide bond – accordingly the binding of the correct aa-tRNA is kinetically favored. This is important for the initial selection of aa-tRNAs, as well as proofreading within the ribosome, to ensure fidelity of translation.[34, 35]

To successfully preserve the flow of information from DNA to the functional protein, another important feature of tRNAs needs to be mentioned: their diverse posttranscriptional modifications. Depending on their position and structure, these modifications create a second layer of information, which extends the limited diversity established by the four canonical nucleobases.[6] This enables tRNA to perform its complex functions while transferring the genetic information. Besides tRNAs,

modifications also occur for other RNA species, and to a lesser extend in DNA. There is a large variety of modifications, which will be discussed in greater detail in the next section.

## 1.3 Post-transcriptional modifications as a second layer of information

Posttranscriptional modifications can exist at different positions of the RNA sequence and with a diverse degree of chemical complexity in all three domains of life (eucarya, bacteria and archaea). They are important components to modulate the functions of different RNA species.

### 1.3.1 Naturally occurring tRNA modifications

More than 100 known naturally occurring RNA modifications have been identified so far. Structural variations are not limited to the nucleobase scaffold but can also occur on the sugar moiety and the phosphate backbone. Chemical alterations range from small groups like methyl, hydroxy or acetyl, over larger residues, like isopentenyl, to amino acid side chains and changed nucleobase scaffolds (= hypermodified modifications, Figure 4).[36] Hypermodifications may require multiple synthesis steps and corresponding enzymes.

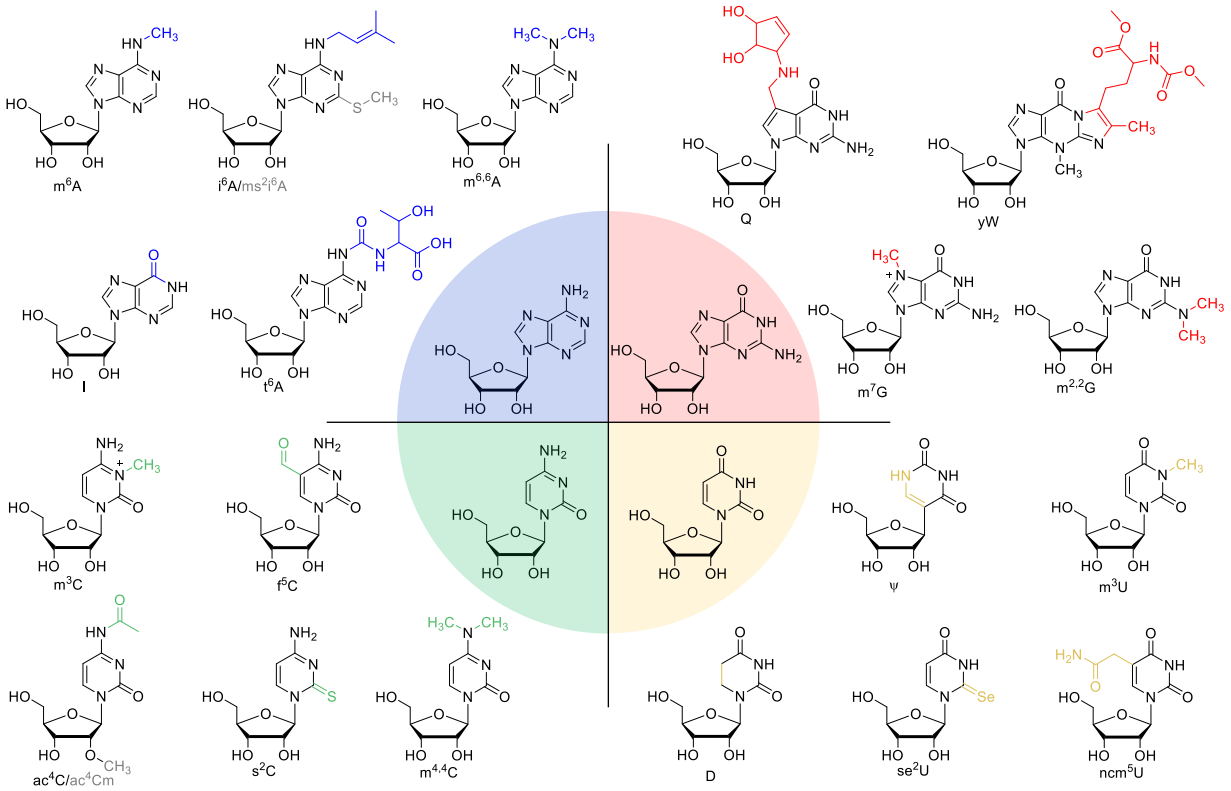



Figure 4. Selection of naturally occurring ribonucleoside modifications. Adenosine modifications are shown in blue, guanosine in red, cytidine in green and uridine in yellow.

The first discovered posttranscriptional RNA modification was pseudo uridine (Ψ) in 1951, an isomer of uridine with a carbon-carbon bond instead of the nitrogen-carbon glycosidic bond.[37, 38] Pseudouridine synthases isomerize uridine by breaking the glycosidic bond. The base rearranges by rotation and forms a new glycosidic bond through C5.[39] Pseudouridine was declared as the fifth

ribonucleotide and turned out to be the most frequently occurring RNA modification.[37, 40] It can be found in multiple RNA species, like tRNA, rRNA, snRNA,[41] and was recently also found in mRNA.[42, 43] In rRNA, Ψ affects translation accuracy and accurate folding.[44] Similarly, Ψ within the anticodon was also proposed to stabilize tRNA structure[45], codon:anticodon interactions[46], and to increase translation accuracy[47] (will be further discussed in chapter 1.3.2), due to improved stacking interactions and more versatile hydrogen bonding properties compared to uridine.[48]

Around 170 RNA modifications in the different coding and non-coding RNA species have been identified so far.[49] Among these, tRNAs are the most extensively modified ones, carrying 111 distinct modifications.[36] A sequencing experiment with tRNAs from the three domains of life revealed that around 12% of tRNA nucleosides are post-transcriptionally modified. This corresponds to an average of 8 modifications per tRNA.[50] Most of them are clustered in the three main loops of a tRNA (D, TΨC and anticodon loop, Figure 5).[36] Especially position 34 and 37 of the anticodon loop are most variably modified in all organisms (Figure 5), often with hypermodifications.[51-53] Simpler modifications like pseudouridine or methylated cytidines ($m^3C$, Cm) occur on position 32 instead.[53]

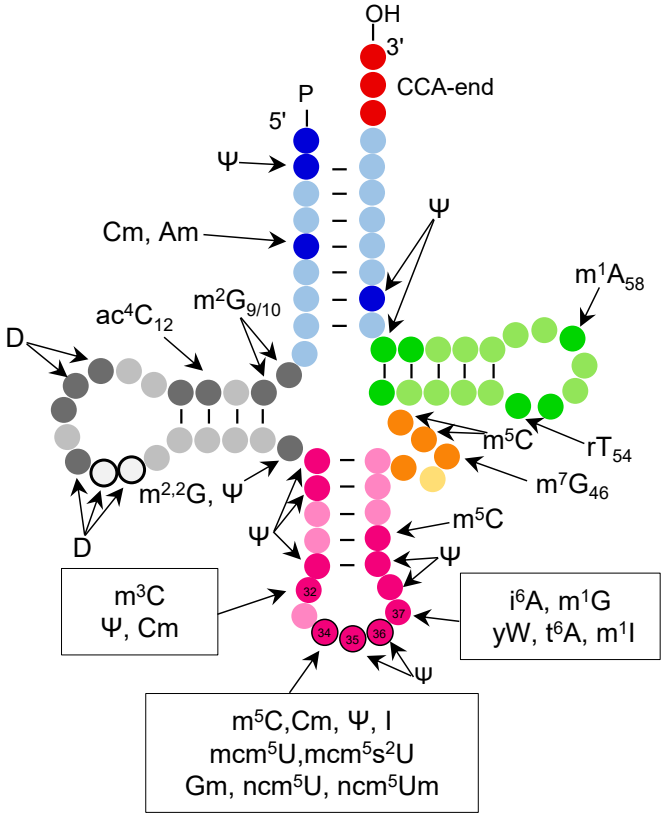



Figure 5. Modifications found on position 32, 34 and 37 of cytoplasmic tRNA (cyt-tRNA, *S. cerevisiae*). Adapted from Ref.[53] with the same color code shown in Figure 3: 3-methylcytidine ($m^3C$), pseudouridine (Ψ), 2'-*O*-methylcytidine (Cm),*N*$^6$-isopentenyl adenosine ($i^6A$), 1-methylguanosine ($m^1G$), wybutosine (yW), *N*$^6$-threonylcarbamoyl adenosine ($t^6A$), 1-methyl inosine ($m^1I$), 5-methylcytidine ($m^5C$), inosine (I), 5-methoxycarbonylmethyluridine ($mcm^5U$), 5-methoxycarbonylmethyl-2-thiouridine ($mcm^5s^2U$), 2'-*O*-methyl guanosine (Gm), 5-carbamoylmethyluridine ($ncm^5U$), 5-carbamoylmethyl-2'-*O*-methyluridine (ncm5Um) – modifications often found on position 32, 34 and 37. Other modifications found in cyt.tRNAs (*S. cerevisiae*) are exemplarily shown. Bold circles are often or always modified in tRNA species, whereas light circles are unmodified in yeast tRNAs. Circles with the black outline in the D loop are additional residues (20a, 20b) in some tRNA species. The anticodon is highlighted with a black contour.[53]

A conserved modification at position 32 of the anticodon loop of tRNAs is $m^3C$, which has only been found in eukaryotes so far.[54] It occurs in cytoplasmic $tRNA^{Ser}$ and $tRNA^{Thr}$ from yeast to humans. In the latter it was also found in mitochondrial-tRNA for $Ser^{(UCN)}$ and Thr.[52, 55-58] Additionally, in higher eukaryotes $m^3C$ also appears in cytoplasmic $tRNA^{Arg}$ at position 32[59, 60], in the variable loop at C47d in $tRNA^{Leu[CAG]}$,[61] C47d in some cytoplasmic $tRNA^{Ser}$ species in addition to C32[61-63], and C20 of $tRNA^{Met[CAU]}$.[63] Besides in tRNA, $m^3C$ was supposed to occur in mRNA.[64, 65] Exemplarily, hypermodifications at position 37 are $i^6A$ and $t^6A$, with an isopentenyl residue at N6 and with the amino acid threonine bound to N6 *via* a carbamoyl linkage, respectively.[66, 67] Both can be further modified with a thiomethyl group at N2, leading to $ms^2i^6a$ and $ms^2t^6a$.[68, 69] The hypermodification $t^6A$ is universally present in almost all tRNAs, which are decoding ANN codons in the three domains of life.[67, 70]

These post-transcriptionally modified nucleosides were shown to either have a direct impact on various functions, stability and structure of tRNA, or an indirect impact, with the help of special reader proteins, and are also related to certain diseases.[71] A selection of these specific interactions for the dynamic regulation of gene expression is outlined below (Figure 6).

### 1.3.2 Functions of post-transcriptional modifications

Modified nucleotides, particularly in the core region of tRNAs (D and T arm), were shown to stabilize the tertiary structure of tRNAs and influence aminoacylation by improving the recognition of ATP and the corresponding amino acid by the synthetase.[51, 72-75] In addition, the specificity of the synthetase to its cognate tRNA can be decreased by the absence of these modifications.[73] Modifications regulate the function of tRNAs and help differentiate between initiator $tRNA^{Met}$ and elongator $tRNA^{Met}$ by ribosylation of position $A_{64}$ inside the T-stem.[76] Modifications in the anticodon-loop can control tRNA cleavage by endonucleases in two ways: They either enhance the cleavage as a prerequisite ($mcm^5s^2U$ at position 34 in yeast $tRNA^{Glu}$)[77] or protect tRNA against stress-induced cleavage (methylations by Dnmt2[78] or NSun2[79]).[80, 81] Consequently, the loss of NSun2-mediated tRNA methylation ($m^5C$) is further linked to neurological disorders in humans.[79]

In the anticodon loop, modifications can generally affect codon-anticodon interactions and modulate misreading by tRNAs[82], expand the decoding capability[83], prevent frameshifting[84, 85] and fine-tune tRNA structure[51] for precise translation of the genetic code[86]. As an example, deamination of adenosine to inosine (Figure 5) at position 34 extends the codon-anticodon recognition, thus expanding the decoding capability, by enabling base pairing with C, A and U.[87, 88] Aside from that, modifications like $t^6A$ or $i^6A$ at position 37 strongly affect the translation efficiency.[89, 90] For instance, $t^6A$ is stabilizing the codon:anticodon interactions. It is forming a cross-strand stack with A1 of the mRNA codon.[91] Similarly, residues at position 32 are fine-tuning decoding accuracy (precise codon-anticodon recognition), due to the stabilization of the anticodon loop by forming interactions

with the residue at position 38.[55, 63, 92] These important regulatory activities can be further fine-tuned by interactions between nucleotides within the anticodon.[52]

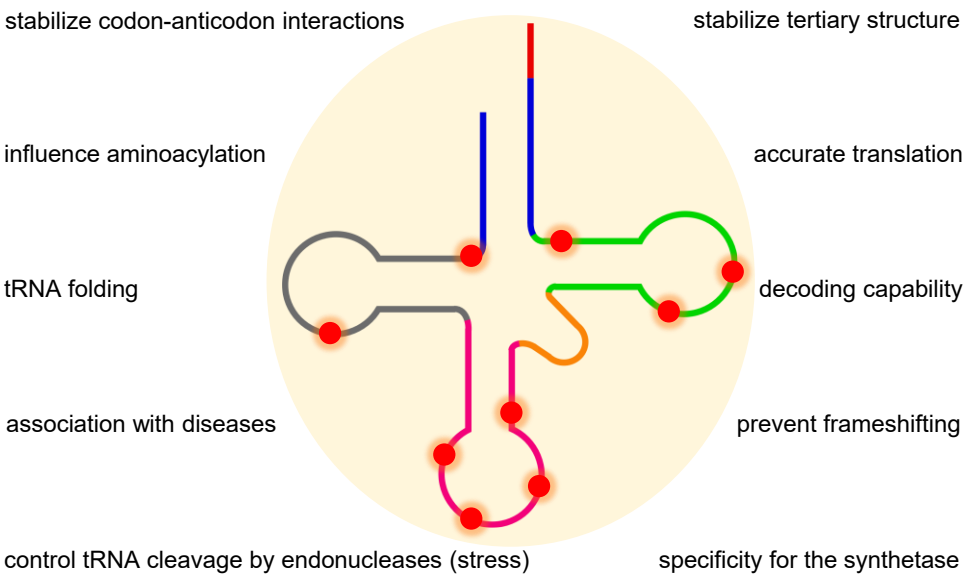



Figure 6. Influence of post-transcriptional modifications on the regulation of translation and association with diseases. Based on Ref.[93]

Most of these ribonucleoside modifications are catalyzed by enzymes,[36] which will be discussed in the following chapter, focusing mainly on $m^3C$ and the interaction with adenosine modifications at position 37, as they were investigated in this thesis. This process is one part of a dynamic interplay of enzymes being able to install ("writer"), remove ("erasers") or bind ("readers") to specific modifications.

### 1.3.3 'Writer' enzymes, interdependencies, and recognition elements

Among all posttranscriptional modifications, methylations are the most abundant ones in tRNAs. RNA methyltransferases with *S*-adenosylmethionine-binding- and methyltransferase-like domains are responsible for the transfer of the functional group.[94, 95] They are commonly referred to as 'writers'. Even though they can markedly differ in their structure, the underlying mechanism for the transfer is supposed to be the same: an acid-base catalysis is performed to transfer the methyl group from *S*-adenosylmethionine (SAM) onto the respective position of the target nucleoside (endo-/exocyclic nitrogen, carbon, ribose oxygen). The basis for substrate recognition is of high interest. Some enzymes do not show strict specifications but perform the transfer of the methyl group as soon as the correct nucleotide is at the desired position. Others are tRNA species-specific, such as the methylation of $m^3C_{32}$ in $tRNA^{Thr}$.[94]

The first methyltransferase for the methylation of cytidine at position 32 was identified in 2011 in yeast *Saccharomyces cerevisiae (S. cerevisiae*) using two approaches: A primer extension assay of tRNAs isolated from methyltransferase knockout (-Δ) strains[60] and a reverse genetic approach

in combination with mass spectrometry (= ribonucleome analysis)[55]. Since the presence of the methyl group at N3 would block primer extension, the known $m^3C$-containing $tRNA^{Thr[IGU]}$ (yeast) substrate revealed that Trm140 (renamed ABP140) is the responsible gene for methylation of $C_{32}$. It is specific for tRNAs encoding Thr and Ser.[60] These findings were consistent with the results obtained from ribonucleome analysis. Furthermore, a coherence between the lack of the modification and impaired translation was predicted, as Trm140-$\Delta$ trm1-$\Delta$ mutants without $m^3C$ and $m^{2,2}G$ showed a growth defect with a translation inhibitor being present.[55] In fission yeast (*Schizosaccharomyces pombe* (*S. pombe*)), methylation of C32 depends on two Trm140 homologs: cytoplasmic $tRNAs^{Thr}$ are substrates of $Trm140^+$, whereas $tRNAs^{Ser}$ are substrates of $Trm141^+$. Homologs of these enzymes were found in higher eukaryotes.[52]

Three methyltransferase-like (METTL) enzymes are responsible for catalyzing RNA methylation in humans. METTL2 and 6 were supposed to mediate modification of $tRNA^{Ser}$ or $tRNA^{Thr}$ and $tRNA^{Arg[CCU/UCU]}$, respectively, whereas METTL8 was expected to install the methyl group for $m^3C$ in mRNA.[64, 96, 97] However, the latter could not be proven by any transcriptome-wide mapping attempts, yet.[57, 98] METTL2 comprises of two variants, namely METLL2A and 2B. The latter appears to be the true homolog.[52] Accordingly, METTL2B was proposed to be the homolog of yeast Trm140, whereas METTL6 the homolog of Trm141. METTL8 was supposed to represent a third homolog.[52, 64, 99] Just recently, it was demonstrated that METTL8 catalyzes the methylation of C32 in mitochondrial $tRNA^{Thr}$ and $tRNA^{Ser}$, which will be discussed in detail later in the thesis.[58] Besides this, in 2020, METTL2A and METTL2B were shown to form a complex with the protein DALR anticodon-binding domain-containing protein 3 (DALRD3) in mammals to methylate C32 in the tRNA-isoacceptors $tRNA^{Arg[CCU]}$ and $tRNA^{Arg[UCU]}$.[97]

The substrate recognition of these modifying enzymes is an important aspect for a better understanding of their functionality. Different recognition elements have been discussed in literature. Some might depend on a unique anticodon-loop[100], or a mutually conserved part of the tRNA[101] as determinant. For some other enzymes, the recognition elements were not observed yet. In yeast, methyltransferase Trm140 (*S. cerevisiae*) shows two distinct recognition modes for the $m^3C_{32}$ synthesis:[99] $tRNA^{Thr}$ substrates are identified by the $G_{35}$-$U_{36}$-$t^6A_{37}$ motif within the anticodon loop. Thereby the modification at position 37 seems to be important, but not obligatory for the methylation of $C_{32}$. In contrast, the specific $tRNA^{Ser}$ substrate recognition depends on seryl-tRNA synthetase (SARS2) and a distinctive V-loop, as well as $t^6A$ and $i^6A$-modifications on position 37.[99]

Within this dynamic network of tRNA modifications and their corresponding enzymes, it has been demonstrated that some modifications can either have a favorable effect on other modifications, or be a prerequisite for their formation.[102] In 2016, a study revealed a C32-A37 ACL circuitry in which $i^6A_{37}$ was a prerequisite for the methylation of cytidine at position 32 in a subset of $tRNA^{Ser}$

with A at position 36 (*S. pombe*).[52] Another study addressed A-to-I editing at position 34 of cytoplasmic $tRNA^{Thr[AGU]}$ (*Trypanosoma brucei* (*T. brucei*)) by hydrolytic deamination of adenosine and found a correlation of two different editing events. The actual editing site at position 34 is stimulated by the prior C-to-U editing at position 32.[103] In fact, it was further shown that methylation of position 32 is a prerequisite for C-to-U editing in the first place.[104] These two examples show the interconnection of modifications as well as the interrelation of processes within the anticodon of tRNAs for the efficient synthesis of anticodon loop modifications.

Some changes within the genes or expression levels of the corresponding enzymes are directly linked to diseases.[105] For example, metll6-knockout mice studies revealed that METTL6 regulates pluripotency and tumor cell growth: depletion of mettl6 leads to impaired pluripotency and reduced cancer cell proliferation.[96] A transition mutation of the DALRD3 gene was found to be the reason for developmental delay and infantile epilepsy.[97] Further diseases that are prone to associate with defects in tRNA modification states are neurological disorders, mitochondrial pathologies, diabetics (T2D), cardiomyopathy (HSD10), lactis acidosis and encephalopathy.[106-109] These coherencies demonstrate the importance of methyltransferases within the cells regarding human pathologies.[105]

Counterparts of the transferase enzymes ("writers") are the enzymes that enable the removal of functional groups from the corresponding modification ("erasers"). This cellular machinery reflects a dynamic RNA modifications circuit that will be discussed in the following section.

### 1.3.4 'Eraser' enzymes (Dioxygenases)

The level of modifications can be further adjusted by removing functional groups, thereby giving the opportunity for regulatory functions. Removal of methyl groups is performed by enzymes called demethylases. In bacteria (*Escherichia coli* (*E.coli*)[110]) AlkB is a DNA repair enzyme, which is responsible for the repair of cytotoxic and mutagenic $m^1A$ and $m^3C$ modifications, resulting from methylating agents in single-stranded DNA (ssDNA). AlkB is an α-ketoglutarate (α-KG)- and Fe(II)-dependent dioxygenase that uses molecular oxygen for the oxidation of alkyl groups.[110-113] Homologues of AlkB can also be found in eukaryotes. The first reversible RNA demethylation was reported in 2011, namely by the fat mass and obesity-associated (FTO) protein. Due to the sequence similarity, FTO can be classified to the non-heme Fe(II)- and 2-oxoglutarate-dependent oxygenases as a homolog of the AlkB family.[114-116] FTO catalyzes the removal of the methyl group from $m^6A$ in RNA by oxidative demethylation (Figure 7).[115]

The mechanism of FTO was shown to include two intermediates for the oxidative demethylation of $m^6A$: *N*$^6$-hydroxymethyladenosine ($hm^6A$ shown in blue, Figure 7) and the further oxidized species *N*$^6$-formyladenosine ($f^6A$ shown in orange, Figure 7). The relatively long half-life times of the intermediates facilitate their detection by LC-MS/MS. They are even present in isolated mRNAs from mammalian cells.[117] It was later shown that FTO also comprises $m^6A_m$, bearing a 2'-OMe group,

within the 5'-cap as a substrate, forming $A_m$.[118] The dynamic methylation status at this position influences the mRNA stability. Besides FTO, ALKBH5 was found to be another demethylase performing oxidative demethylation of $m^6A$ in mammalian mRNA.[119] Contrary to the oxidative mechanism shown before, ALKBH5 is proposed to form $hm^6A$ as a transient intermediate, which directly decomposes to formaldehyde and adenosine.[120, 121] Both are well studied enzymes, also regarding their substrate affiliation.

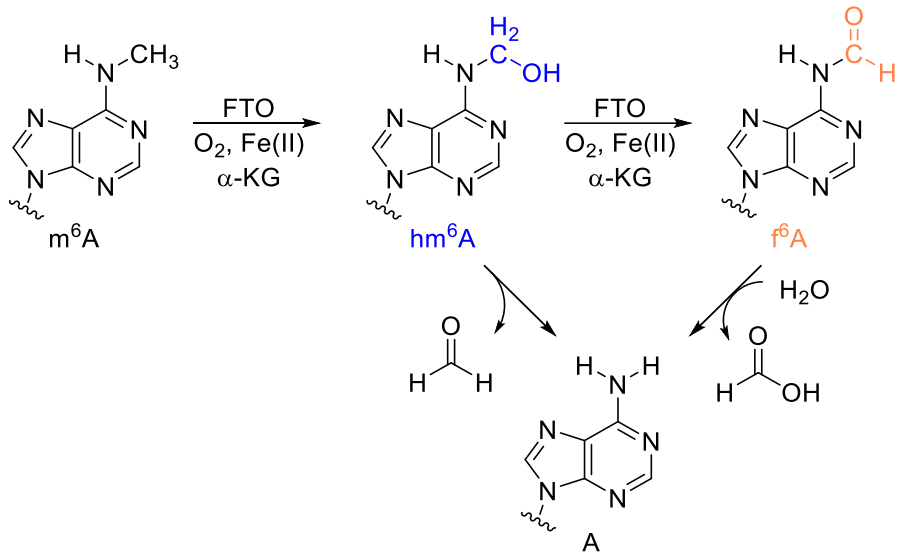



Figure 7. Two-step oxidation mechanism of $m^6A$ with the two formed intermediates $hm^6A$ (blue) and $f^6A$ (orange). Based on Ref.[117]

So far, nine mammalian homologues have been identified: ALKBH1-8 and FTO. Only two of them, namely ALKBH2 and ALKBH3, are able to perform the DNA repair mechanism similar to *E.coli* AlkB to protect the cell from alkylation damage. Both ALKBH2 and ALKBH3 are demethylating $m^1A$ and $m^3C$ (and to a small amount 1-ethyladenosine ($et^1A$)[113]), thus regenerating the canonical nucleobases at DNA lesions.[112, 113] ALKBH2 showed a slightly higher reactivity for $m^1A$, whereas ALKBH3 did for $m^3C$.[113] In accordance to AlkB in *E.coli*, human ALKBH3 prefers single-stranded DNA (ssDNA) and also demethylates $m^1A$ ($mRNA^{65}$) and $m^3C$ from RNA ($tRNA^{122}$) (Figure 8).[123] Contrary, ALKBH2 shows a higher efficiency for double-stranded DNA (dsDNA) and does not recognize RNA modifications as a substrate. In the subcellular localization, despite both being nuclear enzymes, differences occur between these two enzymes. ALKBH2 has been found in nucleoplasm, nucleoli, and replication foci, depending on the stage of the cell cycle. In contrast, ALKBH3 was predominantly found in nucleoplasm and also in cytoplasm, but is excluded from nucleoli.[123] Besides $m^1A$ and $m^3C$, $m^6A$ was also supposed to be a substrate of ALKBH3 in mammalian tRNAs.[124]

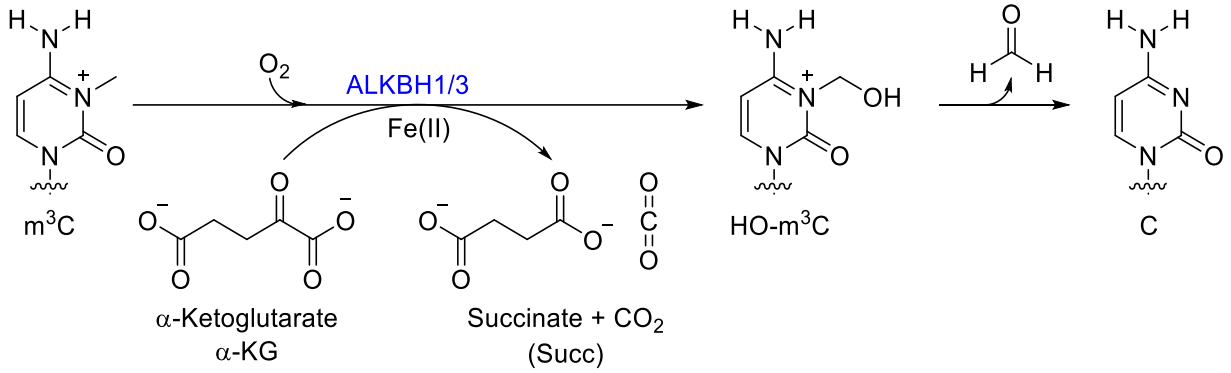



Figure 8. Mechanism for the demethylation of $m^3C$ in RNA with the responsible enzymes ALKBH1 and ALKBH3 leading to the canonical nucleobase C after the release of formaldehyde.[112]

Another enzyme with the ability to demethylate $m^3C$ is ALKBH1. Westbye *et al.* demonstrated that ALKBH1 repairs $m^3C$ in ssDNA and RNA (Figure 8). It was shown to be a mitochondrial protein with the highest sequence homology to AlkB,[125, 126] which exhibits a lower repair activity compared to AlkB, hALKBH2 and hALKBH3.[126] However, contradictory evidences about the cellular localization of the enzyme were reported. A later study from Larsen and coworkers revealed that ALKBH1 is predominantly localized in the nucleus,[127] explaining the diverse functions of ALKHB1 that were gradually reported: a histone dioxygenase activity acting specifically on histone H2A (demethylation),[127] and exhibiting abasic or apurinic/apyrimidinic (AP) lyase activity.[128] In addition, ALKBH1 performs biogenesis of $m^5C_m$ to 5-hydroxymethyl-2'-*O*-methylcytidine ($hm^5C_m$) and $f^5C_m$ at position 34 in cytoplasmic $tRNA^{Leu}$,[129] as well as $f^5C$ of mt-$tRNA^{Met}$ to regulate mitochondrial translation.[125, 129] Furthermore, the enzyme demethylates $m^1A$ in mt-tRNA,[129] $m^6A$ in DNA which plays a role in human tumorigenesis,[130] and $m^3C$ from mRNA in mammalian cells.[65] In general, these enzymes were often shown to be related to cancer cell progression and proliferation depending on their expression level, which makes them promising targets for the treatment of cancer.[65, 122, 124, 131]

Another homologue, ALKBH8, additionally contains a C-terminal methyltransferase domain attached to the AlkB domain and an N-terminal RNA recognition motif (RRM). In mammalian tRNAs, 5-carboxymethyluridine ($cm^5U$) at position 34 in the anticodon loop gets methylated to 5-methoxycarbonylmethyluridine ($mcm^5U$) and further modified by hydroxylation, generating (*R*)- and (*S*)-5-methoxycarbonylhydroxymethyluridine ($mchm^5U$) in $tRNA^{Gly(UCC)}$ caused by the responsible domain.[132, 133] ALKBH8 was only found in cytoplasm.[132]

Contrary to the aforementioned demethylases, ALKBH4 is active on proteins. The enzyme catalyzes the demethylation of the monomethylated lysine (Lys) residue (K84me1) in cytoplasmic actin, thus regulating the actin-myosin interaction.[134]

The remaining two enzymes, ALKBH6 and ALKBH7, are sparsely studied. ALKBH6 was reported to promote repair of DNA damages in pancreatic cancer cells,[135] and the possibility as an $m^6A$-

and $m^5C$-demethylase was described.[136] ALKBH7, located in mitochondria, was shown to demethylate $N^2$,$N^2$-dimethylguanosine ($m_2^2G$, Ile) and $m^1A$ (Leu1) in pre-tRNAs, regulating mitochondrial RNA processing and activity.[137] Contrary to the demethylases shown earlier, which protect the cells after the appearance of DNA lesions, ALKBH7 was shown to facilitate controlled cell death (programmed necrosis) as a result of DNA damage, and mitochondrial dysfunction. This might enable the use of ALKBH7 as a therapeutic for tumor treatment.[138]

Hence, there are still many open questions that need to be explored for these human enzymes, which are involved in important cellular functions for RNA metabolism: What is their responsibility? Where are they exactly located? What is the effect the enzyme level has onto certain cellular aspects and what are the substrate specificity and recognition elements of demethylases in RNAs, especially tRNAs? Therefore, in almost the same manner, it is a requirement to know where naturally occurring modifications are located within an RNA sequence, providing a highly progressing research field with already more than 150 identified posttranscriptional modifications.

### 1.3.5 Detection of modifications

Since naturally occurring modifications are generally less abundant than canonical nucleosides and some functional groups are rather small, their precise detection and structural analysis is quite challenging. In theory, detection methods can be based on different approaches.[139, 140] Chromatography allows for good separation, especially if a great structural variety of modifications like wybutosine features different physicochemical properties among each other, and compared to the canonical nucleosides. For smaller functional groups, digestion of the RNA strand prior to chromatography, capillary electrophoresis, or mass-spectrometry, facilitates their discrimination (e.g. LC-MS/MS). Moreover, organic chemistry can sometimes be used to spot modifications, as some modified nucleosides are forming, for instance, abasic sites directed by the reaction with specific chemicals (e.g. aniline[57]). Alternatively, bulky residues could be attached to interfere with reverse transcription (RT).[139, 140] This correlates with the last method, which relies on differential enzymatic turnovers, since the enzymatic efficiency changes between modified and unmodified substrates (e.g. during RT[104]).[139] The latter two methods (organic chemistry and enzymatic turnover) can be linked to high-throughput sequencing methods.[140] Continuous effort for new detection methods is ongoing, such as antibody-based immunoprecipitation or exploring DNA as a detection tool in form of DNA enzymes and this detection-field will further grow, owing to its importance. Some of these methods will be discussed in greater detail in this section.

#### 1.3.5.1 Hydrazine-Aniline Cleavage Sequencing (HAC-seq)

Various approaches have been described for mapping $m^3C$ modifications. For example, primer extension assays can identify modification sites, since the methyl group at N3 blocks RT.[104] Alk-

Aniline-Sequencing can identify 7-methylguanosine ($m^7G$) and $m^3C$ with single nucleotide resolution.[57] Similarly, with demethylase-thermostable group II intron RT tRNA sequencing (DM-tRNA-seq) $m^1A$, $m^3C$ and $m^1G$ modifications can be analyzed at single-nucleotide resolution, by combining demethylation with reverse transcription followed by Illumina sequencing.[141] Further, LC-MS/MS analysis is possible after oligonucleotide digestion. However, all of these exhibit various drawbacks such as low-throughput or no specificity for $m^3C$, or cannot provide clarity on the exact modification site.[142]

These drawbacks can be overcome by the recently published Hydrazine-Aniline Cleavage Sequencing (HAC-seq, Figure 9) method, which can be used for the transcriptome-wide detection of $m^3C$ within RNA at single nucleotide resolution.[142] It is based on chemical alterations of $m^3C$, as hydrazine selectively attacks the modification selectively under high salt concentrations (3M NaCl). The subsequent addition of aniline leads to position-dependent cleavage of the RNA strand at the modification site. An additional demethylation step (AlkB) after Alk-Aniline treatment removes other methyl groups which could potentially block RT. The resulting 3'-OH ends facilitate adaptor ligation for cDNA library formation. Analysis was initially performed with LC-MS/MS and Northern blot and later linked to next generation sequencing, to use it as a transcriptome-wide detection method. Sequencing results from HAC-seq compared with RNA datasets showed a high agreement with reported $m^3C$ sites.[142]

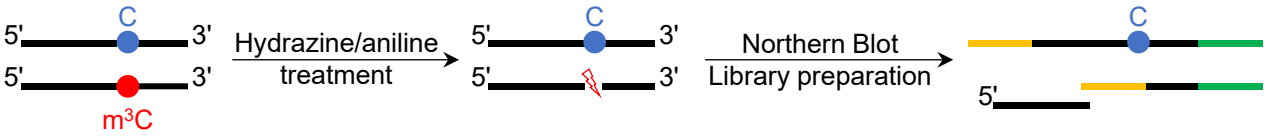



Figure 9. Schematic mechanism of the HAC-sequencing technique based on a chemically induced cleavage at the $m^3C$ site (red). Hydrazine is specifically reacting with $m^3C$ at 3M NaCl concentration. Subsequent aniline treatment is then leading to strand cleavage at the modification site. Results can either be analyzed by Northern blot or linked to sequencing. For the latter, 3'-OH residues allow for library preparation with adaptor ligation (green and orange). Thus, the 3'-fragment and the unmodified reference (blue) are ligated. Figure reproduced from Ref.[142]

As this approach is selective for the $m^3C$ modification, other methods need to be implemented to provide the possibility to analyze multiple modifications throughout the genome/transcriptome. One recently reported tool fulfills these requirements and will be presented next.

#### 1.3.5.2 Nanopore sequencing

In 2014, Oxford Nanopore Technologies (ONT) released the first Nanopore sequencing device MinION as a portable device for long-read single molecule sequencing (TGS). In contrast to next-generation sequencing (NGS, e.g. Illumina sequencing), much longer reading frames can be achieved and single DNA or RNA strands can be directly sequenced, without prior amplification or chemical conversion.[143, 144]

MinION utilizes nanopores, embedded within an electrically resistant membrane and surrounded by ionic solution (Figure 10). A voltage is applied to this membrane that leads to an ionic current, which electrophoretically moves the desired molecule step-wise through the pore.[144-146] This gradual movement (and the separation of a dsDNA to ssDNA) of the single strand is facilitated by a motor protein (polymerase or helicase) bound to the nucleotide chain.[145, 146] As soon as a nucleobase is passing by the narrowest region (= sensing region) within the pore, the current of ions is impeded compared to the unblocked pore. The resulting change in current level reflects a translocated molecule, and is specific for each nucleobase, thus revealing the sequence of the examined nucleotide strand.[144, 146] This method cannot only be used to determine the sequence, but also to find modifications on the single nucleobases. For example, around two years after the release of this device, it was shown that methylated cytosine species could be detected and distinguished in genomic DNA. Canonical cytosine, 5-methylcytosine ($m^5C$), 5-hydroxymethylcytosine ($hm^5C$), 5-carboxylcytosine ($ca^5C$) and 5-formylcytosine ($f^5C$), were differentiated, showing the relevance of this technology for epigenetic investigations, as they are influencing gene regulation as epigenetic modifications.[147, 148]

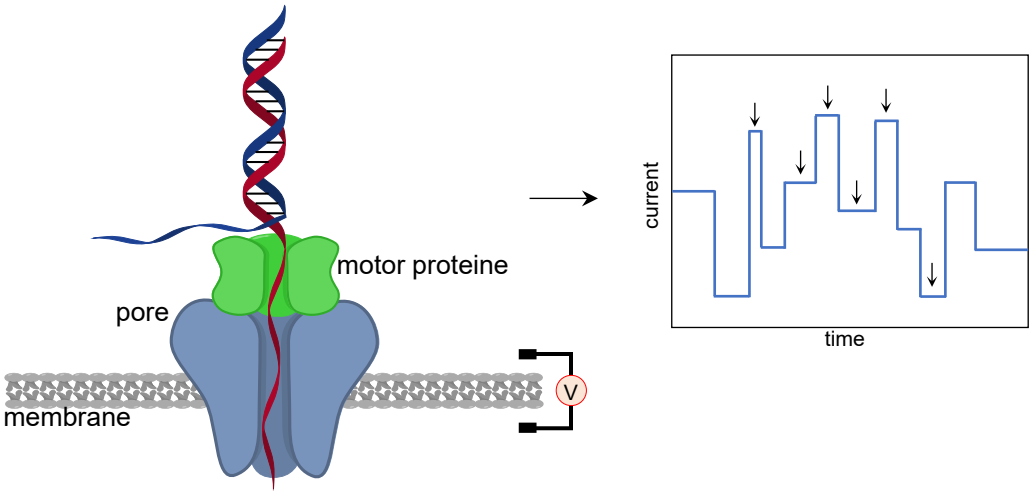



Figure 10. Schematic construction of the ONT device. The nanopore (blue) is located within an electrically resistant membrane (grey) surrounded by ionic solution. The applied voltage leads to an ionic current through the pore, enabling the movement of the single strand (red) towards the other side of the membrane. An attached motor protein (green) ensures the step-by-step motion. The measured ionic current is changing depending on the characteristic of the nucleobase, which is blocking the nanopore, shown in a simplified graph (right), with some current levels highlighted with arrows. Graphic based on Ref.[146]

Just recently, minION was established to directly sequence single full-length tRNAs, using an adapter approach to ensure proper unfolding and the full-length reads. This allows for the direct detection of multiple modifications in RNA without chemical conversion or RT.[149] The RNA sequencing protocol is not as accurate as the one for DNA yet,[145] but there is continuous progress in this field, which might soon lead to better approaches.

Instead of using nanopores as auxiliaries to detect modifications, DNA itself can be used as a tool, in the form of DNA enzymes.

#### 1.3.5.3 DNAzymes

Despite serving as a storage for genetic information, DNA can also catalyze specific reactions. In nature, this is only known from RNAs, the so-called ribozymes. However, DNA can also be made synthetically, to fulfill these kinds of catalytic functions as deoxyribozymes (DNA-enzyme or DNAzyme). Even though they have not been found in nature, catalytically active molecules can be evolved by *in vitro* selection as synthetic enzymes.[150] The first DNAzyme found by *in vitro* selection was identified in 1994 by Breaker and Joyce. It catalyzes the cleavage of the phosphodiester bond of a single ribonucleoside within a DNA sequence, in a $Pb^{2+}$-dependent fashion.[151] Since then, a wide variety of DNAzymes for diverse applications was reported, including DNA-catalyzed RNA cleavage, RNA/DNA ligation, Diels-Alder reactions, DNA phosphorylation or nucleopeptide linkage formation. They can then be used as tools like sensors for environmentally relevant ions or as therapeutic agents (e.g. cleavage of cancer-related genes).[152] One of the most famous and well-studied RNA-cleaving DNAzymes are 8-17 and 10-23 (*in vitro* selection round-clone number; both $Mg^{2+}$-dependent), found by Santoro and Joyce in 1997. The original 8-17 was supposed to cleave only AG-dinucleotide junctions, but later faster variants were shown to more generally cleave NG-junctions (N=A,G,C,U), and became the best characterized RNA-cleaving enzyme.[153-157] The 10-23 enzyme with a catalytic domain consisting of 15 nucleotides was able to cleave RNA substrates with a purine-pyrimidine junction site-specifically.[155]

This RNA-cleavage potential of DNAzymes can now also be utilized for the detection of modifications. Recently, for the first time, cleavage of rRNA samples with 10-23 and variants of 8-17 RNA-cleaving DNAzymes was studied. The reaction was reported to be impaired if 2'-*O*-ribose methylations were present within the rRNA substrate (18S/25S; *S. cerevisiae*) as the 2'-OH group is not available for the cleavage reaction. As consequences, rRNAs that contain these posttranscriptional methylations remain intact and only unmodified regions get cleaved.[158] This was also shown to some extent if pseudouridine was present within the substrate.[158, 159] In another approach for site-specific analysis, after the cleavage of the RNA substrate with the variants of 10-23 and 8-17 DNAzyme, the corresponding modified nucleotide remained at the 5'-end of the fragment. The resulting 5'-OH group can subsequently be phosphorylated with $^{32}P$ for labeling, prior to analysis *via* Stanley-Vassilenko sequencing, including digestion to mononucleotides and analysis by thin layer chromatography. However, only modifications not located at the Watson-Crick face, like $m^5C$, $f^5U$ or pseudouridine, can be analyzed this way.[159, 160] The first RNA-cleaving DNAzyme (8-17 variant) reported in 2018 was found to be sensitive for methylation of adenosine. RNA substrates with $m^6A$ within the DRACH motif were cleaved faster than the unmodified substrate. Even though the DNAzyme is not completely inactive for unmodified substrates, the kinetic differences (5-10-fold accelerated for $m^6A$) allow for the investigation of modification levels. In the same study, another

DNAzyme was found, which could distinguish between canonical adenosine and $m^6A$. The enzyme is active to the unmethylated substrate and inhibited by $m^6A$.[161]

There is still a need for DNAzymes that are highly selective for substrates bearing posttranscriptional modifications, to determine modification states of RNA. Maybe larger residues would facilitate the discrimination of modified over unmodified substrates, due to more favorable interactions within the active site of the enzyme.[162] Furthermore, access to the detection of more posttranscriptional modifications would be highly desirable. Synthesis and characterization of DNAzymes with these properties will be described in this thesis. Early 2022, a new high-throughput sequencing method (DZ-seq) was published.[163] With this method, the activity and cleavage-site of a multitude of deoxyribozymes can be measured simultaneously, and further reveals the sequence of the DNAzyme. This method facilitates the discovery of new deoxyribozymes with a distinct function.

This chapter summarized some important aspects of posttranscriptional modifications, which are creating a second layer of information, to modulate the functions of the different RNA species. There is a variety of enzymes involved, to ensure the dynamic interplay of modifications and their desired functions and impairing this synergy often relates to diseases. The knowledge about nucleoside modifications can further be utilized to generate artificial nucleoside modifications for various applications, which will be described in the next section.

## 1.4 Artificial modifications

A large number of various artificial nucleoside modifications have emerged within the last years, with specific properties for the desired application based on the defined structural geometries of native nucleosides and nucleotides. Examples are fluorescent analogs for visualization and investigation of the dynamic structures and functions of nucleotides, crosslinkers for exploration of DNA protein interactions or even therapeutic interventions – the characteristics of natural nucleosides can be conducive for important investigations by diverse alteration of the scaffold.

### 1.4.1 Fluorescent nucleoside analogs

To get a better understanding of the cellular functions, dynamics of protein-nucleotide interactions, and structural/folding abilities of nucleic acids different methods are available, including nuclear magnetic resonance (NMR), circular dichroism (CD), microscopy, and fluorescence spectroscopy. The latter represents a highly sensitive, rapid and accurate detection technique for real time spectroscopy even in living organism in a non-static fashion.[164, 165] For evaluation of ongoing processes, different quantitative parameters, like absorption/emission wavelength, anisotropy, quantum yield, fluorescence life time and fluorescence intensity can be measured.[166, 167] Common methods often use covalently attached fluorescent dyes *via* a linker as an extrinsic fluorogenic probe (external

modifications[168]). These dyes are commercially available and cover different absorption and emission ranges (e.g. fluoresceine or rhodamine).[167, 169] An alternative is the usage of fluorescent nucleobase analogs that offers several advantages for the investigation of nucleotide dynamics compared to external fluorescent labels. These analogs can directly be incorporated at the location of interest within the nucleotide sequence, resulting in a more sensitive probe (environment-sensitive reporter) at a defined position. Conformational changes and interactions that vary the physical properties can be followed more precisely by a change of the fluorescence signal, with internal modifications positioned close to the examination site. Furthermore, they interfere less with the native structure and protein-enzyme interactions, thus maintaining the functions of the investigated nucleic acids.[165, 170]

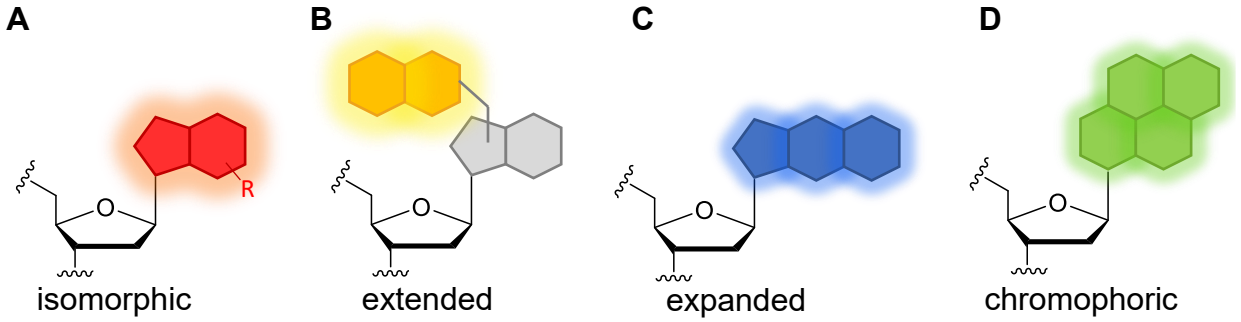



Figure 11. Schematic representation of how the nucleobase scaffold can be modified to create fluorescent nucleoside analogs. A: Small residues or heteroatoms (isomorphic), B: Chromophores attached via a linker (extended), C: The purine or pyrimidine ring scaffold is enlarged by additional rings (expanded), D: The nucleobase is replaced by chromophores (chromophoric).[167]

To create fluorescent nucleoside analogs, the nucleobase needs to be modified, as native nucleosides themselves are in principle non-fluorescent. Non-radiative relaxation pathways/decay[171] seems to account for their low quantum yield[172] ($\Phi_F$<$10^{-4}$).[167] For the design of fluorescent analogs, two main approaches can be considered. In the first one, the overall nucleobase scaffold maintains the characteristic properties of canonical nucleosides (i. e., WC-base pairing, protein interaction, purine/pyrimidine architecture with regular geometries, enzyme incorporation), maintaining the native structure; extended, isomorphic and expanded nucleobase scaffolds are possible (Figure 11). Whereas in the second approach, the structural variations are not limited to retain the nucleobase characteristics (chromophoric design, Figure 11), allowing diverse changes and improved fluorescent properties.[170] Both designs will be described in more detail in the following section. Depending on the desired properties, analogs can be chosen out of these two groups.

#### 1.4.1.1 Canonical design

As mentioned earlier, the main characteristic for canonical analogs is to retain the canonical base pairing abilities and stacking behavior in order to achieve the best possible accordance with the native structure, while providing fluorescent properties (Figure 12A). Only minor structural alteration to the nucleobase scaffold can be implemented, exempted from the WC-face, to enable at

least two WC-hydrogen bonds (Figure 12B) and to ensure stability and proper folding. Based on the structural analogy to the native nucleosides, these chromophores can also be incorporated enzymatically.[170]

Chemical alterations can occur on both the purine and pyrimidine scaffold, ranging from ring fusions (expanded) over conjugated linker extension (extended) to additions of small substituents or heteroatoms (isomorphic). Especially for *in vivo* applications there is a need for selective analysis of the fluorescent nucleobase in the presence of canonical nucleosides. Therefore, the optical properties should be red-shifted, with excitation and emission in the visible region, which can be achieved by enlarging the π-systems or installing residues with heteroatoms. In addition, a high quantum yield ($\Phi_F$) after incorporation into the nucleic acid structure would be favorable and the analogs should act as environment-sensitive reporters.[167, 170]

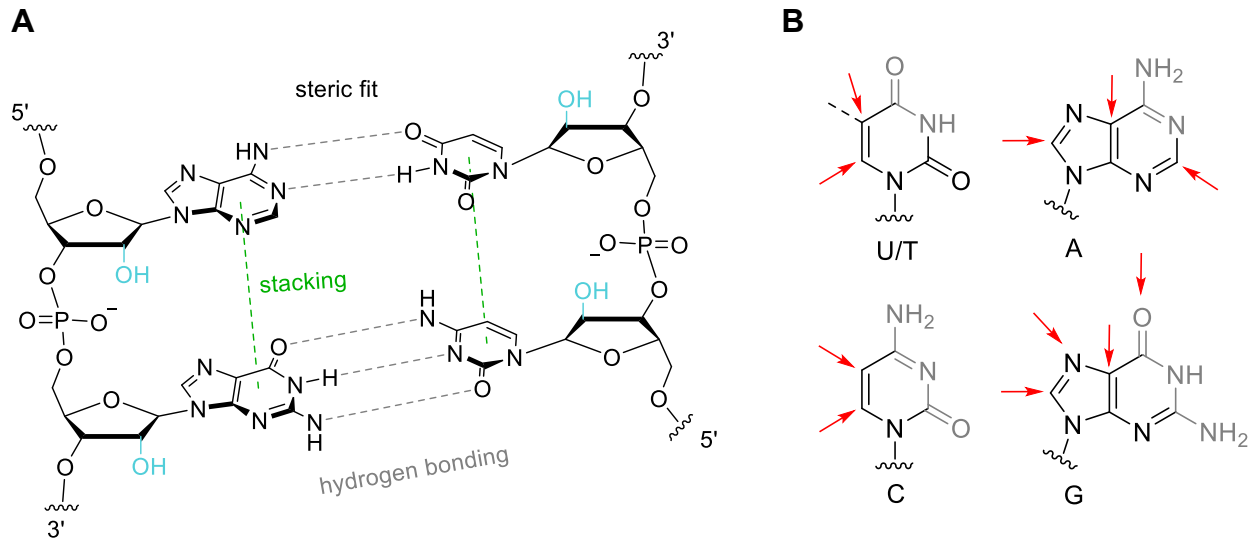



Figure 12. Requirements for designing fluorescent nucleoside analogs to ensure the properties of native nucleosides. A: Main characteristics of canonical nucleobases with the WC-base pairs for DNA and RNA (2'-OH shown in light blue). The purine/pyrimidine architecture ensures the regular geometrie with the nucleobases stacking (green) on top of each other. B: Positions which could be modified to create fluroresent nucleobase analogs, retaining the WC-face (grey), are highlighted with red arrows. Graphics reproduced from Xu *et al.*[170]

In 1969 a fundamental study by Stryer and coworkers revealed the fluorescent properties of 2-aminopurine (2-AP, Figure 13), a constitutional isomer of adenosine.[173] It predominantly forms base pairs with thymine and to a minor extent with cytosine.[174, 175] In contrast to the canonical nucleoside, this chromophore exhibits intrinsic fluorescence under physiological conditions while also having absorption at a longer wavelength, with an absorption maximum at 303 nm and emission at 370 nm (free 2-AP nucleoside). Since this enables selective excitation of the chromophore, Stryer and coworkers proposed that this compound could be a sensitive tool after incorporation into an oligonucleotide, to study for example interactions with proteins by fluorescence.[173] Indeed, 2-AP became one of the most frequently used fluorescent nucleobase analogs, due to the fact that the fluorescence quantum yield (0.68 in phosphate buffer) of the free nucleoside is highly sensitive

to the local environment and decreases strongly when incorporating it into an oligonucleotide strand (caused by stacking interactions[176]).[173] 2-AP has been used for various investigations as a site-specific fluorescent probe to study conformational changes and to monitor interactions.[177-180] However, a recent report in 2014 revealed that the fluorescent properties of 2-AP are actually dependent on hydration:[181] The number of water molecules and the interaction site with 2-AP *via* hydrogen bonds influences the fluorescence lifetime. Contrary, in the gas-phase it is non-fluorescent, contradicting the assumptions of 2-AP being intrinsically fluorescent.[181] Hence, the results of the investigations using 2-AP have to be critically scrutinized.

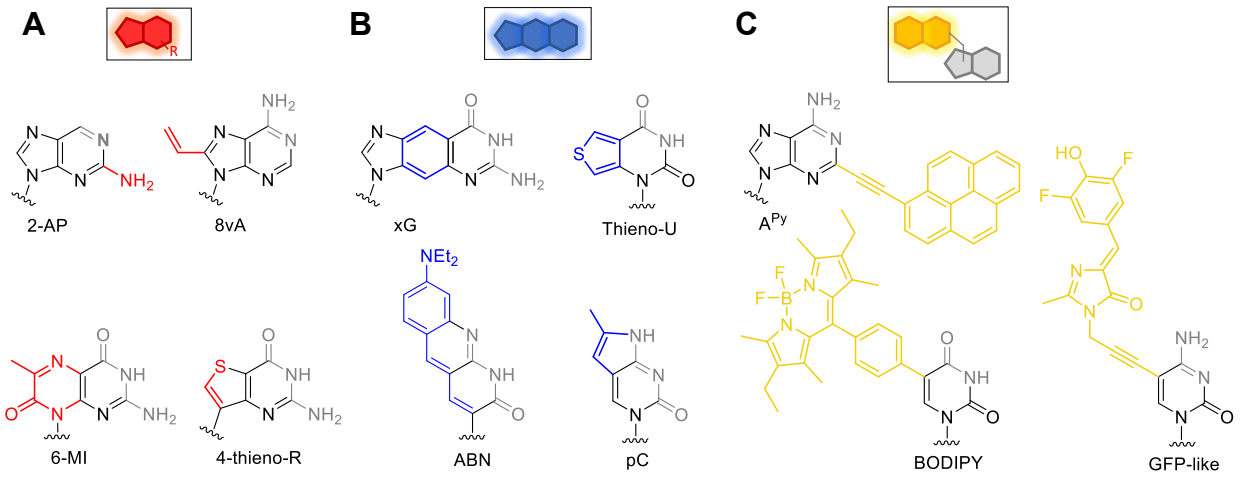



Figure 13. Selection of fluorescent nucleoside analogs that are mimicking canonical nucleobases, while retaining WC-hydrogen bonding by incorporating heteroatoms or small residues (red), expanding the conjugated π-system (blue) or extending the nucleobase with chromophores (yellow) apart from the WC-face (grey).[170]

Since then, a plethora of fluorescent nucleoside analogs were designed, and their photophysical properties have been characterized. One such analogue is 8-vinyl-deoxyadenosine (8vdA), which has a vinyl moiety at position 8 of adenosine (Figure 13). 8vdA was supposed to be an alternative to 2-AP, as the absorbance and emission wavelengths are in the same range. By comparison, 8vdA was shown to be less perturbing, reflected in higher duplex stabilities, and exhibits a higher quantum yield when it is incorporated into oligonucleotides. In addition, 8vdA is a more sensitive analogue, due to a higher molar absorption coefficient.[182] A selection of other nucleosides are shown in Figure 13 divided into the three possible groups of fluorescent nucleobase-modifications. Among them are a pteridine analogue, 6-methyl isoxanthopterin (6-MI), which is fluorescent in the visible range but quenched upon incorporation into oligonucleotides, especially if flanked by purines[183], a highly emissive thiophene expanded uridine derivative (thieno-U)[184], a highly fluorescent pyrrolo-modified C-analogue, with excitation maxima at 272/345 nm and emission at 473 nm (pC)[185], and a pyrene linked to an adenosine, with good quantum efficiencies and an absorption range of 350-440 nm, and emission between 440-470 nm ($A^{Py}$)[186]. As a special case, C-linked 8-(diethylamino)benzo[b][1,8] naphthyridin-2(1H)-one nucleoside (ABN) comprises a push-pull motif within a tricyclic conjugated system, to create a bright organic fluorophore ($\varepsilon$ = 20000 $M^{-1}cm^{-1}$) with

both red shifted absorption (442 nm) and emission (540 nm). When incorporated into an oligonucleotide ABN showed enhanced quantum yields compared to the free nucleoside (from 0.39 to 0.50 (when base paired with A)), only minimally perturbating the duplex structure. ABN is the brightest and most red-shifted example for an fluorescent nucleoside analogue in oligonucleotides reported so far.[187]

These fluorescent nucleoside analogs have been used for multiple applications.[188, 189] One example is the usage of the adenosine-derivative of thieno-U (Figure 13), namely thieno-A ($^{th}$A), as a probe for the deamination reaction of Adenosine Deaminase Acting on RNAs2 (ADAR2). This enzyme is responsible for A-to-I editing in RNAs. The emissive $^{th}$A within the RNA strand was accepted as a substrate and got deaminated to thieno-Inosine ($^{th}$I) by ADAR2, whereupon changes in the fluorescence could be observed. These fluorescent changes allowed monitoring of the reaction progress of RNA editing in real-time.[188, 190] However, if the spectral range should be further expanded, e.g. to circumvent excitation in the UV region, which could lead to cytotoxic photo lesions, and to achieve a brighter analogue, the canonical design cannot be maintained, but larger alterations are needed.[167]

#### 1.4.1.2 Chromophoric nucleobases

Chromophoric nucleobases are structurally not related to canonical nucleobases, since the whole nucleobase is replaced by large, aromatic moieties (Figure 14).[167] This allows for a greater variety of fluorescent properties, especially to shift the emission range even further towards the red (avoiding cytotoxicity and background noise) than the fluorescent nucleobase analogs that were described earlier.[170] In addition, even though the formation of WC hydrogen bonds is not possible due to the missing nucleobase scaffold, they offer other stabilizing effects: Hydrophobic residues like pyrene exhibit much stronger stacking interactions than natural bases, leading to a higher thermal stability, if they are incorporated at an unpaired position at the end of a DNA duplex (Figure 14A).[191] Pyrene-triphosphate (dPTP) was even shown to be more efficiently incorporated by polymerases opposite an abasic site of a DNA strand than natural nucleoside TP, caused by the better stacking properties. Given the fact that the pyrene residue has almost the same size as of a canonical nucleobase, the residue can be correctly accommodated in the abasic site.[192-194] However, the size of the chromophoric nucleobase may affect the efficiency of polymerase enzymes, especially with highly sterically demanding residues, that might not fit well into a double helical structure.[170, 192]

This class of chromophoric nucleosides was initialized by Kool and coworkers in 1996.[195] They synthesized DNA molecules with the nucleobase being replaced by an aromatic residue that has a similar steric demand as canonical nucleobases, but lacks the base pairing ability. While investigating aromatic stacking interactions, they found that pyrene and phenanthrene-substituted DNA

analogs showed promising fluorescent properties for using them as fluorescent probes in DNA, with emission maxima at 395 nm and 370 nm, respectively (Figure 14B).[195] Using two pyrene-nucleoside analogs as fluorescent probes, and with careful design of their relative distances, the interactions between these two residues can lead to the formation of an excimer complex. This results in significant changes of the emission wavelength, shifting from 398 nm (blue) of the pyrene monomer to 490 nm (green) for the newly formed excimer complex. This set-up was shown to be effective for the detection of single-base-mutations.[196]

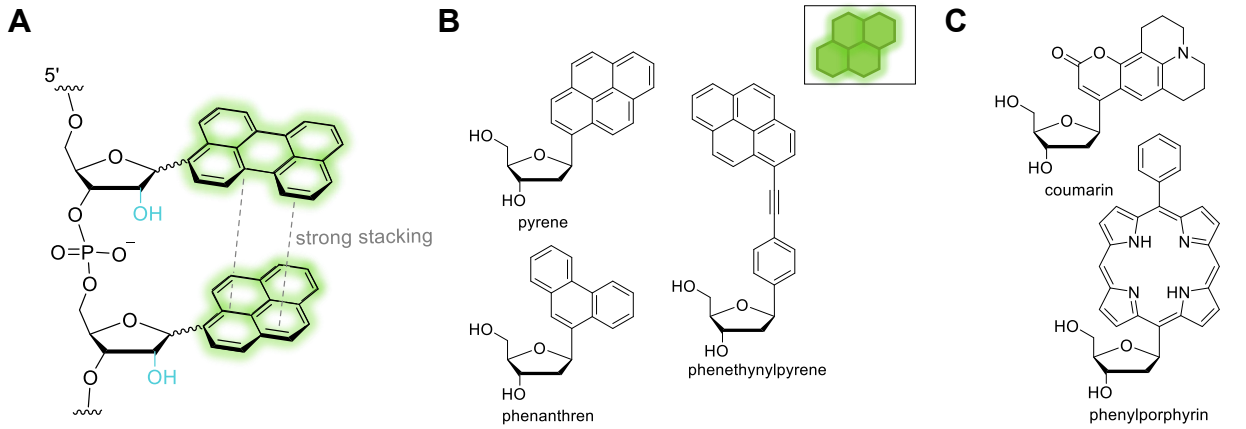



Figure 14. Overview of chromophoric nucleosides. A: Stronger stacking interactions due to the larger aromatic scaffold. Reproduced from Ref.[170] B: Examples of chromophoric nucleosides with different aromatic residues, C: Heterocyclic fluorophores.[170]

The development of members of this new class of fluorescent nucleosides proceeded as also heterocycles can be used to replace the native nucleobase besides hydrophobic aromatic residues. For this, sometimes established dyes with known stacking and fluorescent properties were selected (Figure 14C). For example, the coumarin dye (e.g. exhibiting a large Stokes shift) was reported as nucleobase replacement in 1998,[197] and used for applications, such as investigation of the local dynamics of DNA.[198] Moreover, larger residues like porphyrin derivatives were also linked to the DNA sugar.[199] Neither the fluorescent quantum yield (Φ ~0.1) nor the excitation (~400-408 nm) or emission maxima (~630 nm) of the intercalated porphyrin were strongly affected by the adjacent nucleobases in the oligonucleotides, compared to the free building block.[199]

One interesting application for chromophoric nucleobase analogs was published in 2016,[200] where Kool and coworkers used pyrene modified DNA to examine ALKBH3 (see chapter 1.3.4) demethylation activity. This enzyme is a known anti-tumor target, as it was shown to be involved in tumor formation. The fluorescent probe was incorporated adjacent to an $m^1A$ modification, to quench the emission by photoinduced charge transfer, as $m^1A$ is positively charged (=electron deficient) at neutral pH. Upon demethylation of $m^1A$, this experimental setup led to a fluorescence enhancement, reflecting the activity of ALKBH3. The potency of this assay to monitor ALKBH3 mediated demethylation was even shown *in vivo* using tumor cell lines.[200]

While chromophoric nucleobase analogs help in monitoring the onset cancer, treatment of cancer has been established by methods such as DNA cross-linking agents.[201] The basis of these agents, their evolution and their potential applications beyond antitumor medication, will be discussed in the next section.

### 1.4.2 Crosslinkers

To be able to fulfill the fundamental cellular functions, double stranded DNA (dsDNA) needs to be separated into single stranded DNA (ssDNA), as polymerase requires ssDNA for replication and transcription. The basis for these genetic processes is provided by non-covalent hydrogen-bond interactions, which allows DNA to be arranged in a linear fashion by separation.[6] However, this important ability can be impaired if both strands are covalently connected. This might arise by alkylation agents in nature, if they comprise bifunctional active sites.[202] These agents are able to react with amino acids as well as with the nucleobases of DNA.[201] As DNA lesions, they block replication by the formation of covalent interstrand crosslinks (ISCs),[203] explaining their high cytotoxicity.[202]

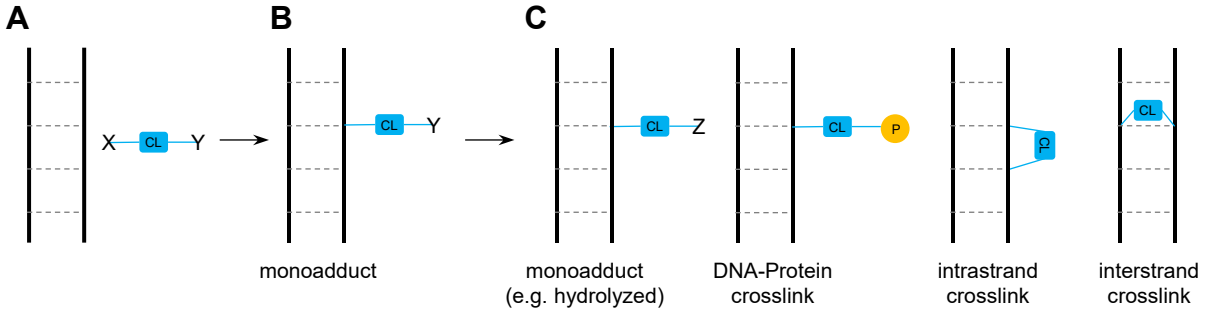



Figure 15. Mechanism of crosslink formation with DNA. A: Association of DNA double helix with bifunctional (X,Y) alkylating agent (blue). B: Covalent monoadduct, after reaction of one reactive group (X). C: The monoadduct can either remain, e.g. by hydrolysis, react with proteins (yellow, DNA-Protein crosslink), within the same DNA strand (intrastrand) or with the opposite strand (interstrand, ISCs). The latter reflects the most toxic one. Graphic based on Ref.[202]

This mechanism includes the formation of a monoadduct at first, which might either remain (e.g. after hydrolysis), or further react intramolecularly within the same DNA strand forming an intrastrand-crosslink, or with the opposite strand forming an ISC, which reflects the most toxic of the alkylated lesions (Figure 15).[201, 202] Besides this homologous reaction, the monoadduct might also react with proteins, resulting in a DNA-Protein crosslink (Figure 15C).[203] However, this toxicity of bifunctional alkylating agents can be targeted, to employ them as drugs for the treatment of cancer.

#### 1.4.2.1 Bifunctional agents for cancer treatment

Three well established classes of agents for tumor treatment are nitrogen mustards, chloroethylnitrosoureas (CENU's) and platinum complexes. Among them is nitrogen mustard, a compound derived from military applications of poison gas in the first and second world war (chemical warfare).[204] Contrary, to its prior use, it was efficiently tested as the first chemotherapeutic agent in

1946 (e.g. leukemia; Nitrogen Mustard Therapy).[205, 206] Based on this, several other DNA-cross-linking-drugs were developed.

The mechanism of action of nitrogen mustard-based agents, like the well-established mechlorethamine and chlorambucil, is the formation of the DNA interstrand crosslink *via* an aziridinium ion (intramolecular cyclization, Figure 16A). This cyclic intermediate is then attacked by the nucleobase.[207] In case of nitrogen mustard, N7 of guanosines are site-specifically alkylated, preferentially within a $^{5'}$d(GNC)$^{3'}$ (N = dA, dC, dG, T) sequence.[208, 209] Repeating the aziridinium-based reaction results in a dialkylated crosslink, which inhibits replication in case of ISCs. This $N^7$dG-$N^7$dG interstrand covalent bond thereby leads to a small distortion of the double helical structure.[201, 207]

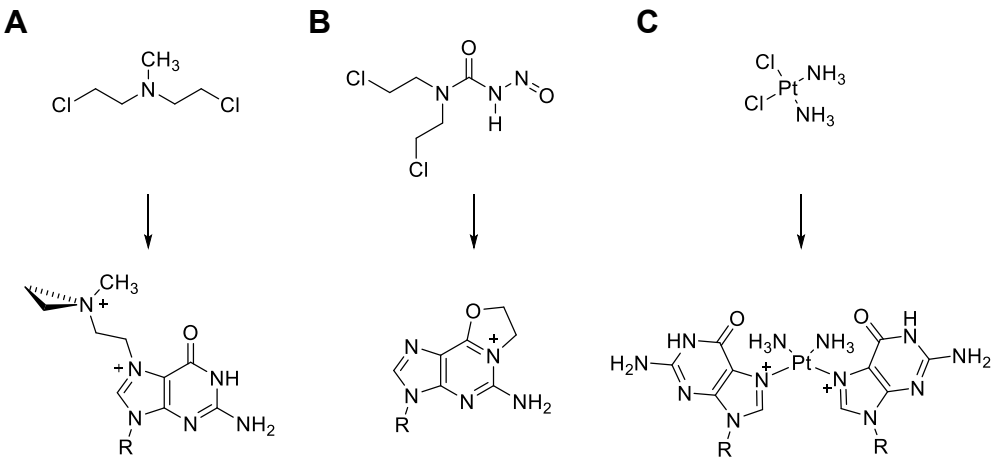



Figure 16. Three examples of bifunctional crosslinking agents. A: Nitrogen mustard (mechlorethamine) and monoadduct with aziridinium ion at N7 of dG. B: Nitrosourea-based (chloroethylnitrosourea) and O6-N5 circularized intermediate. C: *Cis*-diamminedichloroplatinum (cis-platin) with crosslinked dG-dG over N7.[201, 202]

CENU's (Figure 16B) have been developed as chemotherapeutical drugs, mainly for the treatment of malignant brain tumors, since they are able to penetrate the blood brain barrier, and still remain promising candidates for medical attendance. Administration in combination with other drugs might enhance the efficiency.[210-212] The synthetic mechanism, e.g. of CENU's for the ISCs is supposed to start with the alkylation of N1 of deoxyguanosine, followed by the attack of N3 of deoxycytosine on the cyclic intermediate, resulting in a crosslink with an ethane bridge.[213, 214] The last of this three examples is the planar complex *cis*-diamminedichloroplatinum (*cis*-DDP, *cis*-platin, Figure 16C), which is mainly administered for the treatment of testicular cancer. It predominantly (>90%) results in intrastrand crosslinking products (dG-dG and dA-dG), while some interstrand crosslinks are formed between dGs (N7) in 5'GC3' sequence contexts.[215, 216] The formation of these interstrand bonds leads to significant distortions (bending and unwinding) of the DNA double helical structure.[217]

Besides crosslinking between DNAs (intra- or intermolecularly), DNA-Protein-crosslinks also occur (Figure 15C) by exogenous and endogenous agents, like transition metals, UV-light, reactive oxygen species, or bifunctional electrophiles.[218, 219] This is a highly relevant research field for the investigation of DNA-protein interactions, as they are involved in many biological processes.[220] Especially if there is a weak association between DNA and protein, modified nucleosides, which can perform covalent crosslinking reactions themselves were designed to trap these interactions and ensure a site-specific reaction, for a direct investigation of the position of interest.

### 1.4.2.2 Crosslinking nucleosides

To fulfill their important functions within the central dogma of life, specific interactions between DNA and proteins are a prerequisite to enable their precise performance in biological processes, including the regulation of replication, transcription, repair-mechanisms and installation of modifications.[221] To get a better understanding of their functions and recognition elements, artificially induced covalent DNA-protein-crosslinks (Figure 15C) can be used. Thereby, modified nucleosides bearing functional groups, enabling crosslinker reactions, are a useful tool, which can even be designed to be amino acid specific. Example for non-specific crosslink approaches, without the possibility to address one certain amino acid, commonly comprise photo-crosslinking.[222-225] The light-reactive groups for photoaffinity labeling can be attached either to the nucleobase directly (5-bromo/iodo uracil, Figure 17B), or *via* a linker, or to the phosphate backbone (or less common to the sugar moiety).[225] Upon irradiation with the respective wavelength, these light-sensitive functionalities form reactive intermediates like carbenes (e.g. diazirine attached to 5' deoxyuridine *via* an alkyne spacer[222] or phosphate backbone[225]) or radicals (5-iodouridine[223]), and react with amino acids in close proximity of a protein (Figure 17).

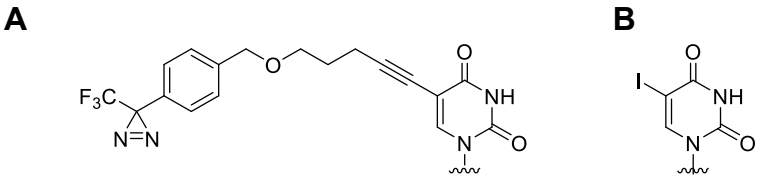



Figure 17. Two examples of photo-crosslinking agents. A: Azirine-functionalized uridine leading to reactive carbenes upon UV-light irradiation. B: 5-Iodouridine forms radicals upon radiation with the respective wavelength as reactive intermediates.

Besides this, several amino-acid-specific methods were published, with different chemical mechanisms. This provides access to more precise and locally confined interactions, but only few examples are available. Mainly position 5 of pyrimidines and position 7 of (deaza-)purines were chosen for modification, since the attached residues were shown to not disturb the DNA structure but are located in the major groove, and these modified nucleosides can efficiently be incorporated by polymerases.[226]

One example is disulfide-crosslinking, where a thiol tethered cytosine selectively forms a disulfide bond with cysteine of proteins (Figure 18A). It is an equilibrium process, in which the efficiency of the bond formation also depends on the strength of the DNA-protein interactions, making them less prone to form imposed complexes.[227] A modified cytosine with the protected two-carbon thiol tether at N4 was shown to selectively bind to cysteine amino acids (mutated H131C) within the active center of AlkB (*E.coli*), forming a disulfide crosslink. With this, structure and function investigations of this demethylase for the damage repair mechanism of $m^1A$ and $m^3C$ DNA-lesions were performed.[228]

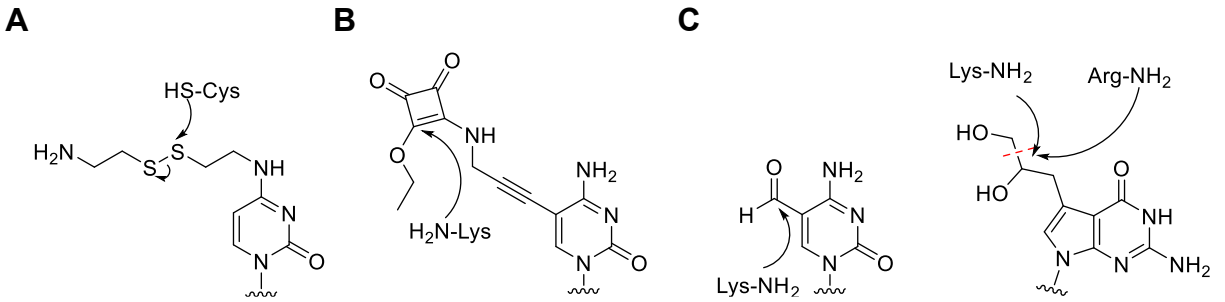



Figure 18. Examples for amino acid-specific crosslinking-nucleobases. A: Thiol tethered deoxycytosine forming a disulfide-bond crosslink with Cys.[228] B: quaramide-linked deoxycytosine reacting with Lys.[229] C: Aldehyde-bearing nucleobases, either naturally occurring[230], or with a masked aldehyde as a diol linker.[218]

Beyond, squaramide scaffolds were utilized for DNA-protein-crosslinking (Figure 18B). It is a structurally rigid, four-membered ring system on the basis of squaric acid, which could potentially form up to four hydrogen bonds.[231] Delocalization of the lone pair of the nitrogen into the cyclo-butenedione ring provides aromaticity to the squaramide ring.[232] 2'-Squaramate-functionalized RNA within the sequence of the acceptor arm of $tRNA^{Ala}$ was selectively crosslinked to the Lys (K305) in the active site of aminoacyl-transferase ($FemX_{Wv}$), even in the presence of other lysine residues.[233] A similar approach was using squaramides, but for base-functionalization of deoxycytosine and the subsequent enzymatic incorporation as a triphosphate.[229] The squaramate-linked cytosine in DNA efficiently crosslinks to Lys-residues of histones, but not to tumor suppressor protein (p53[234]) which does not contain lysine within the active center. Consequently, the close proximity of lysines to the DNA binding site seems to be crucial for the effective formation of the covalent DNA-protein-crosslink (= proximity effect).[229]

Another possibility is the functionalization of aldehydes with amine-containing amino acids like Lys or Arg (Figure 18). Since aldehydes are quite reactive, they often need to be protected for the incorporation into oligonucleotides. Upon crosslinking with the amino acid, either reversible Schiff base or stable amine bonds can be synthesized. For the latter, an additional reducing agent like $NaCNBH_3$ is necessary.[235] Some naturally occurring modifications like $f^5C$ already bear the desired reactive aldehyde (Figure 18C).[230, 236] A precursor of $f^5C$ with a photolabile protection group (4-(2-nitrophenyl)-1,3-dioxane) was incorporated into nucleosome core particles *via* solid phase synthe-

sis.[230] After photolytic *in situ* deprotection (350 nm, 5 min), location-dependent DNA-protein-cross-links with Lys (or Arg) of histones were formed *in vitro*. Thereby, the crosslink with the resulting imine function (Schiff base conjugate) exists in a dynamic equilibrium, which could be transferred into a stable C-N bond by reducing with $NaBH_4$.[230] To further investigate the biological relevance of this reversible crosslink, the formation of these Schiff-base conjugates was analyzed in a cellular context (*in vivo*).[236] In mammalian cells, $f^5C$-histone(H3) crosslinks were found, with $f^5C$ contributing to tissue-specific organization of nucleosomes.[236] Besides naturally occurring aldehyde-modifications, they can also be artificially introduced, like 5-formylthiophene-linked/formylphenyl-linked cytosines,[237] or the 7-deaza-7-(2,3-dihydroxypropyl)-guanine (Figure 18C).[218, 238] The latter was incorporated into oligonucleotides with the diol residue at position 7 of the $N^7$-deazaguanosine, which can be unmasked to the aldehyde *via* oxidation (e.g. $NaIO_4$).[218] The resulting crosslink with Lys and Arg residues of proteins can be stabilized by reductive amination (e.g., $NaCNBH_3$, Figure 19A). Indeed, this resembles the crosslink-adduct, formed by the antitumor agent nitrogen mustard (see chapter 1.4.2.1) at N7 of guanosines, but with the advantage of being DNA sequence-specific. However, this substrate showed low specificity for crosslinking sites within the protein. Different Lys and Arg residues in the DNA binding groove and the active site of AlkB were identified being involved in crosslink formation with the deaza substrate.[218]

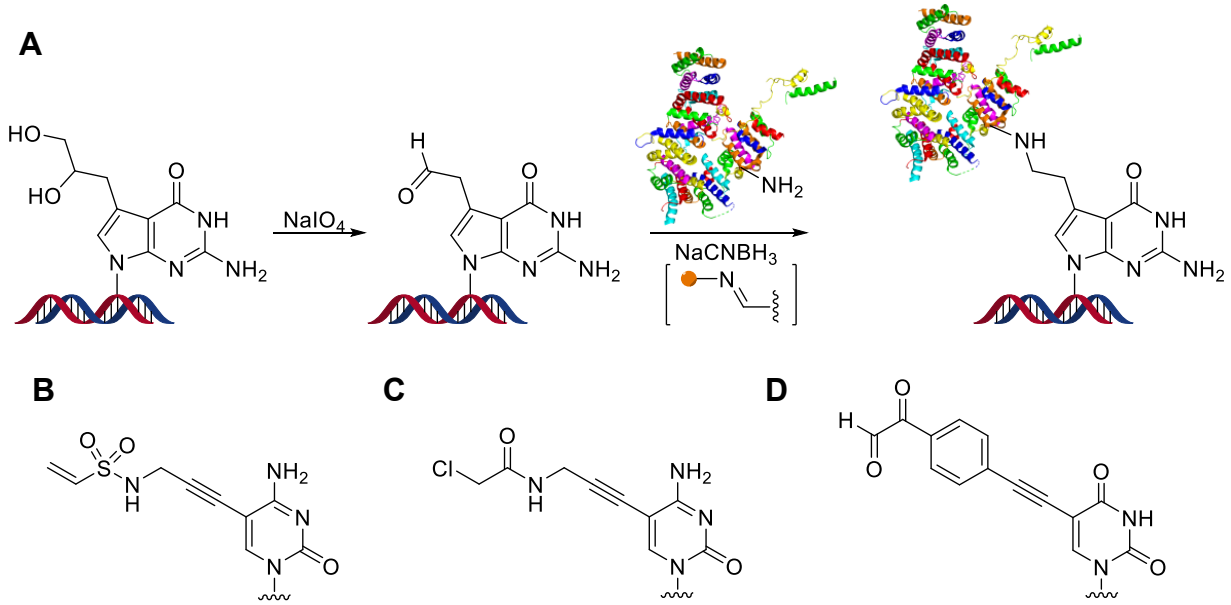



Figure 19. Amino-specific crosslinking agents. A: Mechanism of the masked aldehyde functionalized deaza-guanosine. Oxidation with $NaIO_4$ releases the free aldehyde, which can react with the amine-residue of Lys or Arg of a protein (here as an example: Condensin NDB:5OQN). Reduction of the formed Schiff based (shown in bracket, orange circle represents the protein) with $NaCNBH_3$ is leading to a stable amine linkage. Graphic adapted from Ref.[218] B: Michael acceptor; vinylsulfonamide-linked deoxycytosine reacting with Cys.[239] C: Chloroacetamide-linked deoxycytosine reacting with Cys and His.[240] D: Glyoxal linked uridine forming an imidazolone ring with Arg.[241]

Further examples are Michael acceptors (vinylsulfonamide-linked DNA[239], Figure 19B) for Michael-addition with Cys, reactive chloroacetamide ($dC^{CA}$ or $dA^{CA}$(7-deaza)[240], Figure *19*C) with Cys and His, and glyoxal linked nucleobases ($dU^{GE}$, Figure 19D)[241], which are forming an imidazolone ring

with Arg. As this is still a highly relevant research topic, continuous progress is made to achieve diverse crosslinking agents for the selective DNA-protein crosslinking.[241-245]

This shows the great potential application of modified nucleosides as an analytical tool for tracing diverse biological processes by forming covalent interactions with the proteins of interest. The profound knowledge about the programmable structure of nucleosides facilitates their versatile application. Due to a pandemic at the end of 2019, modified nucleosides obtained a renewed significance in another related research field, in which the interaction of these nucleosides with enzymes was also the crucial feature – for the potential prevention or treatment of viral-diseases by vaccines or antiviral compounds.

### 1.4.3 Antiviral nucleoside analogs

In December 2019, the pathogen severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) led to a global pandemic with serious upper respiratory infections and a fatal disease progression at times.[246] This newly emerged viral strain belongs to the known family *Coronaviridae*, which was initially reported in 1968, due to its special morphology.[247] *Coronaviridae* contains various strains that infect humans, like Middle East respiratory syndrome coronavirus (MERS-CoV), HCoV-229E (299E) and SARS-CoV.[248] All of them consist of a positive-sense RNA genome (26-32 kb)[248] encased by a nucleocapsid (N) protein, which in turn is encompassed by a lipid bilayer (=membrane). In this, the S protein (S), envelope protein (E) and membrane glycoprotein (M) are embedded (Figure 20A).[249] These characteristics make *Coronaviridae* a member of *Nidovirales*, consolidating positive-sense single stranded RNA (+ssRNAs) genomes.[247]

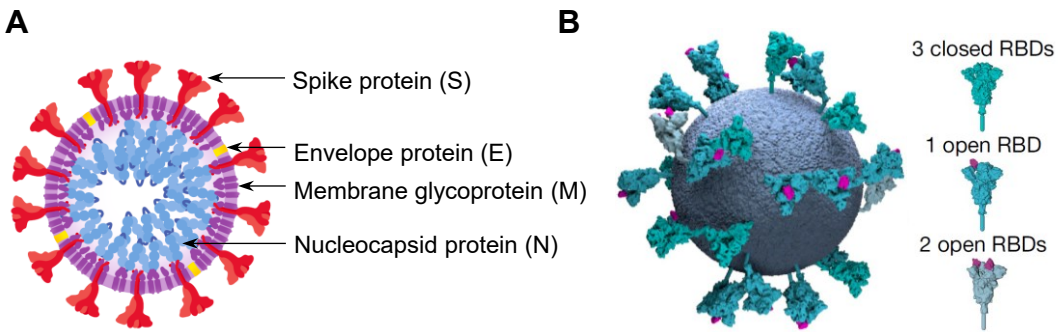



Figure 20. 2D-and 3D-structure of one single virion of SARS-CoV-2. A: Schematic structure of a virion of SARS-CoV-2 with the encapsulated RNA virus with the nucleocapsid protein (N), Membrane protein (M) and Envelope protein (P). Figure edited from Ref.[250] B: Three-dimensional model of a SARS-CoV-2 virion determined by Cryo-EM. The almost spherical particle has a diameter of around 100 nm and is covered with trimeric spike (S) proteins. They bind to the angiotensin-converting enzyme 2 (ACE2) receptor of the host cell facilitating the viral membrane fusion. The receptor-binding domains (RBD, lifted ones shown in red) are located on top of the spike protein. Figure reprinted with permission from Zunlong Ke et al. Copyright Springer Nature 2020.[251]

The virions have characteristic petal-shaped spikes (S) (Figure 20B),[247] leading to the visual appearance of a crown/coronet, explaining the origin of the name (*corona* in Latin).[252] Several outbreaks of members of human *Coronaviridae* have been reported within the last years,[248, 253-255] but this new pathogen strain (SARS-CoV-2) led to a global pandemic with rapid spread causing large

number of deaths. Provoked by this development, there was an urgent need for antiviral compounds or vaccines for the treatment of SARS-CoV2.

### 1.4.3.1 Characteristics of SARS-CoV-2 and viral RNA polymerase

The polycistronic RNA genome (~ 30 kb) is organized in two main subgroups – nonstructural protein (nsp) genes (= replicase region) and several structural and accessory protein genes (Figure 21). It is capped at the 5'-end while the 3'-end bears a poly(A) tail. Additionally, both 5' and 3' of the SARS-CoV-2 genome possess UTR sequences. The viral RNA encodes several, partly overlapping, open reading frames (ORFs), where the replicase region contains two large ORFs: ORF1a and ORF1b. They are responsible for the formation of the polyproteins ppa1 and pp1ab by translation.[256] To enable the expression of pp1ab, programmed -1 ribosomal frameshifting (-1 PRF), directly upstream of the ORF1a stop codon, is a prerequisite. The ribosome backtracks by one nucleotide, resulting in a different reading frame and a prolonged polypeptide (pp1ab).[257, 258] This is facilitated by the slippery site (UUUAAAC) and a H-type RNA pseudoknot (3 helices) within the -1 PRF region, mediating the frameshift process.[259]

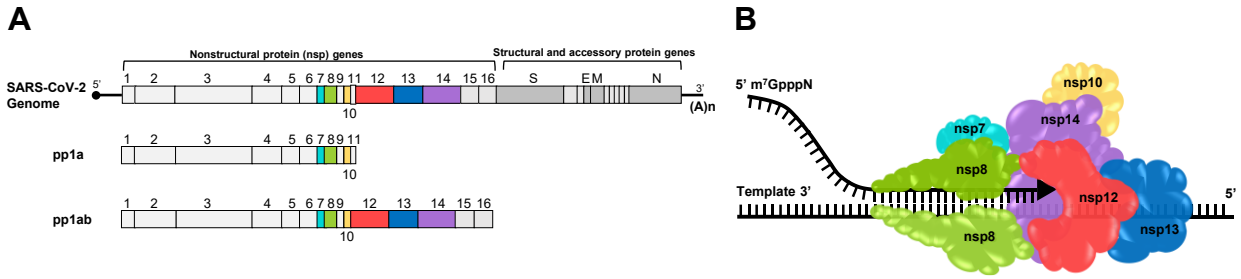



Figure 21. SARS-CoV-2 genome and the RNA-dependent RNA polymerase (RdRp-) complex for viral replication. A: Around two-third of the 30 kb viral genome encodes the two polyproteins pp1a and pp1ab that are cleaved into 11 and 15 nsp, respectively. Contrary, the other part, towards the 3'-end, encodes structural and accessory proteins. B: RNA replicase/transcriptase complex, including the crucial units nsp12 (RdRp), np13 (helicase), and nsp14 (exonuclease). Both figures reproduced from Ref.[256]

These polyproteins are then proteolytically cleaved into individual nsp by proteases (PLpro/nsp3 and CLpro/nsp5).[256] The shorter pp1a comprises 11 nsp, whereas the larger pp1ab is cleaved into 15 nsp, containing proteins that are crucial for viral replication; especially the RNA-dependent RNA polymerase (nsp12), and NTTPase/helicase (nsp13).[258, 260] Additionally, the RdRp complex implies an nsp14 subunit with exonuclease activity to guarantee the fidelity of the viral biosynthesis.[261]

The enveloped, positive-strand RNA virus SARS-CoV-2 uses RNA-dependent RNA polymerase (RdRp) for replication and transcription of the viral genome.[262] RdRp is a complex with different subunits, possessing various interactions with the RNA substrate. Cryogenic electron microscope (cryo-EM) structure of the complex consisting of the recombinant non-structural protein (nsp) subunits nsp12, nsp8 and nsp7 with a self-annealing RNA-substrate (Hairpin) within the active center (= active form) revealed that the enzyme is occupied with more than two turns of the DNA-du-

plex.[262] Nsp12 is the catalytically active subunit with an N-terminal nidovirus RdRp-associated nucleotidyltransferase (NiRAN)-, an interface-, and a C-terminal RdRp-domain (with fingers, palm and thumb subdomains).[262-264] The first turn of the RNA substrate is enclosed by the finger and thumb subdomain upon binding to the nsp12 subdomain. The active site is encompassed by two accessory subunits nsp8 and nsp7 (Figure 22).[262] For incorporation, the incoming ribonucleotide triphosphate (rNTP) binds to the +1 site, whereas the 3'-end of the product strand is located at the -1 site (Figure 22, excerpt).[265] The RNA duplex that is released from nsp12 is flanked by α-helical (nsp8) extensions from the N-terminal regions of the two nsp8 subunits (Figure 22). They are comprised of positively charged amino acids, which may form stabilizing interactions with the negatively charged phosphate backbone of the DNA substrate, thus enabling the processivity of RdRp.[262] Comparison of the RdRp (SARS-CoV-2) with the earlier SARS-CoV polymerase structure showed a high structural similarity.[262, 264]

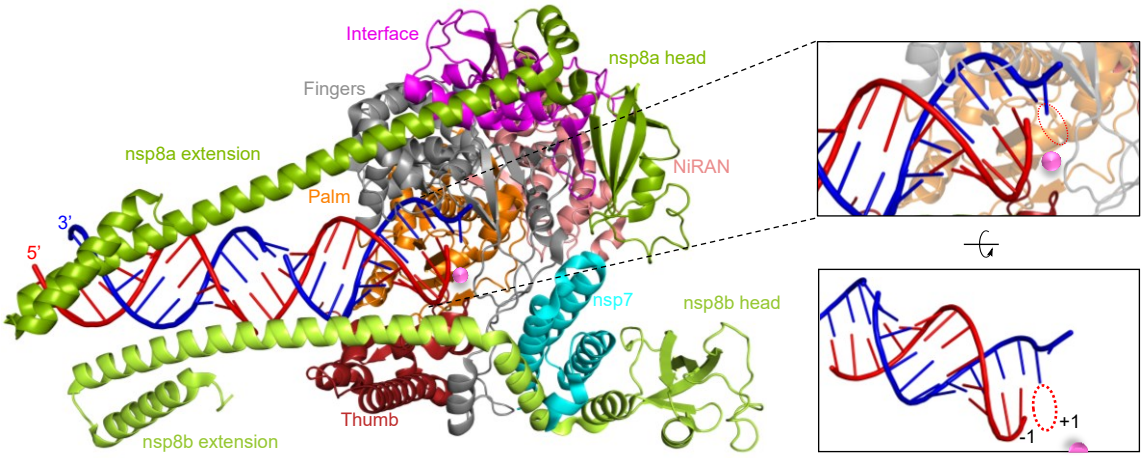



Figure 22. Cryo-EM structure of the viral RdRp. Left: active-form complex of nsp12, nsp8 (green) and nsp7 (cyan) subunits. RNA product is shown in red and the active site indicated by a pink sphere. zoom: active center of RdRp, which is occupied by RNA. +1 site is free for the incorporation of the next rNTP, before translocating towards the -1 site. PDB: 6YYT; adapted from Ref.[262]

The amplified viral RNA genome (gRNA) is then processed by the structural proteins (S, M, E), encoded at the 3' end of the polycistronic RNA (Figure 21) for further distribution.[258]

### 1.4.3.2 Potential target sites

There were various potential treatment approaches under investigation when the pandemic started, as there were no effective drug or vaccine available for Covid-19 so far. The life cycle of the SARS-CoV-2 permits several target sites to be addressed for viral inhibition.[266] It starts with the attachment of the single virions to the ACE2 receptor (S1 subunit) of the host cell with the help of the S-proteins. The viral entry into the host cell can then take place either *via* direct fusion of the cellular and viral membrane by proteolytic cleavage (TMPRSS2) of the S-protein or by endosomal/lysosomal uptake, followed by the processing with cathepsin L. The untrapped viral RNA genome is then available for replication at the RdRp (Figure 23).[266] Thus, there are different potential

target sites against viral infection: from preventing the virion entry to inhibition of the viral replication, some of which will be described now in more detail.

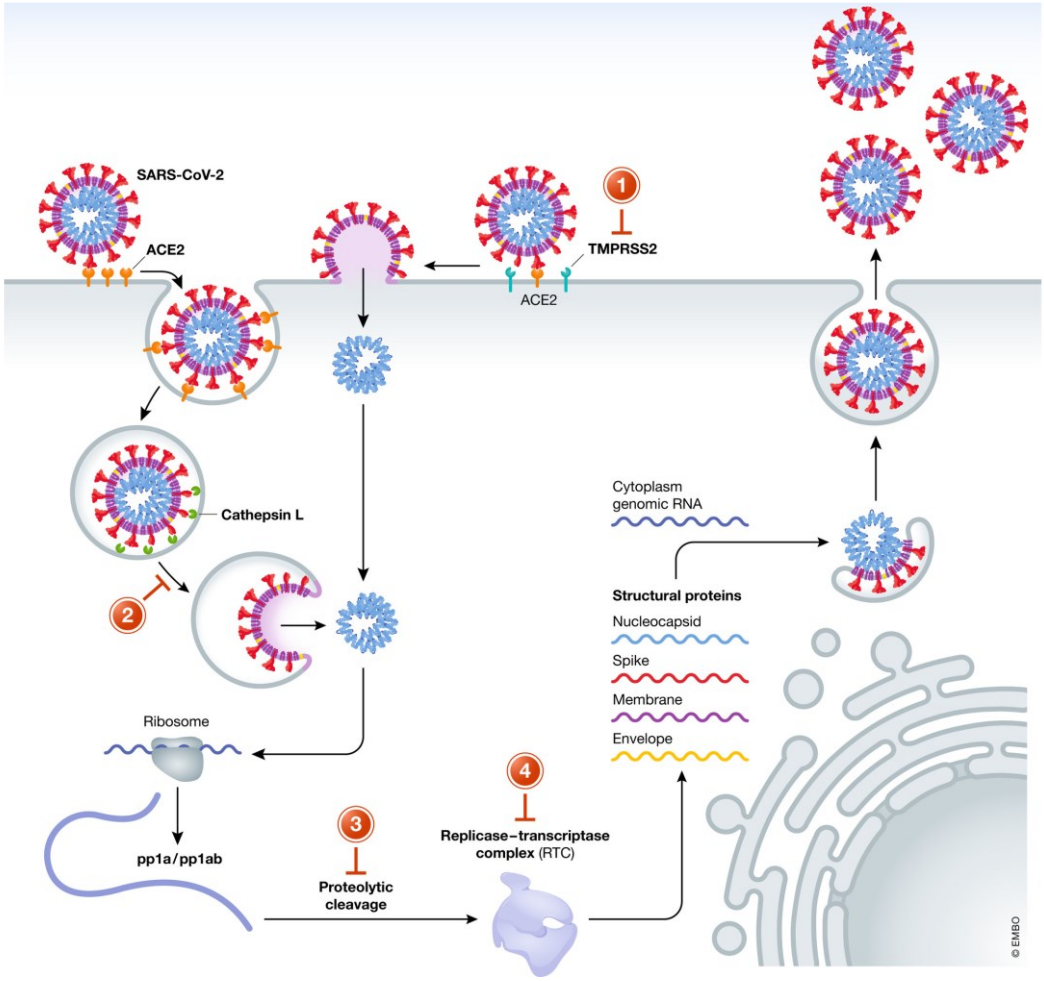



Figure 23. Life cycle of SARS-CoV-2 with highlighted potential target sites, starting with the entry into the cells to direct inhibition of the viral replication for the treatment of Covid-19. Figure taken from Ref.[266]

Monoclonal antibody therapy was supposed to be a helpful clinical treatment for the recovery of infected patients as well as prevention of infection.[267] SARS-CoV-2 shows high similarities with SARS-CoV (79.6% sequence identity[246]), with the latter having the availability of already established SARS antiviral antibodies.[268] 47D11 was the first reported human monoclonal antibody, which was able to neutralize SARS-CoV-2 by binding to a conserved epitope on top of the trimeric spike protein at the RBD. Due to the sequence similarity, this antibody is also effective for SARS-CoV.[269] By targeting this position on the viral surface, interaction with the human angiotensin converting enzyme 2 (ACE2) will be impaired, thus prevents the virus from entering the cells.[246, 269] The interaction of the virus with ACE2 would initiate conformational changes of the spike protein to enable membrane fusion.[251, 270]

Another possibility is to use immunomodulators (interferons or corticosteroids).[271] Type I interferon (IFN-I) response occurs within the cell after viral infections and induces the expression of interferon-stimulated genes (ISGs), which can inhibit viral replication and cause immune response by

promoting immune cell activation.[272] SARS-CoV-2 was shown to be sensitive to the treatment with recombinant human interferons (IFNα/β) *in vitro*. Upon administration, virus titers were clearly reduced by inhibiting viral replication.[273] Corticosteroids are frequently used as drugs for patients with severe disease progression after virus infection (SARS-CoV, MERS-CoV, SARS-CoV-2).[274] Though, recent studies proposed that administration of corticosteroids retards viral clearance, but no conclusive beneficial effect regarding mortality, reduced hospitalization duration or intensive care admission (and adverse effects) were reported. Thus, the treatment with this steroid should be carefully reconsidered.[274, 275]

The last option to mention is to use direct-acting antivirals (DAAs), interfering with viral replication (protease, helicase, polymerase inhibitor).[271] Implemental, viral proteases are used to provide mature active proteins (nonstructural proteins (NSPs)[276]) by cleaving longer precursor proteins (pp1a and pp1ab[276]) that are needed for viral replication.[276, 277] Accordingly, therapeutic usage of viral protease inhibitors could impair the formation of SARS-CoV-2 nsp, and thus prevent viral replication in infected cells.[276] One cleavage product of pp1ab by the viral protease is nsp13,[260] which is a viral helicase and forms a complex with RdRp for viral replication and transcription.[278] It is responsible for unwinding ds-RNA and ds-DNA.[260] The strand separation is necessary for viral replication within the host cell and makes it a viable target for drug development. Different potential inhibitors, like two natural myricetin and scutellarein (for SARS-CoV, impair ATPase activity, but not unwinding),[260] the anti-inflammatory drug Cepharanthine (SARS-CoV-2),[279] or the anti-leprosy drug clofazimine (SARS-CoV-2), which additionally impedes spike-mediated viral cell fusion, have been reported.[278] Besides protease and helicase inhibitors, inhibitors of the RdRp (nsp12) are the third members of DAAs and will be described in more detail in the following section.

#### 1.4.3.3 Antiviral nucleoside analogs as polymerase inhibitors

The viral RdRp is vital for replication and transcription of the viral genome based on the +ssRNA template. Due to its conserved components (e.g. 96% sequence similarity between SARS-CoV and SARS-CoV-2), RdRp of various viruses comprise mutual functions, which might facilitate broad-spectrum antiviral drugs, namely nucleoside analogue inhibitors, to be active against multiple CoVs.[280] Another prerequisite for an effective inhibitor besides the broad-spectrum activity is a high barrier for resistance and to evade the exonuclease activity (nsp14-ExoN) of the RdRp.[271] These antiviral nucleoside analogs with chemical alterations at the nucleobase or sugar scaffold are synthesized as precursors, which will be further processed to the active triphosphates by kinases after cellular uptake. They can then be used as nucleotide mimicries from the RdRp besides the four canonical cellular rNTPs for viral replication and transcription.[271, 281] Here, different inhibition mechanisms to interfere with viral RNA synthesis are possible upon incorporation depending on the structural features of the nucleoside analogue: chain termination, mutagenesis or depletion

of cellular nucleoside stores (Figure 24).[271] During normal replication of the viral +ssRNA, the four canonical rNTPs are used as substrates for chain elongation, leading to the full-length product (Figure 24, upper line).

If an antiviral nucleotide analogue is present, it can get incorporated into the viral RNA instead of the native nucleotide. This might either lead to a direct chain termination, or possibly delayed termination after a limited number of additional nucleotides, which interrupts viral replication. Thus, formation of the full-length product cannot be achieved, indicated by the STOP sign in Figure 24.

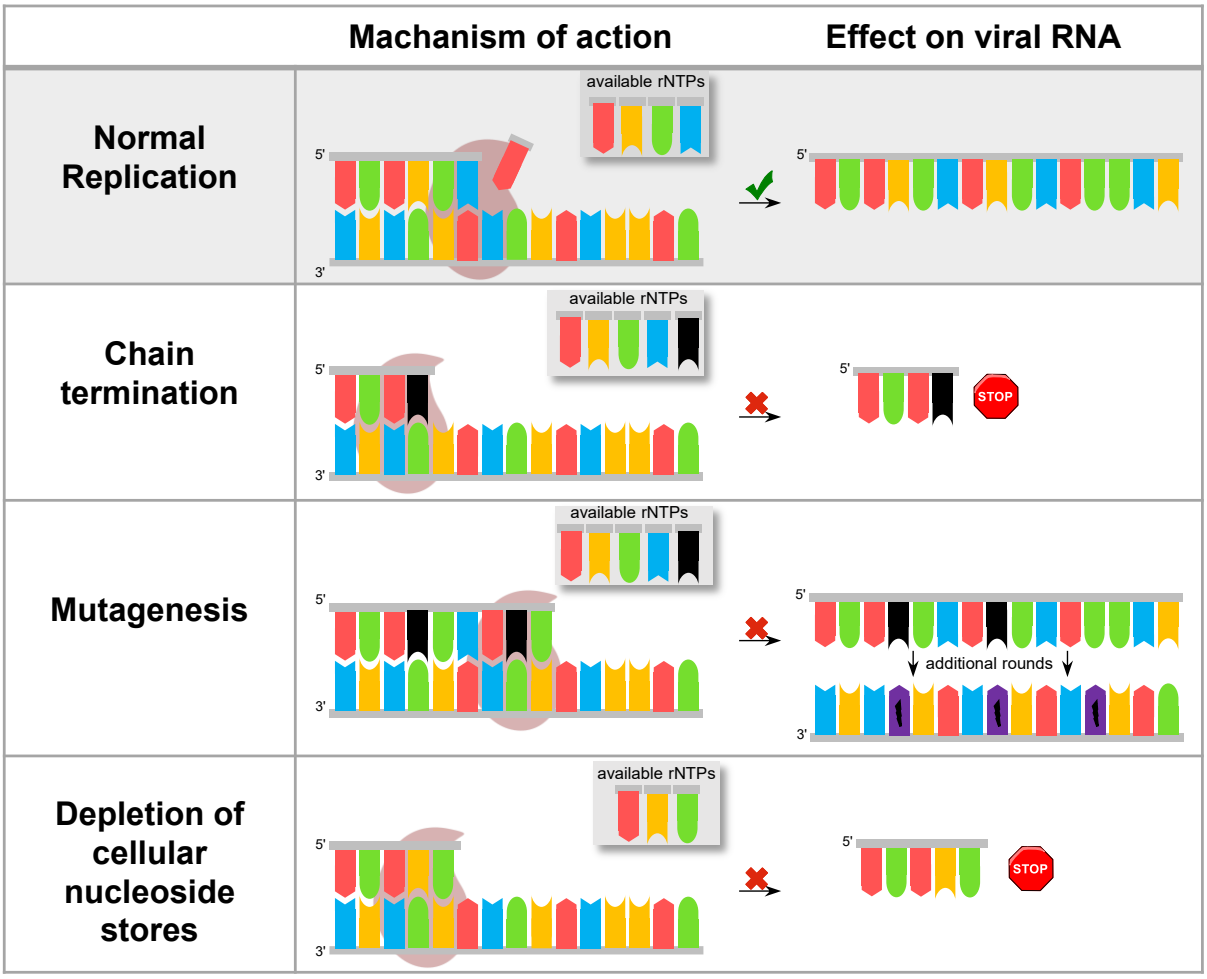



Figure 24. Possible mechanism of antiviral nucleoside analogs for the inhibition of RdRp (light brown). The available nucleoside triphosphates are always shown in the grey box. In the upper line (grey background) the normal replication mechanism is shown, leading to the full-length viral product. Upon incorporation of a nucleoside analogue (black), chain termination occurs, whereby a shortened strand is produced. A similar shortened viral RNA strand occurs by the depletion of one canonical nucleoside triphosphate, which is then not available anymore for incorporation. Some nucleoside analogs (black) are not leading to termination, but to mutagenesis. As a mimic it will get incorporated multiple times, leading to an accumulation of mutations (purple with black flash) with each round. Graphic reproduced from Ref.[271]

One example is the antiviral nucleoside analogue remdesivir (Gilead Sciences, GS-5734, Figure 26), a phosphoramidite prodrug of the adenosine derivative GS-441524, with a 2-ethylbutyl L-alaniate moiety attached to the phosphate group. The advantage of the masked monophosphate prodrug GS-5734 over GS-441524 is that it circumvents the rate-limiting step, namely the first phosphorylation.[282] For almost two years it was the only FDA-approved drug for treating COVID-19

patients.[283, 284] The prodrug is administered intravenously to hospitalized patients[285, 286] and has a broad-spectrum activity against multiple viruses, including CoVs (SARS[287], MERS[288]), filoviruses (Ebola[282, 289]) and pneumoviruses[290]. If the active form of remdesivir gets incorporated into the growing RNA chain, further extension by the RdRp stalls after three additional nucleotides. The reason for this is a translocation barrier caused by the C1'-cyano group of remdesivir. There is a steric clash between the cyano group and the side chain from serine-861 in the nsp12 subunit (Figure 25).

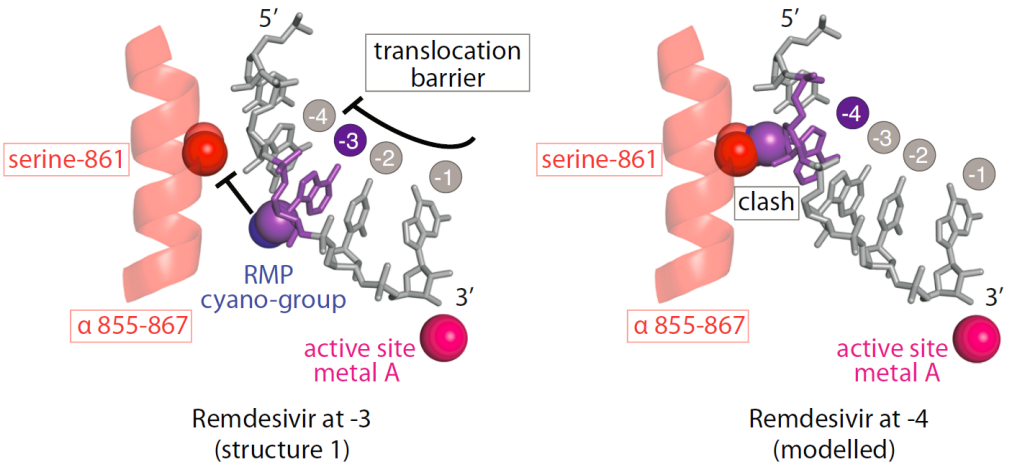



Figure 25. Mechanism of action of remdesivir. The cyano-group at the 1'-position of the sugar, shown in purple, is inducing a translocation barrier when trying to bypass serine-861 of the nsp12 subunit.[265] (Creative Commons Springer Nature).

Still, this mechanism does not cause a complete chain termination, as the stalling of the viral RdRp can be circumvented by high NTP concentrations.[265] Additionally, remdesivir cannot entirely evade proofreading. Nsp14 of RdRp is supposed to be able to remove not only one, but multiple nucleotides from the double stranded 3'-end, presumably with the help of the viral helicase (nsp13[291]), thus further reducing the efficiency of remdesivir.[265, 292] Other examples are galidesivir and aciclovir (Figure 26). Galidesivir (BCX4430), is a non-obligate (belated) chain-terminator inhibiting several RNA viruses like flaviviruses (zika virus[293]), filoviruses (Ebola and Marburg-Virus)[294-296] and CoVs.[297] Docking studies revealed that the inhibitory mechanism relies on the tight binding of Galidesivir to the RdRp of SARS-CoV-2,[298] but lacks conclusive *in vivo* studies. Aciclovir is known for the treatment of the herpes simplex virus (HSV) and varicella zoster virus (VZV).[299] It can be administered as the prodrug valaciclovir, which is selectively metabolized to the active form by viral thymidine kinase in infected cells and then to the triphosphate by cellular enzymes. In contrast to the two aforementioned nucleoside analogs, aciclovir resembles an obligate chain terminator, as it directly blocks chain elongation, caused by the missing 3'-hydroxy group.[299, 300] A combinatorial treatment with favipiravir was suggested to treat SARS-CoV-2.[301] However, a fleximer based on acyclovir only showed activity against human CoVs ((HCoV)-NL63) and MERS, but not SARS.[302]

Alternatively, reduction of intracellular canonical nucleoside levels by nucleoside mimics might occur, resulting in a diminished viral synthesis (Figure 24). One example for such an antiviral nucleoside analogue with a broad-spectrum antiviral activity is ribavirin (1-β-ribofuranosyl-1,2,4-triazole-3-carboxamide). It was reported that administration of ribavirin causes a drop of intracellular GTP levels *in vitro*, as the monophosphate form inhibits host inosine monophosphate dehydrogenase (IMPDH), which is needed for the biosynthesis of guanosine nucleotides. Hence, the mutagenic effect of ribavirin is based on disruption of the viral RNA replication due to missing natural substrates and an increased immune response.[303-305] The antiviral effect of ribavirin was tested for many respiratory RNA viruses, like influenza virus[306], CoVs[307] and hepatitis C virus (HCV[305]).[280] Ribavirin was additionally proposed to act as a mutagen, being able to base pair with cytidine and uridine (Figure 26).[308] Lately it was also tested for SARS-CoV-2, but did not reveal beneficial effects for the clinical outcome.[309] Hence, combinatorial administration with interferon-α is under current investigation.[280] Mizoribine is another example of an inhibitor of IMPDH (Figure 26).[310] Consistent with ribavirin, mizoribine is a selective inhibitor of SARS-CoV, impeding the replication of the virus even more effectively, however not entirely.[311] Recently, it was proposed as a treatment against heat shock protein (HSP60), which seems to translocate upon SARS-CoV-2 infection and increases inflammatory responses.[312]

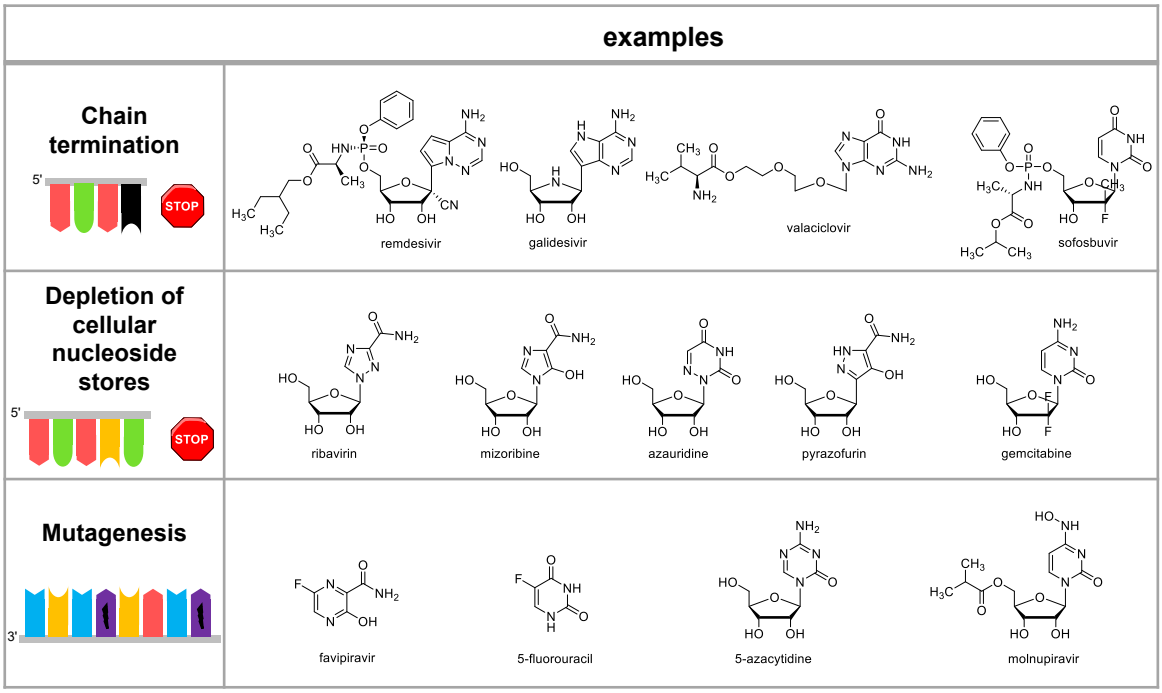



Figure 26. Examples of antiviral nucleoside analogs corresponding to the three different groups, depending on their proposed main mechanism of action.

6-Azauridine and pyrazofurin exhibit a similar mechanism (Figure 26). They show inhibitory activity against the orotidine monophosphate decarboxylase, needed for the *de novo* synthesis of pyrimidine triphosphates, which is supposed to be the underlying mechanism for the antiviral efficiency.[313] *In vitro* docking studies demonstrated 6-azauridine to be a good potential candidate for

inhibiting SARS-CoV-2,[314] but further *in vivo* studies are not yet available. One last example for this class of antiviral nucleoside analogs is gemcitabine (Figure 26), an FDA approved anti-cancer therapeutic (GemzarR),[315] which has also shown broad spectrum antiviral activity against HCV,[316] zika virus,[317] influenza virus,[318] HIV[319] and CoVs (SARS, MERS[320]) and was recently shown to be active against SARS-CoV-2 in cell culture.[321] It interferes with viral replication by inhibiting pyrimidine biosynthesis, presumably by targeting the salvage pathway to synthesize C and/or U,[321] like it was first shown for enteroviruses.[322]

In the last example, the foreign nucleoside mimicry might replace a canonical nucleotide, inducing mispairing (Figure 24). Thereby, mutations that occur are accumulating with each round of replication. This is described as error catastrophe, due to the loss of viral viability (= lethal mutagenesis[323]).[271] Thus, this class of nucleoside analogs does not directly interfere with the polymerase, but affects the resulting viral RNA product. The aforementioned favipiravir (T-705, Figure 26), a pyrazincarboxamide derivative, is a broad-spectrum antiviral compound, selectively inhibiting the viral RdRp of several viruses,[280] including flaviviruses (Ebola[324]), influenza virus[325], norovirus[326] and zika-virus[327]. After metabolism (phosphoribosylation), the active form of favipiravir gets efficiently incorporated by RdRp of SARS-CoV-2 as a purine analogue into the elongating RNA strand, mimicking guanine as well as adenine NTPs.[281] The emerging misincorporations lead to mutations in the resulting viral RNA product, and to further accumulations, as favipiravir seems to evade the proofreading activity.[281] This explains the G-to-A and C-to-U transition mutations found in *in vitro* studies.[328] Thereby, favipiravir seems to act primarily as a GTP analogue. This mechanism was further confirmed by cryo-EM analysis, solving the structure of the pre-catalytic state with favipiravir bound to the replicating polymerase complex.[281, 328] In addition, the antiviral effect of favipiravir as a chain terminator was shown.[328] In contrast, 5-fluorouracil (Figure 26) mimics pyrimidine nucleosides, enabling the incorporation of both GTP and ATP.[329] However, 5-fluorouracil is supposed to be removed from the viral RNA by ExoN proofreading during replication.[330] 5-Azacytidine was shown to be an active mutagen for HIV.[331] A similar mechanism of action is supposed for molnupiravir (EIDD-2801/MK4482), an isopropyl ester prodrug of β-D-$N^4$-hydroxycytidine (NHC, EIDD-1931). It was initially tested as a treatment against influenza virus,[332] but proved to be useful as a broad-spectrum antiviral nucleoside analogue against multiple RNA viruses, like HCV,[333] ebola virus,[334] bovine viral diarrhea virus (BVDV),[333] venezuelan equine encephalitis virus (VEEV),[335] CoVs (SARS[336], MERS) in non-clinical models. Molnupiravir was further tested as a potential drug against SARS-CoV-2 in various studies.[337-339] The mutagenic mechanism of action was first shown in 1974:[340] Investigations of the influence of NHC on bacteriophage Qβ RNA revealed a G-to-A transition within the sequence context. The incorporation of NHC monophosphate leads to these efficient transition mutations, as it is mimicking both UMP and CMP in the template, based on two

tautomeric forms of this compound.[340] The mechanism of action of molnupiravir will be further discussed in the next section (see 3.2.3), as part of this thesis.

1.4.3.4 Molnupiravir

The isopropyl ester prodrug of NHC, molnupiravir, is hydrolyzed and further activated to the triphosphate within the cytoplasm by cellular kinases.[335, 341] Compared to other approved drugs like remdesivir which needs to be administered intravenously, the prodrug is orally bioavailable.[342] The RNA-like building block is accepted by the viral RdRp as an active substrate for the synthesis of the viral genome and is supposed to lead to error catastrophe (Figure 27). A biochemical *in vitro* polymerase assay demonstrated an increase of mutations (G-to-A and C-to-U transition mutations for influenza virus), caused by the incorporation of the NHC-TP instead of CTP, while not acting as a chain terminator.[342] A recent report also showed that NHC is a substrate for the mitochondrial DNA-dependent RNA polymerase (POLRMT), but does not impair mitochondrial function.[343]

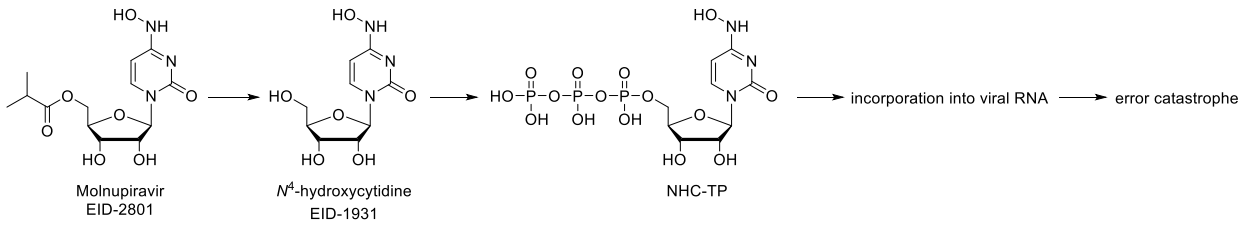



Figure 27. Metabolism of the isopropyl ester prodrug molnupiravir to form the triphosphate, which is leading to error catastrophe. After a rapid hydrolysis in cytoplasm, EID-1931 gets activated by host kinases to the triphosphate, as the active antiviral agent.[338]

The dual functionality of NHC and related substrates like $N^4$-methoxycytidine (NMC) is already known for more than 50 years. There have been various reports, including NMR, IR and crystallography, analyzing the behavior of these monomers as well as their interactions with the corresponding base pairing partners. The equilibrium of the two tautomeric forms, namely amino and imino (Figure 28A), seems to be dependent on the solvent.[344] In both aqueous and non-aqueous media, the imino form is predominant.[345-349] X-ray studies and IR spectroscopy proposed an intramolecular hydrogen bond ($N^4$-OH···N3) to stabilize this tautomeric form,[349, 350] the $N^4$OR moiety orientated syn to N3.[345, 347] With a $K_T$ of ~10 in aqueous media, around 10% of NHC are present in the amino form,[345] whereas 0.5% are in the anti-amino form by estimation.[350]

These experimental results were further validated by theoretical *ab initio* quantum mechanical calculations in 1993 (Figure 28B).[351] In accordance with the previous data, the syn-imino form is the most stable form, followed by its rotamer form (+12.8 kJ/mol). The difference between the syn- and anti-form of the amino form is almost not distinguishable, but both are markedly less stable than the imino form.[351] The same trend was also observed in a recent computational analysis.[352]

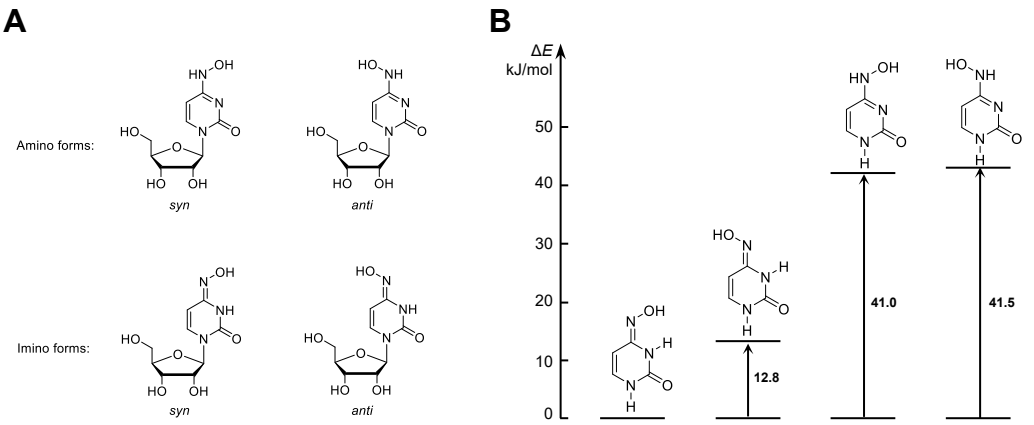



Figure 28. Tautomeric forms of $N^4$-hydroxycytidine. A: Amino and imino tautomers with the corresponding syn- and anti-forms to N3. B: Theoretical calculations propose the syn-imino to be the most stable form. Reproduced from Ref.[351]

In the amino form, NHC should be able to form a Watson-Crick base pair with guanine, as a cytidine mimicry, whereas in the oxime form two base pair formations are possible: either Watson-Crick base pair formation with adenine, or a Wobble formation with guanine (Figure 29). This would be in accordance with prior experimental data for the NHC derivative NMC.[353-355]

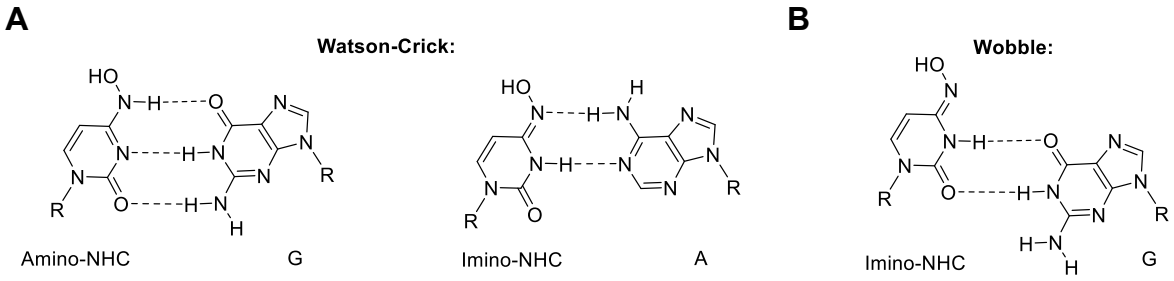



Figure 29. NHC can form different base pairs, depending on the tautomeric form. A: Watson-Crick base-pair formation between the amino form NHC:G or between the imino form NHC:A. B: wobble base pair formation between the imino form NHC:G.

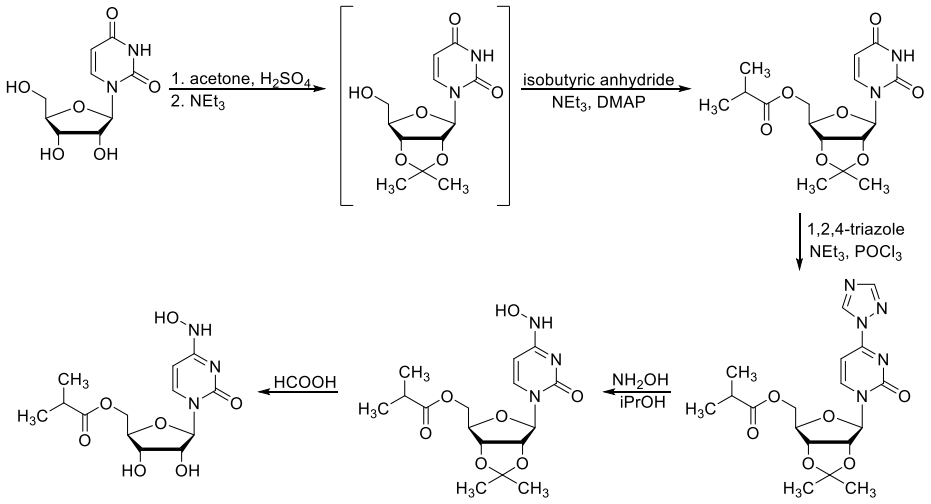
The potency of NHC as an antiviral agent against MERS, SARS-CoV and SARS-CoV-2 was tested using human airway epithelial cells lines.[339] NHC was an effective inhibitor of viral replication for both CoVs (SARS, MERS). Thereby, A-to-G and U-to-C transition mutations were reported for MERS-RNA *in vitro*. Additionally, first *in vivo* tests with NHC were performed to investigate the prophylactic and therapeutic potential for viral inhibition in mouse models and proved to be efficient.[339] The efficacy of molnupiravir as a potential therapy for the pathogenic SARS-CoV-2 was further tested in a ferret model.[337] Therapeutic administration of the drug to infected animals twice a day markedly reduced the viral load in the upper respiratory tract. In addition, molnupiravir blocked the SARS-CoV-2 transmission to untreated contact-ferrets without showing any apparent adverse effects.[337]

By early 2020, molnupiravir was tested in clinical trials in humans by MERCK in collaboration with Ridgeback Biotherapeutics. Phase 1 was performed with healthy participants, as a randomized,

double-blind and placebo-controlled study. After oral administration of different doses of the prodrug, safety, tolerability and pharmacokinetics were analyzed.[338] In conclusion, molnupiravir was well absorbed, showed good plasma concentrations and was well tolerated, and in addition showed no severe adverse events.[338] Due to these promising results, Phase 2-3 (MOVe-OUT) were started in October 2020.[356] The objective of this study was equivalent to phase 1, but involved infected, but non-hospitalized adults with mild to moderate symptoms. Up to 800 mg molnupiravir were administered twice daily for 5 days and the safety, efficacy and pharmacokinetics were analyzed.[356] Based on the results of phase 2, phase 3 focused on participants in the early stage of the infection, not later than 5 days after the appearance of symptoms. With an early treatment (800 mg twice daily for 5 days), antiviral drugs should be most useful to prevent hospitalization and the spread of the virus. Besides, participants should have at least one risk factor which might lead to a severe progression of the virus infection.[356, 357] According to an interim report about phase 3 on October 1, 2021, administration of molnupiravir reduced the risk of hospitalization or death by about 50% compared to placebo. To provide this valuable therapy, the study was discontinued and instead application for emergency use was applied for.[358] Shortly thereafter, the final evaluation revealed only a reduction of the risk of about 30%. Still, with proven efficacy, it was supposed to be advantageous for patients with a high risk of a severe disease progression.[359] On December 23, 2021, The US Food and Drug Administration (FDA) released the Emergency Use Authorization (EUA) to provide molnupiravir as a drug for the treatment of SARS-CoV-2 in early stages of infection for adults with high risk of severe sickness or death if no other FDA approved Covid-19 medication is available or suitable.[360]

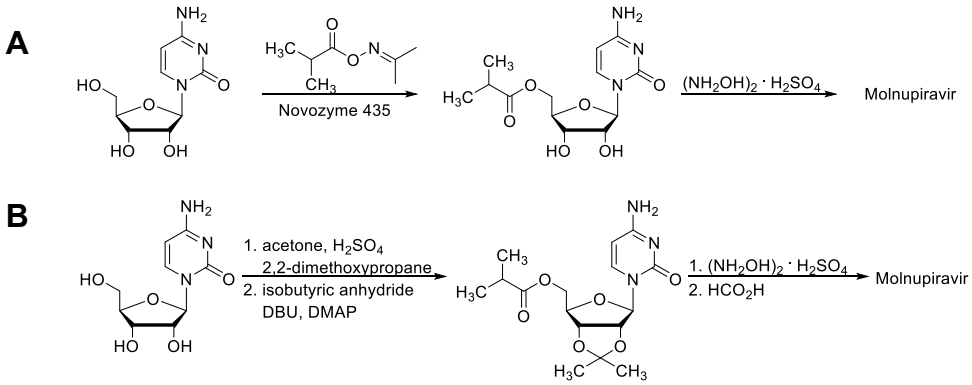
#### 1.4.3.5 Synthesis of molnupiravir

Different methods are available to synthesize the isopropyl ester prodrug of NHC, starting from either cytidine or activated uridine and substituting with hydroxylamine. The initial synthesis for molnupiravir was reported in 2019 as a five-step synthesis starting from uridine and involved the protection of the sugar hydroxy groups (Scheme 1).[361] Acetonide protection of the 2'- and 3'-hydroxy groups, followed by esterification and activation with 1,2,4-triazole facilitated substitution with hydroxylamine. Subsequent acidic deprotection of the acetonide moiety resulted in the final product molnupiravir. However, the overall yield was only up to 17%.[361] One year later, Kappe and coworkers reported an improved route of this patented synthesis pathway, involving less purification and isolation steps, increasing the yield up to 61% by reordering the reaction sequence. During deprotection, two impurities were obtained – free NHC (EID-1931) and 5'-isopropylester-protected uridine, which necessitates to choose the reaction conditions carefully.[362]



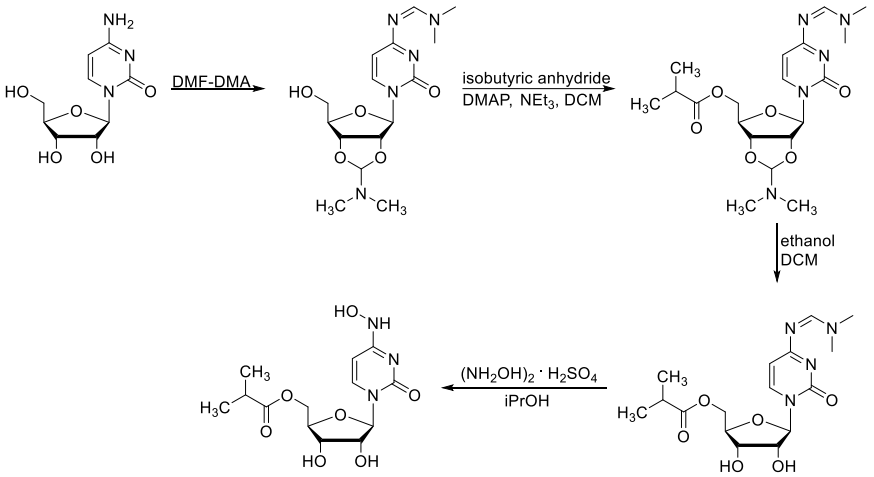

Scheme 1. Initial synthesis pathway for the prodrug molnupiravir starting from uridine. After 5 steps, the overall yield was only up to 17%.[361]

Changing to cytidine as starting material seemed beneficial, due to a lower cost of cytidine compared to uridine, and the opportunity to reduce the number of the synthesis steps.[363] Snead and coworkers reported a two-step synthesis to obtain molnupiravir starting from cytidine, without the need for protection groups (Scheme 2A). Esterification was performed enzymatically (enzyme mediated selective esterification), with isobutyric oxime ester and immobilized lipase (Novozyme 435). Direct transamination was achieved in high yields with the hydroxylamine sulfate in isopropanol, giving an overall yield of 75%.[363] The same group presented a follow up study, in which they circumvented the need for enzymatic esterification, for a less expensive strategy (Scheme 2B).[364] The four steps synthesis route was performed similar to the initial route, but starting from cytidine. Acetonide protection allowed specific 5'-esterification, followed by direct transamination with hydroxylamine sulfate. After deprotection, molnupiravir was isolated with an overall yield of 44%, but also with the occurrence of some impurities under cleavage conditions of the acetonide group.[364]




Scheme 2. Two synthesis pathways from Snead and coworkers starting from cytidine. A: Selective esterification with an immobilized lipase (Novozyme 435).[363] B. Esterification after 2',3'-acetonide protection and subsequent transamination with hydroxylamine.[364]

Shen and coworkers provided a new approach using *N*,*N*-dimethylformamide dimethyl acetal (DMF-DMA) as an alternative protection group instead of the acetonide group (Scheme 3).




Scheme 3. Changing the protection strategy from acetonide to DMF-DMA, followed by 5'-esterification and transamination.[365]

They entitled it as an ‘one-pot’-synthesis, as there is no need for chromatographic purifications, which might provide potential for industrial applications. Molnupiravir was obtained in high quality with an overall yield of 70%.[365]

This summarized the current state of knowledge and provided background information for the discussion of the results described in chapter 3. It highlighted some important aspects of nucleoside analogs, concerning both naturally occurring and artificial nucleosides. The analogs help with the better understanding of cellular functions and biological processes. Further, the analogs can serve as fluorescent tools or structural elements to study DNA-protein interactions and as efficient drugs for antiviral therapy.

# 2. Research Objectives and thesis outline

The goal of this thesis was to develop efficient methods for the preparation of modified RNA and DNA oligonucleotides containing various natural and artificial nucleobase variations. This required the synthesis of phosphoramidite building blocks and nucleoside triphosphates and their incorporation into oligonucleotides. The synthetic nucleic acids were characterized in detail and used in collaborative projects to investigate biological questions related to the analysis of tRNA modification pathways, the mode of action of antiviral nucleosides, and the method development for RNA-protein crosslinking. The general context of these topics was introduced in chapter one. The collaborations involved Prof. Dr. Markus Bohnsack (University Göttingen), Prof. Dr. Patrick Cramer (MPI multidisciplinary sciences, Göttingen) and Prof. Dr. Christian Häring (University Würzburg, Biocenter) and their coworkers, and the results of the respective projects will be presented in four major sections as outlined below (Figure 30).

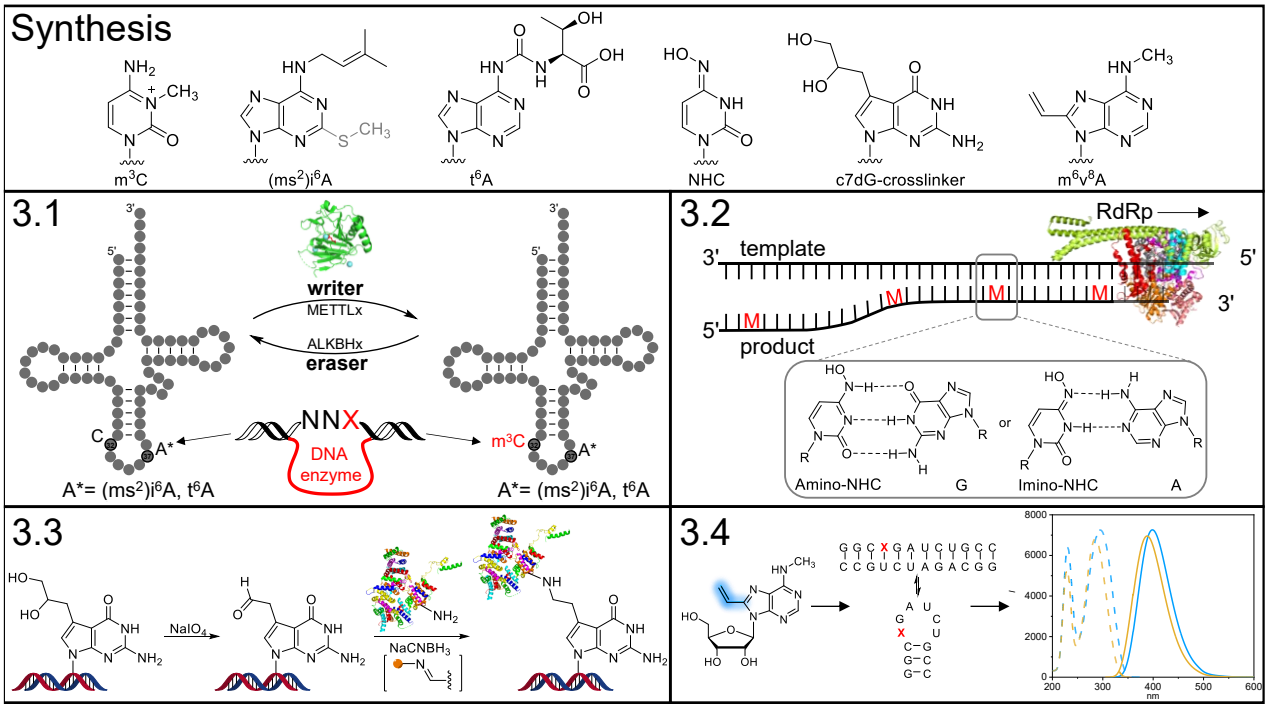



Figure 30. Overview of the four projects worked on in this thesis after the successful synthesis of the building blocks. 3.1: Study of writer and eraser proteins and reader deoxyribozymes. 3.2: Investigating the mechanism of action of Molnupiravir. 3.3: DNA-protein interactions between deaza-Guanosine crosslinker $c^7$dG and Condensin. 3.4: Fluorescent nucleoside analog 8-vinyl-$m^6$A.

The first of these four projects deals with the synthesis of the naturally occurring tRNA modifications 3-methylcytidine ($m^3C$), $N^6$-isopentenyladenosine ($i^6A$), 2-thiomethyl-$N^6$-isopentenyladenosine ($ms^2i^6A$) and $N^6$-threonylcarbamoyladenosine ($t^6A$) as phosphoramidite building blocks. These modifications were used to i) investigate RNA-protein interactions to gain insights into their biological relevance and structural influence on the tRNA and ii) to generate enzymatic tools for the detection of modifications in the form of DNA enzymes. The phosphoramidite building blocks were incorporated into certain positions, either 32 or 37, of the ACL of specific tRNAs by solid phase

synthesis. With these strands, both methylation of cytidine at position 32 with RNA methyltransferase METTL8 and demethylation of $m^3C_{32}$ with ALKBH1 and ALKBH3 were analyzed in collaboration with the laboratory of Prof. Dr. Markus Bohnsack (Figure 30, 3.1). First, the impact of these modifications on the activity of METTL8 was shown and the structural influence of $m^3C_{32}$ on the anticodon stem loop (ASL) structure was to be proven by PAGE, NMR, CD and melting curve analysis. Furthermore, demethylation assays of mitochondrial and cytoplasmic tRNA-ASLs with ALKBH1 and ALKBH3 were performed to spot interdependencies of the modifications and recognition elements of the enzymes. As an additional goal, the already mentioned naturally occurring modifications as well as $N^4$-methylcytidine ($m^4C$) and $N^4$,$N^4$-dimethylcytidin ($m^4{}_2C$) were used to facilitate the implementation of novel tools for the detection of further modified nucleosides. By means of *in vitro* selection, my college Anam Liaqat found RNA-cleaving DNAzymes that show different activities depending on the modification state of the RNA substrate. These DNAzymes can be considered as functional mimics of reader proteins, that recognize the presence of a modification and determine the fate of the respective RNA.

The second project focuses on the antiviral nucleoside analogue Molnupiravir in collaboration with Florian Kabinger from the Cramer laboratory (Figure 30, 3.2). The active substrate $N^4$-hydroxycytidine (NHC) was synthesized as phosphoramidite building block and incorporated into different RNA substrates for biochemical analyses (e.g. RNA extension assays, cryogenic electron microcopy (cryo-EM)) to unveil the mechanism of action of Molnupiravir. Detailed insights about the base pairing abilities of the two tautomeric forms of NHC and the effect on duplex stability should be provided by NMR measurements and melting curve analyses. To achieve this, also a $^{15}N$-labeled version of NHC was synthesized.

The third project deals with the synthesis and incorporation of the crosslinker 7-deaza-7-(2,3-dihydroxypropyl)-guanine for the generation of a covalent bond between DNA and the protein complex condensin. The diol functionality of the crosslinker can be unmasked to an aldehyde, which can react with amino acids Lys or Arg-residues within condensin (Figure 30, 3.3). In collaboration with Julia Locherer from the Häring laboratory the aim is to investigate how the DNA runs through the complex for the structural maintenance of chromosomes, by forming covalent DNA-protein-crosslinks. Therefore, both the phosphoramidite building block and the triphosphate were synthesized. The latter was enzymatically incorporated to verify the functionality of the crosslinker.

The last project addresses the synthesis of a new fluorescent nucleoside analog 8-vinyl-$N^6$-methyladenosine ($m^6v^8A$, Figure 30, 3.4). The free nucleoside as well as the phosphoramidite building block should be synthesized, to compare the fluorescent properties of the monomer and the oligonucleotide containing this fluorescent nucleoside. It should be examined if this fluorescent modification can be used to monitor the enzymatic demethylation activity by fluorescent changes.

# 3. Results and discussion

This chapter comprises an overview of the results obtained during this thesis. Chemical syntheses of the natural or artificial nucleoside modifications as phosphoramidite and/or triphosphate building blocks will be described. This includes newly developed syntheses pathways, and the improvement of previously published compounds. Subsequent incorporation of the substrates by solid phase synthesis or using enzymatic methods will be presented, followed by the respective deprotection and workup conditions for each modification and the unmodified reference strands. Furthermore, the respective biochemical assays, thermodynamic and spectroscopic properties (e.g. NMR, CD) for the different research projects will be evaluated.

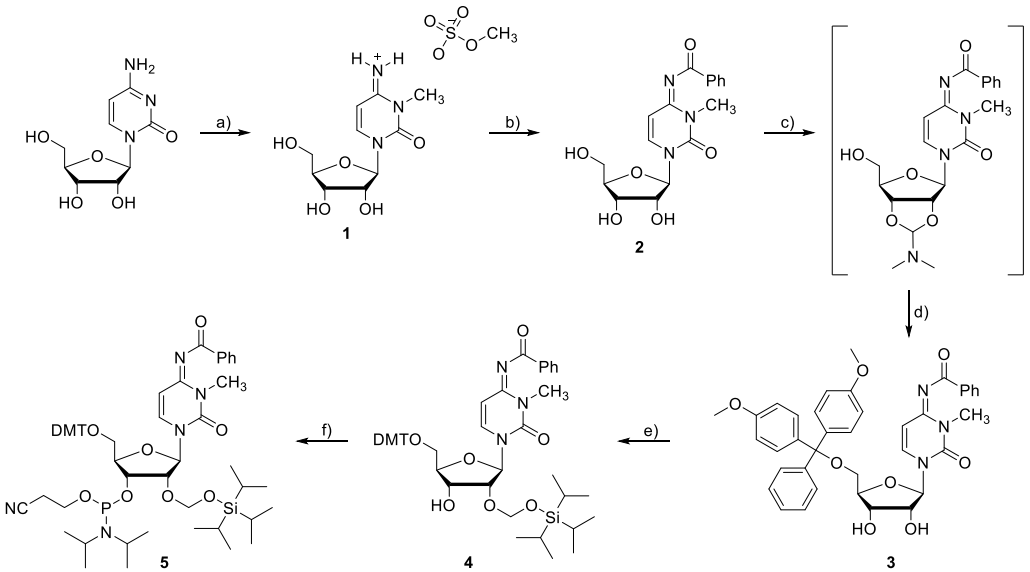
## 3.1 Naturally occurring tRNA modifications

In the first part, several naturally occurring nucleoside modifications were synthesized as phosphoramidite building blocks for the incorporation into RNA strands by solid-phase synthesis. This section contains the investigations of tRNA-ASLs with RNA methyltransferase METTL8 as well as the demethylation with ALKBH1 and ALKBH3. Additionally, the last part deals with *in vitro* selection of RNA-cleaving DNAzymes. The following modifications were obtained for these studies: $m^3C$, $m^4C$, $m^4{}_2C$, $i^6A$, $ms^2i^6A$, and $t^6A$, of which the chemical synthesis will be described first.

### 3.1.1 Synthesis of phosphoramidites of naturally occurring tRNA modifications

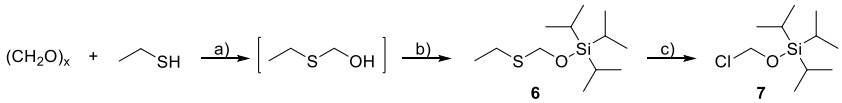
#### 3.1.1.1 Modified cytidines

***3-Methylcytidine* ($m^3C$)**




Scheme 4. Synthesis of the $m^3C$ phosphoramidite building block **5.** a) Dimethyl sulfate, DMF, 37 °C, 30 min (67%) b) 1. TMSCl, pyridine, r. t., 1 h; 2. $NEt_3$, BzCl, 4 h, r. t. (89%) c) DMF-DMA, pyridine, r. t, o. n. d) 1. DMT-Cl, pyridine, r. t., 2 h; 2. MeOH, 10 min (89%) e) 1. $^iPr_2NEt$, $Bu_2SnCl_2$, DCE, r. t., 1 h; 2. TOM-Cl, 80°C, 20 min (49%) f) $Me_2NEt$, CEP-Cl, DCM, r. t., 2.5 h, (91%).

The synthesis of the m$^3$C phosphoramidite **5** was conducted over five steps. Orthogonal protection groups were introduced (Scheme 4) to allow selective deprotection during solid phase synthesis and the subsequent purification. With dimethyl sulfate the methyl group was selectively attached to N3 of cytidine, resulting in the positively charged free m$^3$C nucleoside **1**, which was obtained after precipitation and used in the next step without further purification.[366] For selective protection of N4, the hydroxy groups of the sugar moiety needed to be silylated with TMS-Cl in pyridine first.[367] Subsequent treatment with benzoyl chloride did not lead to any product formation. Instead, triethylamine had to be added before for the deprotonation of the positively charged imine. Mild basic deprotection conditions cleaved the temporary silyl groups and afforded to the benzoyl-protected m$^3$C **2**. The next step would have been the protection of the 5' position with 4,4'-dimethoxytritylchloride (trityl-chloride, DMT-Cl),[368] which is needed for solid-phase synthesis. Under mild acidic conditions (e.g. dichloroacetic acid), a DMT-cation is formed, of which the absorption can be measured at 495 nm, due to its orange color. This allows the monitoring of the coupling efficiencies during solid-phase synthesis.[369] Normally, the tritylation occurs regioselectively at the 5' position, as this is a primary hydroxy group and sterically less hindered compared to the 2' and 3'-OH groups. The sterically demanding trityl group then blocks the 2' and 3' positions and impedes reaction with DMT-Cl. However, for **2** double tritylation occurred, even if the DMT-Cl was added in portions. Thus, the 2' and 3' hydroxy groups had to be protected before the incorporation of the 5'-DMT-group. DMF-DMA, which is forming an unstable 2',3'-acetal, is an advantageous approach, as the *in situ* formed product can directly be converted further with DMT-Cl. After completion of the reaction, the acetal is easily cleaved off under workup conditions with methanol. This protection strategy markedly improved the reaction yield from 38 to 89%. Compound **3** was then activated *in situ* with $Bu_2SnCl_2$, resulting in a cyclic 2',3'-*O*-dibutylstannylidene intermediate (Figure 31A), and treated with (triisopropylsiloxy)methyl chloride (TOM-Cl), resembling a fluoride-labile protection group, to yield the 2'-*O*-silyloxymethylated m$^3$C **4**.




Scheme 5. Synthesis of TOM-Cl (**7**) as a reagent for 2'-silylation, based on a published procedure.[369] a) $NaOH_{cat}$, 40 °C, 1 h b) triisopropylsilyl chloride (TIPS-Cl), imidazole, DCM, 25 °C, 14 h (54%) c) $SO_2Cl_2$, DCM, 25 °C, 1 h (24%).

This reagent had to be synthesized over two steps based on a published procedure,[369] starting with ethanethiol and paraformaldehyde (Scheme 5) under base catalysis. The hydroxy group of the ethylthiomethanol intermediate was treated with triisopropylsilyl chloride (TIPS-Cl) forming the silylated compound **6**, which was further transformed with $SO_2Cl_2$ to the silylation reagent TOM-Cl **7**. The reaction of the nucleoside with TOM-Cl can likewise take place at 2'- or 3'-OH. Besides, only 1.2 equivalents of TOM-Cl were used to avoid additional alkylation. Thus, partially starting material remains together with the two formed isomers, explaining the relatively low yields obtained

for this reaction. Pitsch *et al.* proposed that reaction at 80 °C in 1,2-dichloroethane (1,2-DCE) is beneficial for the alkylation of 2'-OH, whereas, at lower temperatures in other solvents like THF or MeCN, preferentially the 3'-*O*-alkylated product is formed. Additionally, using $^{t}Bu_2SnCl_2$ as an activation agent compared to $^{n}Bu_2SnCl_2$ gives 5-10% higher yields, but is much more expensive.[369]

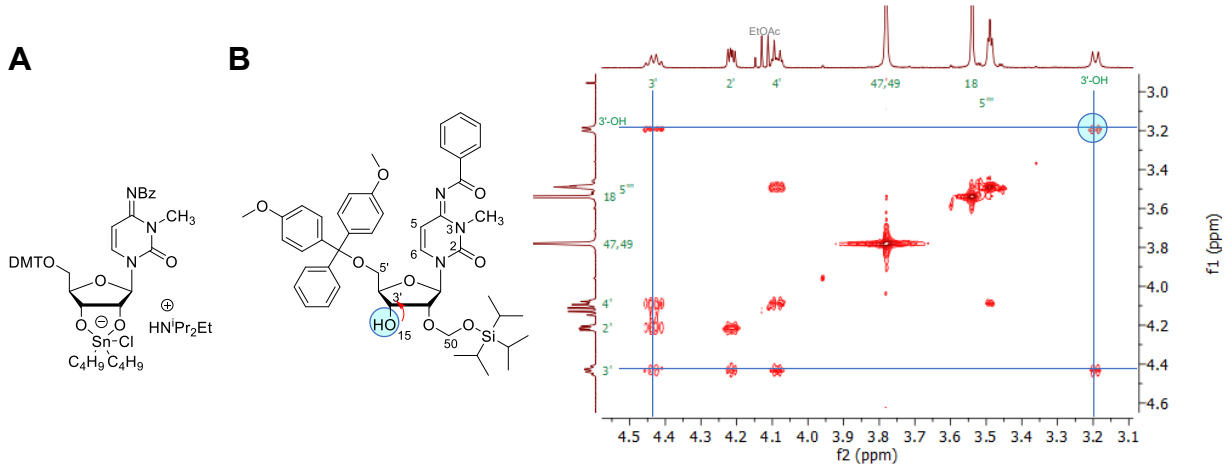
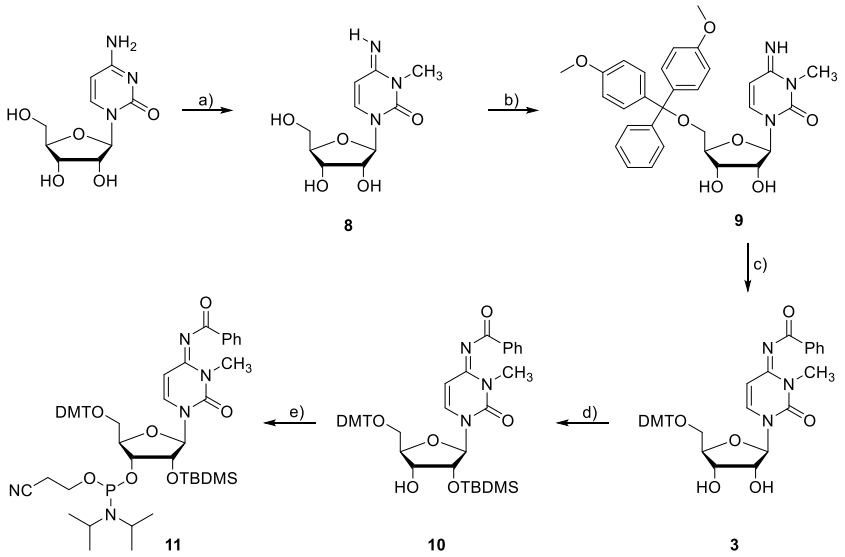



Figure 31. TOM-protection of the m$^3$C nucleoside. A: Activated intermediate, which is formed before the addition of TOM-Cl. B: Excerpt of the COSY-spectra of the 2'-TOM-protected compound **4**. After separation of the two TOM-protected isomers, they can be differentiated by the assignment of the 3'-OH group (light blue circle), by correlations between the 3'-H and the 3'-OH, indicated with dark blue lines.

The two isomers had to be separated by column chromatography and assigned by 2D spectra. They could clearly be distinguished by NMR, as a correlation between the 3'-H and the 3'-OH was observed in the COSY spectra for the desired 2'-TOM-protected nucleoside **4** (Figure 31B, indicated by the blue square) and the respective correlations between 2'-H and 2'-OH for the undesired 3'-TOM side product.

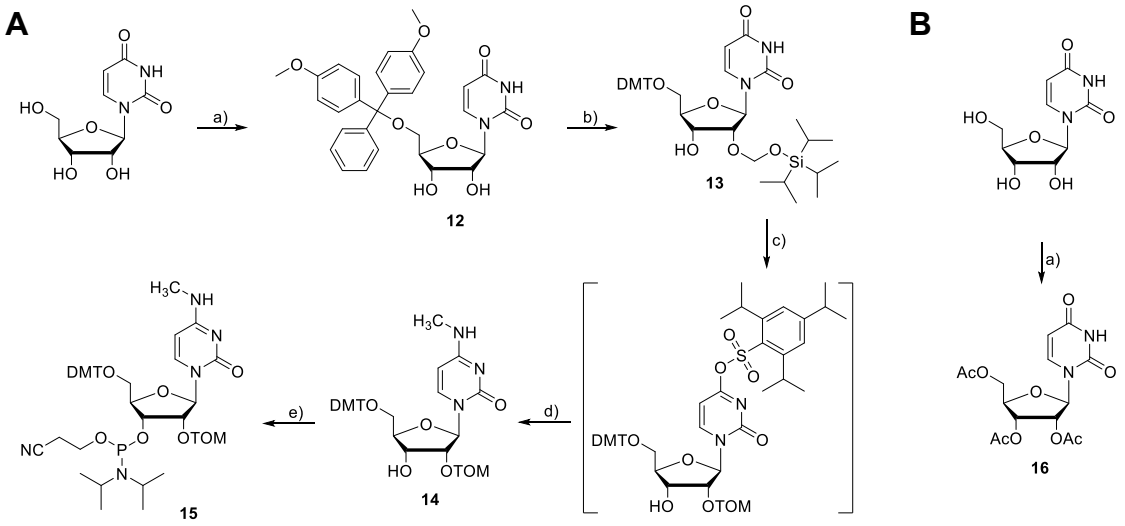



Scheme 6. Alternative synthetic pathway for the synthesis of the m$^3$C phosphoramidite building block **11** reported by Sheng and coworkers.[370] a) MeI, DMF (70%) b) DMT-Cl, pyridine (43%) c) TMS-Cl, pyridine, Bz-Cl (82%) d) TBDMS-Cl, imidazole, DMF (39%) e) CEP-Cl, $^{i}Pr_2NEt$, 1-methylimidazole, DCM (66%).

After isolation of the 2'-TOM-protected **4** (elutes faster from column than the 3'-isomer), the last step was the phosphitylation of the 3'-hydroxy group. This can generally be achieved in two ways: with the less reactive 2-cyanoethyl *N,N,N',N'*-tetraisopropylphosphorodiamidite in the presence of a weak acid like 4,5-dicyanoimidazole (DCI) or with 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite (CEP-Cl) and an organic base. The latter approach was taken, resulting in the orthogonally protected $m^3C$ phosphoramidite **5** with 23% yield over five steps. At the start of this project no report for $m^3C$ phosphoramidite synthesis was available. However, in 2021, a similar synthetic pathway was published by Sheng and coworkers, using TBDMS instead of TOM for 2'-silylation (**10**),[370] but with a lower overall yield (6%), especially because of low tritylation values (Scheme 6). In addition, Micura and coworkers reported the synthesis of N4-acetylated $m^3C$ phosphoramidites in 2022.[371]

### *$N^4$-Methylcytidine* ($m^4C$)

$N^4$-methylcytidine ($m^4C$) was synthesized as phosphoramidite building block **15** according to a published procedure (Scheme 7A)[372] as a reference compound and as the free nucleoside for HPLC analysis. Synthesis of 5'-*O*-DMT and 2'-*O*-TOM protected uridine **13** was accomplished first and facilitated the *in situ* activation of the lactam oxygen with 2,4,6-triisopropyl-benzolsulfonyl chloride (Trisyl-Cl). Subsequent treatment with methylamine in EtOH led to the desired $m^4C$ compound **14** which was finally phosphitylated with CEP-Cl to yield nucleoside **15**. The phosphorous NMR showed an additional peak at around 14 ppm in addition to the two product peaks at around 150 ppm, which belongs to the hydrolysis product of CEP-Cl (phosphonate). Thus, further purification by precipitation had to be performed.

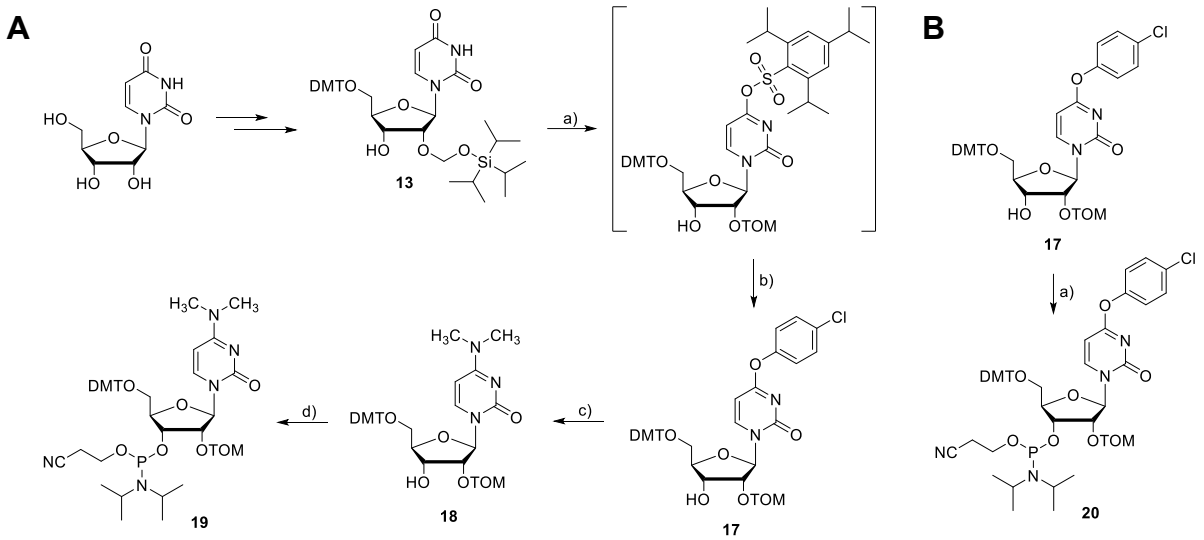



Scheme 7. Synthesis pathway for the formation of the $m^4C$ phosphoramidite **15**, reproduced from Ref.[369, 372] and the initial step for the synthesis of the free $m^4C$ nucleoside. A: a) DMT-Cl, pyridine, r. t., 4 h, (88%) b) 1. $^{i}Pr_2NEt$, $Bu_2SnCl_2$, 1,2-DCE, r. t., 1 h; 2. TOM-Cl, 80 °C, 20 min (32%) c) Trisyl-Cl, DMAP, $NEt_3$, r. t., 1 h d) 8 M $MeNH_2$ in EtOH, r. t., o. n., (79% over c and d) e) CEP-Cl, $Me_2NEt$, DCM, r. t., 3 h, (86%). B: first step of the synthesis of the free $m^4C$ nucleoside based on Ref.[373] a) $Ac_2O$, pyridine, r. t., o. n., (99%).

Therefore, compound **15** was dissolved in a minimal amount of DCM and added dropwise to hexane. After centrifugation and removal of the supernatant, the phosphoramidite **15** was redissolved in DCM and evaporated. Unfortunately, the impurity could only partially be removed, but coupling efficiencies were not impaired during solid-phase synthesis. The free nucleoside was synthesized similarly (Scheme 7B). Acetyl protection of the sugar hydroxy groups with acetic anhydride was followed by activation with Trisyl-Cl and substitution with methylamine. The basic conditions simultaneously deprotected the sugar moiety and yielded the free $m^4C$ nucleoside for HPLC analysis.

### *$N^4$,$N^4$-Dimethylcytidine* ($m^4{}_2C$)

Another naturally occurring cytidine modification is $m^4{}_2C$ (Scheme 8). The synthetic route is very similar to the one for $m^4C$ shown in Scheme 7. As the Trisyl activation group is comparably labile to hydrolysis, the conversion with 4-chlorophenol was leading to a more stable convertible nucleoside.[374] Thereby, the sulfonyl ester was efficiently replaced, even without the need of a protection group at 3'-OH. Nevertheless, the workup with column chromatography had to be performed carefully – the reactant 4-chlorophenol often eluted together with the target compound, observed as additional signals in the aromatic region of the $^1$H-NMR. Eluting with DCM until the reactant (only faintly visible under UV light, but the spots on TCL turned brown upon longer irradiation) was washed out before adding MeOH to the eluent system, gave **17** in good yields and high quality. Subsequent nucleophilic substitution with dimethylamine ($Me_2NH$) in dioxane resulted in the precursor nucleoside **18** in almost quantitative yields. The last step was the phosphitylation with CEP-Cl in the presence of $Me_2NEt$ in DCM to get phosphoramidite **19**.




Scheme 8. Synthetic pathway for $m^4{}_2C$ phosphoramidite **19** and the convertible phosphoramidite **20**. a) Trisyl-Cl, DMAP, $NEt_3$, r. t., 1 h b) 4-chlorophenol, DBU, $Me_2NEt$, DCM, r. t., 2 h, (86%)[374] c) 40% $Me_2NH$, dioxane, r. t., 45 min (98%)[375] d) CEP-Cl, $Me_2NEt$, DCM, r. t., 2.5 h (81%). B: a) CEP-Cl, $^iPr_2NEt$, DCM, r. t., 2 h, (63%).

Direct phosphitylation of the activated compound **17** led to the convertible 4-chlorophenyluridine phosphoramidite **20** (Scheme *8*). It is stable under solid phase synthesis conditions and facilitates post-derivatization of the nucleobase at the exocyclic amino group, by adjusting the deprotection conditions.[376] For example, basic deprotection of an oligonucleotide containing this convertible nucleoside with $Me_2NH$ led to substitution and release of the leaving group – the resulting oligonucleotide contained the modification $m^4{}_2C$ at the position of the convertible nucleoside.

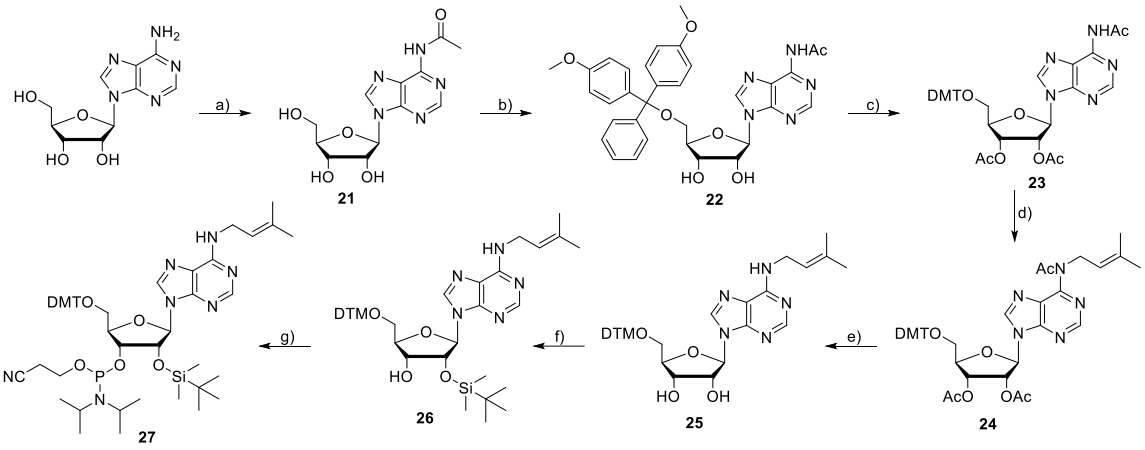
3.1.1.2 Modified adenosines

Besides cytidine modifications, also three naturally occurring adenosine modifications with additional residues at N6 were synthesized. Several methods for the alkylation of N6 of adenosine have been described in literature. They include synthetic pathways with or without the need of preliminary protection of the hydroxy groups of the sugar moiety. However, due to solubility issues, purification of free nucleosides can often be difficult, especially using column chromatography.[377] The following general approaches for $N^6$-alkylation have been described: conversion of C(6)-halogenated purines with amines by nucleophilic substitution ($S_NAr$),[378, 379] N1 alkylation with following Dimroth rearrangement of $N^1$- to $N^6$-alkylated nucleosides under basic conditions,[380] one-pot approach with simultaneous activation using hexamethyldisilazane (HMDS)[381] or 1-*H*-benzotriazol-1-yloxy-tris(dimethylamino)phosphonium hexafluorophosphate (BOP)[382] starting from inosine, reductive alkylation of the respective carboxylic acid amides with lithium aluminium hydride ($LiAlH_4$),[383] or acylation of adenosine and reductive monoalkylation *via* amidine intermediates with sodium borohydride ($NaBH_4$).[384] Furthermore, with the help of tetrabutylammonium bromide ($Bu_4NBr$) as phase transfer catalyst, $N^6$-acetyl-2',3',5'-tris-*O*-(TBDMS)adenosine can be converted with the corresponding alkyl halide to $N^6$-alkylated derivatives.[385] Porcher and Pitsch reported a synthetic pathway starting with an acetyl- and TOM-protected inosine. The carbonyl function of the lactam is activated with 3-nitro-1-*H*-1,2,4-triazole to form a 6-(nitrotriazolyl)-substituted adenosine derivative, which can then be alkylated with the corresponding amine.[386] Elzbieta and Ryszard Kierzek used a post-synthetic approach to functionalize an oligonucleotide containing 6-methylthiopurine or 2-methylthio-6-chloropurine as precursors.[387] Treating these oligonucleotides with the respective amine led to the transformation into the desired $N^6$-alkylated modifications at the position of the precursor. Later, Vitali I. Tararov *et al*. described a method for the regioselective incorporation of the isopentenyl function at N6 starting with $N^6$-acetyl-2',3',5'-tri-*O*-acetyladenosine.[377] They presented two possible ways for the $N^6$-alkylation: 1. the alkylation with alkyl halides using $K_2CO_3$ and 2. reaction with the respective alcohols (Mitsunobu). After deprotection under basic conditions they obtained the free nucleoside of $N^6$-isopentenyl-adenosine. It has to be taken into account that during the synthesis of adenosine with $Ac_2O$ to form a tetraacetate a mixture of different acetylated products occurred. Therefore, the penta-acetyl derivative with an excess of $Ac_2O$ was synthesized, followed by the selective *N*-deacetylation with imidazole in MeOH. The alkylation

can then be achieved either base-mediated with alkyl halides or Mitsunobu. Regioselective reaction at N6 was proven by NMR analysis. Interestingly, by using $N^6$-benzoyladenosine instead of $N^6$-acetyl for alkylation, there was a decrease in regioselectivity, resulting in a $N^6$-/$N^1$-alkylated mixture.[377]

### *$N^6$-Isopentenyladenosine* (i$^6$A)

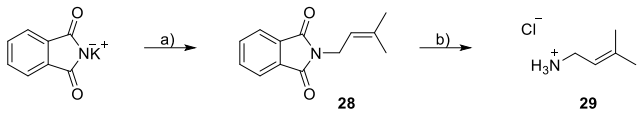
In this thesis, one synthesis pathway of i$^6$A was based on the latter approach (Scheme 9). Tritylation after the basic deprotection of the $N^6$-acetyl-2',3',5'-tri-*O*-acetyl-$N^6$-isopentenyladenosine gave low yields, probably due to the steric hindrance of the alkyl function on N6. Thus, contrary to the report of Tararov *et al.*, DMT-protection of the 5'-OH group of compound d **21** was performed first, after the selective monoacylation at N6 with *in situ* TMS-protection of the sugar hydroxy groups of adenosine. Subsequent acetyl protection of the 2'- and 3'-hydroxy groups of compound **22** enabled the selective incorporation of the isopentenyl group at N6 of **23**. Due to the steric demand of the trityl moiety, no tetra-acetylated compound was observed (no additional acetyl at N6) after the treatment of **22** with $Ac_2O$, which rendered the selective *N*-deacetylation step with imidazole in MeOH described by Tararov *et al.* unnecessary.[377] Basic deprotection of the acetyl groups was performed with $MeNH_2$ in EtOH, resulting in the 5'-*O*-DMT-protected-i$^6$A compound **25**.




Scheme 9. Synthetic pathway for the i$^6$A phosphoramidite over 7 steps. a) 1. TMS-Cl, pyridine, 18 h, 0 °C to r. t. 2. AcCl MeCN, , 1 h, 0 °C to r. t. (78%) b) DMT-Cl, pyridine, 2 h, r. t. (42%) c) $Ac_2O$, pyridine, 1.5 h, r. t. (79%) d) $K_2CO_3$, isopentenyl bromide, DMF, 20 h, 50 °C (22%) e) $MeNH_2$ in EtOH, 16 h, r. t. (70%) f) TBDMS-Cl, $AgNO_3$, THF/pyridine, 19 h, r. t. (46%) g) $Me_2NEt$, CEP-Cl, 2.5 h, r. t. (89%).

For 2'-*O*-silylation TBDMS was chosen, another commonly used silyl-protection group. In contrast to TOM, higher coupling times and lower coupling efficiencies were reported,[388] and it is less stable under basic conditions. The TBDMS group might migrate from 2'-O to 3'-O (2' to 3'-isomerization and vice versa). Loss of the 2'-protection group could lead to strand cleavage or 3'-to-2'-phosphate translocation.[388, 389] On the other hand, this property can also be exploited to enhance the overall yield of the TBDMS protection, by converting the isolated 3'-product into an 2'-3'-equilbrium with

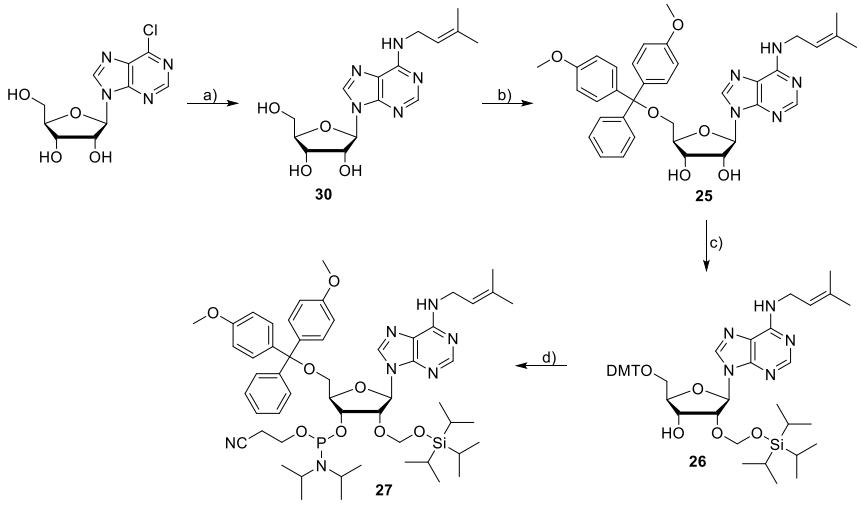
MeOH and triethylamine and re-isolation of the desired 2'-protected compound **26**. Phosphitylation was accomplished with CEP-Cl and $Me_2NEt$ in DCM to yield **27** (Scheme 9).




Scheme 10. Synthesis of 2-isopentenyl amine (**29**), needed for the substitution of the 6-chloropurine riboside.[390] a) potassium phthalimide, prenyl bromide, DMF, 80 °C, 16 h b) 1. $N_2H_4 \cdot H_2O$, EtOH, 50 °C, 2 h; 2. 6 M HCl, 0 °C to r. t., 18 h.

As aforementioned, alkylation can also be achieved by nucleophilic substitution of halogenated purines with amines. This was performed as an alternative route to get the $i^6A$ phosphoramidite **27** without the need of prior protection of the hydroxy groups, which reduced the number of required steps from 7 to 4. Therefore, 3-methyl-2-buten-1-amin hydrochloride (2-isopentenylamine hydrochloride, **29**) had to be synthesized over 2 steps by Gabriel synthesis, starting from prenyl bromide and potassium phthalimide and subsequent hydrazinolysis after precipitation of the intermediate **28** (Scheme 10).[390] Phthalhydrazide, which is formed during the hydrazinolysis, could not completely be separated from the 2-isopentenylamine hydrochloride salt **29**.

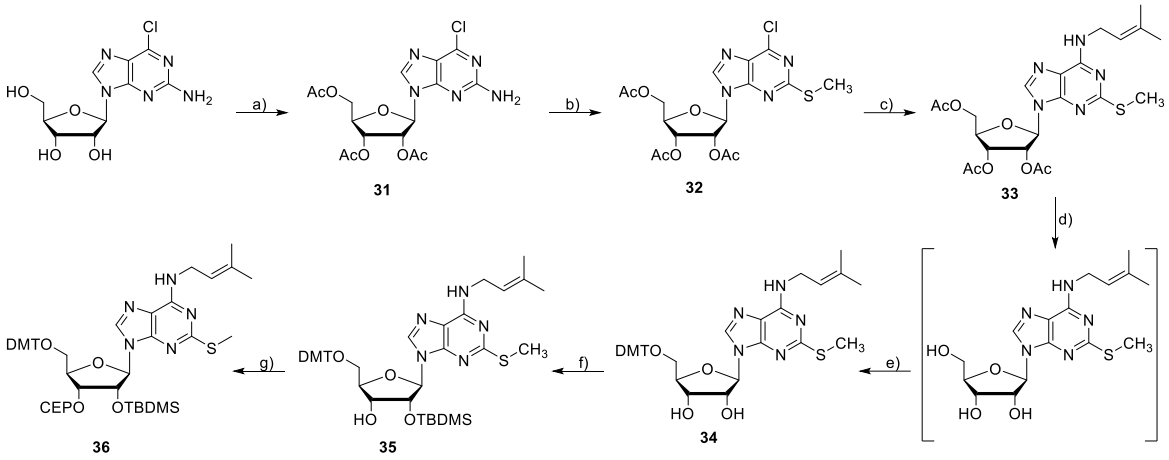
The isolated isopentenyl amine hydrochloride salt **29** was then used for the functionalization of the 6-chloro purine riboside to synthesize the free $i^6A$ nucleoside **30**. After the *in situ* protection of the 2'- and 3'-OH groups with DMF-DMA, the 5'-OH group could selectively be tritylated. Under workup conditions, the acetal was cleaved off and the DMT-protected $i^6A$ nucleoside **25** was obtained. Silylation of 2'-OH and subsequent phosphitylation with CEP-Cl was performed according to the synthetic pathway shown earlier in Scheme 9.




Scheme 11. Alternative synthetic pathway to synthesize the $i^6A$ phosphoramidite **27**. a) 6-chloropurin riboside, isopentenyl amine hydrochloride, $NEt_3$, EtOH, 80 °C (62%) b) 1. DMF-DMA, pyridine, r. t., o.n.; 2. DMT-Cl, pyridine, 2.5 h, r. t. (64%) c) TBDMS-Cl, $AgNO_3$, THF/pyridine, 19 h, r. t. (46%) g) $Me_2NEt$, CEP-Cl, 2.5 h, r. t. (89%).

### *N*$^6$-*Isopentenyl-2-methylthioadenosine* (ms$^2$i$^6$A)

The natural modification i$^6$A can be further derivatized with a 2-thiomethyl group at the nucleobase. It was introduced *via* diazotization of the acetyl protected 2-amino-6-chloropurine riboside **31** with isoamyl nitrite and reaction with dimethyl sulfide, resulting in compound **32**. The reaction time mentioned in the report of Kierzek *et al.* had to be prolonged from 45 min to 2 h, to achieve higher reaction yields.[387]




Scheme 12. Synthetic pathway for the ms$^2$i$^6$A phosphoramidite **36**. a) $Ac_2O$, DMAP, $NEt_3$, MeCN, 0 °C to r. t., 3 h, (83%) b) dimethyl disulfide, isoamyl nitrite, MeCN, 60 °C, 2 h (66%)[387] c) isopentenyl amine, pyridine, $NEt_3$, r. t., 17 h, (81%) d) MeOH, $NEt_3$, r. t., 48 h, e) DMT-Cl, pyridine, 2 h, r. t. (81% over d and e) f) TBDMS-Cl, pyridine/THF, $AgNO_3$, 18 h, r. t. (32%) g) CEP-Cl, $Me_2NEt$, DCM, r. t., 2 h, (87%).

Initially, the procedure of Kierzek *et al.* should be pursued for the next two steps.[387] There, DMT-protection was performed after the deprotection of the acetyl groups, followed by substitution with isopentenyl amine. However, low yields were obtained for the deprotection of the acetyl groups, due to the partial substitution of the chloride group under the basic conditions in MeOH. Reordering the synthesis steps solved this problem and resulted in markedly higher yields. First, the isopentenyl group was introduced at position 6, forming the acetyl protected ms$^2$i$^6$A **33** (Scheme 12). Then the acetyl groups were deprotected with MeOH and $NEt_3$ and treated with DMT-Cl after evaporation without further purification, giving the 5'-*O*-DMT protected ms$^2$i$^6$A compound **34**. In accordance with the i$^6$A phosphoramidite **27**, TBDMS was introduced at the 2'-position (**35**) prior to phosphitylation of the 3'-OH group with CEP-Cl to yield the phosphoramidite building block **36**.

### *N*$^6$-*Threonylcarbamoyladenosine* (t$^6$A)

Another naturally occurring modification, which is conserved in tRNAs, generally responsible for ANN codons in all domains of life, is t$^6$A (see chapter 1.3.1). Like i$^6$A it can be further derivatized to 2-thiomethyl-t$^6$A (ms$^2$t$^6$a) with a thiomethyl group at C2 of adenosine,[68] or N6-methylated t$^6$A (m$^6$t$^6$A),[391] which are already known for more than 40 years (Figure 32A).

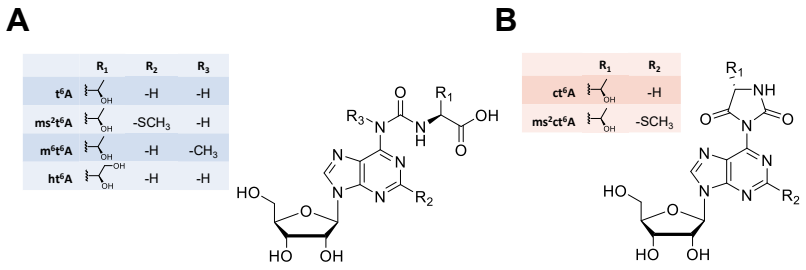
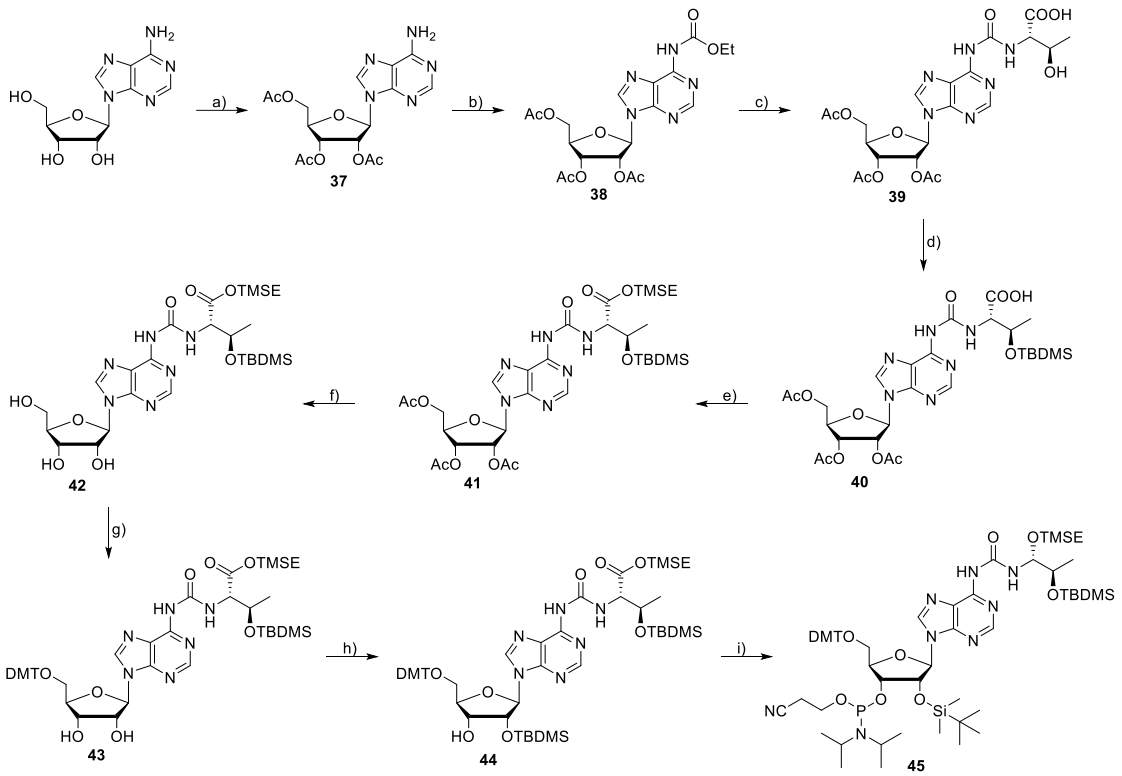



Figure 32. Linear (blue) and cyclic (red) derivatives of $t^6A$. Figure reproduced from Ref.[392] A: $t^6A$, $ms^2t^6A$, $m^6t^6A$, $ht^6A$. B: $ct^6A$ and $ms^2ct^6A$.

Around 10 years ago, cyclic $t^6A$ ($ct^6A$) was identified in *E.coli* tRNAs by Suzuki and coworkers, on account of a milder tRNA isolation method.[393] This modification remained undiscovered for a long time, since it is highly sensitive to hydrolysis under conditions used before for the isolation and analysis of tRNAs.[394] Suzuki and coworkers revealed that $t^6A$ is a hydrolysis artefact of $ct^6A$, which actually appears at position 37 of some tRNAs in *E.coli*, yeast and plants (Figure 32B). CsdL (E1-like enzyme) was found to be responsible for the formation of $ct^6a$ by catalyzing the dehydration of $t^6A$. The amino acid side chain of this $t^6A$ analogue was indicated as an oxazolone ring (cyclic active ester).[393] However, X-Ray analysis and LC/MS comparison later revealed that $ct^6A$ is actually present as a hydantoin ring.[395] The two latest identified analogs were $ms^2ct^6A$[396] and *N*$^6$-hydroxy-$t^6A$ ($ht^6A$),[397] again from Suzuki and coworkers. $ms^2ct^6A$ was found in tRNAs from *Bacillus subtilis*, *T. brucei* and plants,[396] and $ht^6A$ was identified in mt-tRNAs$^{Lys}$ from *Mesocentrotus nudus* (sea urchin).[397]

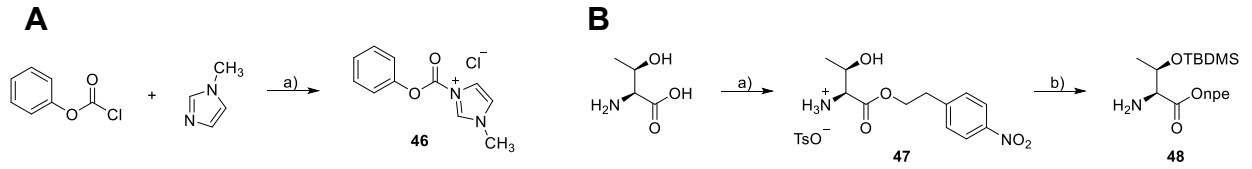
In this thesis, the $t^6A$ and $ms^2t^6A$ modifications were required and thus the phosphoramidites needed to be synthesized. Several syntheses pathways for the $t^6A$ phosphoramidite have already been published in literature: Chheda and Hong converted *O*-benzyl-L-threonine into an isocyanate, prior to the reaction with tri-acetylated adenosine. Unfortunately, these two steps gave low yields, with 39 and 48%, respectively.[398] Similarly, Herdewijn and coworkers were also using an approach with isocyanate. With triphosgene, they converted the tri-acetylated adenosine into the isocyanate, which was then treated with TBDMS- and 2-(4-nitrophenyl)ethanol- (npe-OH) protected L-threonine. This coupling step between the nucleobase and the amino acid, only resulted in 19% yield.[399] Davis and coworkers (Scheme 13) functionalized N6 of the acetyl-protected adenosine with ethylchloroformate to generate the active carbamate, followed by nucleophilic substitution with L-threonine, resulting in sugar protected $t^6A$ nucleoside **39**.[400] Related to the approach of Chheda and Hong, Carell and coworkers performed the protection of L-threonine with TBDSM and npe first, before transferring it onto the adenosine nucleoside.[401] The urea linkage was achieved with 1-*N*-methyl-3-phonoxycarbonyl-imidazolium chloride.[402]



Scheme 13. Synthesis pathway of $t^6A$ based on Ref.[400, 403] a) $Ac_2O$, DMAP, $NEt_3$, MeCN, r. t., 21 h, (65%) b) ethylchloroformate, pyridine, -10 °C to r. t., 18 h, (54%) c) L-threonine, pyridine, 120 °C, 4 h, (75%) d) 1. TBDMS-Cl, imidazole, DMF, 65 °C, 18 h; 2. 2 M $NH_3$/MeOH, 1 min (78%)[404] e) 1. DCC, TMSE-OH, pyridine, r. t., o. n., f) 2 M $NH_3$/MeOH g) DMT-Cl, pyridine h) TBDMS-Cl, imidazole, pyridine i) 2-cyanoethyl *N*,*N*,*N*',*N*'-tetraisopropylphosphorodiamidite, DMAP, DIPEA, THF.

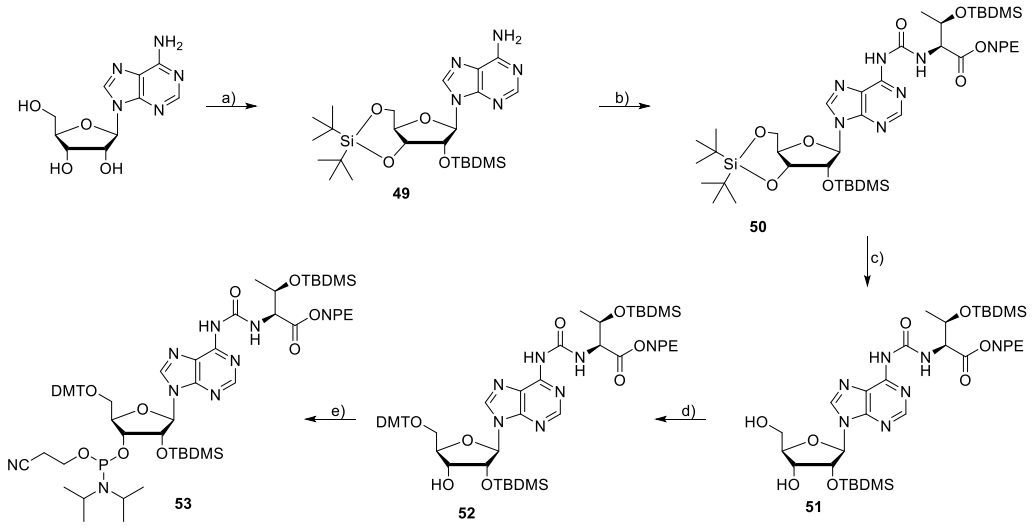
The first attempt to synthesize $t^6A$ was based on the synthetic pathway of Davis and coworkers,[400] performing the first three steps according to the published procedure from Matuszewski and Sochacka.[403] The tri-*O*-acetylated adenosine **37** was obtained after reaction with acetic anhydride in acetonitrile and recrystallization from ethanol. Reaction time should be decreased, to prevent acylation of N6, which slightly reduced the reaction yield. The functionalization of the exocyclic N6 amino group to create a reactive carbamate was achieved with 2-ethylchloroformate. Only 54% of compound **38** were obtained, as this reaction partially resulted in a double substituted compound. Separation of these two compounds was possible by column chromatography. Nucleophilic substitution with L-threonine gave the carbamoylated adenosine **39**, with a reduced reaction time of 4 h instead of 7 h. Purification was achieved by recrystallization from MeOH. The protection of the hydroxy group of the amino acid side chain was performed with TBDMS-trifluoromethanesulfonate (TBS-OTf) and $NEt_3$ in DCM according to literature, with subsequent selective hydrolysis of the labile TBDMS-ester.[400] However the first part of this reaction always resulted in a mixture of single and double TBDMS-protected compounds. Changing to TBDMS-Cl and imidazole in DMF at elevated temperature was more favorable for the double protection, followed by the hydrolysis of the TBDMS-ester, giving **40** with 78% yield. The remaining carboxylic acid of the sugar protected $t^6A$ nucleoside **40** was then activated with *N,N'*-dicyclohexylcarbodiimide (DCC) to further react it with

2-trimethylsilylethanol (TMSE-OH) for protection to generate compound **41**. Unfortunately, only the DCC-ester could be isolated. Even the repeated treatment with TMSE-OH did not lead to product formation, but re-isolation of the intermediate. An alternative might be to use EDC and DMAP instead of DCC.[405] To circumvent the problem of post-protection of the amino acid side chain after the introduction to adenosine nucleoside, the synthetic pathway of Carell and coworkers was chosen as an alternative.




Scheme 14. Preparation of the coupling reagent **46** and the protected threonine **48**.[401] A: *N*-methyl-3-phenoxycarbonyl-imidazolium chloride. a) DCM, 0 °C to r. t., 2 h. B: npe- and TBDMS-protected threonine. a) npe, TsOH, toluene, 120 °C, 4 d, (70%) b) TBDMS-Cl, imidazole, pyridine, r. t., o. n., (74%).

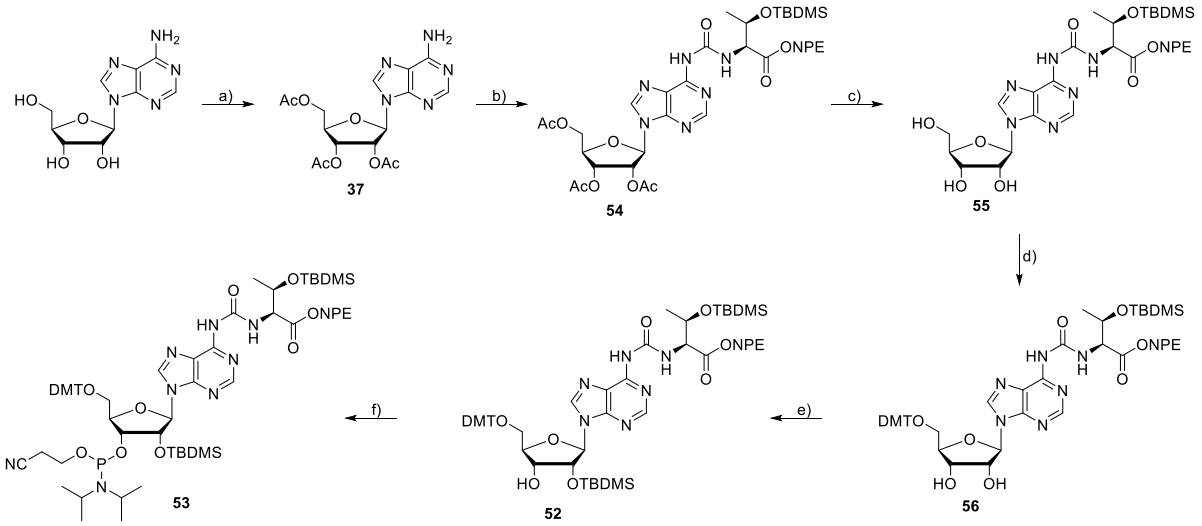
To create the urea linkage between adenosine and threonine, *N*-methyl-3-phenoxycarbonyl-imidazolium chloride **46** was synthesized starting from phenyl chloroformate and *N*-methylimidazole in a one-step synthesis with 98% yield (Scheme 14A). Meanwhile, the carboxylic acid residue of L-threonine was protected with p-nitrophenyl ester (npe) under acid catalysis with *p*-toluenesulfonic acid (TsOH) in toluene to yield **47** (Scheme 14B). The remaining hydroxy group was subsequently treated with TBDMS-Cl to get the npe- and TBDMS-protected threonine **48**.




Scheme 15. Synthesis pathway of $t^6A$ phosphoramidite based on the publication of Carrel and coworkers.[401] a) 1. DTBS-ditriflat, DMF, 0 °C, 45 min; 2. TBDMS-Cl, imidazole, r. t., o. n., (72%) b) 1. activated carbonate **46**, DCM, r. t., 2 h; 2. protected L-threonine **48**, $NEt_3$, DCM, r. t., o. n., (81%) c) $Py{\cdot}(HF)_n$, pyridine, DCM, 0 °C, 2 h (85%) d) DMT-Cl, pyridine, r. t., o. n. (64%) e) CEP-Cl, DIPEA, DCM, r. t., 4 h (59%).

Prior to the attachment of the protected threonine **48**, the 3'- and 5'-hydroxyl groups of adenosine were protected with di-*tert*-butylsilyl-bis(trifluormethansulfonate) (DTBS-ditriflat), followed by the TBDMS-protection of the 2'-OH group in a one pot reaction to give **49** with 72% yield (Scheme 15). Afterwards, the carbamate intermediate was formed by treating the exocyclic amine of **49** with the

activated carbonate **46**. The *in situ* formed carbamate intermediate was further reacted with the protected threonine compound **48**, to get the threonyl carbamoyl adenosine **50** with 81% yield. The cyclic 3',5'-silylether was deprotected with a HF·pyridine complex (Py·$(HF)_n$) in DCM for 2 h, enabling the subsequent protection of 5'-OH with DMT-Cl in pyridine (**52**). The final step was the phosphitylation of **52** with CEP-Cl and DIPEA in DCM, giving the protected $t^6A$ phosphoramidite **53**.

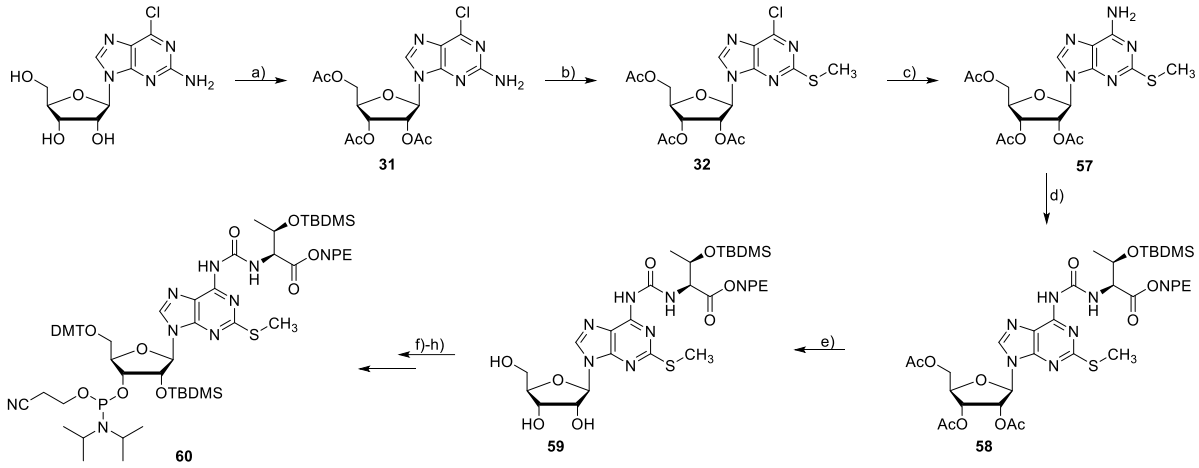



Scheme 16. Alternative synthetic pathway to synthesize the $t^6A$ phosphoramidite **53**. a) $Ac_2O$, $NEt_3$, DMAP, MeCN, r. t., 2 h, (88%), b) 1. activated carbonyl **46**, DCM, r. t., 64 h; 2. protected threonine **48**, $NEt_3$, (63%), c) $NH_3$ in MeOH, r. t., 14 h, (82%), d) 1. DMF-DMA, pyridine, r. t., 17 h, 2. DMT-Cl, pyridine, r. t., 25 h, (60%), e) TBDMS-Cl, $AgNO_2$, pyridine/TFH (1:1), r. t., 17 h (21%), f) CEP-Cl, $Me_2NEt$, DCM, r. t., 2.5 h (86%).

Another alternative synthesis pathway for the $t^6A$ phosphoramidite **53** was performed, circumventing the need of the toxic Py·$(HF)_n$ complex for deprotection of the 3',5'-silylether. Acetyl protection of adenosine was performed, as shown earlier in Scheme 13.[403] The protected threonine amino acid **48** was attached *via* the activated carbonyl reagent **46** in accordance with the report of Carell and coworkers,[401] shown in Scheme 15, with 63% yield (**54**). The acetyl groups were deprotected with ammonia in methanol, resulting in the free $t^6A$ nucleoside **55**. The 5'-OH was further functionalized with DMT-Cl, after the *in situ* formation of the 2'-3'-acetate with DMF-DMA. In contrast to the 3',5'-silylether strategy, lower yields are obtained for the subsequent TBDMS protection of the 2'-OH group of **56**, due the formation of the two isomers. Compound **52** was then converted to the phosphoramidite **53** with CEP-Cl and $Me_2NEt$ in 86% yield. This synthesis pathway (Scheme 16) thus represents a good alternative for the one shown in Scheme 15 from Carell and coworkers. The comparably low overall synthesis yield might get increased if the synthesis is performed in a larger scale, which facilitates the conversion of the 3'-*O*-TBDMS group into a 2'-3'-equilibrium and re-isolation of the correct 2'-isomer (**52**).

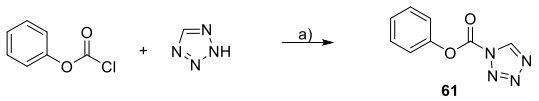
### *N*$^6$-*Threonylcarbamoyl-2-methlythioadenosine* (ms$^2$t$^6$A)

Similarly to the reaction pathway shown in Scheme 16, ms$^2$t$^6$A was supposed to be synthesized. The synthesis started with the introduction of the thiomethyl group at position 2 of 2-amino-6-chloropurine riboside by diazotization *via* an aryldiazonium intermediate, after acetyl protection of the sugar moiety, in analogy to the synthesis of the ms$^2$i$^6$A compound **32** (Scheme 12).



Scheme 17. Planned synthesis pathway for the ms$^2$t$^6$A phosphoramidite **60**. a) $Ac_2O$, DMAP, $NEt_3$, MeCN (85%) b) dimethyl sulfate, isoamyl nitrite, MeCN, 60 °C, (56%) c) 0.4 M $NH_3$ in dioxane, 80 °C, (65%) d) 1. activated carbonyl **46**, DCM, r. t., h; 2. protected threonine **48**, $NEt_3$, e) MeOH, $NEt_3$, f) 1. DMF-DMA, pyridine, r. t.; 2. DMT-Cl, pyridine, r. t. g) TBDMS-Cl, $AgNO_3$, THF/pyridine (1:1) h) CEP-Cl, $Me_2NEt$, DCM.

The acetyl protected 2-thiomethyl-6-chloropurine **32** was converted to the adenosine derivative **57** by 0.4 M ammonia in dioxane at 80 °C using a sealed tube (Scheme 17). The adenosine derivative **57** could be isolated with 65% yield. At room temperature, no conversion was observed. Partially the formation of monodeacylated product was observed (11%), which might occur if the reaction conditions are not completely anhydrous. This side product was again acylated and combined with the product fraction of **57**. In analogy to the synthesis of t$^6$A, the urea linkage should be generated with the activated carbonyl reagent **46** in DCM. However, under these conditions no product formation was observed for the protected 2-thiomethyl-adenosine **57**. Also changing the solvent to 1,2-DCE to enable higher reaction temperatures was not successful.



Scheme 18. One-step synthesis of phenoxycarbonyl tetrazole **61**.[406] a) $NEt_3$, dioxane, 0 °C to 5 °C, 30 min (98%).

Thus, another coupling reagent, phenoxycarbonyl tetrazole **61** (Scheme 18), which was also used in a report of Sochacka and coworkers, was synthesized.[407] However, test reactions using 3 to 10 equivalents of phenoxycarbonyl tetrazole **61** in dioxane at 40 °C to 50 °C did not lead to product formation. By increasing the reaction time or the temperature to 70°C, the formation of several side products was observed on TLC. Maybe the reagent has not been dried enough in vacuum, or the

concentration of the reaction solution was too low, which might lead to a lower reactivity of the carbamylating reagent **61**. Directly using the phenyl chloroformate is not possible due to its high reactivity – *N*,*N*-disubstituted urea compounds might arise as side products during the reaction with exocyclic $NH_2$ group of **57.**[408] Better drying of the reagent **61** and higher concentrations should be tried next to complete the phosphoramidite synthesis.

### 3.1.2 Synthesis of oligonucleotides containing natural modified nucleosides

The phosphoramidite building blocks described in section 3.1.1. were used for the preparation of RNA oligonucleotides by solid-phase synthesis. Solid-phase synthesis commonly starts with a CPG-solid-support where the first nucleobase is already attached. But the solid support can also bear other functional groups, to directly functionalize the 3'-end of the resulting oligonucleotides (Figure 33A, **62**) for example to generate a 3'-phosphate. This can either be used to prevent circularization during ligations or other side reactions during primer elongation experiments. The 5'-end can also be phosphorylated by a similar compound as a phosphoramidite building block (Figure 33B, **63**) to enable ligation experiments.

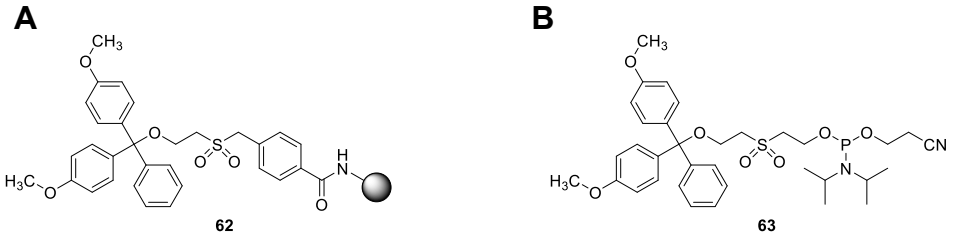
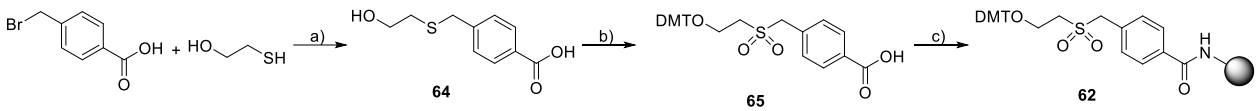


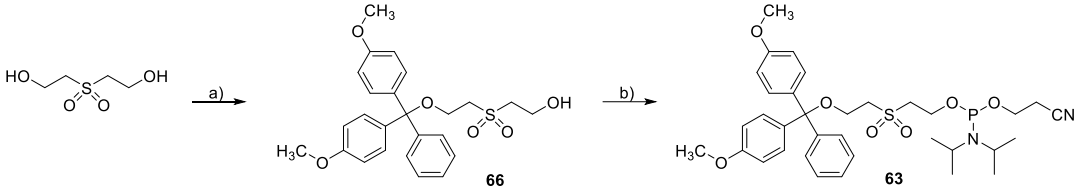

Figure 33. Building blocks to phosphorylate 3'- and 5'-end of oligonucleotides. A: Phosphate-CPG B: Chemical phosphorylation amidite.

To prepare the CPG-support, the precursor had to be synthesized first (Scheme 19). 4-Bromomethylbenzoic acid was treated with 2-mercaptoethanol to get the intermediate **64**, of which the hydroxy group was protected with DMT-Cl to get **65** with 69% yield. This precursor was then loaded onto an amino-functionalized CPG-support using peptide coupling conditions, by activating the benzoic acid of compound **65** with benzotriazolyloxytris(dimethylamino)phosphoniumhexafluorophosphate (BOP) under basic conditions using ($^i$Pr)$_2$NEt in DMF. To prevent side reactions, the amino groups of unreacted CPG-support were protected with capping solution, containing acetic acid anhydride, followed by the oxidation of the thioether with *meta*-chloroperoxybenzoic acid (*m*CPBA) as the last step. Loading density of **62** was determined by absorbance measurement of the trityl cation after treatment with dichloroacetic acid (see chapter 5.1.2.8).



Scheme 19. Phosphate-CPG: a) $NEt_3$, MeOH, r. t., 17 h, (88%) b) DMT-Cl, pyridine, r. t., 19 h, (69%), c) 1. Native amino CPG, BOP, $(^iPr)_2NEt$, DMF, 40 °C, 2 h; 2. $Ac_2O$, 2,6-lutidine, *N*-methylimidazole, THF, r. t., 10 min; 3. *m*CPBA, DCM, r. t., 30 min.

The 5'-*O*-DMT-2,2'-sulfonyldiethanol phosphoramidite **63** was synthesized over 2 steps. One of the two available hydroxy groups of 2,2'-sulfonyldiethanol was protected with DMT-Cl, using 0.7 equivalents of the tritylation agent to prevent double protection (**66**), followed by the formation of the phosphoramidite **63** with CEP-Cl. The CPG solid support **62** was used to functionalize the 3'-end of the synthesized oligonucleotide, whereas the 5'-phosphate amidite **63** was used to phosphorylate the 5'-end.



Scheme 20. Synthesis of the 5'-Phosphate Amidite (*O*-DMT-2,2'-sulfonyldiethanol).[409] a) DMT-Cl, pyridine, r. t., (44%), b) CEP-Cl, $^iPr_2NEt$, DCM, r. t. (80%).

### 3.1.2.1 RNA containing modified cytidines

The $m^3C$ phosphoramidite building block **5** (Scheme 6) was incorporated into RNA strands by solid phase synthesis, either within the position of a tRNA-ASL where the modification also occurs in nature, or into a selection substrate for DNAzymes. Different deprotection conditions after the successful synthesis were tested to find suitable conditions for the removal of the benzoyl protection group and to circumvent the formation of side products: 1) $NH_4OH$/EtOH (3:1), 55 °C, 6 h; 2) $NH_4OH$/EtOH (3:1), 55 °C, 23 h; 3) $MeNH_2$ $H_2O$/EtOH (1:1), 37 °C, 6 h. Anion HPLC revealed an additional peak with higher retention time compared to the main product (Figure 34A). After PAGE purification, the second peak was still present, but to a minor extend than before (Figure 34B). The additional peak was eluting slower than the main peak on anion exchange chromatography and is present under all of the three deprotection conditions. Precise cutting out of the band was necessary to decrease the amount of the additional peak. It is eluting faster on the gel, thus is present in the lower part of the product band. A report of Li and coworkers showed that treatment of a DNA substrate containing $m^3C$ with ammonium hydroxide at 80 °C for 3 h led to the formation of $m^3U$ by deamination.[410, 411] Consequently, the additional peak is most likely the deamination product of $m^3C$ under the basic deprotection conditions. HPLC analysis of the lower part of the gel band revealed that the intensity of the additional peak is markedly increased (Figure 34C). High-resolution electrospray ionization mass spectrometry (HR-ESI-MS) of this sample showed a broad isotope pattern. Comparison with the predicted mass of the oligonucleotide containing $m^3C$ (blue) or

$m^3U$ (orange) indicated the presence of both species, but mainly $m^3U$ containing oligonucleotide (Figure 34D). Both conditions using ammonia gave the correct mass (5743.9) of the modified oligonucleotide. The shorter reaction time (6 h), was preferred over the longer one, to prevent further side reactions. Contrary, using methylamine for basic deprotection is giving an additional mass of +14 (5757.9 g/mol) compared to the correct oligonucleotide. As reported before, using methylamine as deprotection agent for benzoyl-protected cytidines partial transamination might occur, giving $N^4$-methylated side product.[412, 413] Replacing the benzoyl with acetyl protection group was shown to typically avoid this side reaction. However, a recent report by Micura and coworkers in 2022 revealed that for the acetyl-protected $m^3C$-modification transamination occurred during basic deprotection of the oligonucleotide using 40% $MeNH_2$ for 12 min at 65 °C as well.[371] Consequently, ammonia is required for the deprotection of $m^3C$-containing oligonucleotides, and methylamine needs to be avoided.

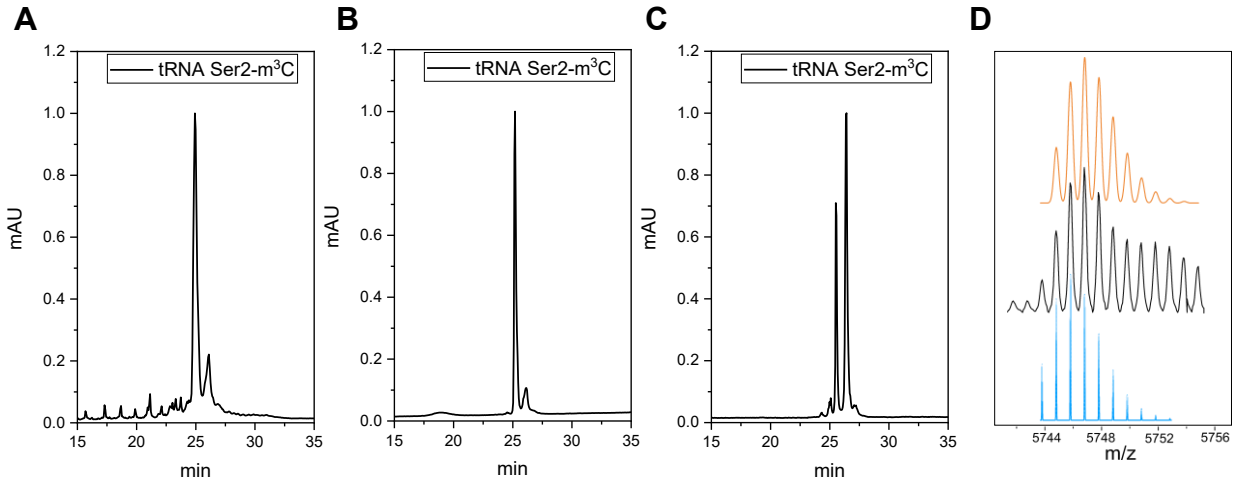



Figure 34. Normalized HPLC chromatograms of the A: crude B: pure and C: down (lower part of the gel-band) oligonucleotide samples of $tRNA^{Ser}$-ASL modified with $m^3C$. D: Partial overlay of the ESI-sample shown in C (black). Simulated spectra of oligonucleotides containing $m^3C$ and $m^3U$ are shown in blue and orange, respectively.

### 3.1.2.2 RNA containing modified adenosines

$I^6A$ and $ms^2i^6A$ containing RNA oligonucleotides were deprotected using either 25% $NH_4OH/MeNH_2$ ($H_2O$, 1:1) at 37 °C for 3 h and 55 °C for 2 h or $MeNH_2$ in $EtOH/H_2O$ (1:1) for 6 h at 37 °C, respectively, without the formation of side products. It has to be noted that labeling attempts with $NaIO_4$ to introduce for example fluoresceine with fluorescein 5-thio-semicarbazide (FITSC) after the oligonucleotide synthesis without an amino linker at 3'-end were not successful for $i^6A$ and $ms^2i^6A$-modified RNA. Oxidation creates a di-aldehyde from the vicinal hydroxy groups at the 3'-terminal end which allows functionalization with fluorescent analogs. It is an RNA specific labeling method and only modifies the 3'-terminal end with the vicinal hydroxy groups. However, $NaIO_4$ reacts with the double bond of the isopentenyl group, making it an unsuitable method for oligonucleotides with $i^6A$ or $ms^2i^6A$.

As the $t^6A$ modification has the protected threonine residue more deprotection steps had to be performed. Corresponding oligonucleotides were treated with 10% DBU in THF at ambient temperature for 2 h to remove the npe-protection group from the carboxylic acid. Cleavage from the solid support and removal of the remaining protection groups was achieved by 25% $NH_4OH/MeNH_2$ ($H_2O$, 1:1) at ambient temperature for 1.5 h.

By solid-phase synthesis, the modifications were incorporated into specific positions within tRNA-ASLs, corresponding to the sequences derived from human cytosolic or mitochondrial tRNAs, where they also occur in nature. Additionally, unmodified reference oligonucleotides, single and double modified ASLs and single point mutations of selected ASLs were synthesized (Table 1).

Table 1. Overview of the synthesized tRNA-ASLs as the unmodified references, single or double modified ASLs and ASLs with single-point mutations (R856-R858, R860) and with a Ser-loop region but Phe-Stem (R859).

| number | ASL | sequence | type | modification |
|---|---|---|---|---|
| R599 | $tRNA^{Ser(UCN)}$ | GUUGG**C**UUGAAACCAGCU | mitochondrial | unmodified |
| R600 | $tRNA^{Thr}$ | CCAGU**C**UUGUAAACCGGA | mitochondrial | unmodified |
| R601 | $tRNA^{Phe}$ | AUACA**C**UGAAAAUGUUUA | mitochondrial | unmodified |
| R764 | $tRNA^{Arg(CCU)}$ | CUGGC**C**UCCUAAGCCAG | cytoplasmic | unmodified |
| R765 | $tRNA^{Thr(AGU)}$ | CCUGU**C**UAGUAAACAGG | cytoplasmic | unmodified |
| R766 | $tRNA^{Ser(UGA}$ | AUGGA**C**UUGAAAUCCAU | cytoplasmic | unmodified |
| R767 | $tRNA^{Arg(CCU)}$ | CUGGC$m^3C$UCCUAAGCCAG | cytoplasmic | $m^3C$ |
| R768 | $tRNA^{Thr(AGU)}$ | CCUGU$m^3C$UAGUAAACAGG | cytoplasmic | $m^3C$ |
| R769 | $tRNA^{Ser(UGA)}$ | AUGGA$m^3C$UUGAAAUCCAU | cytoplasmic | $m^3C$ |
| R604 | $tRNA^{Ser(UCN)}$ | GUUGG$m^3C$UUGAAACCAGCU | mitochondrial | $m^3C$ |
| R605 | $tRNA^{Thr}$ | CCAGU$m^3C$UUGUAAACCGGA | mitochondrial | $m^3C$ |
| R606 | $tRNA^{Phe}$ | AUACA$m^3C$UGAAAAUGUUUA | mitochondrial | $m^3C$ |
| R622 | $tRNA^{Ser(UCN)}$ | GUUGGCUUGA$i^6A$ACCAGCU | mitochondrial | $i^6A$ |
| R783 | $tRNA^{Thr}$ | CCAGUCUUGU$i^6A$ACCGGA | mitochondrial | $i^6A$ |
| R623 | $tRNA^{Phe}$ | AUACACUGAA$i^6A$AUGUUUA | mitochondrial | $i^6A$ |
| R816 | $tRNA^{Ser(UCN)}$ | GUUGGCUUGA$ms^2i^6A$ACCAGCU | mitochondrial | $ms^2i^6A$ |
| R817 | $tRNA^{Thr}$ | AUACACUGAA$ms^2i^6A$AUGUUUA | mitochondrial | $ms^2i^6A$ |
| R818 | $tRNA^{Phe}$ | CCAGUCUUGU$ms^2i^6A$ACCGGA | mitochondrial | $ms^2i^6A$ |
| R824 | $tRNA^{Ser(UCN)}$ | GUUGG$m^3C$UUGA$i^6A$ACCAGCU | mitochondrial | $m^3C$ and $i^6A$ |
| R825 | $tRNA^{Ser(UCN)}$ | GUUGG$m^3C$UUGA$ms^2i^6A$ACCAGCU | mitochondrial | $m^3C$ and $ms^2i^6A$ |
| R836 | $tRNA^{Thr}$ | CCAGUCUUGU$t^6A$AACCGGA | mitochondrial | $t^6A$ |
| R837 | $tRNA^{Thr}$ | CCAGU$m^3C$UUGU$t^6A$ACCGGA | mitochondrial | $m^3C$ and $t^6A$ |
| R856 | $tRNA^{Ser(UCN)}$ | GUUGGCU**A**GA$ms^2i^6A$ACCAGCU | mutated | $ms^2i^6A$ |
| R857 | $tRNA^{Ser(UCN)}$ | GUUGGCUU**A**A$ms^2i^6A$ACCAGCU | mutated | $ms^2i^6A$ |
| R858 | $tRNA^{Ser(UCN)}$ | GUUGGCUUGA$ms^2i^6A$**C**CCAGCU | mutated | $ms^2i^6A$ |
| R859 | $tRNA^{Ser/Phe}$ | AUACACUUGA$ms^2i^6A$AUGUUU | mutated | $ms^2i^6A$ |
| R860 | $tRNA^{Thr}$ | CCAGUCUUGU$t^6A$**G**ACCGG | mutated | $t^6A$ |
| R861 | $tRNA^{Set(AGY)}$ | AAGAACUGCU$t^6A$ACUCAU | mitochondrial | $t^6A$ |

With these oligonucleotides, methylation and demethylation assays, thermal melting curve analysis, NMR- and CD spectroscopy were performed in collaboration with the working group of Prof. Dr. Markus Bohnsack. The coworkers, especially Nicole Kleiber and Nicolas Lemus-Diaz were

analyzing the subcellular localization and substrate specifications of METTL8 with different biomolecular experiments (i.e., methylation assay), using among others the synthesized tRNA-ASLs shown in Table 1, which will be described in the next section. It will be a summary of the experimental results, reported in the paper: *The RNA methyltransferase METTL8 installs $m^3C_{32}$ in mitochondrial $tRNAs^{Thr/Ser(UCN)}$ to optimise tRNA structure and mitochondrial translation* from N. Kleiber, N. Lemus-Diaz and C. Stiller *et al.*.[58] In conformity with the report, for $tRNA^{Ser}$ the codon UCN is stated in this chapter in contrast to the other tRNAs mentioned in this work, where always the respective anticodon is mentioned next to the amino acid ($tRNA^{Ser(UCN)}$ = UGA anticodon).

### 3.1.3 Applications and investigations using synthetic modified RNAs

#### 3.1.3.1 Investigation of writer enzymes: methylation experiments with methyltransferase METTL8

Experiments described in this section were performed in the Bohnsack lab and used the RNAs obtained by solid-phase synthesis as described in section 3.1.2. As mentioned earlier (see chapter 1.3.3), METTL2 and METTL6 were shown to mediate the methylation of C32 of cytoplasmic $tRNA^{Ser}$, $tRNA^{Thr}$, and $tRNA^{Arg[CCU/UCU]}$ in humans. Less information is available for METTL8. To investigate the subcellular localization of METTL8, cell lines for the expression of C-terminally $His_6$-2xGFP tagged METTL8 (METTL8-GFP) were prepared by transfection into HEK293 Flp-In T-Rex cells. Expression was initialized by tetracycline treatment.[58] By this means the longest isoform of METTL8, as well as a shortened version, $METTL8_{21\text{-}407}$, without the mitochondrial targeting sequence (MTS) were generated. These HEK293 cell lines, expressing METTL8-GFP and $METTL8_{21\text{-}407}$-GFP, were analyzed by confocal fluorescence microscopy. For the visualization of mitochondria within the cells the fluorescent dye MitoTracker® Deep Red FM (Abs/Em: 644/665 nm, Thermo Fischer) and for the nuclear localization the blue 4',6-diaminidino-2-phenylindole (DAPI, Abs/Em: 350/470 nm) stain were used. By comparison of the fluorescent signals, METTL8 revealed a mitochondrial localization, since the METTL8-GFP cell lines displayed a dot-like arrangement right at the MitoTracker (See Figure 1 in Manuscript 58). In contrast, the cell lines with the shorter $METTL8_{21\text{-}407}$-GFP protein, omitting the MTS, showed an extensive distribution over the cytoplasm and nucleus by fluorescence, especially concentrating at the nuclear foci (See Figure 1 in Manuscript 58).[58]

Besides the subcellular localization, the substrates of the methyltransferase should be identified as well. With *in vivo* UV cross-linking and analysis of cDNA (CRAC) experiments RNAs in direct contact with METTL8 were detected. Extracted RNAs that were giving a radioactive signal ($^{32}P$) were analyzed by Illumina deep sequencing after copying into cDNA. The majority of the enriched sequences appeared to be mitochondrial RNAs (mitoRNA, 71% of the total reads, See Figure 2 in Manuscript 58), among them particularly mt-tRNAs (76% of mito-RNAs). The normalized sequencing reads were assigned to the 22 mt-tRNAs (See Figure 2 in Manuscript 58) and the enrichment

was plotted by comparing the METTL8-containing sample with the control sample ($His_6$-2xFLAG) without the METTL8 gene. The three most enriched tRNAs were the ones encoding for Ser(UCN), Thr, and Ile. Northern blots after Anti-FLAG immunoprecipitation experiments with the lysates from UV-crosslinked cells as in the CRAC experiments confirmed the binding of those three tRNAs to METTL8. Thereby, the main contacts appear to be formed with the ASL of the tRNAs: The formation of a crosslink between a nucleotide and an amino acid is impeding reverse transcription, thus nucleotide substitutions (sub) are indicating the position of contact (See Figure 2 in Manuscript 58).[58] For $tRNA^{Thr}$ and $tRNA^{Ser(UCN)}$ the substitutions are enriched in the ASL region. Together with the number of sequencing reads profiles for the formation of crosslinks between METTL8-$His_6$-2xFLAG and tRNAs are depicted. Recombinantly expressed His14-MBP-METTL8 (MBP: Maltose-binding protein) was used to perform fluorescence anisotropy measurements to get more insights into the binding affinity of the oligonucleotide samples towards the enzyme as well as specificities for the formation of interactions. Therefore, different fluorescently labeled oligonucleotide substrates were chosen: the $m^3C$ and $i^6A$ modified $tRNA^{Ser(UCN)}$ ASL, the $tRNA^{Met}$ ASL (UCGGGCCCAUACCCCGA), two unstructured RNA oligonucleotides (GUAAAAGUCGUAACAAAGGU and GACAUACUGACCCUUCAAAUA), as well as one DNA oligonucleotide (GATGTGCAGCTGCACATC). As expected from the CRAC experiments, both $tRNA^{Ser}$ and $tRNA^{Met}$ were bound with a low $K_d$ value, reflecting a high affinity. However, also the two unstructured oligonucleotides were behaving similarly. Only the DNA strand was less efficiently bound with a 17-26-fold higher $K_d$ value than the RNA samples (See Figure 2 in Manuscript 58). Thus, METTL8 specifically binds RNA substrates, while being independent on their structure or sequence.[58]

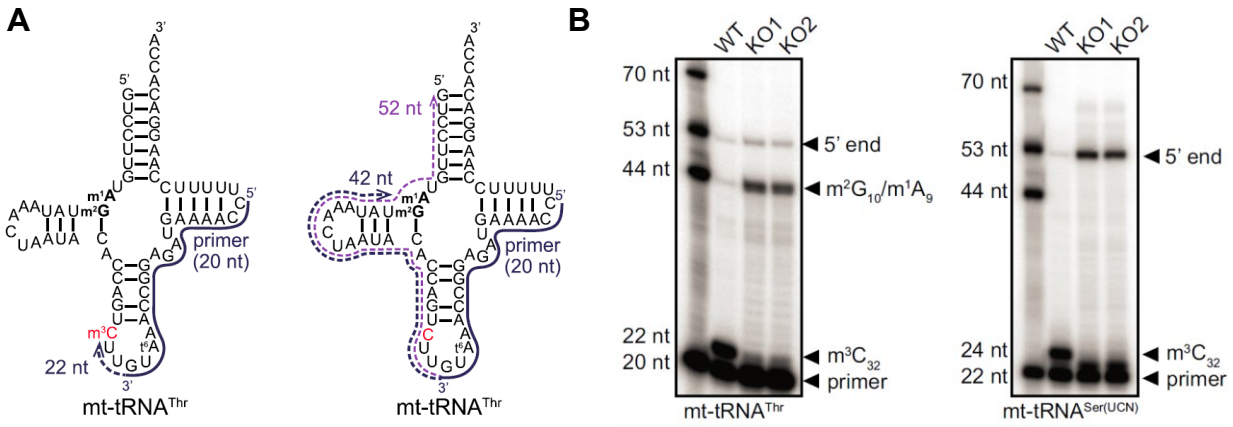



Figure 35. Primer extension experiments with mt-$tRNA^{Thr}$ and mt-$tRNA^{Ser(UCN)}$ to investigate if METTL8 is the responsible methyltransferase for cytidine at position 32. A: Exemplarily shown is the secondary structure of mt-$tRNA^{Thr}$ with $m^3C_{32}$ on the left side and with the unmodified cytidine on the right side. $m^3C_{32}$ (red) would stall the synthesis of cDNA, whereas for cytidine the synthesis would continue until the next impeding modification or 5'-end. The corresponding primers were $^{32}P$-labelled. B: The resulting cDNAs from the primer extension experiments were separated by denaturing PAGE. $^{32}P$-lablled DNA oligonucleotides were used as a size marker (left). WT and the two KO cell lines are shown for mt-$tRNA^{Thr}$ and mt-$tRNA^{Ser(UCN)}$. Figures taken from Ref.[58] Bohnsack lab.

Since the two most enriched tRNAs in CRAC experiments, mt-tRNA$^{Thr}$ and mt-tRNA$^{Ser(UCN)}$, contain m$^3$C at position 32 (see chapter 1.3.1), both of them were tested in methylation experiments with METTL8. For these experiments, two human METTL8 knock-out HEK cell lines (-ΔMETTL8) were afforded by CRISPR/Cas9-mediated mutagenesis, with either one or 15 nucleotide (nt) deletions (KO1 and KO2). Primer extension analysis of RNAs isolated from wild-type (WT), KO1, and KO2 is the means of choice, as the methyl group of m$^3$C causes a noticeable interruption of the elongation by blocking the reverse transcriptase (Figure 35A, left). In the presence of an unmodified cytidine at position 32, the synthesis of the cDNA would not stall (Figure 35A, right) [58]

In agreement with the reported m$^3$C modification at position 32 of mt-tRNA$^{Thr}$ and mt-tRNA$^{Ser(UCN)}$ a strong termination band was observed for both small RNA samples from WT (Figure 35B, marked with m$^3$C$_{32}$). In contrast, almost no stalling at position 32 was observed for the knockout cell lines without METTL8, but at the 5'-end of the tRNAs. In addition, primer extension of mt-tRNA$^{Thr}$ resulted in a distinct stalling band almost at the 3'-end, caused by the modifications m$^2$G$_{10}$ and m$^1$A$_9$. This data strongly supports the presence of m$^3$C at position 32 in mt-tRNA$^{Thr}$ and mt-tRNA$^{Ser(UCN)}$ and indicates that METTL8 might be the responsible methyltransferase for installing the methyl group at C32. The reduced methylation of C32 in the absence of METTL8 was circumvented *in vitro* by treating the isolated small RNAs from KO1 and KO2 cell lines with SAM as a methylation agent in the presence of recombinant His$_{14}$-MBP-METTL8. Primer extension of these samples resulted in the same stalling band as for the WT, due to the restorage of m$^3$C$_{32}$. Performing the same experiment with METTL8 derivatives with a substitution of one amino acid (D230A, D309A) within the methyltransferase domain revealed that the increase of the methylation level could not be established. This data clearly shows that METTL8 is directly responsible for installing the methyl group at C32 for mt-tRNA$^{Thr}$ and mt-tRNA$^{Ser(UCN)}$.[58]

Similar methylation experiments with *in vitro* transcripts of mt-tRNA$^{Thr}$ and mt-tRNA$^{Ser(UCN)}$ using His$_{14}$-MBP-METTL8 and [$^3$H]-SAM were not as efficient in generating m$^3$C$_{32}$. While mt-tRNA$^{Thr}$ could be methylated, only a low amount of mt-tRNA$^{Ser(UCN)}$ got modified. These differences in the ability of installing the methyl group between cellular and *in vitro* transcribed RNAs might arise due the absence of modifications in the *in vitro* samples. This correlation was shown before in a study about a C32-A37 ACL circuitry in *S. pombe*.[52] i$^6$A$_{37}$ was a prerequisite for the methylation of cytidine at position 32 in a subset of tRNA$^{Ser}$ with A at position 36 (see chapter 1.3.3). Accordingly, t$^6$A and (ms$^2$)i$^6$A may be required for methylation of human mt-tRNA$^{Thr/Ser(UCN)}$, respectively, as well.[58]

To analyze this interdependency the mt-tRNA ASLs were synthesized in different modification levels: Unmodified, with a single modification, either t$^6$A/(ms$^2$)i$^6$A on position 37 or m$^3$C on position 32, and the double modified ASL with the respective modifications on position 32 and 37 (Figure

36A). With these ASLs the methylation assay was performed as described before using recombinant $His_{14}$-MBP-METTL8 and [$^3$H]-SAM. For mt-tRNA$^{Thr}$ METTL8 was able to install the methyl group even for the unmodified ASL. Control experiments using the $m^3C_{32}$ modified ASL confirmed position 32 as the position of methylation, as it was not further modified. Using the single modified ASL with $t^6A_{37}$ the methylation efficiency was noticeably increased (Figure 36B). The unmodified mt-tRNA$^{Ser(UCN)}$ ASL resulted in only minimal METTL8-mediated methylation, which was significantly increased using the $i^6A$-modified ASL and even further increased using the $ms^2i^6A$ modified ASL (Figure 36C). Thus, both modifications on position 37 of mt-tRNA$^{Thr}$ and mt-tRNA$^{Ser(UCN)}$ are prerequisites for an efficient installation of the methyl group at position 32 *in vitro*. Further support was obtained by the depletion of the enzymes responsible for modifying A37 to either $t^6A$ (OSGEPL1) or $i^6A$ (TRIT1). Primer extension experiments resulted in a slightly higher read-through for both cells with enzyme depletions caused by siRNAs – the decrease of modification level on position 37 impedes methylation of C32 in turn.[58]

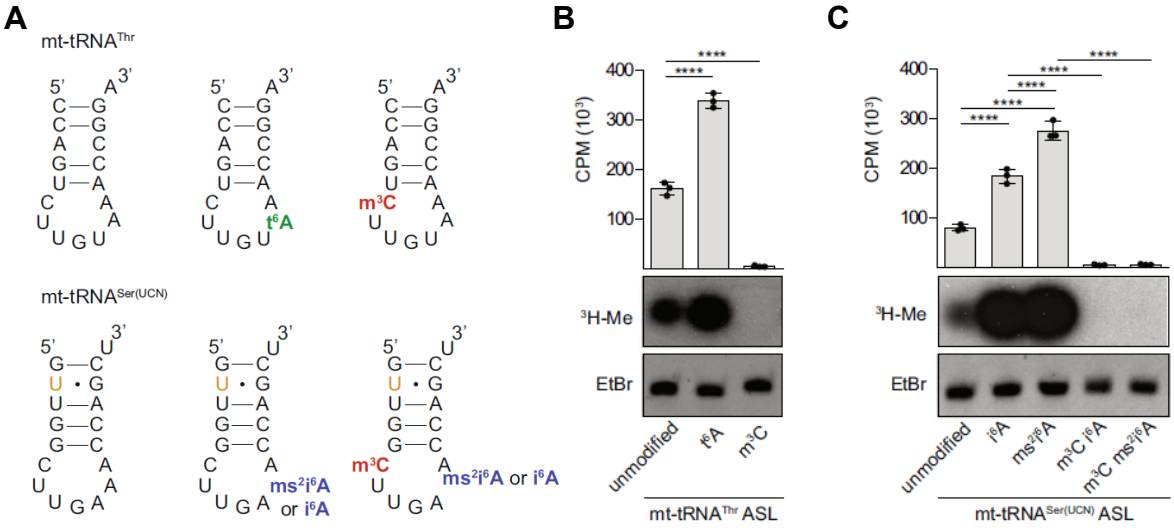



Figure 36. Investigating the interdependencies of the modifications on position 32 and 37 for methylation of C32. A: Hairpin structure of the mt-tRNA$^{Thr/Ser}$ ASLs as unmodified, single, and double modified substrates. $t^6A$ is shown in green, $m^3C$ in red and the $i^6A$ derivative in blue. The uridine marked in orange would be pseudouridine in nature. B/C: Methylation assay using recombinant $His_{14}$-MBP-METTL8 and [$^3$H]-SAM resulted in a significantly higher methylation rate, measured by scintillation counting, if the modifications at position 37 ($t^6A$ or $i^6A$) were present. Bar graphs are given as count per minute (CPM). The ASLs were separated by denaturing PAGE. Detection of methylated RNAs was achieved by autoradiography of $^3H$-$CH_3$ (upper gel image) as well as staining with ethidium bromide (EtBr, lower gel image). B: mt-tRNA$^{Thr}$. C: mt-tRNA$^{Ser(UCN)}$. Figures taken from Ref.[58] Bohnsack lab.

The modification state also influences the binding affinity. Fluorescence anisotropy measurements of $i^6A$ modified mt-tRNA$^{Ser(UCN)}$ ASL revealed only a slightly lower $K_d$ than for the unmodified reference, but a 2.5-fold lower $K_d$ than the double modified ASL, representing a higher binding of $His_{14}$-MBP-METTL8 (Figure *37*). Thus, $m^3C_{32}$ might facilitate the dissociation from METTL8 after methylation of the substrate RNA, but the affinity of METTL8 for the substrate RNA is not markedly enhanced by the presence of $i^6A_{37}$. This highlights that the sequence and also modification on position 37 of the RNA substrate is not determining for binding of METTL8, but the requirement for installation of the methyl group at C32.[58]

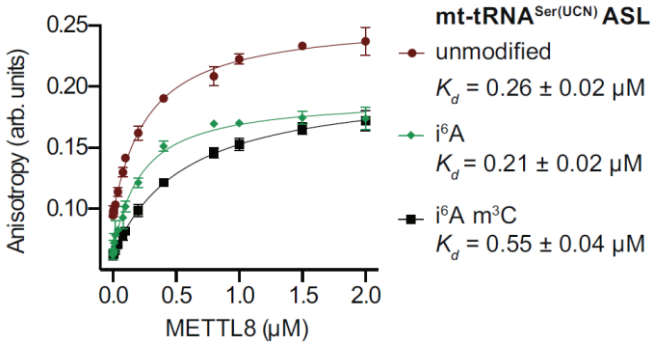



Figure 37. The average of three independent fluorescence anisotropy measurements of unmodified (brown), single i$^6$A-modified (green) and double i$^6$A and m$^3$C modified (black) mt-tRNA$^{Ser(UCN)}$ ASLs are depicted to investigate changes in affinity of His$_{14}$-MBP-METTL8 for fluorescently labeled substrate RNA depending on the modification state. Figure taken from Ref.[58] Bohnsack lab.

In addition to the unmodified, single, and double modified mt-tRNA$^{Thr/Ser}$ ASLs (Figure 36A) ASLs with single point mutations within the loop region were used to identify nucleotides required for substrate recognition by METTL8 (Figure 38). With these mutated RNA substrates methylation experiments were performed using His$_{14}$-MBP-METTL8 and [$^3$H]-SAM.

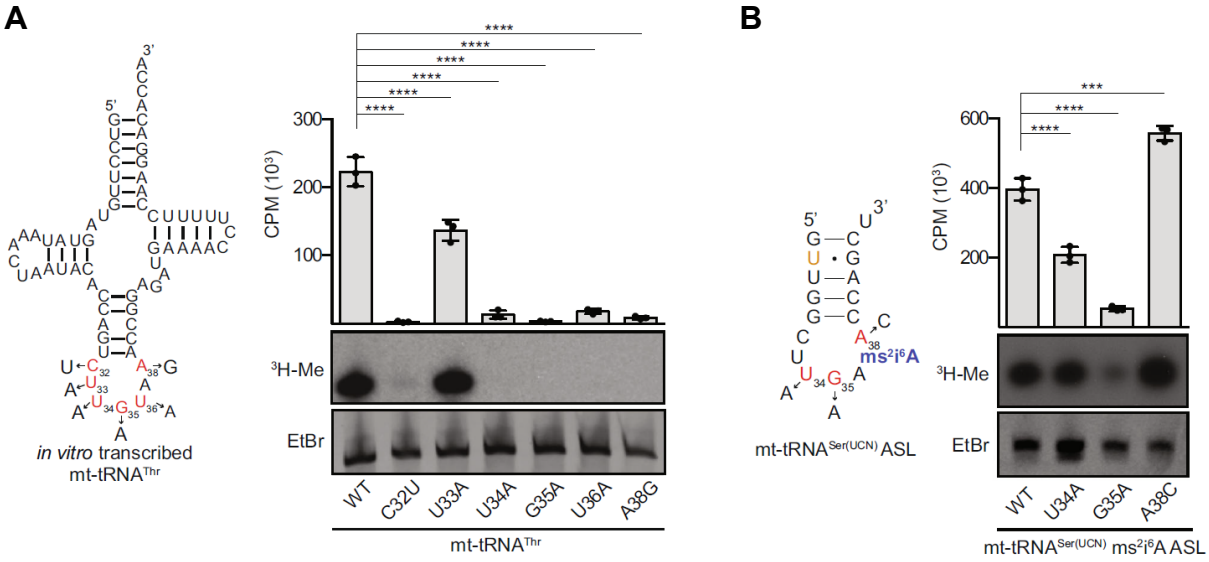



Figure 38. Single nucleotides within the ASL were exchanged to determine the effect on the methylation level of C32 by Mettl8. A: Six *in vitro* transcripts of mt-tRNA$^{Thr}$ with substitutions on position 32, 33, 34, 35, 36 and 38 were chosen for *in vitro* methylation assays using recombinant His$_{14}$-MBP-METTL8 and [$^3$H]-SAM; the methylation was measured by scintillation counting. Bar graphs are given as count per minute (CPM). The ASLs were separated by denaturing PAGE. Figures taken from Ref.[58] Bohnsack lab.

A decrease in METTL8-mediated methylation at C32 was obtained by substitution of position 32, 34, 35, 36, and 38 (A38G substitution was observed in patients with mitochondrial disorders, preventing t$^6$A$_{37}$ formation[414]) of *in vitro* transcribed mt-tRNA$^{Thr}$. The U$_{34}$G$_{35}$U$_{36}$t$^6$A$_{37}$A$_{38}$ motif appears to be necessary for recognition by METTL8 (Figure 38A), in accordance with the recognition mode of Trm140 (*S. cerevisiae*) for tRNA$^{Thr}$ which identifies the G$_{35}$U$_{36}$t$^6$A$_{37}$ motif.[99] Also, in this case, the modification on position 37 is relevant, but not a prerequisite. Only a minor effect was observed for U33A (Figure 38A).[58] Mutated mt-tRNA$^{Ser(UCN)}$ ASLs were chemically synthesized, as the (ms$^2$)i$^6$A modification on position 37 is required for efficient methylation. Position 34, 35 and 38 of the mt-

tRNA$^{Ser(UCN)}$ ASL, containing ms$^2$i$^6$A$_{37}$, were substituted (Figure 38B). The U34A and to a higher extent the G35A substitution noticeably decreased the METTL8-mediated methylation of C32. In contrast, A38C increased the methylation compared to WT.[58] The same trend was already observed by Mao et al. in 2021:[415] while the methylation of human cytoplasmic tRNA$^{Ser(GCU)}$ was completely retarded by U33G, G34A, U36A, and A37C substitutions, the A38C exchange significantly increased the modification level of C32. Furthermore, seryl-tRNA synthetase (SARS) is essential for methylation by METTL6.[415] In contrast, mt-SARS2 was not needed for C32 methylation of mt-tRNA$^{Ser}$ by METTL8.[58]

For tRNA$^{Thr}$ and tRNA$^{Ser(UCN)}$ the nucleotides within the loop region of the ASL are important to be recognized as a substrate RNA by METTL8. Both tRNAs have U34 and G35 in common, whose exchange by other nucleotides has a markedly negative impact on substrate recognition. This might explain why other mt-tRNAs with C32 and a modified adenosine at position 37, but without a U34 and G35, are not accepted as substrates for METTL8-mediated methylation in *in vitro* assays. To determine if besides the nucleotides in the loop region, nucleotides of the stem are required for the recognition as well, *in vitro* transcripts of mt-tRNA$^{Asn}$ were synthesized, where either the loop or the whole ASL is replaced by the corresponding sequence of mt-tRNA$^{Thr}$ (Figure 39A).

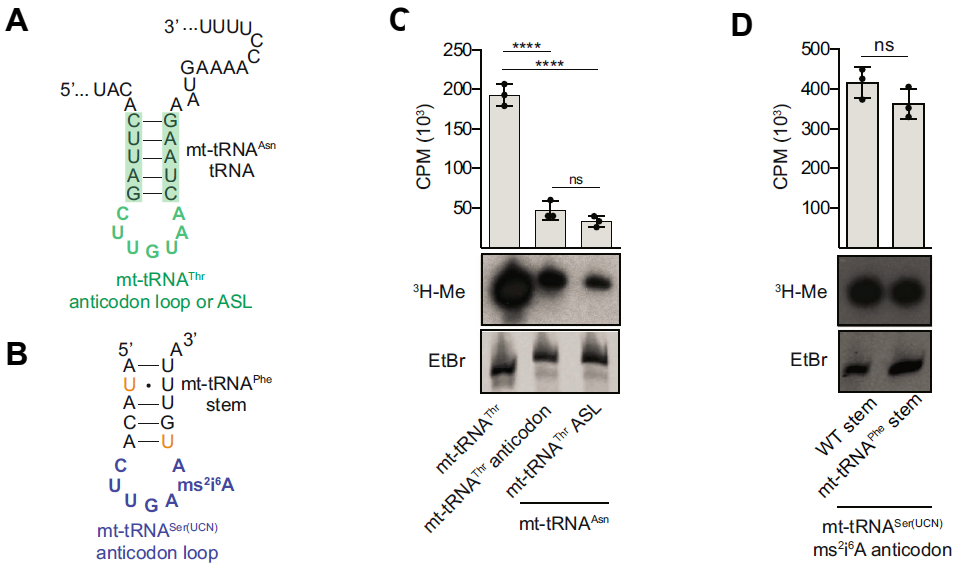



Figure 39. *In vitro* methylation assay with chimeric tRNAs to determine recognition elements. A: *In vitro* transcript of mt-tRNA$^{Asn}$ with the ASL or loop region of mt-tRNA$^{Thr}$. B: mt-tRNA$^{Phe}$ stem with mt-tRNA$^{Ser(UCN)}$ loop. C: Evaluation of the methylation assay, performed with the chimeric tRNA shown in A. D: Evaluation of the methylation assay, performed with the chimeric tRNA shown in B. Figures taken from Ref.[58] (Bohnsack lab).

Those transcripts were still methylated, but with much lower efficiency than mt-tRNA$^{Thr}$ (Figure 39C). Similarly, mt-tRNA$^{Phe}$ ASL containing the mt-tRNA$^{Ser(UCN)}$ loop was synthesized with ms$^2$i$^6$A$_{37}$ (Figure 39B). This chimeric ASL, as a combination of Phe and Ser tRNAs, was equally methylated as mt-tRNA$^{Ser(UCN)}$ (Figure 39D).[58] Thus, for both mt-tRNAs$^{Thr/Ser(UCN)}$ the stem regions do not seem to be involved in substrate recognition. However, mt-tRNA$^{Thr}$ was hypothesized to require additional elements besides the ACL for efficient METTL-8 mediated methylation, as the modification

level was noticeably decreased.[58] In summary, the modifications at position 37 and some nucleotides within the ACL are required for the specific determination of tRNA substrates by METTL8.[58]

### 3.1.3.2 Structural and functional implications of anticodon stem-loop modifications in mt-tRNA$^{Thr/Ser}$

As described in chapter 1.3.2 modifications within the ACL can generally affect functions like codon-anticodon interactions, expansion of decoding capability, and fine-tuning tRNA structure for a precise translation of the genetic code. The influence of the resulting $m^3C_{32}$ modification by METTL8 onto the tRNAs$^{Thr/Ser(UCN)}$ will be further described in the next section. The effect of the $m^3C_{32}$ modification on the structure and stability of mt-tRNA$^{Thr/Ser(UCN)}$ was examined by native PAGE, NMR, thermal melting curve analysis, and CD spectroscopy and was also part of the published report: *The RNA methyltransferase METTL8 installs $m^3C_{32}$ in mitochondrial tRNAs$^{Thr/Ser(UCN)}$ to optimise tRNA structure and mitochondrial translation*, from N. Kleiber, N. Lemus-Diaz and C. Stiller *et al.*.[58] The results of the structural investigations will be summarized in this chapter.

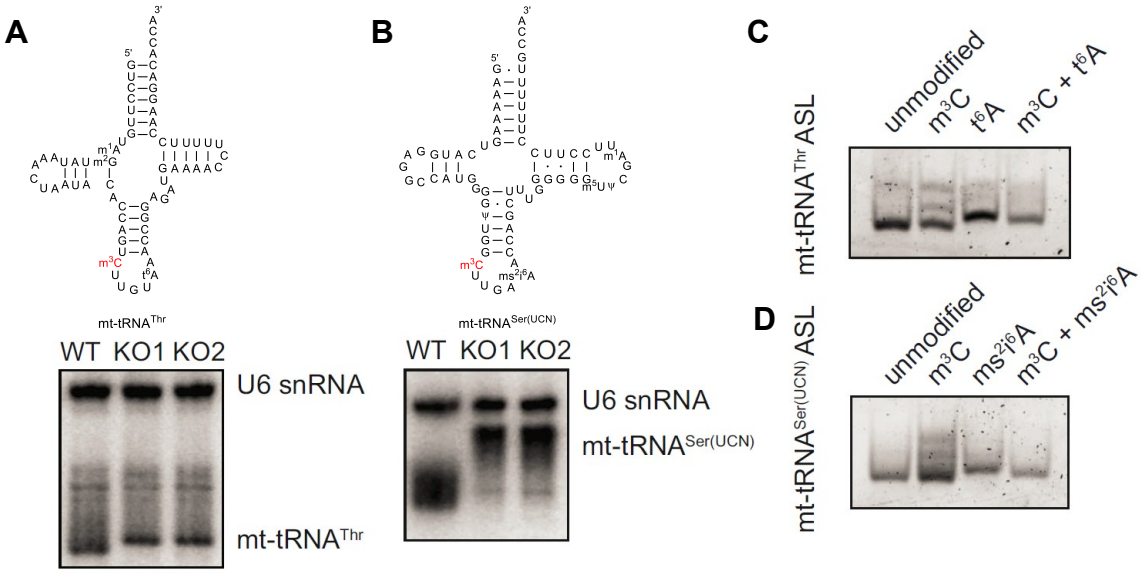



Figure 40. Examination of the appearance of conformational changes depending on the modification level. A: Secondary structure of mt-tRNA$^{Thr}$. Separation of RNAs (mt-tRNA$^{Thr}$ marked in the lowest lane) isolated from WT and METTL8 KO cell lines by native gel electrophoresis. B: Secondary structure of mt-tRNA$^{Ser(UCN)}$. The native page of mt-tRNA$^{Ser(UCN)}$ reveals a larger shift between WT and KO than for Thr. Bohnsack lab. C/D: Native gel electrophoresis of chemically synthesized ASLs with different modification states. C: mt-tRNA$^{Thr}$; unmodified, m$^3$C, t$^6$A and double modified ASLs. D: mt-tRNA$^{Ser(UCN)}$; unmodified, m$^3$C, ms$^2$i$^6$A, double modified ASL. Figures taken from Ref.[58]

Separation of the full-length RNAs isolated from WT and METTL8 KO cell lines by native gel electrophoresis resulted in a markedly distinct migration behavior (Figure 40A/B). The mt-tRNA$^{Thr/Ser(UCN)}$ samples obtained from the KO cell lines (without m$^3$C) ran slower compared to WT. This result is much more pronounced for mt-tRNA$^{Ser(UCN)}$ (Figure 40B) than for mt-tRNA$^{Thr}$ (Figure 40A). Still, upon installation of the methyl group at C32 conformational changes occur for both tRNAs, reflected by retardation of the migration. The different shifts in migration of mt-tRNA$^{Thr}$ and mt-tRNA$^{Ser}$ let assume that other regions apart from the ASL might be structurally influenced by

the presence of $m^3C$ as well. This could be related to other modifications in the tRNA structure and their interplay with $m^3C$.[58] The importance of modifications within the different regions of the tRNA with effect on structure, stability, folding, and function was demonstrated in chapter 1.3.2. These aspects might be further fine-tuned by interactions between other modified nucleotides. As mt-tRNA$^{Thr}$ and mt-tRNA$^{Ser(UCN)}$ have different modifications in different locations (Figure 40A/B), this might result in different interactions with $m^3C$ and thus affect distinct structural changes.

While varying the modification state, the same experiments were also repeated with the chemically synthesized ASLs to get more detailed information about the structural impact of the respective modifications of positions 32 and 37 and if the structural changes emerge in the ASL. The mt-tRNA$^{Thr/Ser(UCN)}$ ASLs with $m^3C$ on position 32 displayed primarily the same migration as the unmodified ASLs, with minor species/fractions being retarded. Similarly to the cellular tRNAs (WT and KO cell lines), the migration of mt-tRNA$^{Thr/Ser(UCN)}$ ASLs containing the respective modification on position 37 is reduced and the migration was again slightly accelerated by the additional incorporation of $m^3C_{32}$ (Figure 40C/D). Thus, depending on the modification state of the substrate and fully modified oligonucleotide conformational changes of mt-tRNA$^{Thr/Ser(UCN)}$ occur.

Conformational changes due to the presence of modifications within the ASL might also lead to different stabilization of the hairpin structures, which can be analyzed by thermal melting curve measurements. The UV-melting experiment takes advantage of the temperature dependency of the oligonucleotide absorbance: the absorbance is increasing upon heating, due to denaturation and the associated changes in the stacking of the nucleobases, referred to as the hyperchromic effect. The temperature gradient should be max. 0.5 °C/min to reach thermodynamic equilibrium, leading to a reversible melting profile after two cycles (two heating/two cooling) between 10 and 90 °C.[416] A sigmoidal curve shape is obtained by plotting the absorption or the hyperchromicity (change in the absorption) vs. the temperature. The inflection point can be assigned with the higher and lower baseline, corresponding to the associated and denatured form, respectively. The determined inflection point gives the melting temperature ($T_m$) of the measured complex.[416] $T_m$ is defined as the temperature, where half of the nucleic acid secondary structure is folded, and the other half unfolded (Figure 41).[416] If duplex structures are formed as a bimolecular reaction (molecularity = 2), the $T_m$ values are concentration-dependent (Figure 41). Contrary, the formation of an intramolecular complex is concentration independent resulting in the same $T_m$ value for different sample concentrations.[416] Thereby, the melting temperature can further be influenced by the ionic strength of the salt content of the melting curve buffer as well as the pH of the solution.[416, 417] The calculations and fits to determine the melting temperatures, as well as the thermodynamic parameters $\Delta G$, $\Delta S$ and $\Delta H$ from the van't Hoff plot, based on a two-state equilibrium without further intermediates being formed, will be explained in chapter 5.3.3.

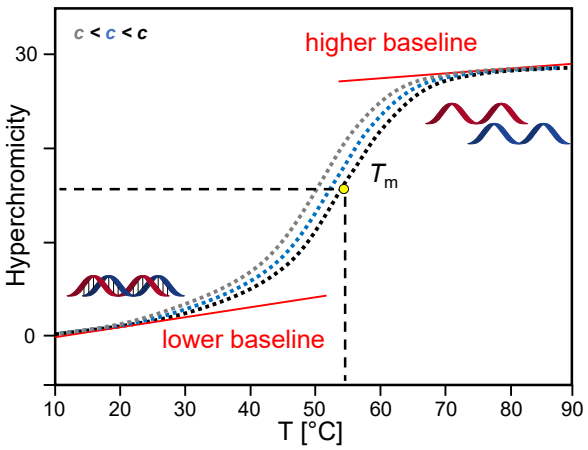



Figure 41. Schematic representation of a thermal melting curve with a sigmoidal curve shape, where the inflection point (yellow) corresponds to the melting temperature ($T_m$). A concentration dependency is shown with three curves, of which grey has the lowest and black the highest concentration. A higher concentration is leading to a higher $T_m$. The higher baseline (red) represents the denatured, and the lower baseline (red) the associated form of the complex.

The thermal melting point of the synthesized ASLs of mt-tRNA$^{Ser(UCN)}$ is obtained at the inflection point of the sigmoidal curve shown in Figure 42. It is given as the average of two heating and cooling ramps in a range of 10 and 90 °C with a 2 µM oligonucleotide concentration. As the ASLs exist as hairpin structures, one concentration is sufficient for $T_m$ determination (monomolecular process). Differently modified mt-tRNA$^{Ser(UCN)}$ ASLs were chosen to examine the influence of the modifications at positions 32 and 37. Compared to the unmodified mt-tRNA$^{Ser(UCN)}$ ($T_m$= 54.9 °C) the melting temperature of the singly (ms$^2$)i$^6$A modified ASL ($T_m$= 53.3/53.9 °C) is slightly reduced (Figure 42). The stability is again minimally enhanced upon the incorporation of additional m$^3$C$_{32}$ ($T_m$= 54.0/54.3 °C) resulting in the double modified mt-tRNA$^{Ser(UCN)}$ ASL (Figure 42). Overall the changes in thermal stabilities were relatively small.[58] These results of the single modified hairpin structure are in line with an earlier report by Kierzek and Kierzek in 2003 about the thermodynamic stability of *N*$^6$-alkyl and 2-methylthio-*N*$^6$-alkyladenosines.[418] They analyzed the steric and electronical effects of *N*$^6$-substituents of naturally occurring and artificially synthesized *N*$^6$-alkylated modifications on duplex and hairpin stability. Within the duplex, they were placing the modified nucleotides on internal (tandem and separated) and terminal positions opposite to uridine as well as a 3'-dangling end. For the hairpin sequence, they chose the ASL of tRNA$^{Trp}$ from *E.coli*, which comprises ms$^2$i$^6$A at position 37 in nature.[418] The thermal stability of the different duplex structures containing the *N*$^6$-alkylated nucleosides is reduced compared to the unmodified oligonucleotides, presumably caused by steric hindrance of the residue at N6. Interestingly, the naturally occurring i$^6$A and m$^6$A and more pronounced ms$^2$i$^6$A, ms$^2$m$^6$A, and ms$^2$A at the 3'-dangling end (ACAUGUX)$_2$) increased the stability of the duplex structures, whereas other *N*$^6$-alkylated derivatives (i.e. 1-methylbutyl; the methyl group at the α-carbon increased the destabilization) decreased the stability compared to the unmodified reference.[418] Similarly to the duplex structures with internal and terminal *N*$^6$-alkylated nucleosides, the modifications within the ACL (ms$^2$i$^6$A, i$^6$A, m$^6$A, ms$^2$m$^6$A and ms$^2$A) of tRNA$^{Trp}$ were destabilizing the hairpin structure in comparison to the unmodified hairpin,

but the differences in thermodynamic stability of the hairpin structures were rather small,[418] in accordance with the obtained results of the singly modified mt-tRNA$^{Ser(UCN)}$ ASLs.[58] However, structures containing $N^6$-alkylated derivatives without thiomethyl group were slightly more stable than the modifications with this additional thiol group.[418] Both experiments, with mt-tRNA$^{Ser(UCN)58}$ and tRNA$^{Trp58,\ 418}$, showed a lower thermodynamic stability of the hairpin structures with (ms$^2$)i$^6$A at position 37 in comparison with the unmodified reference.

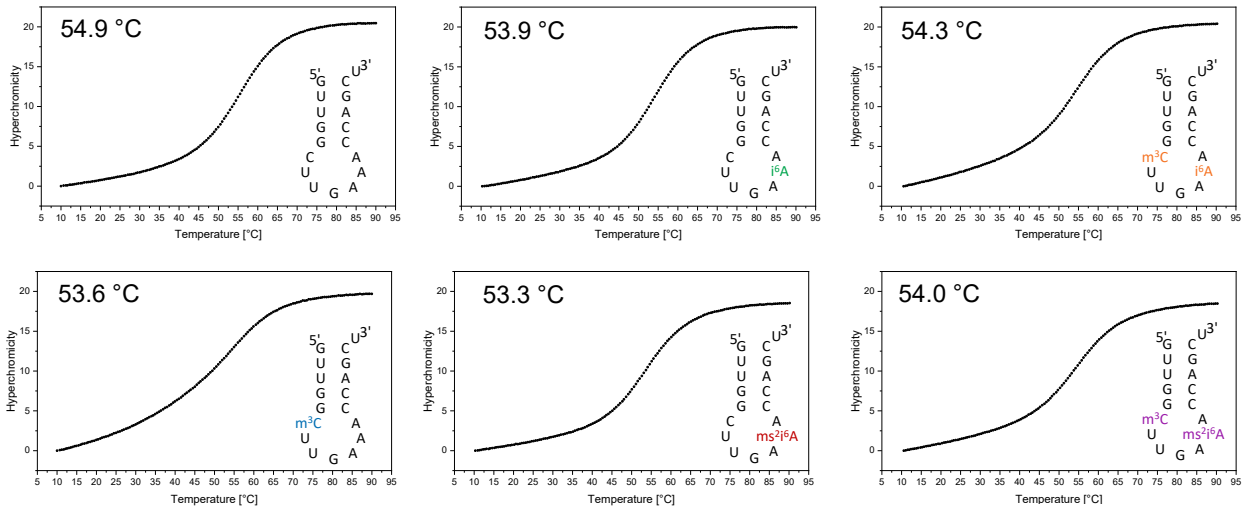



Figure 42. Thermal denaturation experiments of the unmodified (black) as well as i$^6$A$_{37}$ (green), m$^3$C$_{32}$/i$^6$A (orange), m$^3$C (blue), ms$^2$i$^6$A (red) and m$^3$C/ms$^2$i$^6$A (magenta) containing ASLs (from left to right). 2 µM RNA oligonucleotide in 10 mM sodium phosphate buffer (pH 7) containing 100 mM NaCl were used for the six measurements. The graphs are plotted as hyperchromicity at 250 nm versus temperature. In total 5 cooling and heating ramps were measured, of which the first one (20 to 90 °C) is left out of the calculations as an annealing step. The first heating ramp of the two repetitive cycles is shown as an example. The characteristic melting profile is sigmoidal curve due to the transition of the hairpin to ssRNA and the eventuate change in absorption. The average melting temperature, obtained from the inflection points of the four ramps, is shown in the left corner. Figure reproduced from Ref.[58]

Further support for conformational changes caused by varying the modification level was achieved by NMR spectroscopy. $^1$H NMR spectra of 200 µM RNA samples in 10 mM sodium phosphate buffer (pH 7) and 100 mM NaCl in $H_2O/D_2O$ (9/1) were recorded at 10 °C, 17 °C, 25 °C, and 33 °C. Of those, the 10 °C measurements of the unmodified mt-tRNA$^{Ser(UCN)}$ as well as single and double modified ASL, containing i$^6$A and m$^3$C$_{32}$, are exemplarily shown for comparison in Figure 43. Imino protons that are involved in base pairing are obtained as signals between 10 and 14 ppm. Thus, the six obtained signals correspond to the base paired region of the hairpin stem, whereas the loop region does not generate $^1$H signals, caused by high exchange rates with the solvent. The i$^6$A modified ASL showed minor chemical shift changes compared to the unmodified reference, which were increased if m$^3$C$_{32}$ is present in addition. Especially the chemical shift (e.g. peak at 13.3 ppm) and the line-width of the central resonances of the double modified ASL deviate compared to the unmodified and single modified ASLs (Figure 43A): At 10 °C the central resonances exhibit a higher intensity (e.g. peak at 12.2 ppm) and a narrower line width but show varying line broadening upon changing from 10 °C to 25 °C, due to different solvent exchange rates (Figure 43B). Highlighting the signal at 12.2 ppm, it appears to be exceptionally stabilized compared to the other two

ASLs. The positive charge of $m^3C$ might be responsible for the stabilization due to electrostatic interactions,[58] similar to the favorable effect of metal ions, like $Mg^{2+}$, onto the structure and function of RNA by compensating the negatively charged phosphate backbone.[419] The NMR experiments clearly showed the conformational changes of the ASL structure upon the addition of $m^3C$.

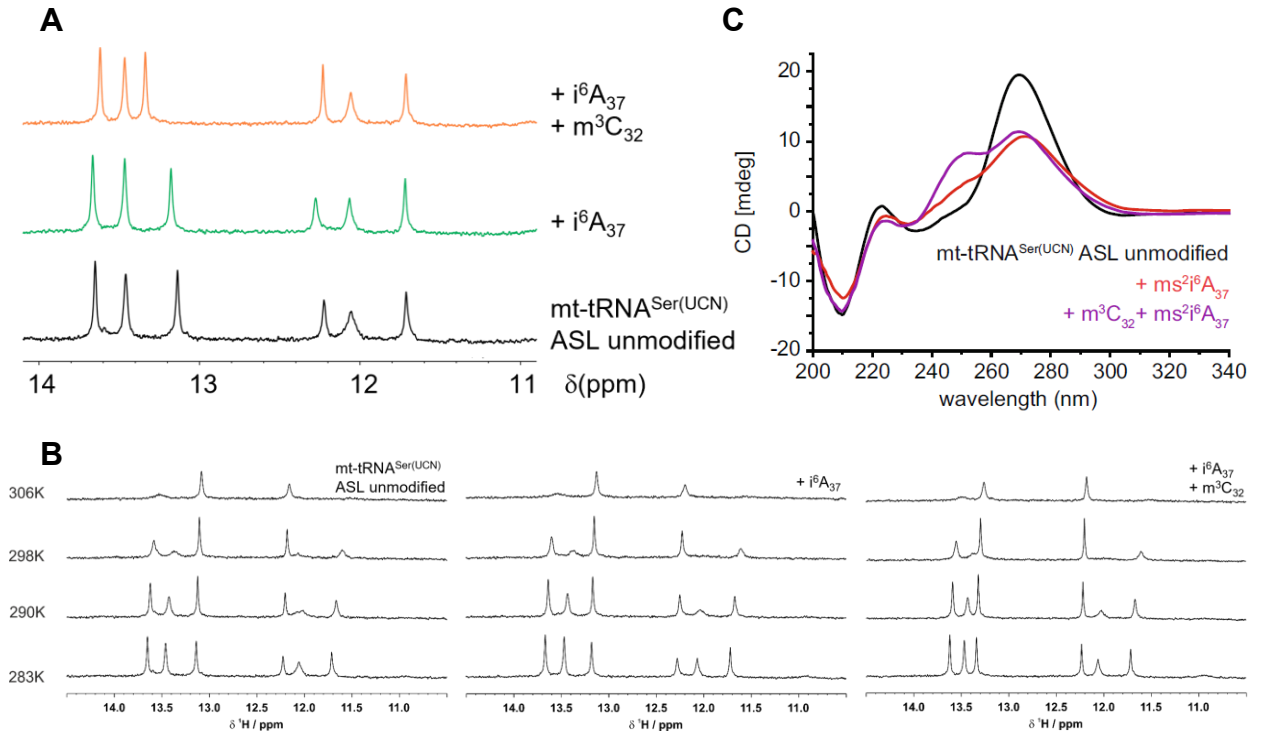



Figure 43. Evidence for conformational changes by NMR and CD spectroscopy. A: NMR of unmodified mt-tRNA$^{Ser(UCN)}$, as well as singly $i^6A_{37}$ and double $i^6A_{37}$ and $m^3C$ modified ASLs with 200 µM RNA using 10 mM sodium phosphate buffer (pH 7) containing 100 mM NaCl in $H_2O/D_2O$= 9/1. The samples were heated to 95 °C for two minutes followed by rapid cooling on ice. The imino region between 11 and 14 ppm shows six signals corresponding to the imino protons of U, A and G involved in base pairing, thus of the stem region of the hairpin. Figure reproduced from Ref.[58] B: Temperature series of the three mt-tRNA$^{Ser(UCN)}$ ASLs measured at 10 °C, 17 °C, 25 °C and 33 °C. C: The graphs of the unmodified mt-tRNA$^{Ser(UCN)}$, singly $ms^2i^6A_{37}$ and doubly $ms^2i^6A_{37}$ and $m^3C$ modified ASLs by CD spectroscopy displayed a broadening and the appearance of a second shoulder upon addition of $m^3C$ to the singly modified ASL. Figure taken from Ref.[58]

The last experiment to gain information about structural changes depending on the modified nucleotides in positions 32 and 37 was CD spectroscopy. It is a powerful method to determine the secondary structure and conformational aspects of chiral molecules like proteins and nucleic acids.[420] Alterations in the curve shape allow for the detection of conformational changes that might also arise due to different temperature or pH conditions.[420] CD spectroscopy relies on the principle of differential light absorption of left- and right-handed circularly polarized light by optically active substances (Figure 44).[421] This circularly polarized light from a xenon lamp is passed through the cuvette containing the chiral molecule and the remaining light intensity is measured by a detector.[422] The Cotton effect (positive or negative) occurs close to the absorption band of the chiral molecule since the optical rotatory dispersion (ORD) of optically active molecules changes rapidly with the wavelength in the region of absorption while crossing zero at the absorption maximum.[423] If the optical rotation initially decreases with the decreasing wavelength it is called the negative Cotton effect and vice versa (Figure 44).[424]

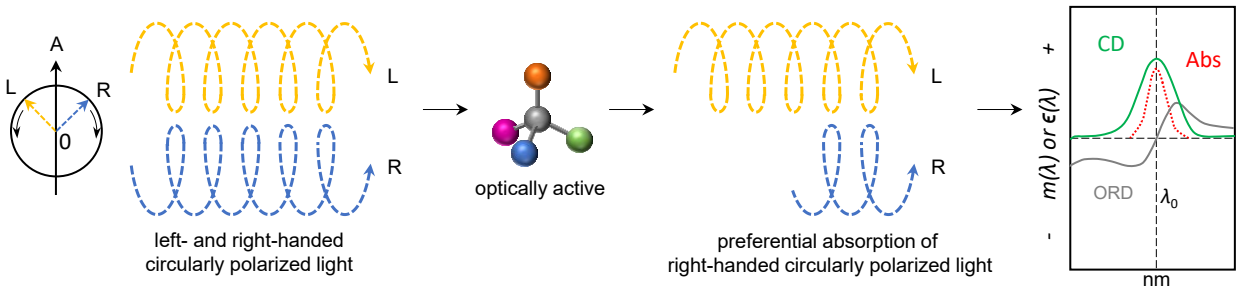



Figure 44. Basis of CD spectroscopy is the differential light absorption of left- (orange) and right-handed (blue) circularly polarized light by chiral molecules (optically active substance). The difference between these two absorptions is resulting in the specific CD signal (green). Thereby, the cotton effect always appears in vicinity to the absorption wavelength (red), displayed by the change of the ORD (grey, positive Cotton effect) with the molar rotation (m(λ)) or molar extinction coefficient (ϵ(λ)). Figure reproduced from Ref.[422, 423]

The CD signal is given as the difference in the absorption of left- and right-handed light, resulting in either positive or negative signals. The signal is obtained according to

$$CD = \Delta A(\lambda) = A \cdot L \cdot \lambda - A \cdot R \cdot \lambda = [\epsilon \cdot L \cdot \lambda - \epsilon \cdot R \cdot \lambda] \cdot d \cdot c = \Delta\epsilon \cdot d \cdot c \qquad (1)$$

with the absorption ($A$), the extinction coefficient of the respective left- or right-handed circularly polarized light ($\epsilon_{R/L}$, in $M^{-1}cm^{-1}$), the concentration ($c$, in mg/mL) of the chiral substance and the path length ($d$, in cm). The resulting CD is plotted against the measured wavelength ($\lambda$) either as molar circular dichroism ($\Delta\epsilon$, equation 1) or ellipticity (θ, in millidegrees (mdeg)): [422, 425]

$$\Delta A(\lambda) = \theta/32.98 \qquad (2)$$

5 µM oligonucleotide samples in 10 mM sodium phosphate buffer (pH 7) containing 1 mM $MgCl_2$ were annealed for 2 min at 95 °C followed by rapid cooling on ice before the measurements. The spectra were measured twice and the average of both measurements from 200 to 340 nm were plotted in Figure 43C after baseline correction by subtracting the spectra obtained from the buffer. The unmodified ASL of mt-tRNA$^{Ser(UCN)}$ shows one major peak at 269 nm (black; Figure 43C) as a positive ellipticity band. The intensity of the single ms$^2$i$^6$A modified ASL was lowered from around 20 to 11 mdeg and its maximum was slightly red-shifted to 271 nm. Additionally, a small shoulder appears at 250 nm. If besides ms$^2$i$^6$A$_{37}$ m$^3$C is added at position 32, this shoulder is significantly more distinct – two peaks at 250 and 270 nm emerged with similar intensity to the single modified ASL.[58] It was shown before, that CD signals of nucleic acids are influenced by the nucleobase stacking geometries.[426, 427] Thus the variations in the ellipticity of the modified samples in Figure 43C reflect changes in base stacking.[58] The same CD experiments were also performed using 10 mM sodium phosphate buffer (pH 7) containing 100 mM NaCl (as in the thermal denaturation experiments) and with additional $MgCl_2$, but the CD spectra mostly coincided.

In addition the influence of the point mutations G35A and A38C (see Figure 38) were analyzed by CD spectroscopy as well. In accordance to the spectra shown in Figure 43C 5 µM oligonucleotide samples in 10 mM sodium phosphate buffer (pH 7) containing 1 mM $MgCl_2$ were annealed for 2

min at 95 °C followed by rapid cooling on ice before the measurements to ensure proper folding.[58, 428] The point mutations, G35A and A38C were analyzed in comparison with the unmodified and $ms^2i^6A_{37}$ modified mt-tRNA$^{Ser(UCN)}$ (Figure 45A/B), respectively, to relate the structure with the functions. The overall shape and intensity of the CD spectra for the unmodified mt-tRNA$^{Ser(UCN)}$ and the mt-tRNA$^{Ser(UCN)}$ containing G35A and $ms^2i^6A_{37}$, remained. Both tRNAs were shown to be weak substrates for the METTL8-mediated methylation, which is reflected in a similar structural appearance (Figure 45A). In contrast, the intensity of the CD spectra of mt-tRNA$^{Ser(UCN)}$ and the ASL with the point mutation A38C, who both contain $ms^2i^6A_{37}$, is reduced due to broadening of the signal at around 270 nm (Figure 45B). Those two tRNA ASLs represent good targets of METTL8 and coincide in the overall shape of the CD spectra. The same trend for both groups was also observed on native PAGE (Figure 45C): The unmodified tRNA$^{Ser(UCN)}$ ASL and the one containing the G35A mutation and $ms^2i^6A_{37}$ from Figure 45A migrate similarly, but faster than the tRNA$^{Ser(UCN)}$ ASL containing the $ms^2i^6A_{37}$ modification and in addition the A38C (green) substitution from Figure 45B.[58]

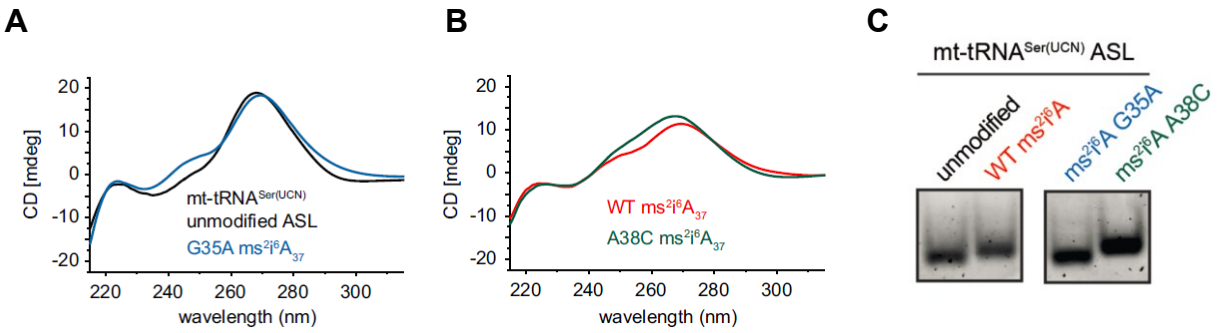



Figure 45. Examination of the structural aspects of the WT, as well as mutated ASLs of mt-tRNA$^{Ser(UCN)}$. A: CD spectroscopy of mt-tRNA$^{Ser(UCN)}$ and G35A substituted substrate. B: CD spectroscopy of mt-tRNA$^{Ser(UCN)}$ A35C substituted substrate, both with $ms^2i^6A_{37}$. C: native PAGE of the WT and mutated ASLs. Figures taken from Ref.[58]

Altogether, these experiments, including NMR analysis, CD spectroscopy, thermal denaturation experiments, and native page separation, gave several indications for the structural impact of $m^3C_{32}$ alone and in combination with the respective modifications $t^6A$ and $(ms^2)i^6A$ at position 37, caused by changes in base stacking and electrostatic interactions.[58] Even though they are present in the loop region, they are affecting and fine-tuning the mt-tRNA$^{Thr/Ser(UCN)}$ ASL structure noticeably.[58] This raised the question if these structural changes are accompanied with effects on the function of the ASLs.

### 3.1.3.2 Effect on aminoacylation and mitochondrial translation

To examine if the loss of $m^3C_{32}$ impairs cellular functions, different experiments were performed by Nicole Kleiber *et al.*.[58] First indications were obtained by monitoring the cell growth of WT cell lines in comparison with KO1 and KO2 cell lines, lacking METTL8, whereby the latter showed a slowed down growth compared to WT (Figure 46A). The absence of $m^3C_{32}$ in the KO cell lines seems to have a detrimental influence on cell growth.

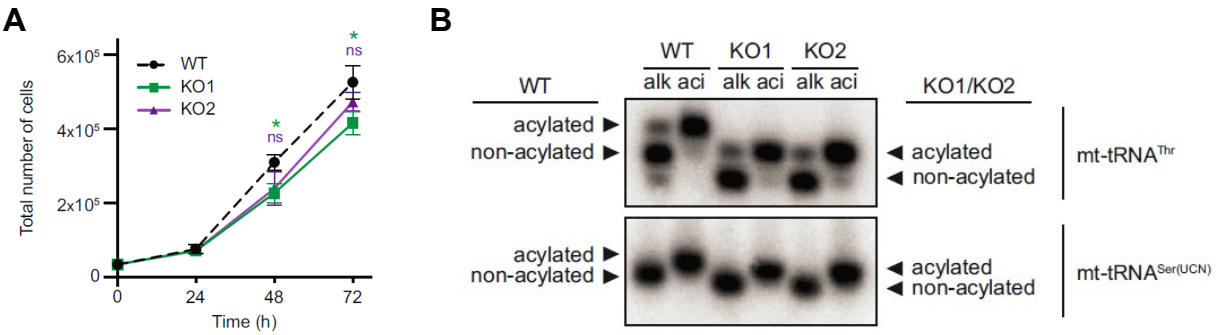



Figure 46. Examining the influence of the absence of $m^3C_{32}$ on cellular functions of mt-tRNA$^{Thr/Ser(UCN)}$. The cell growth of METTL8 KO1 and KO2 cell lines is impaired compared to WT (HEK293), illustrated by a lower number of cells (three independent experiments). B: Isolating tRNAs under acidic conditions preserves the covalently attached amino acid. A control sample was treated with basic conditions to remove the acyl group. The different migration behavior of deacylated and acylated tRNAs obtained from WT and METTL8 KO cell lines were analyzed by acidic PAGE; alk = after alkaline treatment, aci = isolated under acidic conditions, without further treatment. Figures taken from Ref.[58]

However, no effect on aminoacylation for METTL8 KO cells was observed, one of the most important aspects of tRNAs during translation, which could have explained the influence on cell growth. Isolating the tRNAs from WT and KO cell lines under acidic conditions enabled the detection of acylated substances (Figure 46B). Both mt-tRNA$^{Thr/Ser(UCN)}$ from the cells lacking METTL8 were efficiently aminoacylated.[58] Also the next step, the association with the mitoribosome after successful aminoacylation, was not impaired for the METTL8 KO cell lines. The conformational changes of mt-tRNA$^{Thr/Ser(UCN)}$ due to the absence of $m^3C_{32}$ did not affect this part of the translation process.[58] Conspicuously, even though that aminoacylation and association with the mitoribosome of mt-tRNA$^{Thr/Ser(UCN)}$ from cell lines lacking $m^3C$ on position 32 were unaffected, the level of some of the de novo synthesized mitochondrial proteins was decreased compared to WT (See Figure 7, Manuscript 58).[58] Thus, the reported structural changes (chapter 3.1.3.2) upon installation of $m^3C_{32}$ by METTL8 seem to fine-tune the mt-tRNA$^{Thr/Ser(UCN)}$ ASLs to promote efficient mitochondrial translation.[58] This is further supported by a recent report by Gunter Meister and coworkers.[429] They examined the dependency of codons on the METTL8 level using ribosome profiling. In this way, they could identify codons located at the A, P, and E sites of the mitoribosome. Particularly, the level of codons encoding for Ser (mt-tRNA$^{Ser(UCN)}$) is increased in the P site for KO cell lines, lacking METTL8, reflecting a longer period of residence of the ribosome. In contrast, overexpression of METTL8 is leading to a prolonged accommodation of the ACC codon (mt-tRNA$^{Thr}$) to the A site of the ribosome. They assumed that specific codons of mt-tRNAs that are substrates of METTL8 are varyingly strong affected by the METTL8 level. This leads to altered levels of the mitochondrial proteins by translation.[429]

Together with the collaboration partners vital new information about the RNA methyltransferase METTL8, the interdependencies of the modifications within the ACL, and their structural impact could be elucidated. Both human mt-tRNA$^{Thr}$ and mt-tRNA$^{Ser(UCN)}$ are substrates of the mitochondrial matrix protein METTL8, earlier supposed to be responsible for the installation of $m^3C$ in

mRNAs.[64] The selection of the correct substrate is achieved by recognition elements within the ACL, like U34 and G35, which are present in both tRNAs. The efficient installation of the methyl group at position C32 is further promoted by the respective modification on position A37, of which ($ms^2$)$i^6A$ is a prerequisite for the methylation, while $t^6A$ is enhancing the reaction. These features are leading to the observed selectivity for mt-$tRNA^{Thr/Ser(UCN)}$ as METTL8 targets even though METTL8 was shown to crosslink to more mt-tRNA ASLs by CRAC experiments. A variety of experiments were able to demonstrate the structural influence of the single ($m^3C_{32}$) and double modified mt-$tRNA^{Thr/Ser(UCN)}$, as well as of U34 and G35. Besides the impact of $m^3C$ to fine-tune the tRNA structure, also the functional aspects have been analyzed. Both mt-$tRNAs^{Thr/Ser(UCN)}$ isolated from METTL8 KO cell lines were still aminoacylated and associated with the mitochondrial ribosomes, but the mitochondrial translation was slightly impaired, reflected by a decrease in the synthesis of some mitochondrial proteins. This shows how important the effect of posttranscriptional modifications with their respective writer enzymes can be on the continuation of the mitochondrial functions and how extensive the consequences can be. As already mentioned earlier, the pathogenic A to G mutation on position 38 of $tRNA^{Thr}$ prevents the formation of mitochondrial $t^6A$ leading to mitochondrial disorders.[414] And a recent report from Gao and coworkers revealed the contribution of METTL8 on R-loop regulation, a DNA-DNA:hybrid structure, via a SUMOylated nuclear RNA-binding protein complex. This hybrid structure affects genome stability adversely, further promoting tumorigenesis.[430] The obtained results in collaboration with Nicole Kleiber *et al.* thus contributed to a better understanding of the methyltransferase METTL8 and the involved modifications within the ASL as part of the fundamental research necessary to gain further knowledge of cellular processes, especially related to diseases. This also includes the interplay with the corresponding eraser enzymes, namely the demethylation enzymes ALKBH1 and ALKBH3, to further examine which enzymes are taking part in the dynamic modification pathway of tRNAs.

### 3.1.4 Investigation of eraser enzymes: the dioxygenases ALKBH1 and ALKBH3

As described in chapter 1.3.4 the level of modifications can be further adjusted. Posttranscriptional modifications e.g., after the installation of a methyl group by writer enzymes, can be removed, thereby giving the opportunity for regulatory functions. Removal of methyl groups is performed by enzymes called demethylases (eraser). Human ALKBH1 and ALKBH3 belong to the α-ketoglutarate (α-KG)- and Fe(II)-dependent dioxygenases as part of a family of nine mammalian homologs (ALKBH1-8 and FTO) that are catalyzing the oxidation of the alkyl group. Human ALKBH3 protects the cell from alkylation damage ($m^1A$ and $m^3C$) by a DNA repair mechanism,[112, 113] and also demethylates $m^1A$ from mRNA and $m^3C$ from tRNA.[65, 122, 123] ALKBH3 has been found in the nucleoplasm and also in the cytoplasm, but is excluded from nucleoli.[123] ALKBH1 is also able to demethylate $m^3C$ in ssDNA and RNA in vitro, and it is a mitochondrial protein responsible for the conversion of $m^5C$ to $f^5C$ in mt-$tRNA^{Met}$,[128, 125, 129] but the subcellular localization in the nucleus was observed as well.[127]

To examine the activity of ALKBH1 and ALKBH3 on various tRNA targets and modification states, cytoplasmic tRNAs were synthesized in addition to the mitochondrial tRNA described before (see Table 1). Similar questions as for METTL8 should be addressed: Is the ASL sufficient for the recognition of the target substrates by the enzyme and are the respective modifications on position 37 prerequisites for the demethylation of $m^3C_{32}$? Additionally, the specificity of both proteins should be investigated on mitochondrial and cytoplasmic tRNAs.

Chen *et al.* already reported the demethylation of $m^1A$ and $m^3C$ from tRNAs by ALKBH3 *in vivo* and *in vitro* in 2019.[122] By immunofluorescence, western blot analysis, and staining by Flag antibody after overexpression of ALKBH3 the primary cytoplasmic localization could be determined in various cell lines. To further unveil which RNAs are bound to this protein, crosslinking and immunoprecipitation sequencing (CLIP-Seq) were performed, revealing tRNA as the highest enriched group of the input RNAs. Of this, the major fraction have been $tRNA^{Gly(GCC)}$, $tRNA^{Gln(CTG)}$, $tRNA^{His(GCG)}$, $tRNA^{Gly(CCC)}$, $tRNA^{Ser(GCT)}$, and $tRNA^{Lys(CTT)}$. Thus, ALKBH3 was able to bind to tRNAs *in vivo*. *In vitro* assays with tRNAs isolated from HeLa and HEK *293T* cells identified $m^1A$ and $m^3C$ as the only modifications that were demethylated by ALKBH3. In ALKBH3 KO cell lines (i.e. HeLa cells that do not express ALKBH3), the level of $m^3C$ was significantly enhanced for cytoplasmic $tRNA^{Ser(GCT)}$, $tRNA^{Leu(CAA)}$, $tRNA^{Arg(CCT)}$ and $tRNA^{mMet(CAT)}$. Overall, ALKBH3 was able to regulate $m^1A_{58}$ (Gly), $m^3C_{32}$ (Arg, Ser) and $m^3C_{47}$ (Ser) in tRNAs.[122] Similar experiments were performed for ALKBH1 by Liu *et al.*.[431] Primarily tRNAs were enriched from all RNA species in CLIP sequencing, as ALKBH1 contains a tRNA-binding domain motif. The methylation levels of known modifications were analyzed by LC-MS/MS of tRNAs isolated from HeLa cell lines treated with recombinant ALKBH1. Only the level of $m^1A_{58}$ was decreased, but not other modifications like $m^7G$, $m^3C$, or

$m^5C$, identifying $m^1A$ as the main substrate of ALKBH1. This was further supported by LC-MS/MS experiments, in which ALKBH1 was either deleted (KO) or overexpressed in HeLa cell lines. The level of $m^1A$ was correspondingly increased (~6%) or decreased (~16%) in total tRNA, respectively, thus gets regulated by ALKBH1.[431]

Another feature of ALKBH1 was reported by Haag *et al.*. The $m^5C$ modification in mt-$tRNA^{Met}$, initially installed by the methyltransferase NSUN3 at the wobble position C34, is further functionalized by oxidation with ALKHB1, resulting in 5-formylcytidine ($f^5C$).[125] A reduced level of either of the two enzymes, responsible for modifying $C_{34}$ in mt-$tRNA^{Met}$, impairs mitochondrial translation *in vivo*. Both modifications are essential for the function of mt-$tRNA^{Met}$ during initiation and elongation of translation in human cells.[125] The ability of ALKBH1 to oxidize or demethylate $m^3C_{32}$ has not been shown. Even though demethylases are already the focus of extensive studies, the precise target spectrum, recognition elements and the cellular functions have not fully been elucidated yet. More effort is needed to unveil fundamental cellular processes regarding the dioxygenase family.

Demethylation experiments were planned to elucidate some of the open questions in the dynamic network of $m^3C$. Demethylation assays were performed with the cytoplasmic and mitochondrial ASLs with different modification states. The proteins ALKBH1 and ALKBH3 were kindly provided by Prof. Dr. Markus Bohnsack and Nicole Kleiber. For the analysis by HPLC experiments, 200 pmol of the ASL was incubated with cofactors α-KG and ammonium iron (II) sulfate, with Ribolock as RNase inhibitor and the respective Enzyme (ALKBH1/3) in oxidation buffer (with and without $MgCl_2$). The reaction was stopped by digesting the protein with Protein kinase K and processed by phenol:chloroform:isoamyl (PCI) alcohol extraction, to separate the oligonucleotide from the protein. After precipitation with ethanol, the oligonucleotide sample was dissolved in water and the reactions were injected into HPLC to analyze them by anion-exchange chromatography as well as ESI-MS or MALDI. The exact compositions will be shown in chapter 5.2.5.

In the case of successful demethylation, a shift towards higher retention times was expected, due to the loss of the positively charged methyl group of $m^3C_{32}$. This was first tested with the cyt-$tRNA^{Ser(UGA)}$ in the absence and presence of $m^3C_{32}$, obtaining a clear shift of the retention times (~0.5 min) by anion exchange chromatography. Following, the demethylation assay was performed with the three cyt-$tRNAs^{Arg(CCU)}$, cyt-$tRNAs^{Thr(AGU)}$, and cyt-$tRNAs^{Ser(UGA)}$ using ALKBH1, ALKBH3, or the purification Tag MPB as negative control (His14-MBP tagged proteins) for each tRNA, respectively. After the workup, the oligonucleotide samples were injected into anion-exchange HPLC and plotted after normalization. The ratio between the two peaks was calculated by integrating the area under the curve and dividing the obtained peak of the $m^3C$ modified ASL by the demethylated ASL ($m^3C$-ASL/C-ASL). This ratio is shown as bar graphs and should estimate the efficiency of the demethylation activity (Figure 47).

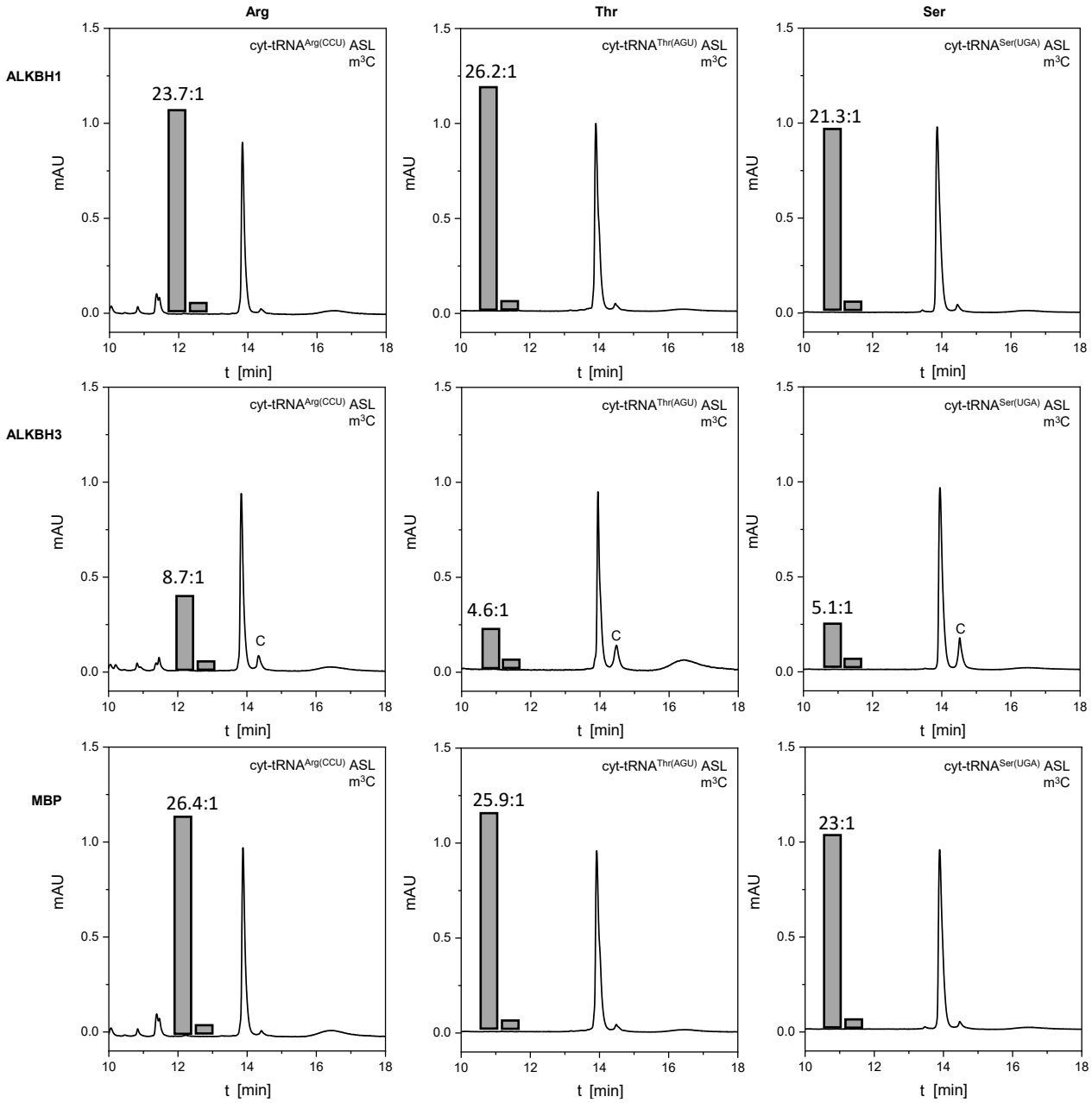



Figure 47. Demethylation assay with three cytoplasmic tRNAs, Arg (CCU), Thr (AGU) and Ser (UGA). Each of them was either incubated with ALKBH1, ALKBH3 or only the MBP tag together with the cofactors α-KG and ammonium iron (II) sulfate and analyzed by anion-exchange chromatography. The experiments were performed once by incubating the reaction for 1 h. The demethylated reaction product is marked with a C (a minor peak occurs for all samples due to byproduct formation occurring during oligonucleotide purification). The ratio of the obtained peaks is given in grey bar graphs. They were calculated by dividing the peak area of the methylated by demethylated oligonucleotides. A smaller ratio corresponds to a higher activity of the enzyme, leading to demethylation.

For cyt-tRNA$^{Arg(CCU)}$ some minor additional peaks were observed at shorter retention times that might be caused by some degradation or incomplete purification. These peaks were not considered for the calculations of the ratio between m$^3$C-ASL and C-ASL. Like the observed activity of ALKBH3 for cyt-tRNAs by Chen *et al.*, ALKBH3 was able to recognize and partially demethylate the three cyt-tRNAs, thus the ASLs seem to be sufficient for recognition. This was identified by the appearance of an additional peak with a longer retention time, due to the removal of the methyl group (Figure 47). By comparison of the three cyt-tRNAs, the ASLs of tRNA$^{Thr}$ and tRNA$^{Ser}$ appear

to be slightly better substrates for ALKBH3 than tRNA$^{Arg}$. For the same reactions using ALKBH1 or MBP as a negative control, no further peak was observed. The small additional peaks, which can be seen in the HPLC chromatograms were already present after the deprotection of the oligonucleotide after solid phase synthesis, most likely due to the formation of m$^3$U (see chapter 3.1.2.1), which could not be separated. Accordingly, the calculated ratio of the cyt-tRNAs treated with ALKBH1 and MBP is noticeably higher than the samples treated with ALKBH3. The ASL of cyt-tRNAs might not be sufficient for the recognition of ALKBH1, which could be tested using the whole tRNA structures. But also before, only the efficient demethylation of m$^1$A in tRNAs was shown by ALKBH1, which is indicative of the specificity of ALKBH3 for m$^3$C in cytoplasmic tRNAs.

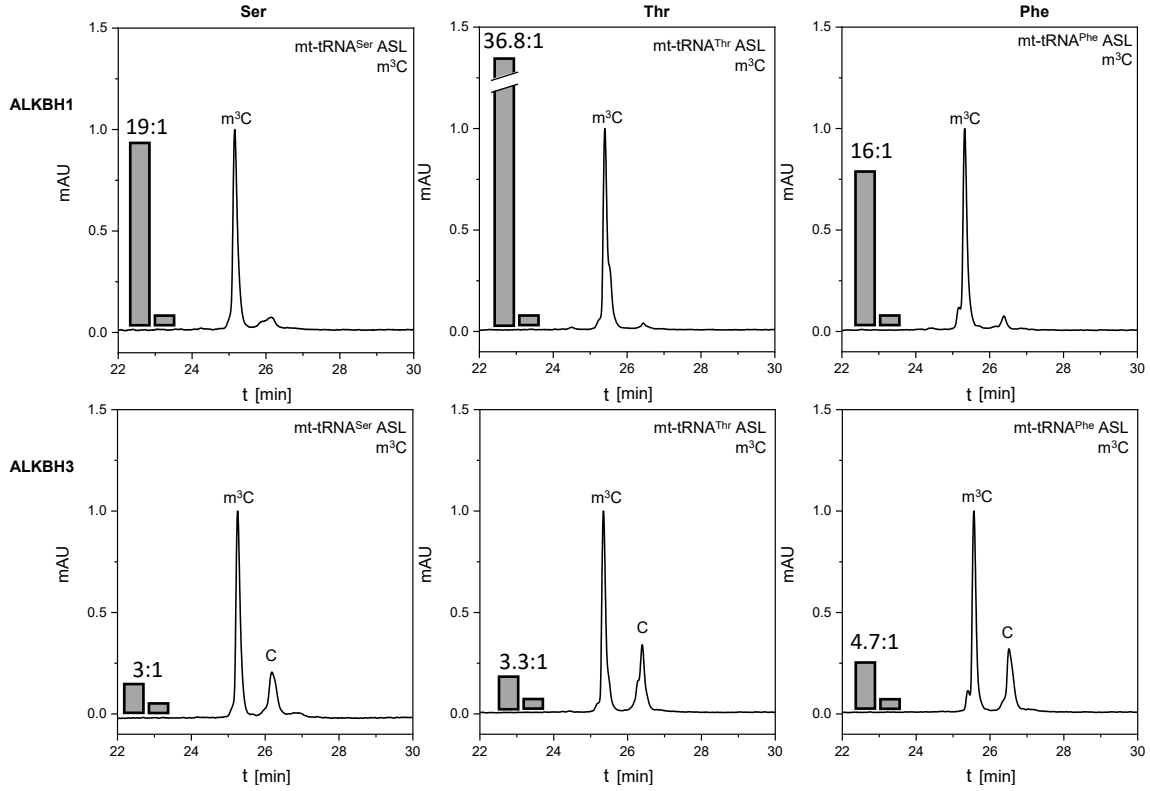



Figure 48. HPLC chromatograms of the mt-tRNA$^{Thr/Ser/Phe}$ ASLs (R604/R605/R606) after treatment with the respective protein together with the cofactors α-KG and ammonium iron (II) sulfate. The upper line shows the result for ALKBH1 and the lower line for ALKBH3. The experiments were also performed with MBP-tag as negative control and were in accordance with the data for ALBKH1 (data not shown here).

To further investigate the substrate specificity of ALKBH3 as an m$^3$C demethylase and to examine if ALKBH1 as a mitochondrial protein can demethylate m$^3$C in mt-tRNAs, the demethylation assay was performed in accordance with the samples of the cyt-tRNAs. For the mt-tRNAs, the same oligos as for the methylation experiments with METTL8 (see chapter 3.1.3) were chosen. In addition, mt-tRNA$^{Phe}$ was synthesized, which does not contain m$^3$C$_{32}$ in nature, to compare the selection characteristics of substrates by ALKBH1 and ALKBH3. The mt-tRNA$^{Thr/Ser/Phe}$ ASLs were incubated with ALKBH1, ALKBH3, and MBP-tag with the cofactors α-KG and ammonium iron (II) sulfate for 1 h, injected into HPLC, and plotted after normalization (Figure 48). The ratio m$^3$C-ASL/C-

ASL is given as grey bar graphs. Surprisingly, ALBKH3 exhibited demethylation activity for all three mt-tRNAs$^{Thr/Ser/Phe}$ ASLs resulting in an additional peak with a higher retention time (Figure 48).

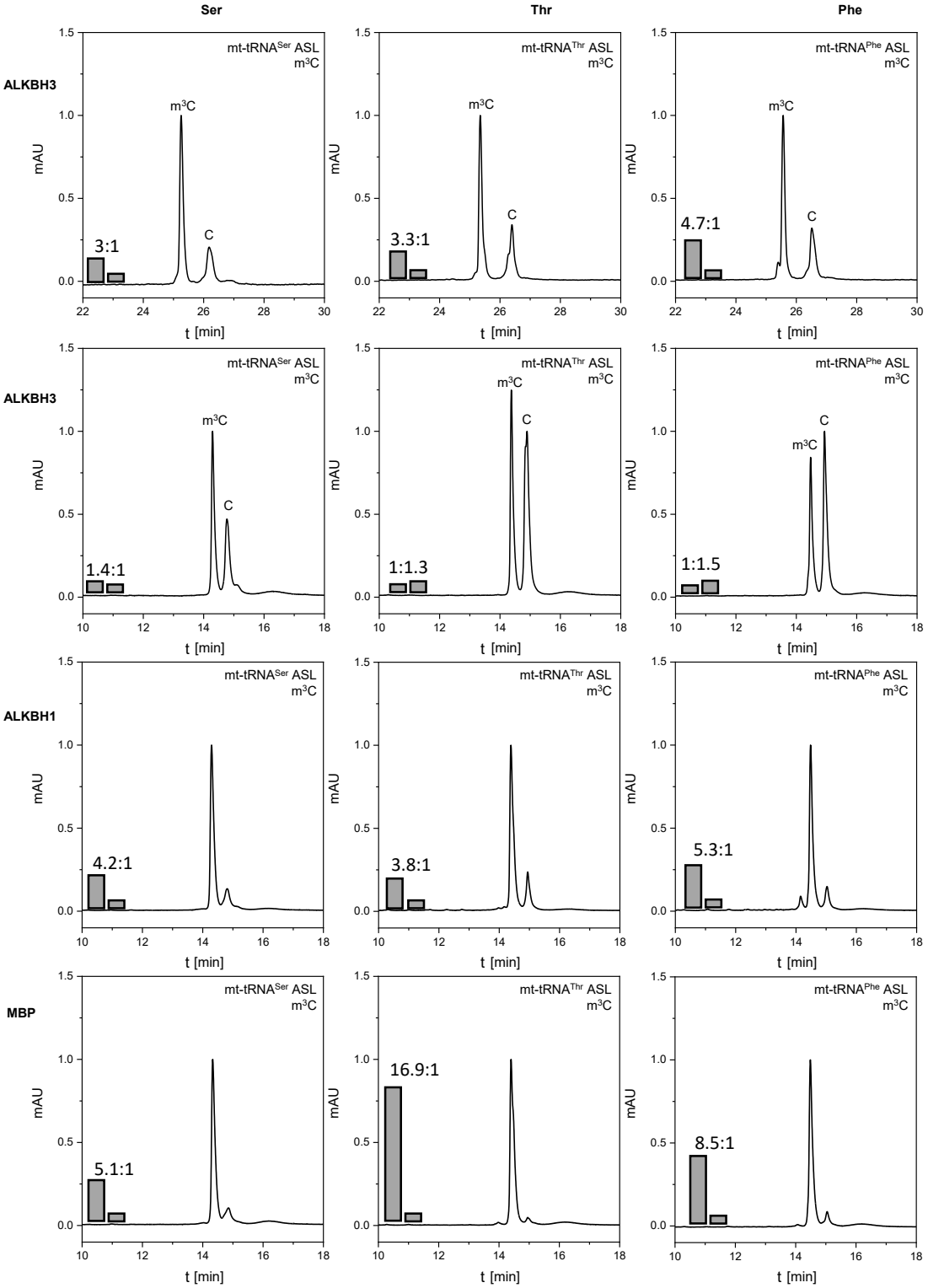



Figure 49. Comparison of the old (upper row) and new batch (second line) of ALKBH3 on mt-tRNA$^{Ser/Thr/Phe}$ ASLs (R604/R605/R606). The difference in retention time is due to the usage of two different column sizes for anion-exchange chromatography: Dionex DNAPac PA200 4x250 for the old batch of ALKBH3 and 2x250 for the new batch of ALKBH3. The new batch of ALKBH3 is giving a different ratio of m$^3$C-ASL/C-ASL and appears to be more active. The treatment of mt-tRNA$^{Ser/Thr/Phe}$ ASLs with ALKBH1 (third line) also led to a slight increase of the additional peak with higher retention times compared to the ASLs treated with the MBP-tag.

The ASL appeared to be sufficient for the recognition and even though the substrates have been mt-tRNAs and ALKBH3 was only found in the cytoplasm and nucleus so far, the treatment with ALKBH3 led to the partial removal of the methyl group. Curiously, mt-tRNA$^{Phe}$ does according to present knowledge not naturally contain the $m^3C_{32}$ modification at this position but rather a conserved cytidine. It was inserted artificially into the synthetic ASL, but it was still accepted as a substrate for ALKBH3, reflecting an arbitrary selection of the ASL targets. The sequence of the ASL does not seem to be determining for substrate recognition in cellular processes. If the fact that ALKBH3 can demethylate $m^3C_{32}$ in mitochondrial tRNAs is physiologically relevant has to be further investigated. Therefore, it needs to be demonstrated that ALKBH3 like ALKBH1 also has a mitochondrial localization, which was indicated in a protein database.[432] In accordance with the demethylation results of cyt-tRNA ASLs with ALKBH1, also the mt-tRNAs ASL were no substrates for the demethylase, which would be indicative of the specificity of ALKBK3 for $m^3C_{32}$.

However, after the first round of demethylation assay using ALKBH1 and ALKBH3 on mt-tRNA$^{Thr/Ser/Phe}$ ASLs, a new batch of each protein was purified by Nicole Kleiber. With this, the latter experiments with mt-tRNA$^{Thr/Ser/Phe}$ were repeated, to compare the activity of the proteins. The activity of the new batch appeared to be significantly higher, but the trend of the demethylation remained (Figure 49). Also, the ratios for the demethylation experiments with ALKBH1 changed but were less pronounced than for ALKBH3. Compared to the samples treated with the old batch of ALKBH1 (Figure 48, upper line) and to the negative control treated with the MBP-tag the proportion of the second peak with the higher retention time was slightly increased, especially for mt-tRNA$^{Thr}$. The question remains if ALKBH1 might be able to partially demethylate $m^3C_{32}$, but with much lower efficiency than ALKBH3. Further experiments would need to be performed, to evaluate if this higher ratio is indeed due to the demethylation of $m^3C_{32}$. The measurements of the cyt-tRNA ASLs shown in Figure 47 were also repeated with the new batch of ALKBH3. The ratio was decreased by 1.6-2-fold due to the higher activity of the protein. All of the HPLC chromatograms of the demethylation assays shown from now on have been performed with the new batch of ALKBH1 and ALKBH3.

An overview of the expected retention profiles of the mt-tRNA$^{Ser(UCN)}$ ASLs with different modification states is shown in Figure 50A and B. As expected, the ASLs without the methyl group at C32 show a shift to higher retention times (Figure 50B) compared to their $m^3C$ modified ASL reference. Upon incorporation of single modified ($ms^2$)$i^6A$ a similar shift was observed, which was much more pronounced for $ms^2i^6A$ than for the single modified ASL without the thio methyl group. The resulting HPLC chromatograms were evaluated by integration of the oligonucleotide peaks at 260 nm and giving the ratio of the resulting peaks ($m^3C$-ASL/C-ASL) shown as grey bar graphs next to the chromatogram. The $m^3C_{32}$ containing mt-tRNA$^{Ser(UCN)}$ was used in the demethylation assays using ALKBH1 and ALKBH3, respectively. Triplicates were measured and plotted after normalization (Figure 51), to verify the reproducibility of the results shown in Figure 49.

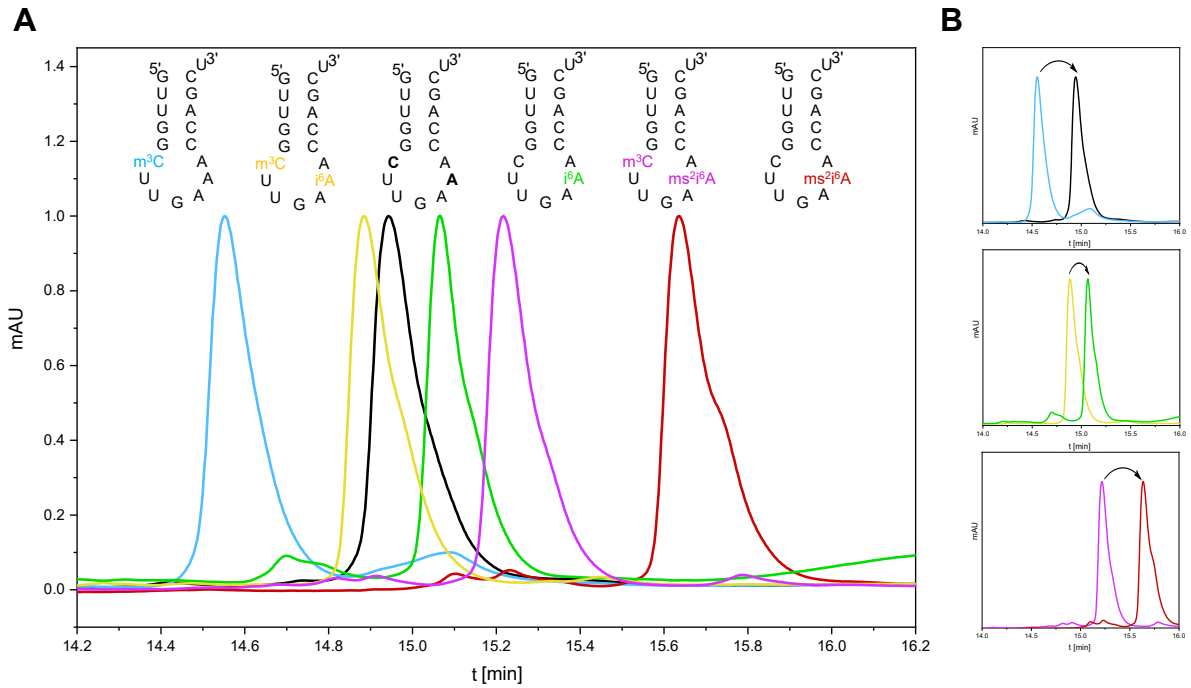



Figure 50. Anion exchange chromatography of mt-tRNA$^{Ser(UCN)}$ ASLs. A: Overview of the retention times of the differently modified ASLs compared to the unmodified one (black). The structures of the ASL are shown on top of the HPLC traces. B: excerpts of the respective ASLs before and after potential demethylation, leading to a shift of the peak to higher retention times, indicated by the arrows.

To summarize, ALKBH3, of which the demethylation activity for the removal of m$^3$C was shown for cyt-tRNAs, led to a distinct formation of a second peak for the mt-tRNAs with almost equal distribution after incubation for 1 h (Figure 51). In contrast, even though the measured ASL is from mt-tRNAs, the demethylation using ALKBH1 as a mitochondrial protein appeared not to be successful.

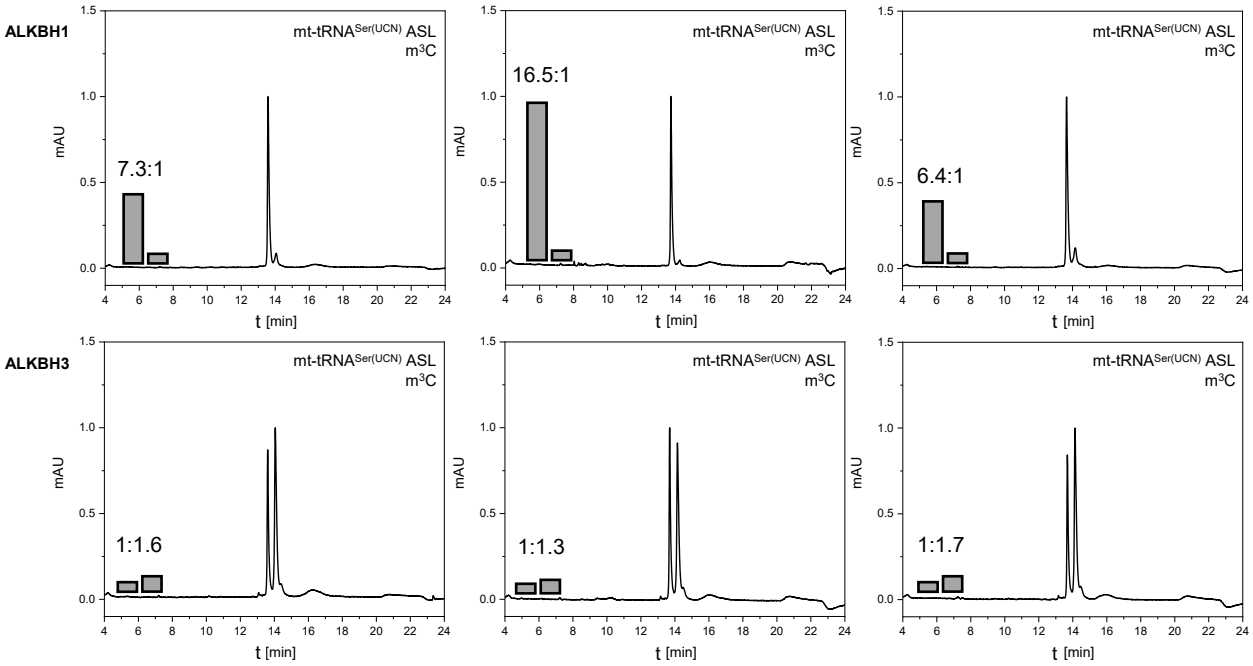



Figure 51. Triplicate measurements from left to right of mt-tRNA$^{Ser(UCN)}$ with m$^3$C$_{32}$ (R604) using ALKBH1 in the upper row, and ALKBH3 in the lower row. The samples were incubated with the respective enzyme for 1 h.

As mentioned earlier for the demethylation assay with the cyt-tRNAs the small additional peak which can be observed in the chromatogram is due to the formation of $m^3U$ during deprotection.

In chapter 3.1.3 it was shown that the methylation activity of METTL8 was dramatically dependent on the presence of the respective modification ($t^6A$/$(ms^2)i^6A$) on position 37. To investigate if this circuitry is also applying to the contrary reaction, the removal of the methyl group of $m^3C_{32}$, the demethylation assay was performed with the $m^3C$ and $(ms^2)i^6A$ modified mt-tRNA$^{Ser}$ ASL (Figure 52). ALKBH1, ALKBH3, and MBP-tag were used for these experiments to examine if the activity of either ALKBH1 or ALKBH3 might be enhanced by the presence of $(ms^2)i^6A_{37}$. However, ALKBH1 is still not able to demethylate $m^3C$ mt-tRNA$^{Ser}$. For ALKBH3, the formation of the additional peak was observed, but with a much lower ratio compared to Figure 51. The structural changes of the ASL for the double modified mt-tRNA$^{Ser}$ ASL described in chapter 3.1.4 seemed to disfavor the activity of ALKBH3, in contrast to the beneficial effect on the methylation activity of METTL8.

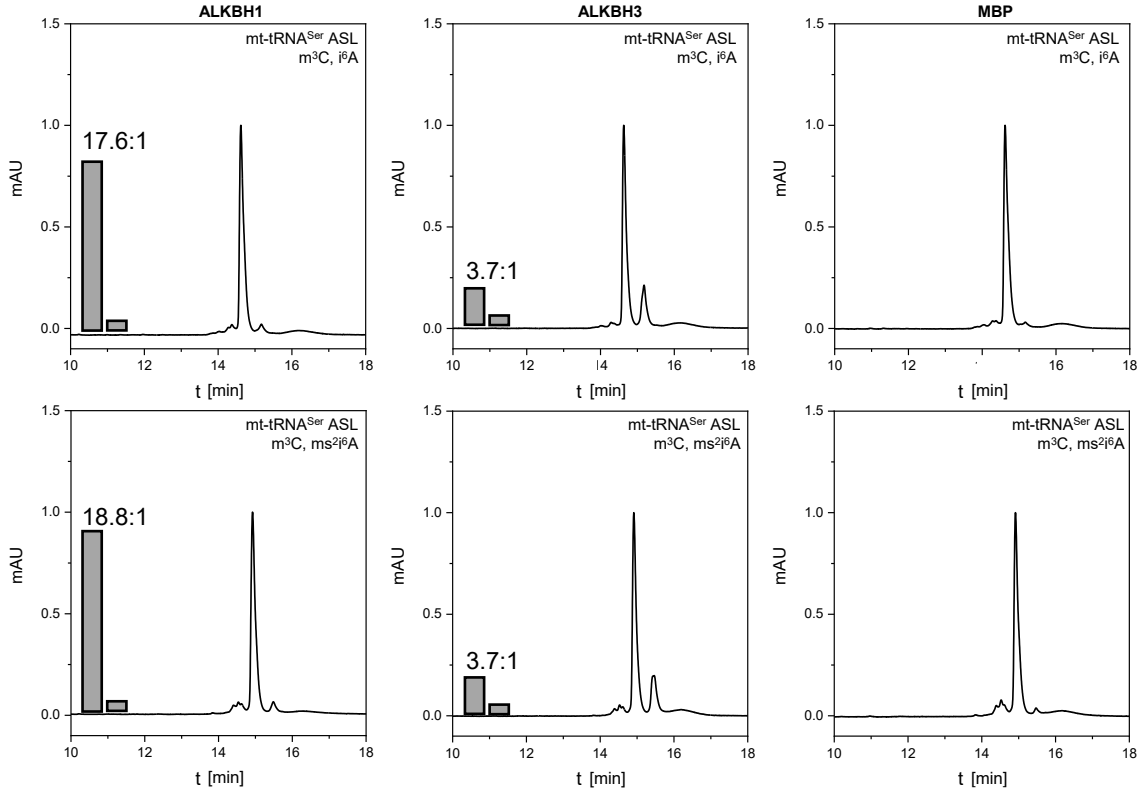



Figure 52. Demethylation assay with double modified mt-tRNA$^{Ser}$ (R824, R825) using ALKBH1, ALKBH3, and MBP-tag. The $m^3C$-ASL/C-ALS ratio is given as grey bar graphs.

To investigate, if ALKBH1 might interact with other modifications within the tRNA ASL, control experiments were performed using ALKBH1 and MBP-tag, respectively, with the mt-tRNA$^{Ser/Thr/Phe}$ ASLs without $m^3C$, but with $i^6A$ on position 37. They were incubated together with α-KG and ammonium iron (II) sulfate for 1 h and purified. For both mt-tRNA$^{Ser}$ and mt-tRNA$^{Thr}$ no additional peak formation was observed, neither in the ALKBH1 nor the MBP-tag reaction (Figure 53). However, for mt-tRNA$^{Phe}$, a small peak with a slightly reduced retention time was observed.

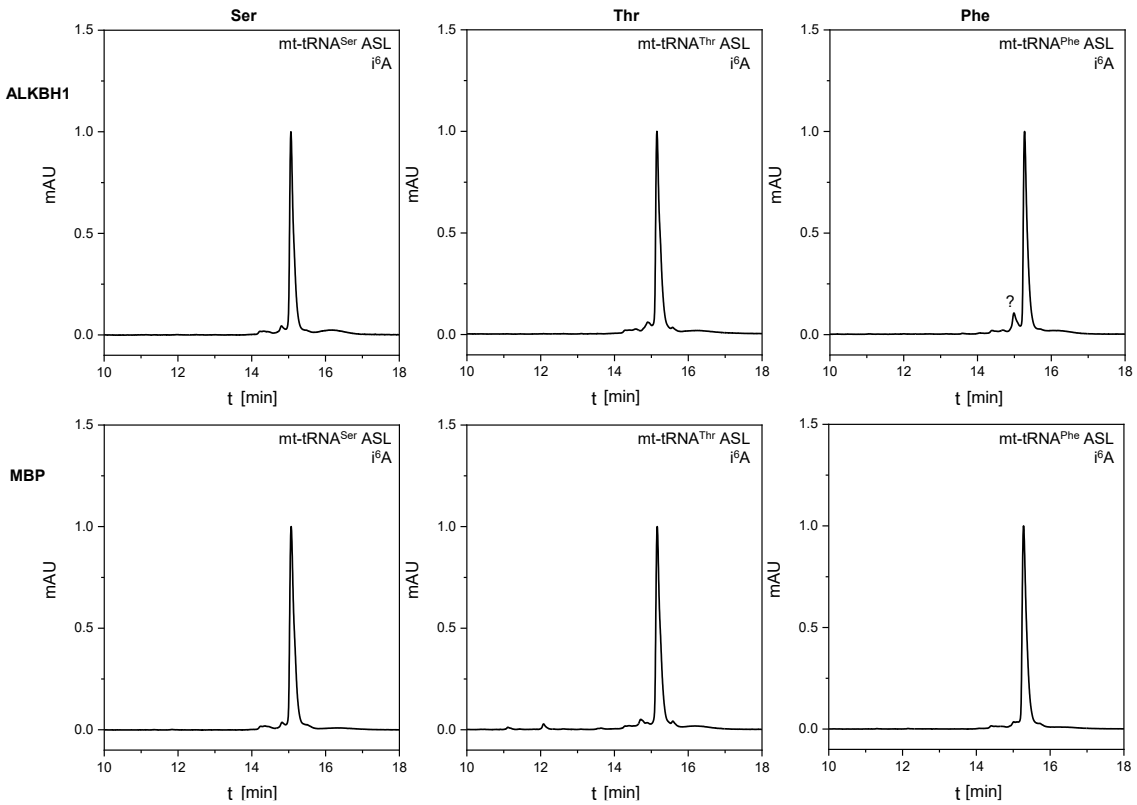



Figure 53. Demethylation assay with $i^6A_{37}$ modified mt-tRNA$^{Ser/Thr/Phe}$ ASLs (R622/R783/R623) using ALKBH1 (upper row) and MBP-tag (lower row, negative control). mt-tRNA$^{Phe}$ showed one minor species with a shorter retention time, marked with a '?'.

This additional peak was not obtained in the control sample where the mt-tRNA$^{Phe}$ was incubated with MBP-tag. Still, as there were some minor impurities with lower retention times than the main peak, even in the untreated sample, mt-tRNA$^{Phe}$ was further purified by reversed Phase (RP) HPLC to ensure unambiguous results. RP-HPLC purification was performed with triethyl ammonium acetate buffer (1M, pH 7). The impurity at around 19 min could be separated from the desired oligonucleotide at 22 min (Figure 54A/B). The fractions were collected and evaporated, and the purity was confirmed by anion-exchange chromatography (Figure 54C).

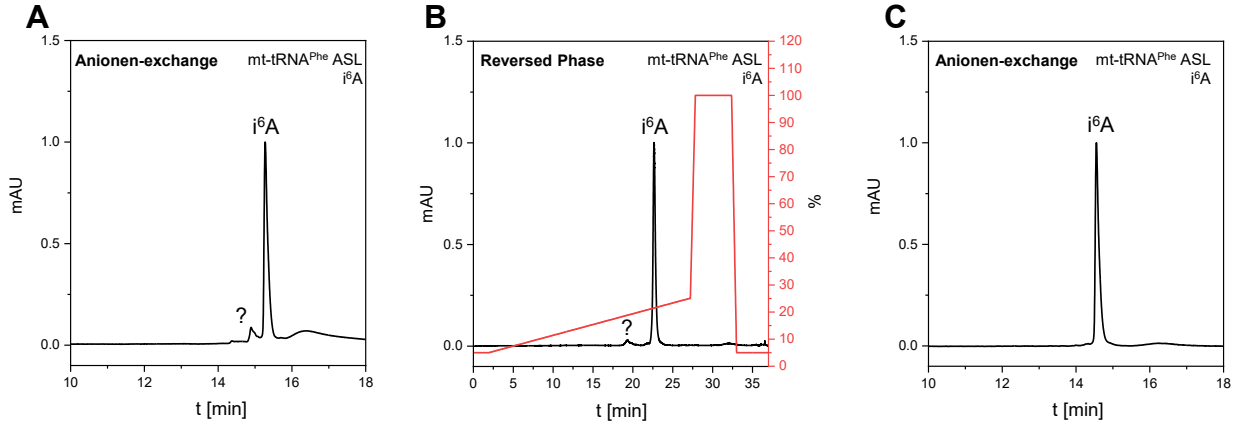



Figure 54. Normalized HPLC chromatograms of mt-tRNA$^{Phe}$ (R623) measured on anion exchange and reversed phase HPLC. A: mt-tRNA$^{Phe}$ chromatogram from anion exchange chromatography after oligonucleotide purification with a small impurity (?) in front of the main peak. B: The sample shown in A was purified by RP HPLC to remove the impurity (?). C: Purity of the collected fraction during RP HPLC was proven by anion exchange chromatography.

Since the impurity could be removed from the mt-tRNA$^{Phe}$ ASL by RP HPLC further experiments were performed with the purified sample. First, the general method of the demethylation assay was repeated to prove that the additional signal that appeared in front of the main peak of the single i$^6$A modified mt-tRNA$^{Phe}$ in Figure 53 occurred due to the treatment with ALKBH1. As a negative control, the general method was performed accordingly, but without the addition of the protein (Figure 55A). No additional peak was visible in the HPLC chromatogram. Upon treatment of the i$^6$A modified mt-tRNA$^{Phe}$ with ALKBH1 a distinct peak with a shorter retention time than the unreacted oligonucleotide appeared (Figure 55B). A prolonged incubation time to 2 h did not lead to an enhancement of the peak with shorter retention time. However, the addition of another portion of protein after 1 h of incubation, and the subsequent longer incubation up to 2 h led to a significant increase of the additional peak (Figure 55C).

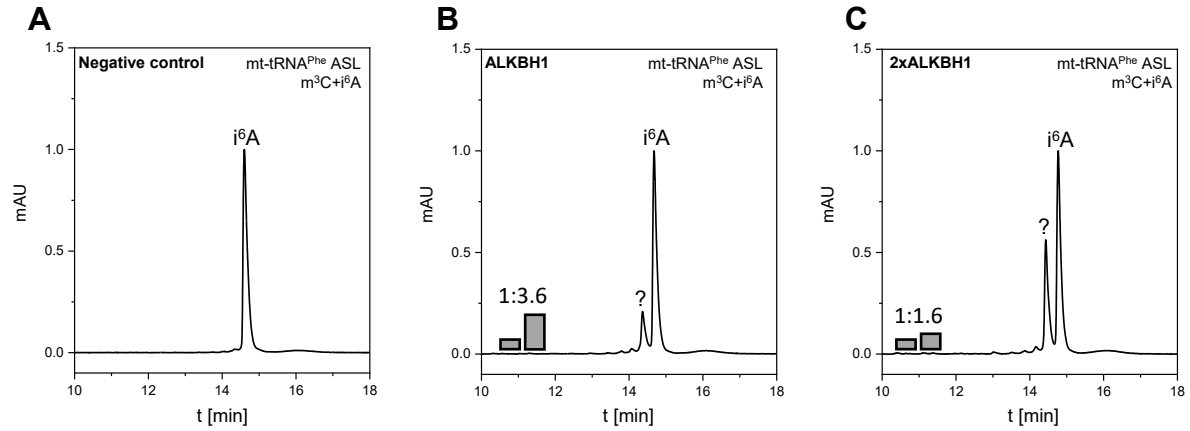



Figure 55. Anion exchange chromatography of i$^6$A modified mt-tRNA$^{Phe}$ (R623). The ratio of the i$^6$A-ASL and the new arising peak is given as grey bar graphs. A: negative control: the demethylation assay was performed without the addition of protein. B: the general method of the demethylation assay was performed with ALKBH1 and an incubation of 1 h. C: the general method of the demethylation assay was performed, but an additional amount of enzyme was added after 1 h incubation and stopped after an additional hour.

To examine which reaction was taking place with the i$^6$A modified mt-tRNA$^{Phe}$ upon treatment with ALKBH1, the demethylation assay was performed on a larger scale to perform ESI-MS measurements.

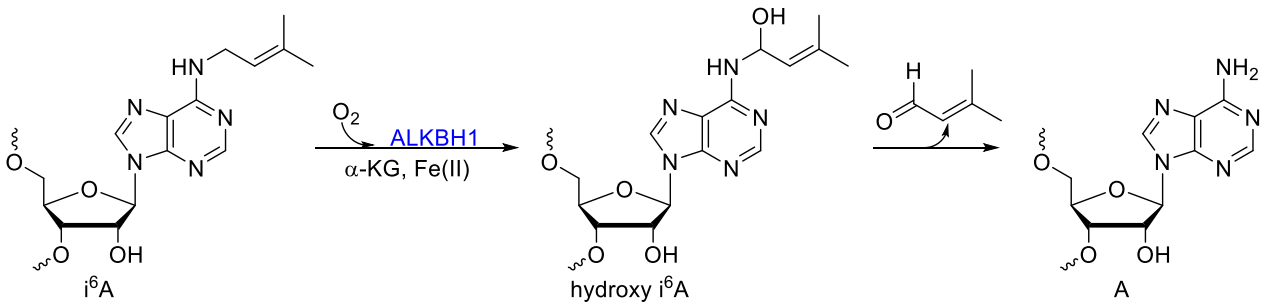



Figure 56. Proposed mechanism to remove the isopentenyl group at N6 by ALKBH1 via oxidation.

The shift towards shorter retention times would be indicative of the removal of the isopentenyl group comparing it with the resulting shifts in the retention times of the mt-tRNA$^{Ser}$ ASL in Figure

50 depending on the modification state. This could happen by oxidation of the isopentenyl group, most likely at the $CH_2$ group right next to N6 (Figure 56). The unstable hemiaminal will decompose to adenosine and the respective isopentyl aldehyde. Since the conversion of the $i^6A$ modified mt-$tRNA^{Phe}$ ASL upon treatment with ALKBH1 was not that efficient, the main product found in the mass spectrum belonged to the starting material (Figure 57). The measured spectra (Figure 57, black, 5774.85 g/mol (deconvoluted), 640.6 (9-)) is shown overlapping with the simulated spectra (green) for the $i^6A$ modified mt-$tRNA^{Phe}$ ASL, which is the most abundant mass. Additionally, with very low abundance, the ASL with oxidized $i^6A$ (blue) was found, as well as the ASL where the isopentenyl group was removed (magenta). ALKBH1 seems at least partially be able not to perform demethylation but a dealkylation reaction (Figure 56). Unexpectedly, the mass peak with the second highest intensity had an offset of 329, which would be in agreement with the loss of one adenosine.

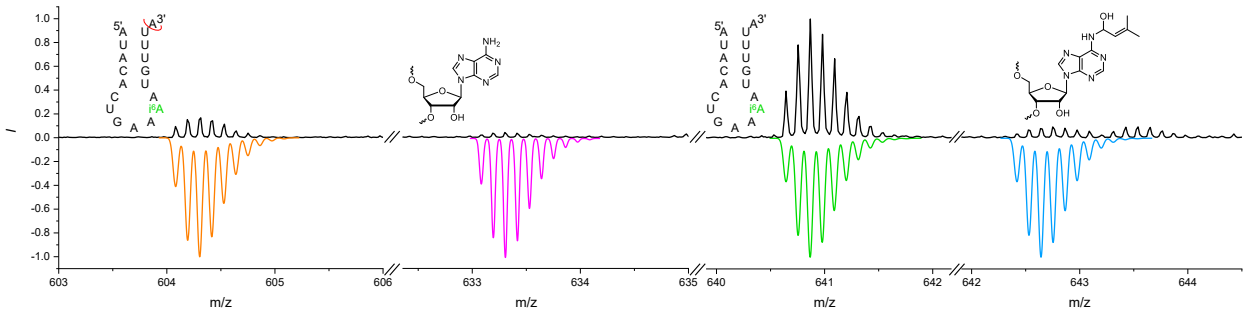



Figure 57. ESI-MS spectra of the 9-fold negatively charged (-9) $i^6A$ modified mt-$tRNA^{Phe}$ ASL after the treatment with ALKBH1. The black line is the measured sample, and the inverted peaks are simulated pattern: without one adenosine (orange), without the isopentenyl group (magenta), $i^6A$ modified mt-$tRNA^{Phe}$ ASL (green), oxidized $i^6A$ (blue).

Indeed both 5' and 3' ends of mt-$tRNA^{Phe}$ ASL comprised adenosines. It appears that ALKBH1 exhibits a 3'-5' exonuclease activity, resulting in 5'-phosphorylated adenosine and a 3'-OH residue at the remaining oligonucleotide, in accordance with known nucleases like nuclease P1. These three additional products could not be found in the mass spectra of the untreated $i^6A$ modified mt-$tRNA^{Ser}$ ASL. Moreover, an RNase inhibitor (Human Placenta) was added during the incubation with the protein, to prevent degradation by nucleases. This confirms the occurrence of the additional peaks due to the treatment with ALKBH1. In 2010, Robert Hausinger and coworkers reported the unexpected findings of a second activity of ALKBH1: besides acting as α-KG an iron (II) dependent demethylase, ALKBH1 was able to cleave DNA at AP sites by a lyase mechanism *in vitro*.[433] The latter activity appeared to be independent of the cofactors and might take place at a different active site. It seems like ALKBH1 can perform more reactions than just demethylation *in vitro*, but the *in vivo* role still has to be determined. In nature, prenyltransferases and cytokinin oxidases are responsible for the installation and the removal of the allylic prenyl group, respectively. Just recently, in 2020, Chen *et al.* reported the first chemical approach to selectively deprenylate the $i^6A$ modification with oxoammonium cations.[434]

Additional tests were performed to see if the activity of ALKBH1 and ALKBH3 could be further fine-tuned. As metal ions like $Mg^{2+}$ can affect the structure and function of RNA, the absence of this metal ion was tested on the outcome of the demethylation assay. Therefore, the oxidation buffer in which the demethylation assay takes place was prepared without $MgCl_2$. Some of the aforementioned ASL were tested with this new oxidation buffer. By comparison of the demethylation results obtained by incubating the $m^3C$ modified mt-tRNA$^{Ser}$ ASL (R604) with ALKBH3 and the Mg-free oxidation buffer, a strong enhancement of the C-ASL was observed in HPLC (Figure 58). Only a small portion of the $m^3C$-containing ASL remained. Since there is no magnesium present, the ASL seems to undergo small structural changes which may allow better interaction with ALKBH3 and thus an increase in demethylation activity. Contrarily, the same experiment using ALKBH1 instead led to the same results as under the standard conditions (Figure 51) as no demethylation occurred. To see how the effect would be for mt-tRNA$^{Ser}$ ASLs with different modification levels, the reaction was performed with single $i^6A$ (R622) and the double $m^3C$ and $(ms^2)i^6A$ modified ASL (R824/R825).

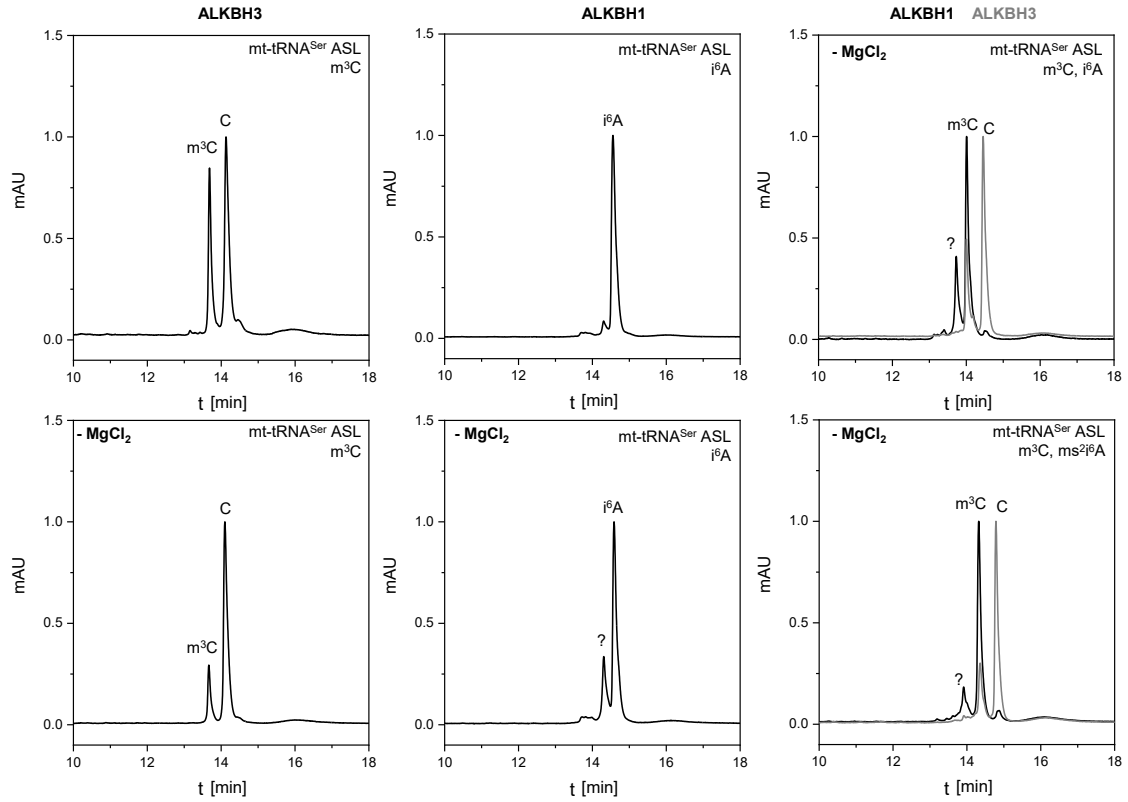



Figure 58. Comparison of the obtained results using the standard oxidation buffer vs. the oxidation buffer without $MgCl_2$ (marked with -$MgCl_2$) for the reaction with ALKBH1 and/or ALKBH3. For the $m^3C$ and $(ms^2)i^6A$ modified mt-tRNA$^{Ser}$ ASL the reaction with ALKBH1 is shown in black and with ALKBH3 in grey. The additional peak which elutes faster compared to the double modified ASL is marked with question mark, as the product was not determined yet.

Interestingly, for the $i^6A$ modified tRNA$^{Ser}$ ASL an additional peak with a faster retention time appeared, similar to the $i^6A$ modified mt-tRNA$^{Phe}$ ASL (Figure 53), which did not show up using the standard oxidation buffer (Figure 58). The absence of $Mg^{2+}$ seemed to cause structural changes, that might allow ALKBH1 to bind and catalyze the reaction to the $i^6A$ modified mt-tRNA$^{Ser}$ ASL,

which was not possible before using $MgCl_2$. Analogous behavior was also obtained for the double-modified mt-tRNA$^{Ser}$ ASLs. With ALKBH1 no reaction and with ALKBH3 minor demethylation was observed under the standard conditions (Figure 52). Changing to the Mg-free oxidation buffer, the peak with shorter retention time appeared after incubation of the m$^3$C and (ms$^2$)i$^6$A modified mt-tRNA$^{Ser}$ ASLs with ALKBH1 in agreement with the results shown before for the single i$^6$A modified tRNA$^{Ser}$ ASL. The demethylation of m$^3$C by ALKBH3 was markedly increased (Figure 58) as well in the absence of $MgCl_2$. The physiological relevance is not accountable, but it shows the incredible influence of the conditions in which the demethylation assay takes place and suggests a change in activity, dependent on metal ion concentration, which might also reflect cellular adjustment. This could help to understand the *in vivo* role of ALKBH1 and ALKBH3.

To validate the compositions of the emerging peak which is eluting faster for the i$^6$A modified mt-tRNA$^{Ser}$ ASL after the incubation with ALKBH1 and the Mg-free oxidation buffer (Figure 58), the reaction was performed in a bigger scale and analyzed by ESI-MS. In accordance with the mass spectra of the i$^6$A modified mt-tRNA$^{Phe}$ ALS (Figure 57) the most prominent species which was found belonged to the unreacted oligonucleotide (5797.92 g/mol), but also the ASLs with oxidized (5813.91 g/mol) as well as removed isopentenyl group (5729.87 g/mol) were observed, too (Figure 59). Under different conditions (with and without $MgCl_2$), ALKBH1 seemed to be able to remove the isopentenyl group from the i$^6$A modified mt-tRNA$^{Phe/Ser}$ ASLs. However, a smaller mass that could belong to an ASL with one nucleotide being removed did not occur. Whether the ASL is not a substrate for the potential exonuclease activity of ALKBH1 or if structural circumstances did not allow for the activity of the protein could not be further elucidated. In general, this interesting possibility of a new feature implemented by ALKBH1 should be more precisely investigated.

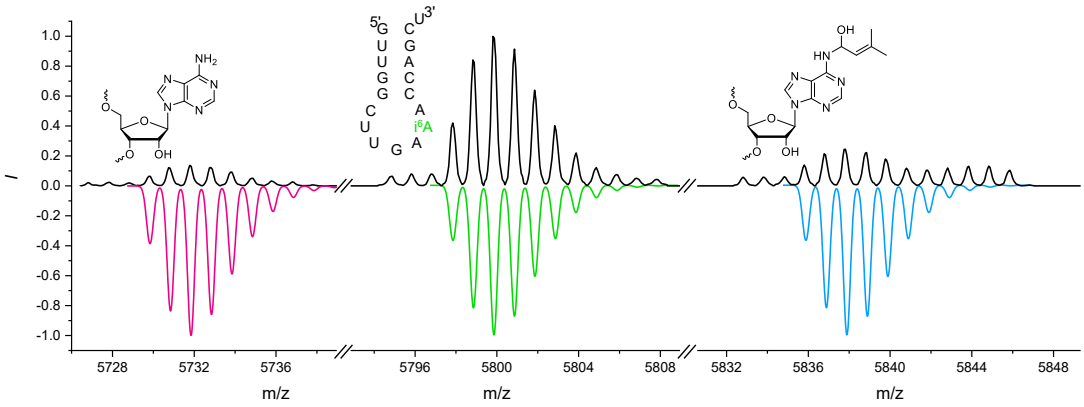



Figure 59. Deconvoluted mass spectra (ESI-MS) of i$^6$A modified mt-tRNA$^{Ser}$ ASL (R622, see Figure 58) after incubation with ALKBH1 in Mg-free oxidation buffer. The measured spectra is shown in black, whereas the simulated patterns are shown in magenta (without the isopentenyl group), green (i$^6$A modified mt-tRNA$^{Ser}$ ASL) and blue (oxidized i$^6$A).

The obtained results of the different demethylation assays are preliminary data, which need to be further validated. Especially the appearance of the faster eluting peak after incubation with ALKBH1 either with standard oxidation buffer or Mg-free oxidation buffer on ASL that does not

contain $m^3C$ but only $i^6A$ gave some insight into possible other activities of ALKBH1 besides being a demethylase (Figure 60A). Although ALKBH1 did not accept $m^3C$ as a substrate, it appeared to catalyze the oxidation of the isopentenyl group instead, similar to the oxidation of $m^5C$ to $f^5C$ in the presence of α-ketoglutarate and Fe(II).[125] A higher conversion of the $i^6A$ modified ASL to yield more of the faster-eluting band would make it easier to analyze the proceeding reactions *in vitro*. The $tRNA^{Ser/Phe}$ ASLs might not be sufficient for the precise recognition by the enzyme, in contrast to the recognition of the mt-$tRNA^{Met}$ ASL for the oxidation of $m^5C$ to $f^5C$ by ALKBH1.[125] Perhaps, further recognition elements outside the ASL are needed. For further investigations longer tRNA-fragments should be included as well. Performing the incubation without the cofactors would show if the hypothesized exonuclease activity is behaving similarly to the proposed endonuclease activity of ALKBH1 on DNA substrates.[433] Also the potential removal of an isopentenyl group has not been reported so far. Admittedly the intensities of the oxidized species of the ASLs where the isopentenyl group is absent are fairly low in the mass spectra (Figure 57, Figure 59), but it could be shown for two mt-$tRNAs^{Phe/Ser}$ ASL. Moreover, the activity of ALKBH3 acting on mt-tRNA ASLs even though the protein has not yet been reported to have mitochondrial localization was another unexpected result (Figure 60B). To prove if this activity might be physiologically relevant, a detailed examination of the cellular localizations of ALKBH3 would be required first. These results prepared the way for further important experiments, to learn more about the dynamic network of methylases and demethylases to regulate cellular functions.

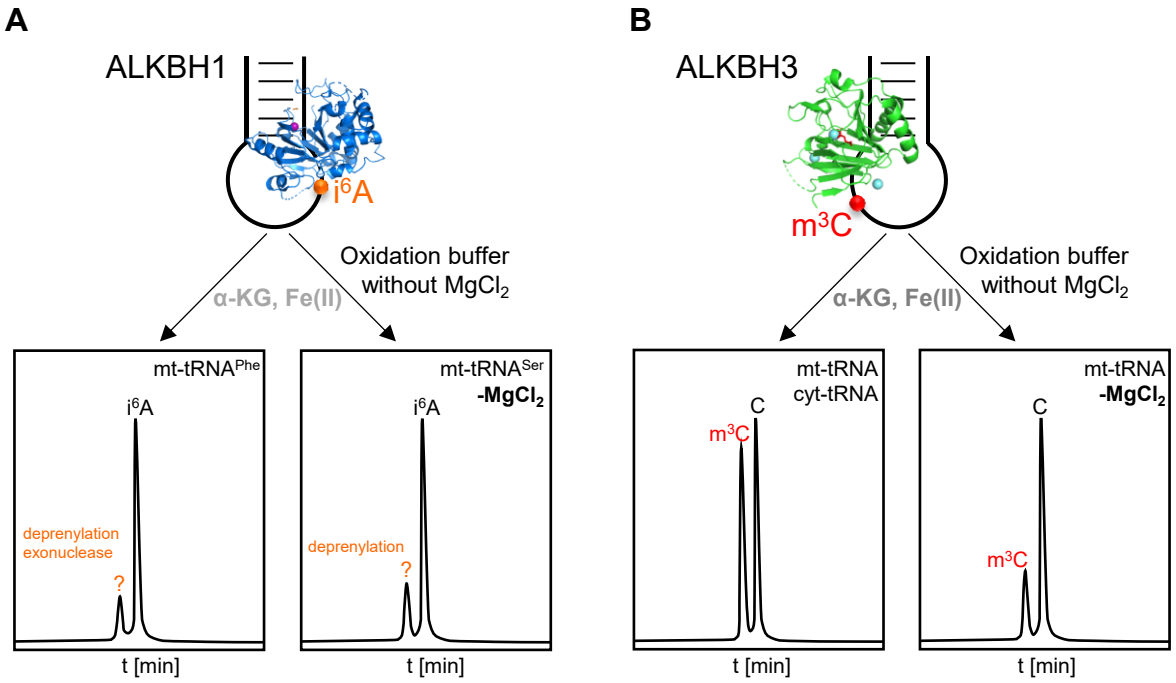



Figure 60. Overview of the most eminent results. A: The incubation of the $i^6A$-modified mt-$tRNA^{Phe}$ (R623) with ALKBH1 (pdb:6ie3[435]) led to the formation of a faster eluting peak in front of the $i^6A$ oligonucleotide. The $i^6A$ modified mt-$tRNA^{Ser}$ (R622) was not reacting under these conditions, but if $MgCl_2$ was left out from the oxidation buffer. Oxidation of the isopentenyl group (deprenylation) as well an exonuclease activity seemed to occur. B: ALKBH3 (pdb:2iuw[436]) was able to demethylate $m^3C$ on mt-$tRNA^{Ser/Thr/Phe}$ ASLs (R604, R605, R606) (and ct-$tRNA^{Arg/Thr/Ser}$ (R767, R768, R769)) resulting in an oligonucleotide with longer retention time (C) during anion-exchange chromatography.

### 3.1.5 In vitro selection of RNA-cleaving DNAzymes for the detection of modifications

To explore these human enzymes, described in the previous chapters, and their influence on cellular functions for RNA metabolism it is equally important to be aware of the localization of naturally occurring modifications. Since naturally occurring modifications are generally less abundant than canonical nucleosides and some functional groups are rather small, their precise detection and structural analysis are quite challenging. With the aforementioned synthesized $m^3C$ and $i^6A$ phosphoramidites, as well as some of their derivatives, my colleague Anam Liaqat performed different gel-based *in vitro* selection experiments to identify DNAzymes which can cleave RNA strands dependent on their modification state (Figure 61).[162, 437]

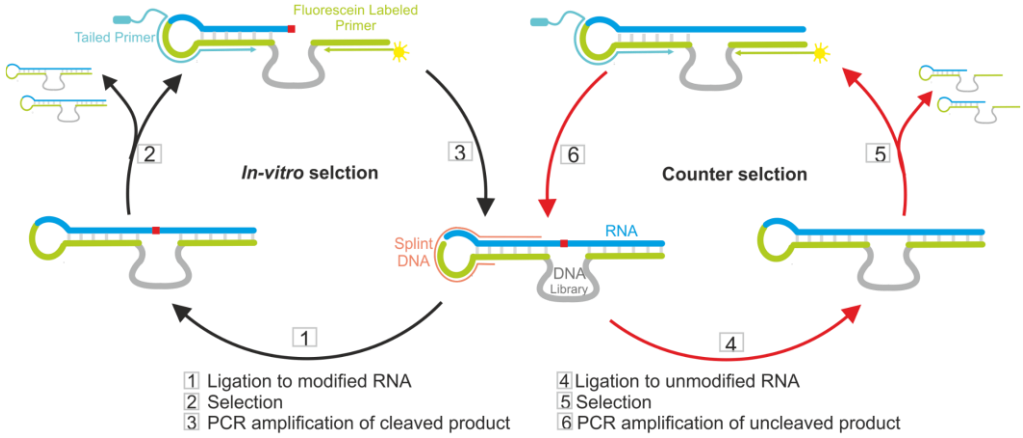



Figure 61.Schematic *in vitro* selection of RNA-cleaving deoxyribozymes.[437] The *in vitro* selection sensitive for $i^6A$ (X) had an unpaired N=X nucleotide (red). 1) ligation of the modified RNA (blue) to the DNA library (grey), which is labeled with fluorescein (yellow). 2) The ligated product is incubated with $MgCl_2$, resulting in cleaved and non-cleaved fractions. The cleaved fraction is separated by PAGE and isolated. 3) The active species are amplified by PCR. 4) For the counter selection the unmodified version of the RNA substrate is ligated to the DNA library. 2) The ligation product is incubated with $MgCl_2$, resulting in cleaved and non-cleaved fractions. They are separated by PAGE and the uncleaved fraction is isolated. 6) The active species are amplified by PCR. Figure taken from.[162]

These RNA-cleaving deoxyribozymes can subsequently be used for the detection of RNA modifications to investigate the modification sites of target RNAs as a detection method described in chapter 1.3.5.3. She found DNAzymes specific for the $i^6A$ modification as well as DNAzymes, which were able to distinguish the three naturally occurring cytidine isomers $m^3C$, $m^4C$, and $m^5C$.[162, 437]

The aim of the gel-based *in vitro* selection was the identification of modification-specific DNA catalysts.[162, 437] Therefore, the 5' phosphorylated RNA substrate with the desired modification was ligated to an $N_{20}$ DNA library using T4 DNA ligase with a splint oligonucleotide. These RNA-DNA hybrids were incubated under desired conditions to initiate intramolecular cleavage reaction. Using denaturing PAGE, the cleaved fractions were isolated representing active DNAzymes. These active DNA species were then amplified by PCR, using a fluorescently labeled primer, and again ligated to the modified RNA substrate. This cycle was repeated until enrichment of active species can be detected by fluorescent measurements. To enhance the selectivity of the DNAzymes to

cleave modified RNA, negative selection rounds (counter selection, Figure 61) were performed in alternative rounds. Therefore, the unmodified RNA substrate was ligated to the DNA library and the RNA-DNA hybrids were incubated with the selection buffer and $MgCl_2$ for the cleavage reaction. Only the DNAzymes that are not active under these conditions were isolated, thus leaving out the enzymes which can cleave both modified and unmodified RNA substrates. To identify DNAzymes that prefer unmodified over modified RNA substrates for the cleavage reactions, the *in vitro* selection was changed the following: It starts with the unmodified RNA as the positive selection, whereas the counter selection was carried out with modified RNA, to enhance the selectivity of DNAzymes that can cleave unmodified RNA and otherwise inhibited by the modification.[162, 437]

In 2018, M. Sednev *et al*. already reported an m$^6$A-sensitive RNA cleaving deoxyribozyme, but the preference for the modified RNA substrate was not sufficiently high.[161] Only a 5-10 fold faster cleavage rate was achieved for m$^6$A modified RNA compared to unmodified RNA and a 150-fold higher cleavage rate for unmodified over modified RNA by VMC10. The methyl group of m$^6$A seemed to prevent the formation of catalytically active DNA. To examine whether a larger residue at N6 may lead to better discrimination between modified and unmodified RNA, due to stabilizing interactions, i$^6$A was chosen for further experiments. By *in vitro* selection, three classes of DNAzymes were identified and characterized with different substrate requirements for their endonuclease activity during 15 rounds of selection. In those rounds, it was alternated between positive and negative selections to increase the specificity for the desired target – unmodified or modified.[162]

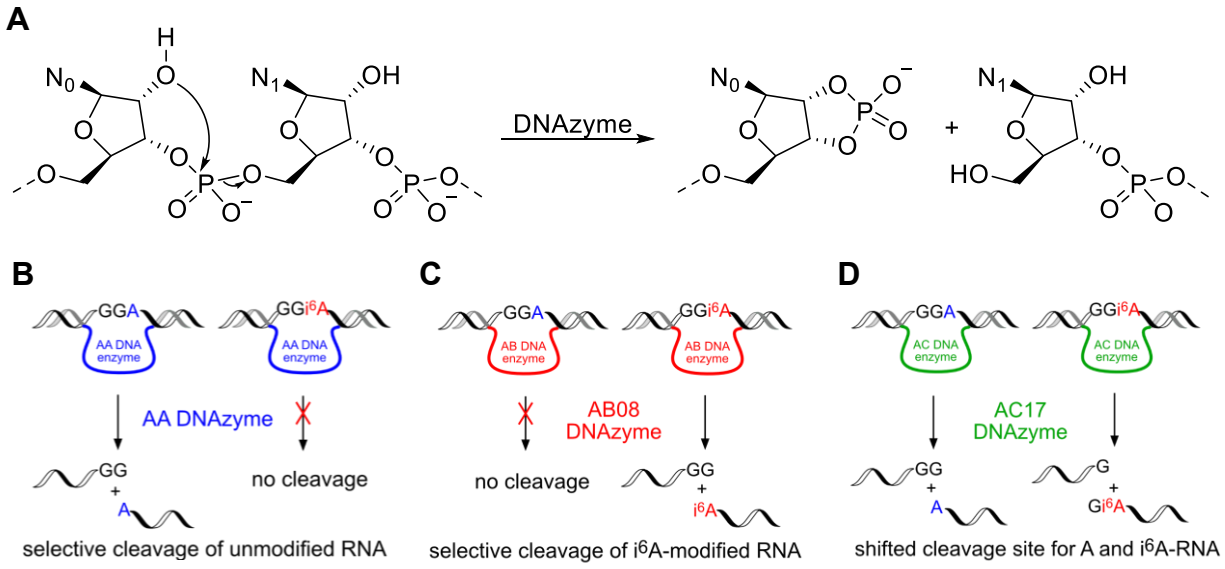



Figure 62. *In vitro* selection results of unmodified and i$^6$A containing RNA substrates. A: The cleavage of the phosphodiester backbone is catalyzed by the DNAzyme (transesterification). B: The AA DNAzyme efficiently cleaves unmodified RNA but is inhibited by i$^6$A. C: The AB08 DNAzyme is activated by i$^6$A. D: The AC17 exhibits two different cleavage sites, dependent on the modification level. Figures taken and reproduced from Ref.[162]

The DNAzyme catalyzed the cleavage of the phosphodiester backbone (Figure 62A). For AA DNAzymes, this reaction is inhibited by $i^6A$ at position 17 and only unmodified RNA is cleaved (Figure 62). G16 was determined as the cleavage site of AA17 by alkaline hydrolysis (RNase T1 digestion ladder) and HR-ESI-MS of both fragments. In contrast to VMC10, which was found to be only slightly inhibited by $m^6A$, $i^6A$ strongly inhibited the cleavage activity of all three AA DNAzymes, even though the modification is not directly located at the cleavage site for two of the endonucleases, AA07, and AA14. The sterically more demanding isopentenyl residue seems to amplify this inhibitory effect compared to the methyl group of $m^6A$.[162] Contrary, the AB08 DNA enzyme cleaved the RNA substrate right next to the $i^6A$ modification with 80% cleavage yield after 6 h at pH 7.5 with 20 mM $MgCl_2$), while no activity was observed for unmodified RNA. The cleavage site was proven by HR-ESI-MS of the $i^6A$-containing fragment. Such a remarkable selectivity of DNAzymes for modified substrates has not been reported before – $i^6A$ was a prerequisite for the cleavage activity of AB08. Moreover, AB08 is selective for $i^6A$ and no cleavage activity was observed for $m^6A$-modified RNA substrates.[162] The AC17 DNA enzyme was the third class of identified endonucleases and was able to cleave both unmodified and $i^6A$ containing RNA substrates but at two different cleavage sites, respectively. The two cleavage products showed a distinct migration on the analysis gel. DNAzyme cleaved the unmodified RNA substrate at G16, whereas it cleaved modified substrate one nucleotides upstream at G15. AC17 had high sequence compliance with the VMC10 enzyme. AC17 can be transformed to VMC10 after a single C12T point mutation. By comparison with the $m^6A$ containing RNA substrate, only minimal cleavage yield was observed at G16, but the endonuclease had no catalytic activity to cut at position G15. Thus, the shift to the cleavage site at G15 was selective for $i^6A$. By deletion mutations of AC17, another mutant was analyzed, AC17-ΔA20, which was selective for RNA substrates that contain $i^6A$. AC17-ΔA20 cleaved the modified RNA substrate one nucleotide upstream compared to AB08 (Figure *63*).[162]

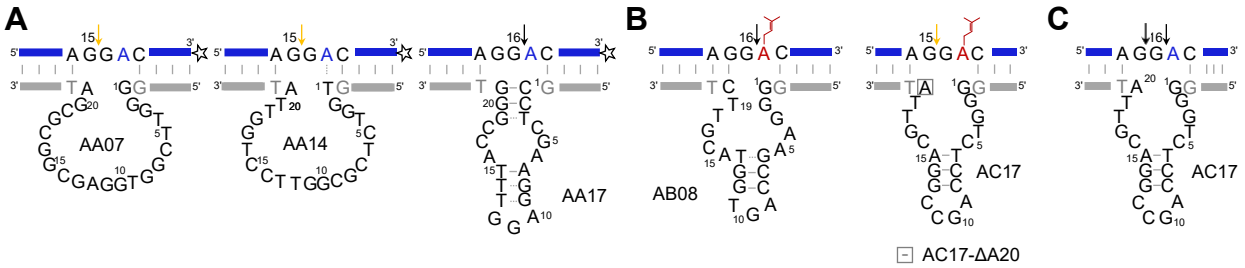



Figure 63. Overview of the different classes of DNA enzymes found during *in vitro* selection. A: The AA DNAzymes cleave unmodified RNA but are inhibited by $i^6A$. B: The AB08 DNAzyme is activated by $i^6A$. AC17-ΔA20 is also selective for i6A, but the cleavage site is shifted compared to AB08. C: The AC17 exhibits two different cleavage sites, dependent on the modification level.

The three classes of DNA enzymes cleave RNA substrates in a distinct manner with regard to the natural modification $i^6A$. The precise mechanism, of how the DNA enzymes recognize and catalyze the cleavage of their respective targets is not known yet but might be unveiled by crystal structure

analysis. However, the isopentenyl group leads to more distinct energetic and kinetic levels compared to $m^6A$, due to the larger and more hydrophobic residue, which facilitates better discrimination between modified and unmodified RNA substrates. Unfortunately, no natural tRNA with $i^6A$ at position 37 could be tested with the DNAzymes, as they require a guanosine (G16) next to the modification, which has not been found in nature yet.

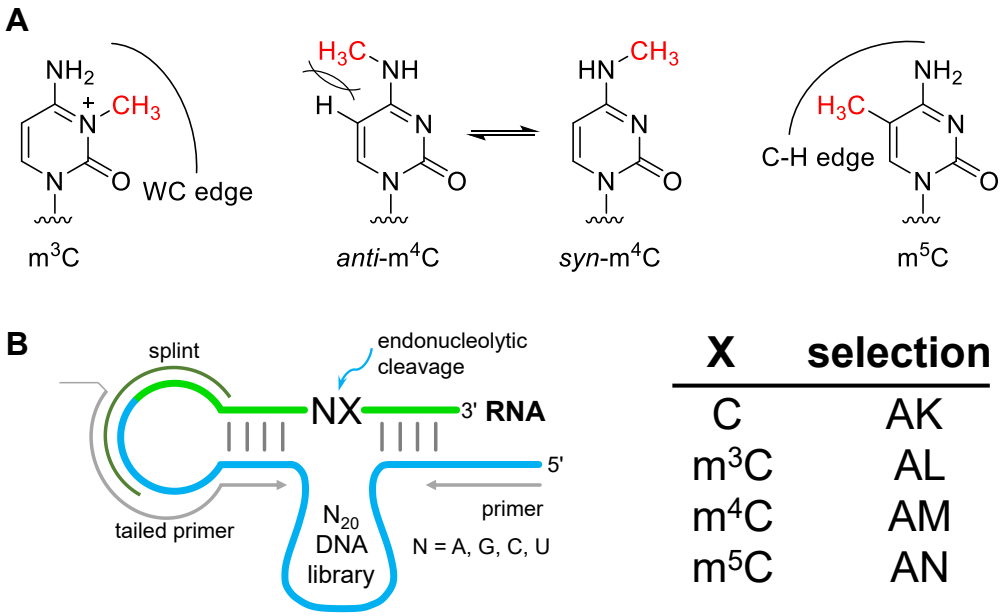



Figure 64. $m^XC$ *in vitro* selection. A: Structure of the naturally occurring methyl cytidine isomers and their respective steric demand and effect on WC base pairing. The methyl group is shown in red. Methylation at N3 introduces a positive charge to the molecule and, similar to the *syn* conformation of $m^4C$, impairs WC base pairing. *Anti*-$m^4C$ enables WC base pairing but is destabilized by steric repulsion between the methyl group (red) and 5-H. The methyl group of $m^5C$ is apart from the WC face and thus does not interfere with WC base pairing, but might interact with adjacent nucleobases by base stacking via the C-H edge.[437] B: Schematic representation of the *in vitro* selection library preparation with the splint (dark green) for ligation and the primer binding sites for amplification. To achieve universal DNA enzymes that can cleave any NC dinucleotide, a degenerated nucleotide right next to the modification or the unmodified cytidine in the RNA substrate was chosen. The four different selections, selective for one modification or the canonical C were named AK, AL, AM, and AN. Figure reproduced from Ref.[437]

Another *in vitro* selection project from Anam Liaqat addressed the three natural methyl cytidine isomers $m^3C$, $m^4C$, and $m^5C$ as well as the unmodified cytidine.[437] Preferably selective DNA enzymes should be identified, that only catalyze the cleavage of one of the mentioned isomers. Due to their different localization of the methyl group, they feature different structural properties, which may lead to varying secondary structures. $m^3C$ and the *syn*-form of $m^4C$ impair WC base pairing, whereas the *anti*-form of $m^4C$ permits WC base pairing, but experiences steric repulsion. In contrast, the localization of the methyl group of $m^5C$ allows base pairing, but the methyl group may have contact with adjacent nucleobase over the C-H edge (Figure 64A). This could facilitate the discrimination of the isomers by the DNA enzymes. To generate a more versatile DNA enzyme for $m^XC$ detection, a degenerate nucleotide next to the target cytidine was used (Figure 64B). This should facilitate to potentially cleave the RNA substrate between all four possible dinucleotides (NC). For the gel-based *in vitro* selection four RNA substrates containing $m^3C$, $m^4C$, $m^5C$ or unmodified cytidine, respectively, were ligated to DNA with an $N_{20}$ library. In accordance with the $i^6A$

*in vitro* selection, the negative selection round for the modified RNA substrates was performed with the unmodified RNA substrate and vice versa for the unmodified substrate.[437]

To test the specificity of the isolated DNAzymes, the cleavage assay with the enzymes had to be performed with the sixteen possible RNA substrates. Exchanging the nucleotide one position upstream to the target cytidine with the four canonical nucleotides A, G, C and U (Figure 64B) facilitated the investigation of all possible NX junctions, for the four cytidine targets. The DNA enzymes found in the AK, AL, AM, and AN selections presented good selectivity for their respective methyl cytidine.

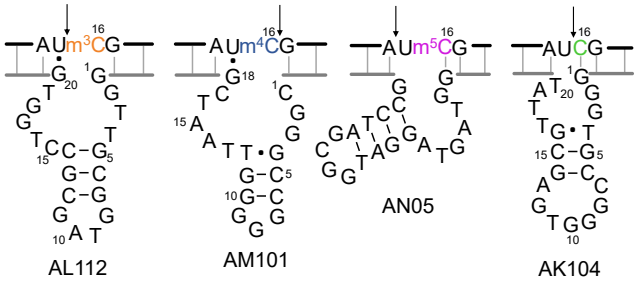



Figure 65. Predicted secondary structures of one example each, obtained from AL (orange), AM (blue), AN (magenta) and AK (green) selection. The cleavage site is indicated with an arrow. Figure reproduced from Ref.[437]

One universal DNAzyme was obtained from the AL selection, named AL112, which was able to cut the m$^3$C containing RNA substrate at all four possible NC junctions. Even though this generality could not be found in the other selections, almost all combinations could be covered with a variety of deoxyribozymes. Especially the UC dinucleotides were cleaved by AL112, AM101, AN05, and AK104 (Figure 65) with good discrimination between the respective modification and the canonical cytidine. The DNAzymes for the modified RNAs preferably cleaved the corresponding modified substrates adjacent to the UC junction, while AK104 was selective for the unmodified RNA substrate and inhibited by the three modified RNAs.[437] The cleavage sites of these four DNAzymes were determined by a gel-based assay and HR-ESI-MS of the fragments, obtained after the cleavage of the phosphodiester backbone (Figure 62A). AL112 and AK104 cut the RNA substrate upstream of the modification site at U15 and AN05 one nucleotide upstream of the modification site at A14, whereas AM101 cut downstream of the modification (Figure 65). The latter could even differentiate between m$^4$C and m$^{4,4}$C. AM101 had a similar cleavage rate for the dimethylated modification as for the unmodified RNA substrate. Accordingly, the oxidized species of m$^5$C, namely hm$^5$C and f$^5$C were analyzed using AN05, to see if the enzyme can distinguish between the different variants. On the one hand, almost no discrimination was noted between m$^5$C and hm$^5$C, as the cleavage rate of the hydroxylated cytidine derivative by AN05 was only slightly lower compared to m$^5$C. On the other hand, the cleavage efficiency of f$^5$C was comparable to unmodified RNA. Thus, AN05 could not distinguish between m$^5$C and hm$^5$C, as well as between f$^5$C and the RNA containing the unmodified cytidine. The DNA enzyme AK104 efficiently cut unmodified RNA substrates

between the AC, GC, and UC junction, right upstream of the target cytidine, but was inhibited by all three modifications within the shown sequence context. However, this trend was changed within the CC context: AK104 achieved around 30% cleavage rate for RNA containing $m^3C$ but was impaired to cut the unmodified RNA substrate. Additionally, the cleavage site of the $m^3C$ containing RNA was shifted on nucleotides upstream. The selectivity for the unmodified RNA substrate as well as the cleavage site directly upstream to C16 was only given in the AC, GC, and UC dinucleotide sequence context. The specificity of the deoxyribozymes was proven by a cleavage experiment, where RNA substrates with the three modifications in the UX context were present simultaneously, but diversly labeled.[437] The FH14 ribozyme was used to introduce fluorescent *N*$^6$-(6-aminohexyl)-ATP derivatives[438]: 6FAM for $Um^3C$, Cy3 for $Um^4C$, and ATTO-647N for $Um^5C$. A mixture of these three RNA substrates was incubated with the respective DNAzymes AL112, AM101 and AN05 (Figure 66A). The enzymes preferentially cleaved their cognate targets, with only minor additional cleavage bands for $Um^4C$-, and $Um^5C$-RNA by AL112 observed on PAGE (Figure 66B).

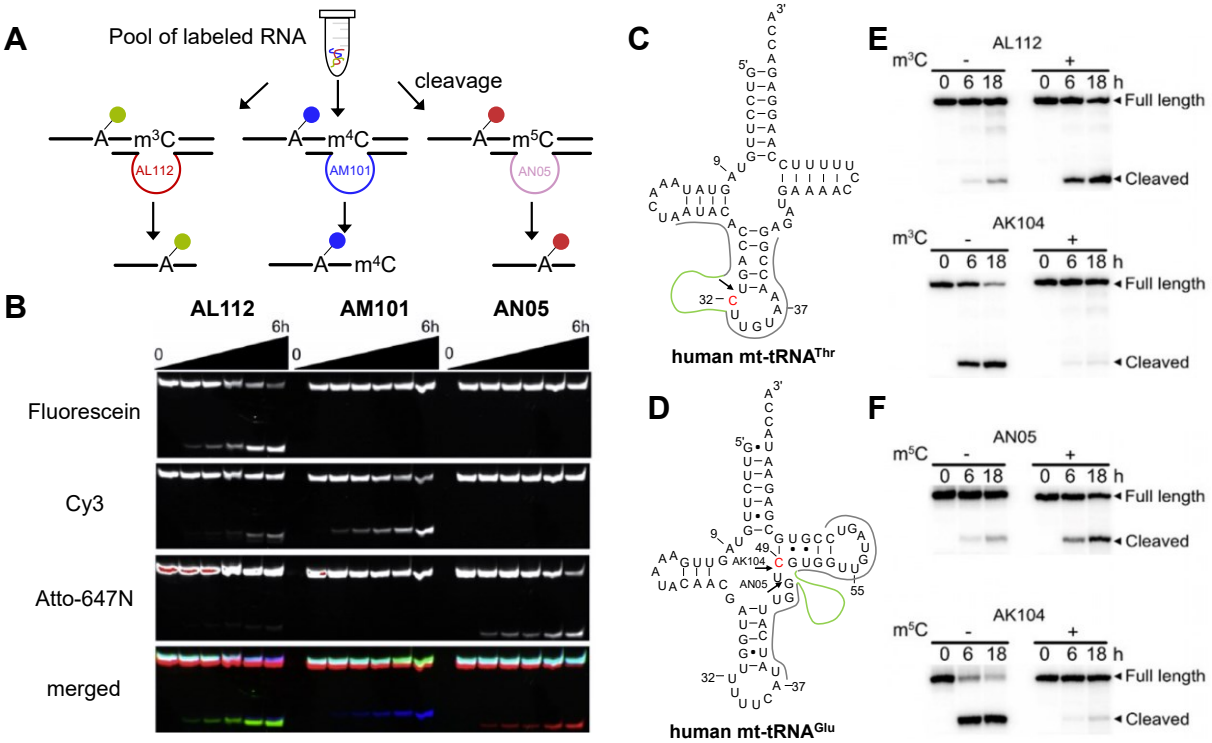



Figure 66. A: The $Um^3C$-, $Um^4C$-, and $Um^5C$-RNAs were labeled with 6FAM (green), Cy3 (blue), and ATTO-647N (red), respectively, mixed in an equimolar ratio and incubated with an excess of each DNA enzyme. The cleavage fragment is shown below. B: As a result, selective RNA cleavage is observed on denaturing PAGE. In black and white, the respective fluorescent channel is shown, whereas in the last row the three images are merged and depicted as false-color representation. C. secondary structure of mt-tRNA$^{Thr}$, treated with AL112. D: mt-tRNA$^{Glu}$, treated with AN05 and AK104. The target cytidine is shown in red, and the cleavage site is indicated with an arrow. The binding arms of the DNAzymes are shown in grey and the catalytic core in green. E: page analysis of mt-tRNA$^{Thr}$ treated with AL112 and AK104. F: Page analysis of mt-tRNA$^{Glu}$ treated with AN05 and AK104. Figure reproduced from Ref.[437]

These experiments demonstrated, that this set of deoxyribozymes can be used for the detection and discrimination of the methyl cytidine isomers in UC motifs.[437] Therefore, the deoxyribozymes

AL112, AN05, and AK104 were tested on human mt-tRNA$^{Thr}$ and mt-tRNA$^{Glu}$ with naturally occurring $m^3C$ on position 32 and $m^5C$ at C49, respectively (Figure 66C,D). The whole tRNAs with the single modification were synthesized as three RNA fragments and connected by splint ligation using T4 DNA ligase. The binding arms of the DNAzymes were designed for the respective tRNA to cut the substrate upstream of the target cytidine. The unmodified tRNAs were efficiently cleaved by AK104 (>90% after 18 h[437]), while the presence of the modifications impaired the reaction (Figure 66E,F). Similarly, AL112 and AN05 revealed activity for the modified mt-tRNA$^{Thr}$ and mt-tRNA$^{Glu}$ respectively, but less efficient than their selection substrates. No activity was observed for the unmodified tRNAs, which confirms the selectivity of AL112 for $m^3C$ and AN05 for $m^5C$ even for larger RNA substrates with a more complex secondary structure. Both enzymes can be used for the detection of these cytidine modifications. These experiments nicely showed the influence of the small methyl group on the activity of deoxyribozymes, also dependent on the localization of the methyl group. Either inhibition (AK104) or acceleration (AL112, AM101, and AN05) of the catalytic cleavage activity can be achieved.[437] The AL112 deoxyribozyme was used for the collaboration with the laboratory of Prof. Dr. Markus Bohnsack to verify $m^3C$ levels in mt-tRNA$^{Thr/Ser}$ for the METTL8 project, summarized in chapter 3.1.4.[58]

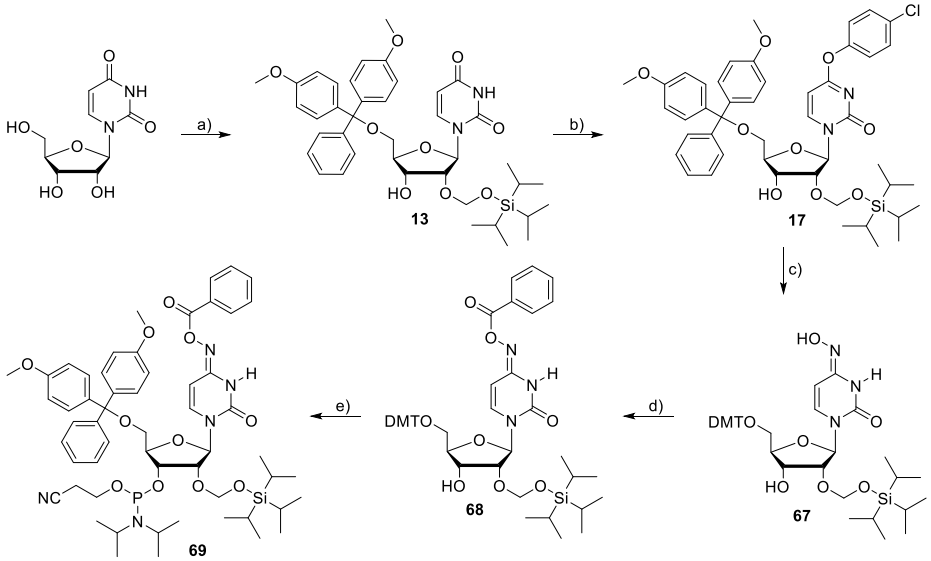
This chapter summarized the results obtained working with the natural modifications including $m^3C$, $m^4C$, $(ms^2)i^6A$, and $t^6A$, in collaboration with the laboratory of Prof. Dr. Markus Bohnsack, as well as with my colleague Anam Liaqat. The knowledge about nucleoside modifications can further be utilized to generate artificial nucleoside modifications for various applications, like antiviral nucleoside analogs. Due to their programmable nature and their predictable behavior, they can be a useful tool for the prevention of viral diseases by addressing viral RNA-polymerases.

## 3.2 The NHC antiviral nucleoside analog against SARS-CoV-2

The Covid-19 pandemic caused by the SARS-CoV-2 virus underlined the great potential of antiviral nucleoside analogs for the treatment of viral infections. In collaboration with the laboratory of Prof. Dr. Patrick Cramer and his coworkers Florian Kabinger and Jana Schmitzová, the mechanism of action of the nucleoside analog molnupiravir was examined. The state of knowledge on the chemistry of molnupiravir has been summarized in chapter 1.4.3.4-1.4.3.5. Parts of these results have been published in Nature structural & molecular biology in 2021 (*Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis*), which will be summarized in the next section.[439] The $N^4$-hydroxycytidine modification will be abbreviated as NHC or M in the text, but as X in the sequences.

### 3.2.1 Synthesis of the NHC phosphoramidite building block

To start the biochemical investigations, the respective phosphoramidite building block of $N^4$-hydroxycytidine had to be synthesized (Scheme 21). For the introduction of hydroxylamine at N4, the convertible nucleoside 4-chlorophenyluridine was prepared according to Scheme 8 (chapter 3.1.1).[374, 376] By substitution with $NH_2OH{\cdot}HCl$ in 1,2-DCE, the 5'-*O*-(4,4'-Dimethoxytrityl)-$N^4$-hydroxy-2'-*O*-(triisopropylsilyloxy)methylcytidine **67** was obtained with 64% yield. The substitution reaction performed directly with the Trisyl activated compound led to lower yields, since this compound is less stable, and resulted in hydrolyzed by-products. The second to last step was the benzoyl protection of the hydroxylamine functional group at N4, following the procedure of J. A. Piccirilli and coworkers.[440] In this report, they synthesized the 5'-*O*-DMT-2'-*O*-TBDMS compound starting from a 5'-3'-silylated uridine and they assigned the benzoyl protection group to the amine, leaving a free hydroxy group.




Scheme 21. Synthesis of the NHC phosphoramidite **69**. a) 1. DMT-Cl, pyridine, 4 h, r. t., (88%); 2. $^{i}Pr_2NEt$, $Bu_2SnCl_2$, 1,2-DCE, 1 h, r. t.; 3. TOM-Cl, 80°C, 20 min, (32%) b) 1. Trisyl-Cl, DMAP, $NEt_3$, r. t., 1 h; 2. 4-chlorophenol, DBU, $Me_2NEt$, DCM, r. t., 2 h, (86%) c) $NH_2OH \cdot HCl$, MeCN, DMAP, $NEt_3$, 21 h, r. t., (64%) d) $(BzO)_2O$, DCM, DMAP, $NEt_3$, 4 h, r. t., (80%) e) 2-Cyanoethyl *N,N,N',N'*-tetraisopropyl phosphoramidite, 4,5-DCI, DCM, 4°C – r. t., 3 h, (40%).

However, after the successful synthesis it could not be unambiguously confirmed if the benzoyl group of the benzoylated 5'-*O*-(4,4'-dimethoxytrityl)-$N^4$-hydroxy-2'-*O*-(triisopropylsilyloxy)methyl-cytidine **68** was attached to the hydroxy or the amino group even by 2D NMR spectra. The assignment was only possible after $^{15}N$ labeling of NHC and will be described in chapter 3.2.7. The conversion into the phosphoramidite building block **69** was performed with 2-Cyanoethyl *N*,*N*,*N'*,*N'*-tetraisopropyl phosphoramidite, as it exhibits a lower reactivity but a higher selectivity compared to CEP-Cl.[440] With CEP-Cl, J. A. Piccirilli and coworkers could not obtain the correct product. They hypothesized that it is presumably due to phosphitylation of the N4-hydroxy group. Unfortunately, even by using the tetraisopropyl phosphoramidite reagent, double phosphitylation was obtained. By addition of the phosphitylation reagent in a single batch, the undesired side product was mainly formed. The reaction conditions had to be chosen carefully. The reaction mixture was cooled to 4°C and the phosphitylation reagent was added in portions. The reaction progress was monitored by TLC, to ensure that the reaction could be stopped if double phosphitylation occurred. Following this procedure, the side product was only formed partially. The remaining starting material was re-isolated and a 40% yield of phosphoramidite **69** was obtained.

### 3.2.2 Incorporation of NHC into RNA oligonucleotides by solid phase synthesis

The NHC phosphoramidite building block was incorporated into different RNA substrates by solid phase synthesis. They were purified by PAGE and analyzed by HPLC and HR-ESI-MS (Table 2). J. A. Piccirilli and coworkers reported a ~10% conversion of $N^4$-hydroxycytidine into cytidine that either occurred during the synthesis or the deprotection of the RNA strands, indicated by the mass spectra of the purified oligonucleotides.[440]

Table 2. Overview of the synthesized oligonucleotides containing NHC, as well as one unmodified primer strand with the corresponding formula and mass. NHC shown as X (red), FAM = fluorescein, Alk = alkyne linker, nt = nucleotides.

| name | 5'-sequence-3' | nt | formula | Mass calc. | Mass found |
|---|---|---|---|---|---|
| R893 | GGGXACUGCGUAp | 12 | $C_{115}H_{144}N_{48}O_{86}P_{12}$ | 3944.5221 | 3944.5297 |
| R909 | XACUGCGUAGGCUCA | 15 | $C_{142}H_{178}N_{56}O_{104}P_{14}$ | 4764.6688 | 4764.6760 |
| R878 | CAXUGCGUAGGCUCA | 15 | $C_{142}H_{178}N_{56}O_{104}P_{14}$ | 4764.6688 | 4764.6797 |
| R894 | UGAGXCUACGCAGUG | 15 | $C_{143}H_{178}N_{58}O_{104}P_{14}$ | 4804.6749 | 4804.6684 |
| R904 | FAM-CUGCGUAG | 8 | $C_{106}H_{122}N_{34}O_{64}P_8$ | 3142.5316 | 3142.5177 |
| R905 | Alk-UGAGCCUACGCAGUXp | 15 | $C_{148}H_{188}N_{56}O_{110}P_{16}$ | 5004.6640 | 5004.6685 |
| R905 | FAM-UGAGCCUACGCAGUX | 15 | $C_{172}H_{206}N_{60}O_{113}P_{15}$ | 5383.8282 | 5383.8124 |
| R906 | UGAGCCUACGCAGUXp | 15 | $C_{142}H_{179}N_{56}O_{107}P_{15}$ | 4844.6351 | 4844.6432 |
| R906 | UGAGCCUACGCAGUX | 15 | $C_{142}H_{178}N_{56}O_{104}P_{14}$ | 4764.6688 | 4764.6568 |

The synthesized oligonucleotides containing the NHC modification were deprotected with 25% $NH_4OH$/EtOH (3/1) at 55 °C for 6 h. The bands obtained during denaturing PAGE had to be cut out precisely, to separate the product oligonucleotide from minor fractions of the resulting by-product. The side product was moving slightly faster on PAGE and thus was present in the lower part of the product band. Consequently, the upper part of the band contained the pure product, whereas

the lower part contained a mixture of product and by-product. This was proven by anion exchange chromatography in combination with HR-ESI-MS. The HPLC chromatogram of the upper part of the band revealed one sharp peak (Figure 67A). In contrast, the lower part featured a double peak (Figure 67B), which was also reflected in the respective mass spectrum (Figure 67C). Two isotope patterns were observed: The one with the higher mass (4764.66 g/mol, red) belonged to the correct NHC containing oligonucleotide. In contrast to the report of J. A. Piccirilli and coworkers, the smaller mass (4749.66 g/mol), which was found as the by-product, appeared to be a uridine at the position of the modification and not a cytidine.

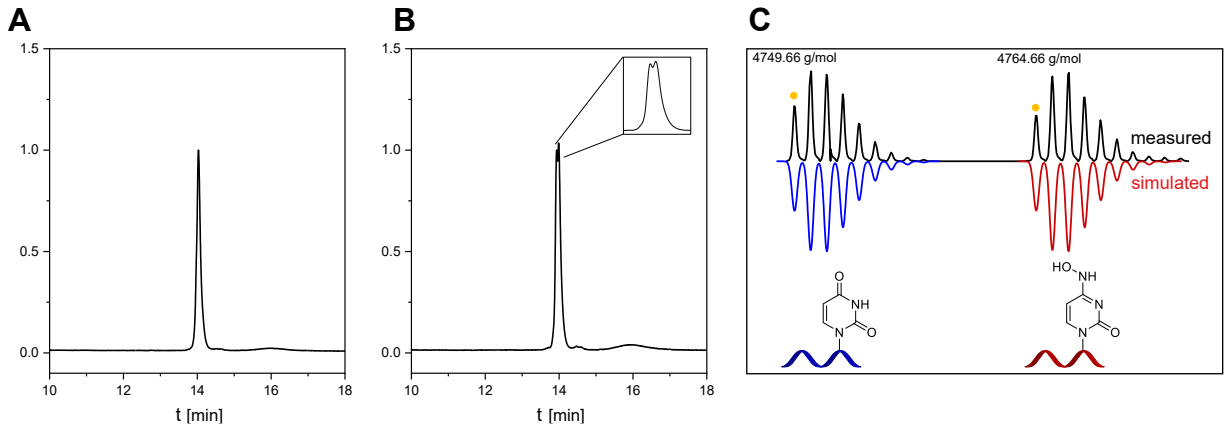



Figure 67. Example of one oligonucleotide containing NHC R878. A: Normalized anion-exchange chromatogram of the upper part of the band from PAGE purification. One sharp band is obtained. B: A double peak is found for the lower part of the band, reflecting a mixture of product and by-product (normalized). C: Measured mass spectrum (ESI-MS) of the sample shown in B is depicted in black. The simulated spectrum of the product is shown in red, and the simulated mass, with a uridine instead of NHC, is shown in blue. The monoisotopic mass is indicated with a yellow dot.

Together with the collaboration partners at the Cramer laboratory, the pure oligonucleotides were used to analyze the biochemical and structural characteristics of the drug candidate molnupiravir to examine the mechanism of the NHC-induced RNA mutagenesis.[439]

### 3.2.3 RNA elongation and thermal denaturation experiments

The collaboration partner performed RNA elongation assays with the NHC trisphosphate (MTP), representing the active form of molnupiravir, to see if the recombinant SARS-CoV-2 RdRp can use the triphosphate as a substrate for RNA synthesis.[439] Four variations of the template strand were used to provide all possible canonical nucleotides at the +1 position (Figure 68A). Opposite to this position, the incorporation of MTP had to be validated. To monitor the elongation, fluorescein was attached to the 5'-end of the product strand. Opposite to templating G and A (+1), MTP was incorporated (Figure 68B), equally to CTP and UTP, respectively. To calculate the efficiency of the incorporation of MTP compared to the cognate nucleotides, time-dependent RNA elongations were performed. The incorporation efficiency of MTP opposite to G was slightly reduced compared to CTP (Figure 68C), but markedly for MTP opposite A compared to UTP (Figure 68D). As described in chapter 1.4.3.4, NHC was supposed to be able to form a base pair with either guanine in the

amino form or with adenosine in the oxime form. The formation of the two tautomeric forms explains the enabled incorporation of MTP, as mimicry of cytidine and uridine.[439, 441]

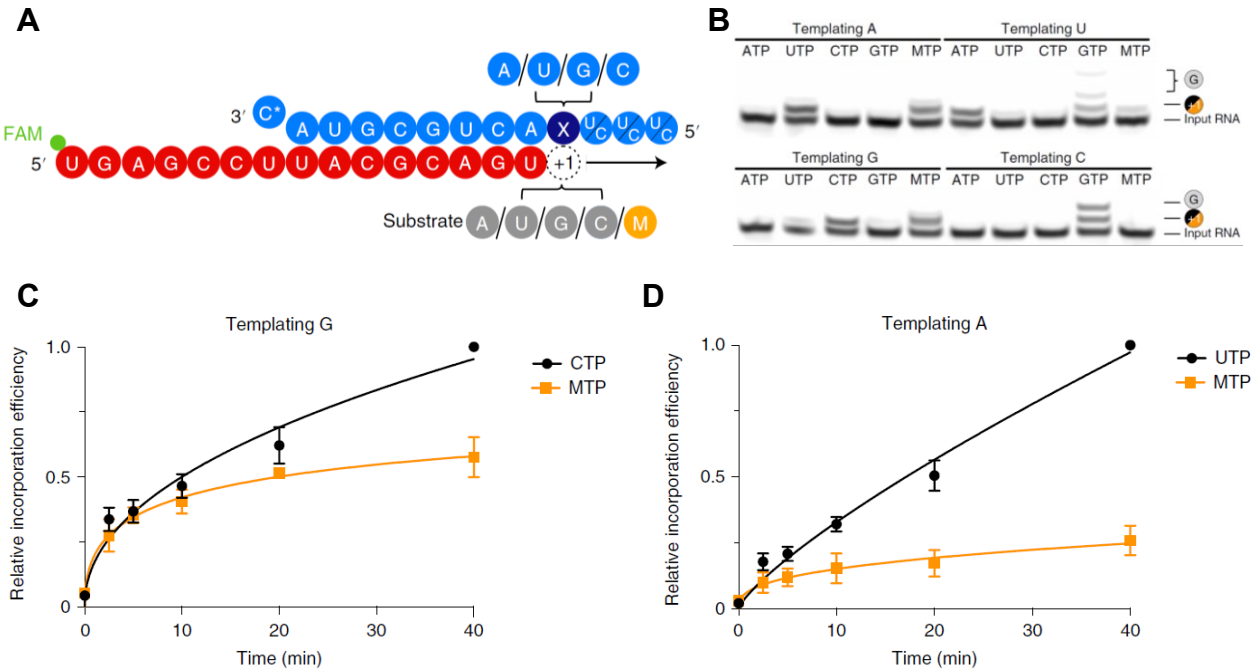



Figure 68. RNA elongation experiments with the recombinant SARS-CoV-2 RdRp. A: Four different RNA template-product duplexes were used, with A, U, G, and C at the position +1 (dark blue) in the template strand (light blue). Opposite to this, MTP (orange) and the canonical nucleotides (grey) should be incorporated. The product strand (red) contains a 5' fluorescent label (green) to monitor the elongation. The 3' end of the template strand contains a dideoxy-C (C*) to prevent elongation of the template. B: The results of the elongation assays were analyzed on PAGE. With A or G as the templating nucleotide (+1), MTP was incorporated into the product strand, instead of C or U. C/D: Evaluation of the time-dependent incorporation of MTP opposite to templating G and A, respectively. The average of three measurements was plotted. C: Relative incorporation efficiency compared to CTP incorporation. D: Relative incorporation efficiency compared to UTP incorporation. C/D: Data were collected by the collaborators in triplicates and plotted as the average with standard deviation. Figures are taken from Ref.[439]

In contrast to nucleoside analogs that act as chain-terminators, the RNA extension is not impaired after the incorporation of NHC into the product strand, and the same full-length products were obtained as for the extension using cognate CTP and UTP.

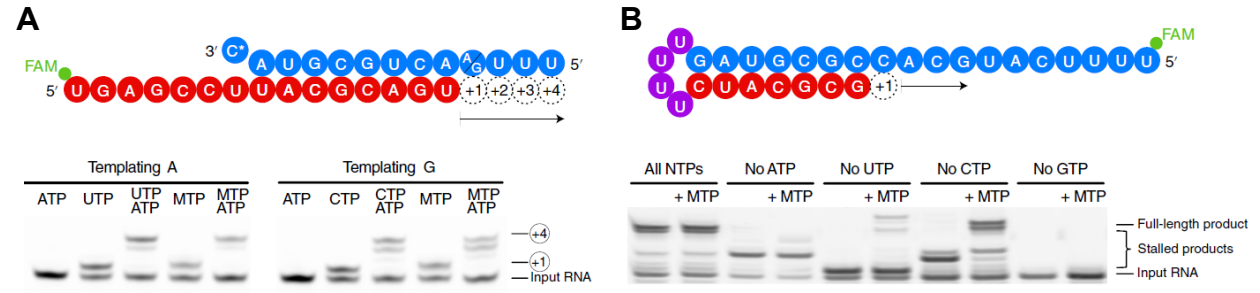



Figure 69. NHC incorporation did not stall the extension of the RNA product by RdRp. A: Three more nucleotides could be further elongated (arrow indicated the direction) after the templating nucleotide at position +1. The RNA product strand is fluorescently labeled (FAM, green) at the 5'-end. The 3'-end of the RNA template is blocked with C*. Full-length product is achieved with MTP instead of CTP or UTP in addition to ATP, shown in the gel-image. B: The template guides the elongation of 11 nucleotides, which is not stalled if MTP is used instead of CTP or UTP together with the remaining NTPs. Stalling is only observed if a cognate NTP is left out. Compensating UTP with MTP appeared to be less efficient than replacing CTP by MTP. Figures are taken from Ref.[439]

This was first demonstrated by the elongation of the RNA product opposite to a template that allows three more nucleotides (nt) to be incorporated after the position +1 ( *Figure 69*A). For another scaffold that enables the incorporation of in total 11 nt, the RdRp also accomplished full-length RNA using MTP ( Figure *69*B). If molnupiravir is administered as a prodrug, the viral RNA polymerase should be able to use it as a substrate and synthesize long NHC-containing viral RNAs.[439]

To examine what happens with the synthesized NHC-containing RNA product, it should be used as a template for RNA elongation experiments. Therefore, one oligonucleotide (R893, Table 2) produced with the NHC phosphoramidite **69** was chosen as a template and annealed to a fluorescently labeled RNA product, locating NHC at the templating position +1 (Figure 70A). Opposite to NHC, both GTP and ATP could be incorporated into the product strand, leading to a full-length product upon the addition of CTP. Whereas the extension with MTP, CTP, or UTP was not possible (Figure 70B). The installation of G or A opposite to NHC could reflect M-GTP and M-ATP base pair formation within the active center of the viral polymerase, due to different tautomeric forms of NHC.

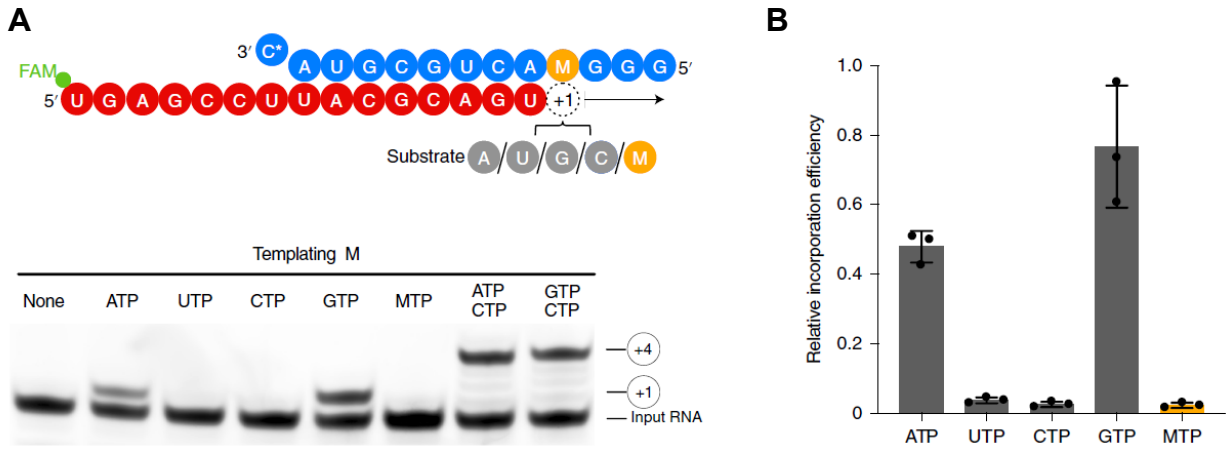



Figure 70. A: The designed template-product duplex placed NHC (M) at the templating position +1. A, U, G, C, M were tested as incoming NTPs with the recombinant RdRp of SARS-CoV-2. The RNA product strand is fluorescently labeled (FAM, green) at the 5'-end. The 3'-end of the RNA template is blocked with C*. The results are shown in the gel image below: Both ATP and GTP could be incorporated opposite to M, but not CTP, UTP and MTP. B: The experiments shown in A were measured in triplicates and the evaluation is shown as bar graphs. The incorporation efficiency was calculated for all NTPs relative to the CTP incorporation opposite to a templating G. Error bars are given as standard deviation (±).[439] Figures are taken from Ref.[439]

Additional insights into the base pairing behavior of NHC were obtained by measuring the thermal melting temperatures of RNA duplexes that contain the proposed M-G and M-A base pairs at an internal and terminal position, as a representation of different stages of viral replication. The UV thermal melting curves shown in Figure 71 represent the state of viral replication right after the incorporation of GTP or ATP opposite to NHC in the template strand by RdRp, leaving a 4 nt overhang. The resulting duplex structures with G-M and A-M base pairs are only slightly destabilized compared to the unmodified reference with a G-C base pair. This was additional support for the proposed G-M and A-M base pair formation in the active center if RdRp uses the NHC-containing RNA as a template.

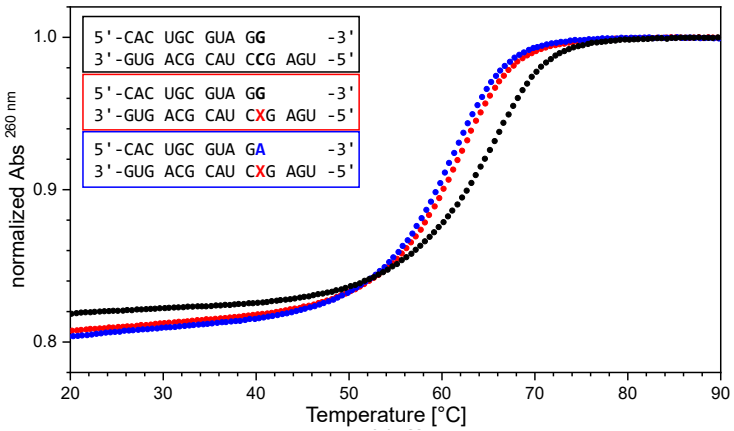



Figure 71. One example of the obtained UV melting curves (ramp 3, 260 nm) of RNA duplexes, which resemble the replication state right after the incorporation of G or A opposite to NHC (X) or C with X-G (red) or X-A base (blue) pairs. The unmodified reference is shown in black. Each duplex was measured with 20 µM duplex concentration in 10 mM sodium phosphate buffer (pH 7), containing 100 mM NaCl. Both modified RNA duplexes, X-G ($T_m$= 61.2°C) and X-A ($T_m$= 60.6°C) are only slightly destabilized compared to the unmodified RNA ($T_m$= 64.7°C). Figure is taken from Ref.[439]

Accordingly, the resulting RNA product will be extended with the correct or incorrect nucleotides, leading to mutations within the viral sequence. The destabilization appeared to become more pronounced if the modified base pair is localized in a more internal position (Table *3*). These correlations will be discussed in more detail in chapter 3.2.5.

Table 3. Results of the UV thermal melting experiments observed at 260 nm. [a] R = G or A, Y = C or X (NHC)

| sequence<br>R – Y | 5'-CACUGCGUAGR[a]-3'<br>3'-GUGACGCAUCYGAGU-5' | 5'-CACUGCGUAGRCUCA-3'<br>3'-GUGACGCAUCYGAGU-5' | 5'-CAYUGCGUAGGCUCA-3'<br>3'-GURACGCAUCCGAGU-5' |
|---|---|---|---|
| G – C | 64.7 °C | 77.6 °C | 77.6 °C |
| G – X | 61.2 °C | 71.2 °C | 73.2 °C |
| A – X | 60.6 °C | 68.2 °C | 73.8 °C |

These indications for the M-G and M-A base pair formations as basis for the mutagenesis were further validated by structural analysis of the RNA duplexes within the active center of the viral polymerase by cryo-EM.

### 3.2.4 Structural basis of NHC-induced mutagenesis

As mentioned before, the two modified RNA duplex structures shown in Figure 71 represent the state of viral replication right after the incorporation of GTP or ATP opposite to templating M by the viral polymerase. Similar RNA duplex scaffolds were prepared by annealing R893 (Table 2) to a hairpin RNA duplex to form RdRp-RNA complexes which should result in M-G and M-A base pairs at the -1 position after translocation within the active center of RdRp (Figure 72) and subjected to cryo-EM analysis. The resolved structures of the RdRp-RNA complex with the M-G and M-A base pair at position -1 had 3.3 Å and 3.2 Å resolution, respectively. Both structures represent the post-translocation state after the incorporation of either G or A opposite to M inside the template strand located at position -1, with a free NTP-binding site at position +1 (Figure 72B).[439]

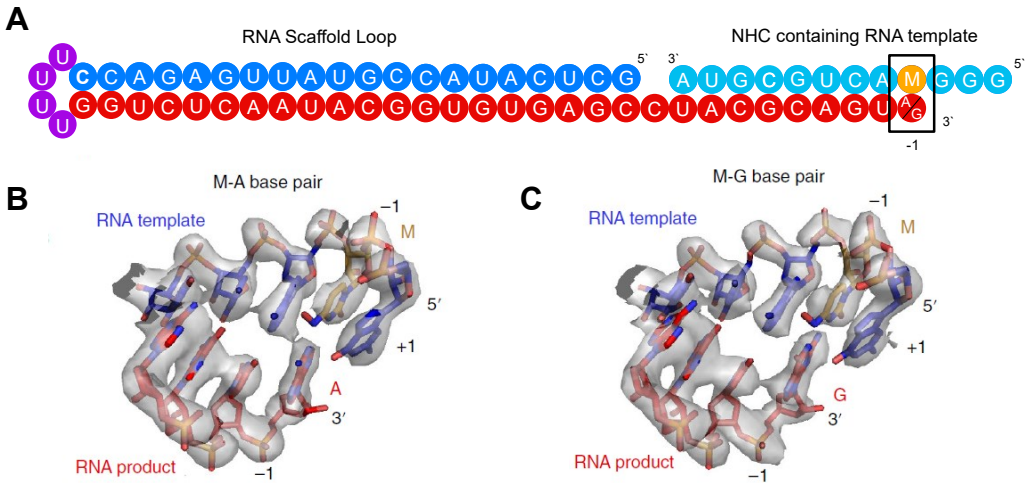



Figure 72. A: The NHC containing RNA template was annealed to a hairpin RNA duplex, with either G or A at the 3'-end of the product RNA opposite to templating M (orange), leaving a 3 nt overhang at the template strand. B: Structure of the RNA duplex in the active center of RdRp, with the M-A base pair in the post-translocation state (-1). The NTP-binding site of the RNA product (red) at position +1 is not occupied. The RNA template is shown in red. C: Structure of the RNA duplex in the active center of RdRp, with the M-G base pair in the post-translocation state (-1). Figures taken from Ref.[439]

The detailed densities of the M-A and M-G base pairs at the translocation state (-1) are shown in Figure 73. At 3 Å resolution, G and A nucleobases could be readily differentiated. The respective base pairs modeled into the cryo-EM densities were consistent with the predicted A-M and G-M base pair formation, and were feasible due to the amino and imino tautomeric forms of NHC, although the base-pair geometries seemed not quite optimal.[351, 439, 441] NHC adopted both tautomeric forms, but the resulting base pairs are not equally well established as canonical WC base pairs. These results were in line with the slightly lower incorporation efficiency of MTP in the RNA elongation assays (chapter 3.2.3).

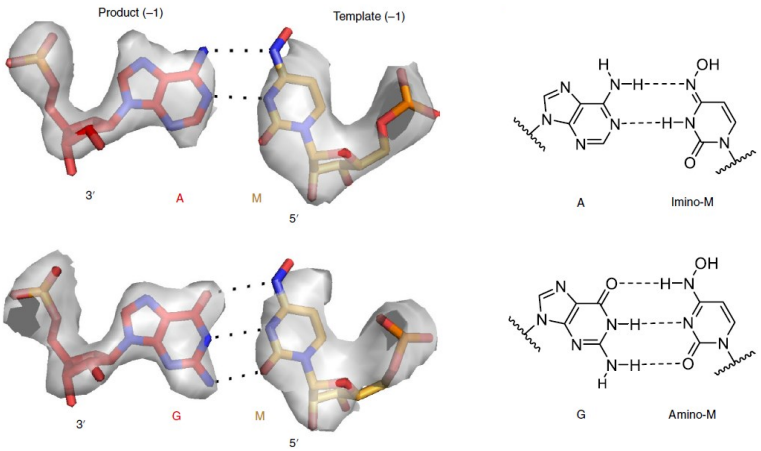



Figure 73. Excerpt of the densities obtained for the A-M (top line) and G-M (bottom line) base pairs in the translocation state (-1) by cryo-EM. On the right, the respective base pair structures on the basis of the two tautomeric forms of NHC are shown. Figure taken from Ref.[439]

Based on the obtained biochemical results, a two-step model for the NHC-induced mutagenesis of SARS-CoV-2 was proposed (Figure 74). RdRp can take MTP, the active form of molnupiravir, as a substrate for viral replication, instead of CTP and UTP. NHC will be incorporated into the - gRNA or -sgRNA product on various positions opposite to G and A of the +gRNA template strand

(Figure 74A). This resulting NHC-containing RNA can be used as the -gRNA template in the next step to synthesize +gRNA or positive subgenomic mRNA (+smgRNA) (Figure 74B). During the extension of the product RNA, opposite to NHC either the correct or incorrect nucleotide will be installed, resulting in mutations of the viral sequence.[439] The modification is not stalling the elongation and with each round of replication, NHC leads to an accumulation of mutations – the viral polymerase is not able to generate functional viruses anymore, also referred to as error catastrophe.[271, 439] This proposed mechanism is consistent with the reported appearance of G-to-A and C-to-U transition mutations in *in vitro* and *in vivo* studies, caused by misincorporation of NHC,[339, 342] and the broad-spectrum antiviral activity of molnupiravir for multiple corona viruses.[332-334]

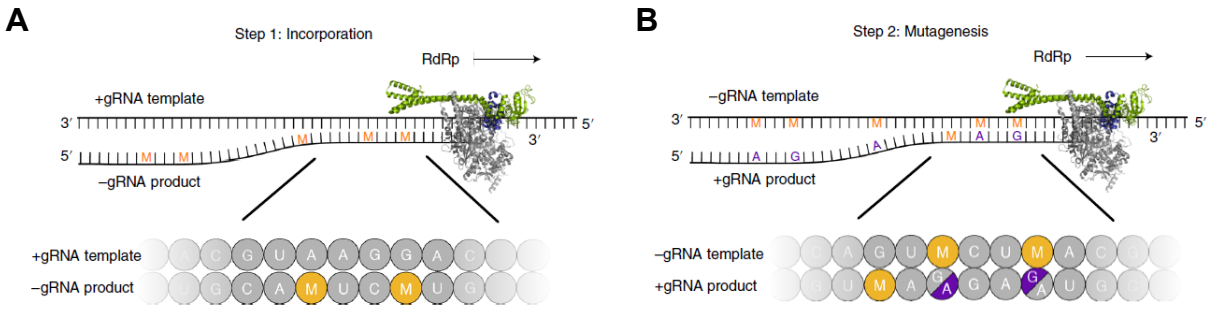



Figure 74. Based on the obtained biochemical results a two-step model for the molnupiravir-induced RNA mutagenesis was proposed. A: MTP can be incorporated into negative-strand (sub)genomic RNA (-gRNA/-sgRNA) by the viral RdRp as a C and U mimicry opposite to G and A in the positive-strand genomic RNA template (+gRNA). B: In the next step the NHC-containing -gRNA product can be used as a template for the formation of +gRNA and subgenomic mRNA (+sgmRNA). During this synthesis, either the correct or incorrect nucleotide (G/A) can be incorporated opposite to M, leading to mutations (indicated in violet) in the viral sequence, making it non-functional. The direction of extension by SARS-CoV-2 RdRp is indicated by an arrow. Figures are taken from Ref.[439]

The study revealed the ability of the M-G and M-A base pair formation in the active center of the viral polymerase, based on the appearance of the tautomeric forms of NHC, and proposed a two-step model for the NHC-induced RNA mutagenesis. This was shown by biochemical experiments as well as structural analysis of RdRp-RNA complexes. Thereby, the potential drug candidate molnupiravir seems to evade the ExoN proof-reading activity, which is an important requirement to be employed for the treatment of SARS-CoV-2.[256]

At about the same time as our study, Gordon *et al.* reported similar biochemical data to show the mechanism of action of molnupiravir based on different tautomeric forms,[442] thus corroborating our proposed NHC-induced mutagenesis. They suggested the NHC-TP substrate to be mainly present as the amino form, mimicking cytidine, as it is more efficiently incorporated opposite to templating G compared to templating A and validated it by steady-state-kinetic measurements. NHC-containing template, however, showed almost equal incorporation of GTP or ATP opposite to NHC. Though, the incorporation of GTP opposite to NHC led to a slight inhibitory effect for further elongation of the strand indicating and sub-optimal G:M base pair conformation. This could be avoided by an increased NTP concentration.[442] Based on these observation a more detailed investigation about the base pairing geometries, e.g. by NRM was sought.

Despite these encouraging results, some reports raised concerns, thus requiring further validation. Sticher *et al.* reported in 2020 that the mitochondrial DNA-dependent RNA polymerase (POLRMT) accepted MTP as a substrate and incorporated it into the growing RNA strand in primer extension assays *in vitro*.[343] This was shown to result in mitochondrial toxicity for some nucleotide analogs, which appeared not to be the case for molnupiravir. Still, they reported the NHC analog to be cytotoxic.[343] Furthermore, Swanstrom and coworkers hypothesized that molnupiravir might be a substrate of ribonucleotide reductase (RNR) during the activation to the active triphosphate form after the cellular uptake of the prodrug. This would convert them into 2'-deoxy-NHC (dNHC), which could get incorporated into host DNA causing mutations.[443] Despite the great potential as an anti-viral nucleoside analog, the interplay of molnupiravir and the $N^4$-hydroxycytidine-TP with human cellular polymerases should be more precisely investigated in future experiments.

The obtained results prove that NHC can form base pairs with either G or A depending on the two tautomeric forms: amino and imino. To learn more about the precise base pairing ability of NHC and the influence of NHC on the stability of the secondary structure with regard to its dual functionality, the modification was incorporated into different sequence contexts. With these synthesized oligonucleotides, thermal denaturing experiments, NMR, and CD-spectroscopy were performed, and the data is discussed in the next chapter.

### 3.2.5 Thermal denaturation experiments on the Dickerson Drew Sequence

NHC was incorporated at different positions of the Dickerson Drew (5'-CGCGAAUUCGCG-3') sequence (DD), replacing either cytidine or uridine, to get the respective NHC-G and NHC-A base pairs. Since it is a self-complementary sequence, both duplex and hairpin formation are possible, and the resulting duplex contained the modification twice. Additionally, a sequence with a G-U wobble, a U-A-containing sequence (see chapter 3.2.11), and an unmodified G-C sequence were synthesized to cover all possible base pairs, that can be formed with the different tautomers of NHC (Table 4).

Table 4: Synthesized sequences with the corresponding high-resolution ESI-MS-Data. X = NHC

| Name | 5'-sequence-3' | nt | formula | Mass calc. | Mass found |
|---|---|---|---|---|---|
| **R922** | CGCGAAUUCGCG | 12 | $C_{114}H_{143}N_{46}O_{82}P_{11}$ | 3808.55476 | 3808.55557 |
| **R976** | CGUGAAUUCGCG | 12 | $C_{114}H_{142}N_{45}O_{83}P_{11}$ | 3809.53878 | 3809.52670 |
| **R961** | CGXGAAUUCGCG | 12 | $C_{114}H_{143}N_{46}O_{83}P_{11}$ | 3824.54968 | 3824.53865 |
| **R962** | CGCGAAXUCGCG | 12 | $C_{114}H_{144}N_{47}O_{82}P_{11}$ | 3823.56566 | 3823.58965 |
| **R911** | CGCGAAUXCGCG | 12 | $C_{114}H_{144}N_{47}O_{82}P_{11}$ | 3823.56621 | 3823.56799 |
| **R910** | CGCGAAUUXGCG | 12 | $C_{114}H_{143}N_{46}O_{83}P_{11}$ | 3824.55022 | 3824.54427 |
| **R963** | CGUGAAUUCGXG | 12 | $C_{114}H_{143}N_{46}O_{83}P_{11}$ | 3824.54968 | 3824.54806 |

First, thermal denaturation experiments were performed using five different duplex concentrations: 20, 10, 5, 2 and 1 µM, while analyzing them at 260 and 280 nm. The sequences with NHC on

position 3 (R961) or 11 (R963) (NHC-G base pairs at position 2 and 3), as well as the sequences with the G-U (R976) wobble on position 3, are forming duplex structures like the unmodified reference (R922). The unmodified oligonucleotide (5 μM) showed an up to 17.1 °C higher melting temperature at 260 nm and 19.4 °C at 280 nm than R961 (NHC-G at position 3), R963 (G-M at position 2) and R976 (U-G at position 3; Table 5). Clearly, the modification and the G-U wobble had a destabilizing effect on the duplex structure. A syn-anti isomerization of the hydroxy group of NHC needs to take place to enable base-pairing, which results in a loss of rotational entropy (Table 6). Nonetheless, the two NHC-modified sequences R961 and R963 exhibited with $\Delta T_{m[5\mu M]}$ = 2.4 °C and $\Delta T_{m[5\mu M]}$ = 1.8 °C a higher melting temperature, respectively, compared to R976, supporting the proposed formation of WC-like NHC-G and NHC-A base pairs. The structural evaluations by cryo-EM revealed that the base pair geometries were not optimal,[439] which might also explain the lower thermal stability of the duplexes. The two modified and wobble base pairs occur twice in the duplex structure due to the self-complementary sequence as mentioned before. This additionally enhances the destabilizing effect. The NHC-G and NHC-A have only a subtle difference in values.

Table 5: Melting temperatures measured at 260 and 280 nm. Red (X=NHC), orange (G/U wobble).

| $T_m$ | 260 nm | | | | | 280 nm | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Conc [μM]** | **20** | **10** | **5** | **2** | **1** | **20** | **10** | **5** | **2** | **1** |
| R922 (CGCGAAUUCGCG) | 67.2 | 65.3 | 63.4 | 60.7 | 59.0 | 68.4 | 67.0 | 65.9 | 64.5 | 64.1 |
| R976 (CGUGAAUUCGCG) | 49.0 | 47.9 | 46.3 | 44.4 | 43.1 | 49.0 | 47.9 | 46.5 | 44.7 | 43.5 |
| R961 (CGXGAAUUCGCG) | 51.1 | 49.4 | 48.1 | 46.6 | 45.5 | 51.3 | 49.8 | 48.5 | 47.2 | 46.2 |
| R963 (CGCGAAUUCGXG) | 51.9 | 50.3 | 48.7 | 47.0 | 45.7 | 52.1 | 50.6 | 49.1 | 47.6 | 46.5 |

If the modification is at the same position as the G-U wobble, the higher stability of R961 compared to R976 appears to be enthalpically driven ($\Delta H_{R976\text{-}R961}$ = -10.6 kcal/mol; Table 6). When NHC was on position 11 (more terminal) the enthalpy value of R963 is equal to R976, but it is slightly entropically favored ($\Delta S_{R976\text{-}R963}$ = +5.2 cal/mol·K). In contrast, R963 is entropically favored compared to R961 due to the more terminal position. As a result, a comparable free energy of R961 and R963 is displayed, indicating two equal G-M base pairs. Surprisingly, the unmodified Dickerson Drew sequence R922 gives the lowest enthalpy but the more favorable/more entropically stabilized entropy values at 260 nm, even though it has a noticeably higher melting temperature (Table 6). Comparing the values obtained at 260 nm with 280 nm, there are distinct changes in the thermodynamic data for the unmodified sequence R922. At 260 nm it seemed to be entropically favored, whereas at 280 nm it appeared to be strongly entropically disfavored ($\Delta S_{280\text{-}260nm}$= -217.7 cal/mol·K). For enthalpy, it was on the contrary ($\Delta H_{280\text{-}260\ nm}$ = -74.6 kcal/mol). Two conformations seemed to be present for R922. Nevertheless, the free energy for R922 always had the highest value compared to R961, R963, and R976, at both 260 nm and 280 nm. In contrast, there are only small changes for the thermodynamic values of R976, R961, and R963, but the overall trend stays the same. The melting curve experiments for the unmodified reference were measured in triplicates to verify these distinct changes between 260 and 280 nm.

Table 6: Thermodynamic data obtained from thermal denaturation experiments by van 't Hoff plots; X=NHC.

| Thermodynamic data | $\Delta H_{260}$ kcal/mol | $\Delta S_{260}$ cal/mol·K | $\Delta G_{260}$ kcal/mol | $\Delta H_{280}$ kcal/mol | $\Delta S_{280}$ cal/mol·K | $\Delta G_{280}$ kcal/mol |
|---|---|---|---|---|---|---|
| **R922 (CGCGAAUUCGCG)** | -81.1 | -215.5 | -16.9 | -155.7 | -433.2 | -26.6 |
| **R976 (CGUGAAUUCGCG)** | -101.9 | -293.8 | -14.4 | -110.6 | -320.2 | -15.2 |
| **R961 (CGXGAAUUCGCG)** | -112.5 | -324.5 | -15.8 | -124.1 | -359.7 | -16.9 |
| **R963 (CGCGAAUUCGXG)** | -101.1 | -288.6 | -15.2 | -112.0 | -322.1 | -16.0 |

Based on the hyperchromicity (Figure 75), R922 has the lowest value (22%) at 260 nm compared to R961 (24%), R963 (24%), and R976 (28%). That reflects an overall lower number of base pairs, indicating hairpin formation at lower concentrations. As opposed to 260 nm, R922 had the highest value (47%) among the four RNAs at 280 nm. This finding was in line with the distinct changes in thermodynamic properties for R922 upon changing to 280 nm (Table 6), indicating the presence of two conformations. Interestingly, R961 and R963 showed comparable values in hyperchromicity at 260 nm as well as 280 nm, clearly showing that the two sequences with the modification are forming stable NHC-G base pairs and thus duplex structures even though the melting temperature is decreased compared to the unmodified sequence. The modification favors duplex formation also at lower concentrations in contrast to the unmodified sequence. Consequently, a duplex structure is more favorable for the modification than being inside the stem of the hairpin. The wobble-containing sequence showed a higher hyperchromicity than both NHC-modified sequences at 260 and 280 nm.

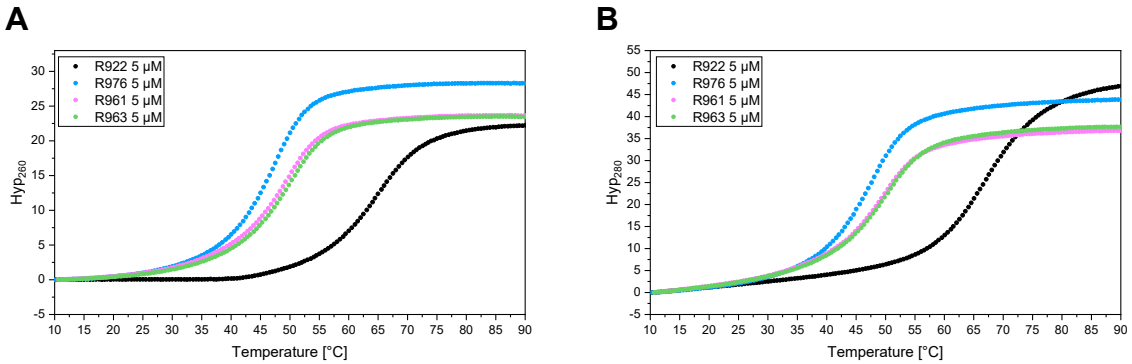



Figure 75. Hyperchromicity of R922 (black), R961 (magenta), R963 (green) and R976 (blue) at 5 µM duplex concentration (ramp 3). A: at 260 nm, B: at 280 nm.

Taking a closer look at the unmodified sequence, the overall $\Delta T_m$ difference decreased from 8.2 °C to 4.3°C from 260 to 280 nm, respectively. By contrast, this difference stayed almost unchanged for the modified sequences.

For R922, the melting temperatures were less concentration dependent at 280 nm. As seen in Figure 76, the slope for R922 at 280 nm is distinctly flatter than at 260 nm, indicating that the melting is almost concentration independent at 280 nm. At 280 nm, primarily the melting of GC base pairs can be followed, since the $\Delta\varepsilon$ of (GC) is considerably higher than that of (AU).[444] R922 seems to prevail, in principle, as a duplex structure with a minor hairpin fraction. The monophasic

behavior seen in Figure 75 might be because of similar $T_m$ values for the unmodified hairpin and duplex structures. Otherwise, with different melting temperatures, a biphasic transition should be observed in the melting curves. Between 40 and 55 °C, for the unmodified sequence, a small minor second transition might be divined at 260 nm. The slopes of the modified and G-U sequences were comparable and were not changing between 260 and 280 nm, since the melting temperatures are not dependent on the wavelength as shown before. To sum up, the NHC modification is stabilizing the duplex structure compared to the G-U sequence, but both are less stable than the unmodified sequence. Due to the destabilizing effect of NHC, localization within the short stem of the hairpin is not permitted and they are primarily forming duplex structures.

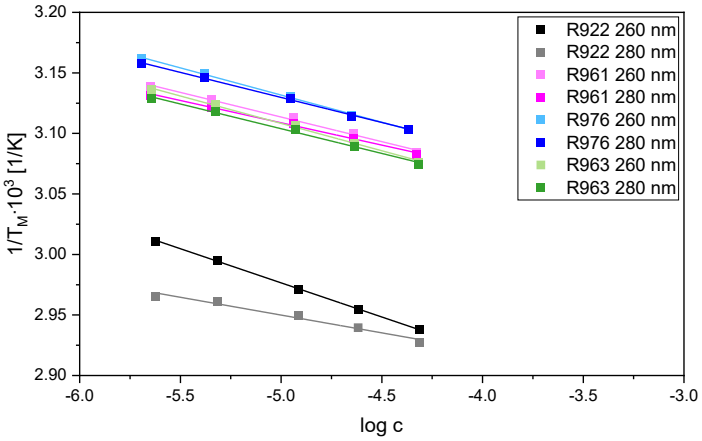



Figure 76. Van't Hoff profiles of the oligonucleotides at 260 and 280 nm. The slope visualized the concentration dependency of the melting temperature. The melting of R922 is less concentration dependent at 280 nm (grey) than at 260 nm (black). For the NHC-modified (magenta and green), and G-U-containing (blue) DD sequence, the concentration dependency is almost equal at both wavelengths (represented by dark and light colors).

So far, base pairs at position 2 and 3 of the Dickerson Drew sequence have been modified. Upon modifying other more central positions, the melting profiles changed. For R910 (G-M at position 4/9), with the modification replacing a central cytidine at position 9, the curves still showed a monophasic behavior at higher concentrations but with a 3.5 – 4.3 °C lower melting temperature than R961 and R963 (Figure 77A,B). With lower concentrations, biphasic transitions emerged at 260 nm. For R962 and R911, where both the uridines at position 7 and 8 were replaced by NHC, no monophasic behavior was observed at 260 nm at any sample concentration (Figure 77C,D). This biphasic melting resulted from a duplex-hairpin equilibrium of the self-complementary Dickerson Drew sequence, explaining the lower hyperchormicity compared to the duplex forming sequences (Figure 77A). The low-temperature transition corresponds to the conversion of a duplex into two hairpin structures, a concentration-dependent process (Figure 77C). The high-temperature transition, however, corresponds to the melting of hairpins to single strands, which is concentration-independent (exemplarily shown as insets for R962, Figure 77C).[444] Accordingly, two melting temperatures were obtained, evaluated by forming the derivation of the melting curve. The obtained maximum represents the melting temperature, and the data are summarized in Table 7.

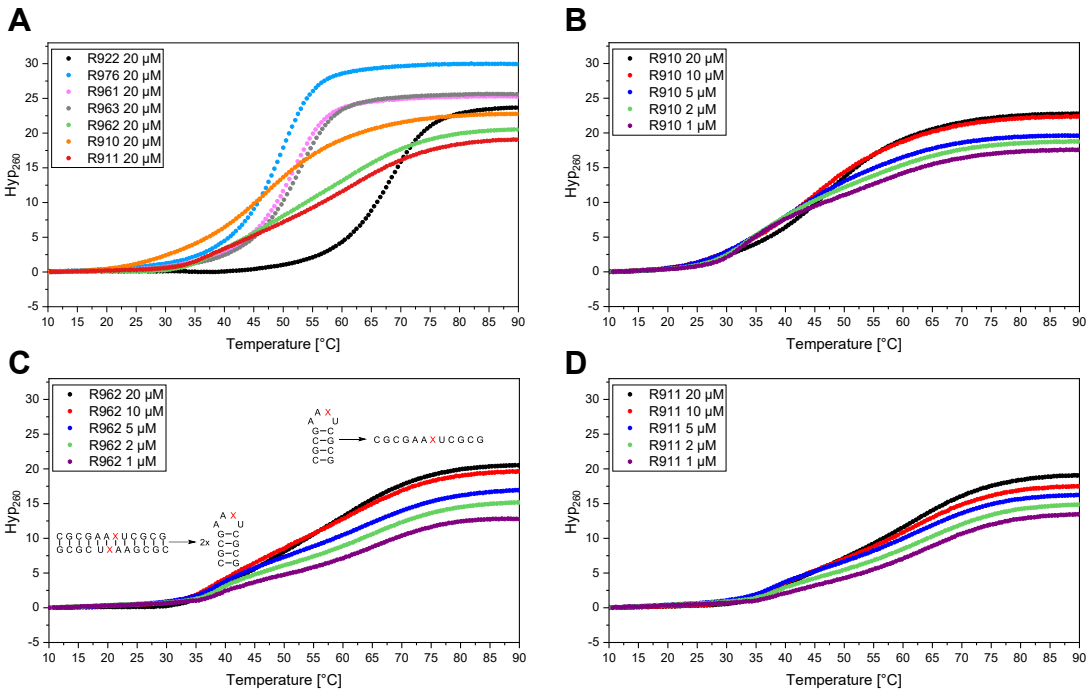



Figure 77. Thermal denaturation experiments by UV spectroscopy. Hyperchromicity vs. Temperature. A: Overview of all the different Dickerson Drew Sequences (20 µM) at 260 nm. B-D are R910, R962, and R911, respectively at 260 nm. They mainly showed a biphasic melting profile, due to a duplex-hairpin-equilibrium, exemplarily shown as an inset in C.

At 280 nm, mainly the melting of G-C base pairs can be observed as mentioned earlier. Accordingly, the transition at 280 nm represents the melting of the hairpin-stem since this only constitutes G-C base pairs. This process is, as expected for the melting of hairpin structures, concentration-independent (Figure 78). The transition at 280 nm can sometimes also be observed at lower concentrations of R910, R911, and R962 as a second transition at 260 nm (Table 7*).* Those three sequences, with the modification at a more internal position, seemed to predominantly form hairpin structures, especially at lower concentrations.

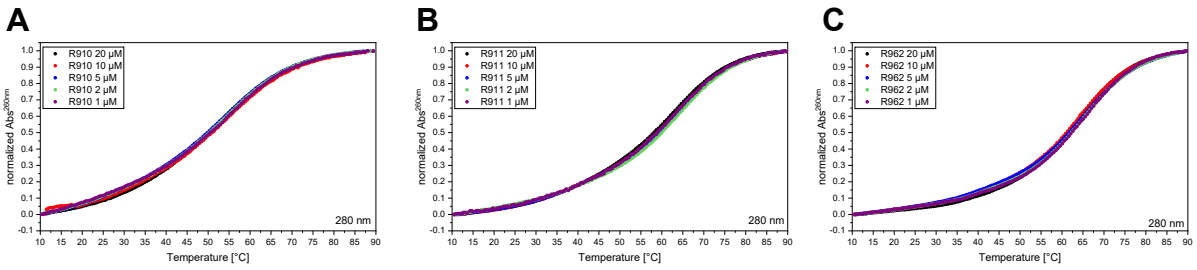



Figure 78. normalized absorbance vs. temperature for A: R910, B: R911, and C: R962, showing the concentration-independent melting at 280 nm.

The melting temperatures for the low-temperature transition (duplex to hairpin) were between 33.7 and 39.9 °C at 2 µM concentration. The high-temperature transitions were between 55.4 and 65.0 °C for R910, R911, and R962, respectively (Figure 77, Table 7).

Hence, the duplex-hairpin equilibrium can be shifted towards either side, depending on the localization of the modification in an equal manner opposite to A or G.

Table 7: Overview of melting temperatures for 20 µM and 2 µM at 260 and 280 nm.

| | $T_m$ [20 µM] | | $T_m$ [2 µM] | |
|---|---|---|---|---|
| | 260 nm | 280 nm | 260 nm | 280 nm |
| **R922 (CGCGAAUUCGCG)** | 67.2 | 68.4 | 60.7 | 64.5 |
| **R976 (CGUGAAUUCGCG)** | 49.0 | 49.0 | 44.4 | 44.7 |
| **R961 (CGXGAAUUCGCG)** | 51.1 | 51.3 | 46.6 | 47.2 |
| **R963 (CGCGAAUUCGXG)** | 51.9 | 52.1 | 47.0 | 47.6 |
| **R910 (CGCGAAUUXGCG)** | 47.6 | 54.0 | 35.7 (*55.3) | 55.4 |
| **R911 (CGCGAAUXCGCG)** | 62.4 | 64.4 | 38.7 (*63.8) | 64.6 (*39.0) |
| **R962 (CGCGAAXUCGCG)** | 61.7 | 64.0 | 39.9 (*64.5) | 64.1 (*39.2) |

In 10 mM sodium phosphate buffer, pH 7.0 containing 100 mM NaCl, *$T_m$ value of the second transition. Black: sigmoidal fitting of thermal melting curve, grey: 1. derivation of thermal melting curve

### 3.2.6 Electrophoretic mobility shift assay

Despite the conformity of the three sequences to form hairpin structures, R910 showed considerable differences in the shape of the melting curves and the resulting melting temperatures. The $T_{m[20\ \mu M,\ 280\ nm]}$ of the high-temperature transition was about 10 °C, and the low-temperature transition was about 5 °C lower compared to R962 and R911. For R911 the modification was placed more towards the center of the sequence. Thus, the formed stem of the hairpin was constituted of 4 G-C base pairs. In contrast to R911, the modification in R910 was placed at the last of these G-C base pairs. We postulate that the NHC-G base pair inside the stem is destabilized, leading to a hairpin with only 3bp-stem, without the modification being involved in base pairing.

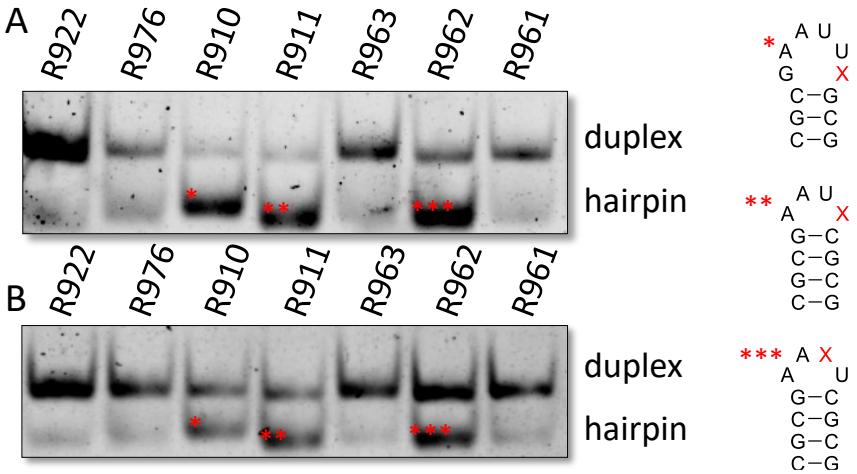



Figure 79. Electrophoretic mobility shift assay on native PAGE. A: Annealing for 2 min at 95°C followed by rapid cooling on ice leads to primarily duplex formation for R922, R976, R963 and R961, and hairpin formation for R910, R911 and R962. B: Annealing for 2 min at 95°C followed by slowly cooling to room temperature leads to a duplex-hairpin mixture for R910, R911, and R962. Red asterisk indicates the formation of different hairpin structures (3bp vs. 4bp stem). The samples are in 10 mM sodium phosphate buffer, pH 7.0 containing 100 mM NaCl. 6 µM annealing concentration; X=NHC.

This assumption is supported by a gel shift assay, where two different annealing procedures were tested for the unmodified, the G-U-containing and the four NHC-modified Dickerson Drew sequences (Figure 79). Annealing for 2 min at 95°C and rapid cooling on ice leads to the primary formation of hairpin structures for R910, R911, and R962.[445] Thereby, a small shift was observed between R910 and R911/R962, indicating a minor structural change. This would be in line with a

3 bp- vs. 4 bp-stem structure (red asterisk, Figure 79A). R910 appeared to form a bigger loop, leaving the potential G-M base pair out of the stem region. Both R911 and R962 formed a 4bp stem, as the modification was in the loop region, which was in line with the differences in the melting temperature for R910 and R911/R962. Annealing for 2 min at 95 °C and slow cooling to room temperature within 30 min was changing the hairpin-duplex ratio towards the duplex structure for these 3 sequences (Figure 79B). For R922, R976, R961, and R963, the duplex structure is dominant regardless of the annealing procedure. Thus, the NHC modification in a more terminal position favored duplex formation, whereas a more central position favored hairpin formation excluding the modification from base pairing. This also led to the formation of two varying hairpin structures comprised of three or four base pairs within the stem.

### 3.2.7 NMR analysis of the duplex forming Dickerson Drew Sequences

To further evaluate the influence of the modification on structural changes, NMR samples with ~200 µM were prepared for each of the Dickerson Drew sequences containing 10 mM sodium phosphate buffer (pH 7) and 100 mM NaCl in $D_2O/H_2O$ (9/1). $^1$H-NMR and NOESY spectra were recorded and assigned by Dr. Irene Bessi.

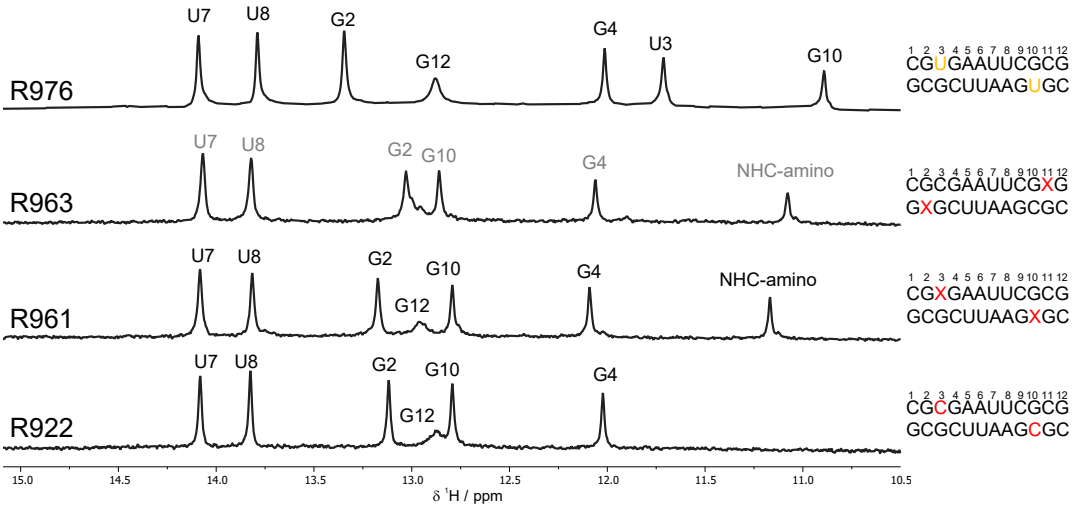



Figure 80. Imino region of the 1D $^1$H NMR spectra of the Dickerson Drew sequences that form duplex structures. NMRs were measured at 283 K in water/$D_2O$ (9:1) containing 10 mM sodium phosphate buffer (pH 7) and 100 mM NaCl and DSS (internal standard) at 600 MHz. In the lowest line of the unmodified R922, the six expected signals were detected. The assignment is shown on top of the signals. Interestingly, for the modified sequence and additional signal between 11.0 and 11.2 ppm appeared. For comparison, the wobble containing sequence was also measured, which showed 7 signals in the topmost line. The peaks were assigned by 2D NOESY (by Dr. Irene Bessi). The assignment of R963 was assumed based on R961 and is therefore shown in grey; X=NHC.[446]

$^1$H-NMR spectra of all Dickerson Drew sequences shown before were recorded over a temperature range of 283-313 K. In addition, imino, anomeric, and aromatic regions of the duplex-forming sequences (R922, R961, R963, and R976) were analyzed and assigned by 2D NOESY NMR. With the help of the imino-region indications on the base pairing behavior, their number and base pair pattern, and structural changes can be obtained. The unmodified R922 12mer should show six signals in the Watson-Crick imino-region between 12 and 14 ppm (4 times G-C and 2 times U-A),

due to the symmetry of the Dickerson Drew sequence. Indeed, six signals were obtained, confirming the formation of a duplex structure which was expected due to the high sample concentration (Figure 80, 200 µM). Five of the signals showed up as sharp signals with almost equal intensity and one as a broad signal with low intensity. The latter can be easily assigned to the G12. The terminal G-C base pair is the weakest and therefore, it only shows a small signal (~12.9 ppm, Figure 80) which disappeared upon heating. The assignment of the other signals was also in accordance with earlier NMR studies on the Dickerson Drew Sequence. [447, 448] In line with the unmodified sequence, R961 and R963 (X-G) show six signals between 12.0-14.2 ppm (Figure 80). While for U7 and U8 no shift was observed, the signals for G2, G10 and G4 were slightly shifted. The modification appeared to influence adjacent base-pairs leading to minor structural changes. Interestingly, an additional signal appeared between 11.0-11.2 ppm for the modified sequences R961 and R963, which was not visible for the unmodified sequence. Consequently, this additional signal might derive from the NHC modification.

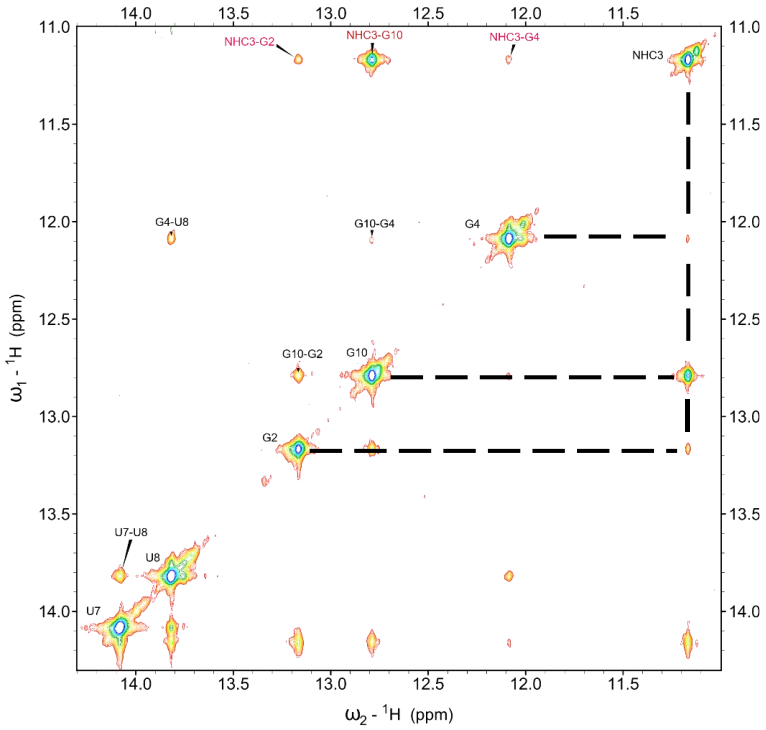



Figure 81. NOESY of the R961 duplex measured at 283K, 600 MHz, with a mixing time of 200 ms. The obtained cross peaks of the additional signal at 11.2 ppm are indicated by dashed lines. Cross peaks with medium intensity were observed to G2 and G4, and a strong cross peak to G10, opposite to the modification. This confirmed that the additional signal at 11.2 ppm belonged to the NHC modification.

The NOESY-spectrum of R961 showed cross-peaks between the additional signal at 11.2 ppm and G2 and G4 with medium intensity and a strong cross-peak to G10 (Figure 81). This confirms that the additional signal belongs to the NHC-modification. Two explanations were possible for the appearance of the signal with a proton chemical shift of 11.2 ppm: 1. The NHC modification is not forming a C-G like base pair but a U-G wobble instead, which would fit to the detection of the strong cross-peak from NHC to the G opposite to it (Figure 81). NHC would adopt the imino form to enable

wobble base pair formation. 2. The N4 amino proton of the modification is exceedingly shifted upon base pairing and/or the influence of the hydroxy group, which would also be in line with the observed cross-peak with G10. Such exceptional downfield shifts of amino signals compared to cytidines involved in WC base pairing have been reported before.[353, 449] Both cases were possible, due to tautomerization of NHC (Figure 82). For the G-U dodecamer R976 seven signals were detected in agreement with R961 and R963 but the imino signals were distinctly shifted. Corresponding with the two NHC modified RNAs, R976 had a signal close to the one belonging to NHC, at 10.9 ppm. However, according to the assignment, this signal belongs to G10, across from U3 (Figure 80).[446]

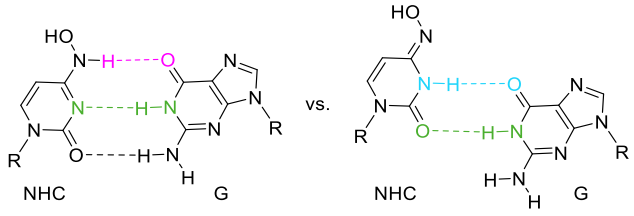



Figure 82. Two possibilities of the NHC-G base-pair in the amine (left) or oxime (right) form.

To get more detailed information about the base pairing behavior of NHC in R961 and to unambiguously assign the additional proton signal at 11.2 ppm, $^{15}$N-NMR experiments are a useful tool, by putting the isotope label directly at N4 of NHC. $^{15}$N NMR experiments are very sensitive since the natural abundance of $^{15}$N (0.37)[450] is only around 1/3 of $^{13}$C. Compared to the major $^{14}$N isotope (spin = 1), $^{15}$N (spin = 1/2) spectra have the advantage that they can be interpreted more easily due to sharper bands. To circumvent the low natural abundance, $^{15}$N can be introduced synthetically using $^{15}$N labeled reagents.[451] With the help of the nitrogen chemical shifts of these labeled molecules, isomeric structures, like tautomers, can be differentiated. The corresponding shifts of two isomeric pairs are relatively large.[452]

The $^{15}$N label was synthetically introduced with $^{15}NH_2OH$ by substituting 5'-*O*-DMT-2'-*O*-TOM-*O*4-chlorophenyluridine (Scheme 21) to get the $^{15}$N4-NHC variant. With this labeled building block, the structure of $N^4$-hydroxy cytidine could be determined. Looking at $^{1}$H-$^{15}$N-HSQC and HMBC, the two tautomeric forms, namely imino or amine form could be differentiated. Since there is a high abundance of $^{15}$N, one intense peak would have been expected in the $^{15}$N-HSQC if a hydrogen would be bound to the $^{15}$N. Instead, two peaks with low intensity appeared in the spectrum, with shifts of -93.29 and -262.89 ppm (Figure 83). The peak at -93.29 ppm was also present in the $^{1}$H-$^{15}$N-HMBC. Due to the high accumulation of $^{15}$N, a cross peak of the $^{2}$J-coupling between OH and N4 is visible. Considering the chemical shift, N3 seems to be present in the amide, and N4 in the imino form, the latter without a hydrogen attached. That is in accordance with the reported nitrogen chemical shifts for amides and oximes,[453] and also with the calculated energetically more favorable imino structure.[454] Additionally the carbons of C4 and C5 have doublet signals in the $^{13}$C spectrum, caused by the coupling with $^{15}$N.

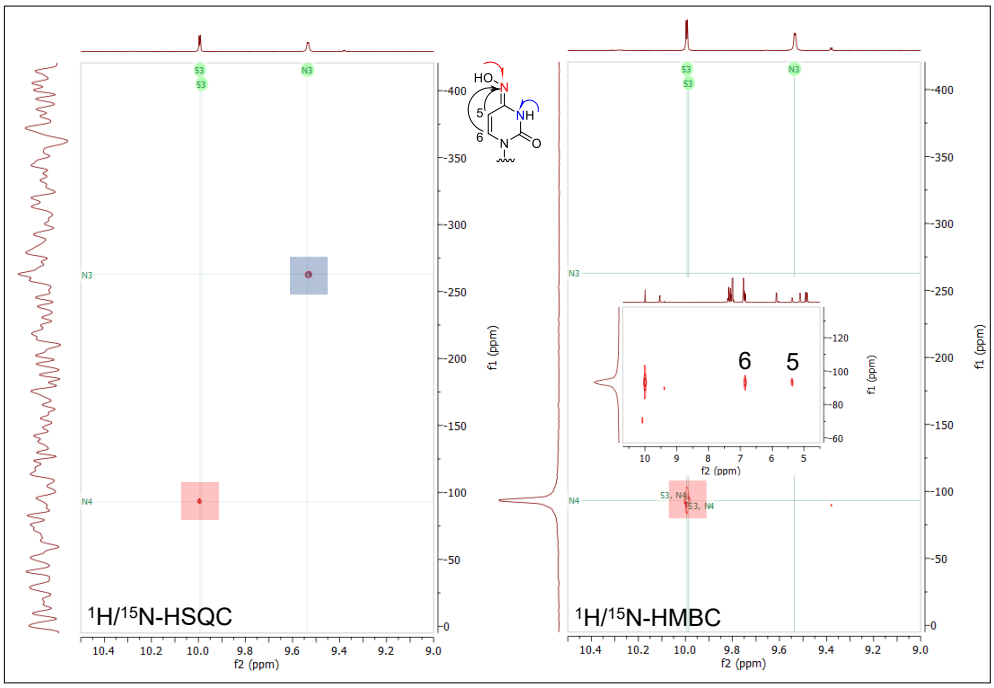



Figure 83. Comparison of 1H-$^{15}$N-HSQC (left) and $^{1}$H-$^{15}$N-HMBC (right) spectra with $^{15}$N labeled 5'-O-DMT-2'-O-TOM-NHC (**67-$^{15}$N**) referenced to nitromethane. Marked in blue is the shift for N3 in the amide form. Marked in red is the shift for the $^{15}$N labeled N4, which is present in the imino form, without a hydrogen attached. The observed cross-peak corresponds to the $^{2}$J-coupling between OH and N4 (red arrow). The hydrogen is directly attached to the $^{15}$N (blue arrow).

In the next step, the hydroxyl group of NHC was protected with benzoic anhydride. The benzoyl protection could either take place at the hydroxy group or at N4. Earlier reports often described NHC to be present in the amino form,[362, 455] and that the nitrogen gets protected.[440]

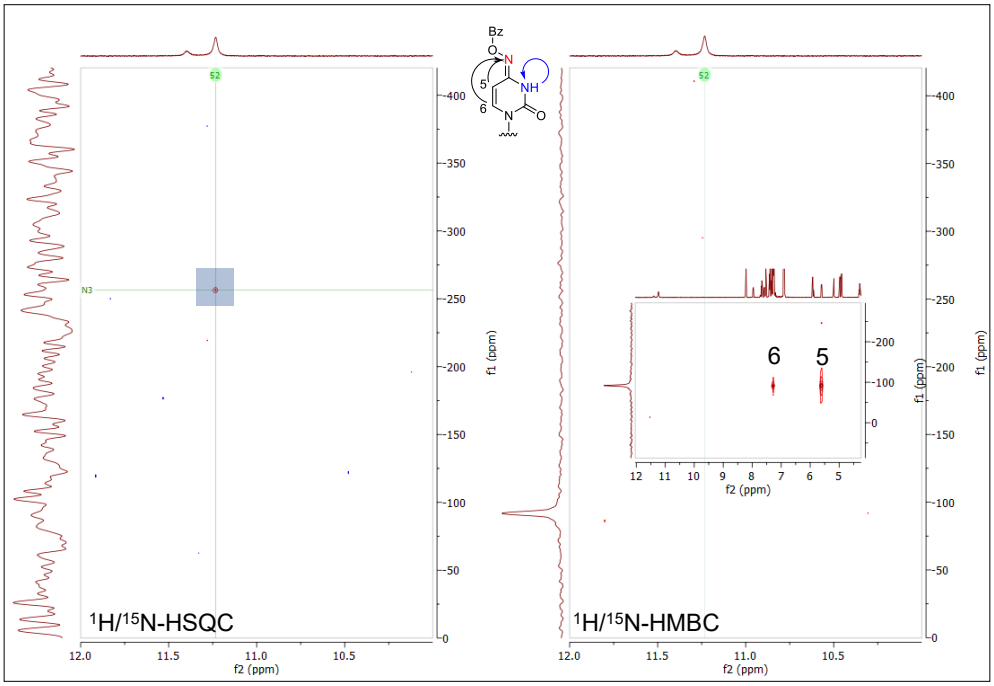



Figure 84. Comparison of $^{1}$H-$^{15}$N-HSQC (left) and $^{1}$H-$^{15}$N-HMBC (right) spectra with $^{15}$N labeled and benzoyl protected NHC (**68-$^{15}$N**) referenced to nitromethane. Marked in blue is the shift for N3 (-256.2 ppm) in the amide form. Since the OH group was protected, no cross peak for N4 was observed. Only the coupling of H-5 and H-6 (see small inset, right) to a peak at -91.65 ppm was detected, which corresponds to the imino form of N4.

However, the cross peak in the $^{1}$H-$^{15}$N-HSQC of the hydroxy group to the $^{15}$N4 vanished and only a single peak for N3-H with a low intensity at -256.2 ppm still remained. This was also observed in the $^{1}$H-$^{15}$N-HMBC. Only the cross peaks of H-5 and H-6 to N4 at -91.65 ppm were still visible, which matches the imino form (Figure 84). Taking into account the collected data, the benzoyl protection took place at the hydroxy group with hydroxy cytidine in the imino form, contrary to the report of J. A. Piccirilli and coworkers.[440]

Table 8: Synthesized sequences with $^{15}$N-NHC with the corresponding high-resolution ESI-MS-Data.

| Name | 5'-sequence-3' | nt | formula | Mass calc. | Mass found |
|---|---|---|---|---|---|
| **R993** | CGXGAAUUCGCG | 12 | $C_{114}H_{143}N_{45}{}^{15}NO_{83}P_{11}$ | 3825.55695 | 3825.54671 |
| **R994** | CGCGAAXUCGCG | 12 | $C_{114}H_{144}N_{46}{}^{15}NO_{82}P_{11}$ | 3824.54891 | 3824.56269 |

The $^{15}$N-NHC phosphoramidite building block was incorporated into the Dickerson Drew Sequence after phosphitylation, following the same procedure as the unmodified NHC building block, replacing either cytidine or uridine (Table 8), equivalent to R961 and R962. Different 1D [$^{1}$H-$^{15}$N] experiments were measured with the $^{15}$N-labeled Dickerson Drew sequence R993, the labeled analog of R961, by Dr. Irene Bessi (Table 8). For a proton coupled to a $^{15}$N, which would be the case if the NHC is present in the amino form, a specific $^{1}$J-coupling is expected with an average of 90 Hz.[456]

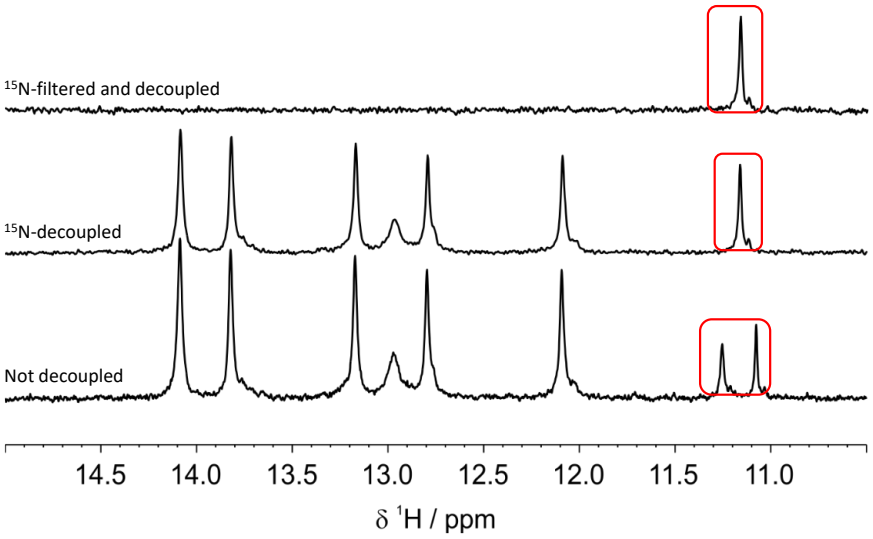



Figure 85. Imino region of the $^{1}$H-NMR spectra with the $^{15}$N4-labeled dodecamer R993, as the labeled analog of R961. The bottom line shows the $^{1}$H-spectrum without $^{15}$N decoupling. The middle line is the $^{1}$H-spectrum with $^{15}$N decoupling. And the top line is with $^{15}$N filter and $^{15}$N decoupling. The signal of interest is marked with a red box.

To observe this coupling, a $^{1}$H spectrum without $^{15}$N decoupling was recorded. Indeed, a splitting of the signal at 11.2 ppm was observed with a coupling constant of 106 Hz, consistent with a one-bond $J_{N\text{-}H}$ coupling (Figure 85, bottom spectrum, marked with a red box). This split signal is merged with $^{15}$N-decoupling and only one signal remained at 11.2 ppm (Figure 85 middle line). The $^{15}$N filtered, with a transfer delay for $J_{N\text{-}H}$ set to 90Hz, and $^{15}$N decoupled 1D 1H spectrum of R993 (Figure 85 top spectrum) showed one peak at 11.2 ppm. This was in agreement with the signal obtained for the $^{15}$N-decoupled spectra. Only protons which are directly bound to a $^{15}$N, resulting

in a coupling of around 90 Hz, can be detected with this experiment. Thus, the signal at 11.2 ppm of the modified Dickerson Drew R993 was unambiguously assigned to the hydrogen bound to the $^{15}$N at position 4 of NHC. The amino form of NHC is present in the dodecamer when base pairing with G in solution, forming a WC-base pair. This modified base pair resulted in an exceptional downfield shift of the amino signal of NHC. Similar downfield shifts have mainly been reported for N3 protonated cytidines involved in Hoogsteen base pairing in C$^+$•G-C triplexes before. In here, the shift is mainly caused by the protonation of cytosine.[457-459] Additionally, a contribution of the hydrogen-bonding to the chemical shift of the N4-amino proton was hypothesized.[457] Even in a canonical G-C base pair, the non-hydrogen amino proton of the cytosine is slightly upfield-shifted compared to the hydrogen-bonded proton.[460] A smaller downfield-shift of the N4 amino signal (~ 9 ppm) was observed in a 2AP-C mismatch, where the amino proton is hydrogen-bonded to the N1 of 2AP.[461] Another example of such a distinct downfield shift was reported by Fazakerley *et al.*[353] They detected the amino proton of N$^4$OMe in a DNA WC base pair with G at 11.25 ppm. Depending on the nucleobase opposite to the modification, the tautomeric equilibrium changed. In a base pair with A, N$^4$OMe was predominantly found in the imino conformation. However, when forming a base pair with G, both tautomers were simultaneously observed in an equilibrium, resulting in a WC- and Wobble-conformation with the amino and imino form of N$^4$OMe, respectively.[353] The NHC amino signal of the modified Dickerson Drew sequence was found at a similar chemical shift, but only the WC base pair was observed and no wobble conformation. These examples showed the advantage of $^{15}$N NMR experiments, due to their sensitivity for the hydrogen-bond environment to the resulting shift of the signals. We presume that the distinct downfield-shift for the NHC-N4-amino signal is caused by base pairing with G and enhanced by the adjacent hydroxyl group, which has not been reported before for the NHC modification.

The $^{15}$N label further allowed the determination of the $^{15}$N chemical shift of the exocyclic amino group of NHC in the NHC-G base pair. The chemical shift of N4 from a canonical cytidine was reported in the BioMagResBank (BMRB) to be at around 98 ppm (indirectly referenced to DSS).[462] The $^1$H-$^{15}$N HSQC experiment of R993, which was indirectly referenced to DSS, revealed a correlation between the proton signal at 11.2 ppm and a nitrogen resonating at 156 ppm (Figure 86). Thus, the N4 chemical shift of NHC and a canonical cytidine seemed to significantly differ from each other. A similar downfield shift of the N4 signal was observed within the C$^+$•G-C triplexes mentioned before, with a nitrogen chemical shift of around 114 ppm for the positively charged cytidine.[449] Furthermore, NMR spectra of another N4-modified free nucleoside, $^{15}$N-labeled *N*$^4$-acetylcytidine, measured in $d_6$-DMSO also showed a downfield shift of around 50 ppm of the exocyclic nitrogen signal of the modification compared to a canonical cytidine.[463] This markedly downfield shift of the N4 of NHC seemed to be reasonable.

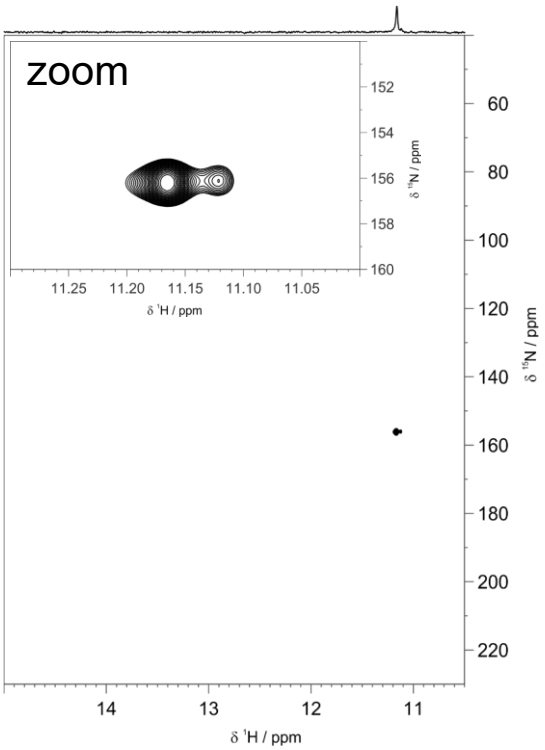



Figure 86. $^{15}$N HSQC spectrum of R993, containing the $^{15}$N labeled NHC at position 3, recorded at 283 K. The detected cross-peak between the proton signal at 11.15 ppm with the nitrogen signal at 156 ppm is shown enlarged (= zoom). A second conformation, which also contains a $^{15}$N label, can be observed right next to the main peak.

However, a minor secondary conformation was observed right next to the main peak in the $^{15}$N HSQC. The presence of this second species was already observed in the NOESY of the NHC-containing R961 duplex. The low abundance (<5%) of this conformation makes it difficult to assume what the second species might be, but it also had to contain a $^{15}$N with the same base-pairing pattern as the R993 duplex, excluding the wobble-conformation. Both, the amino proton as well as the N4 nitrogen signal of the NHC modification were markedly shifted in a Watson Crick base pair with G compared to a canonical C-G base pair.

### 3.2.8 Imino-water exchange rates

Dr. Irene Bessi determined imino-water exchange rates for R922, R976, and R961 (Figure 87) and the amino exchange rates for the NHC modification at position 3 by NMR for the signals obtained in the $^{1}$H imino region (Figure 80). Since the latter was the only amino proton, it could not be compared to the data obtained for the imino protons. The exchange rate of imino protons with the solvent can be reduced compared to a free nucleobase by hydrogen bonding and/or shielding from solvent.[464, 465] The exchange with water of hydrogen bonded imino protons occurs in the open state during the opening-closing kinetics of the WC base pair.[465] Based on the exchange rates information about the relative stability of the oligonucleotide structure and the respective base pairs can be obtained.[466]

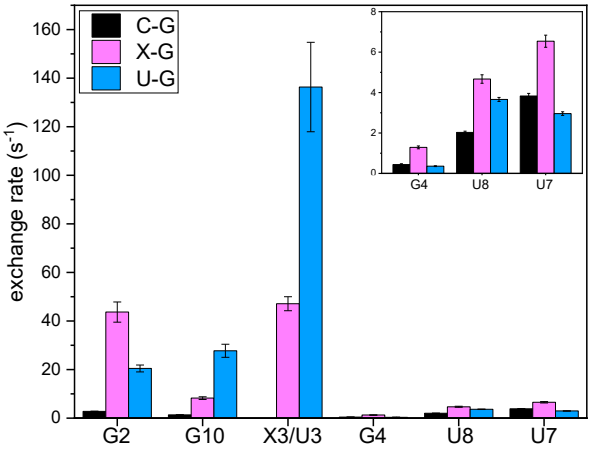



Figure 87. Imino-water exchange rates of the signals obtain in the $^{1}$H imino region at 298K for R922 (black), R961 (magenta) and R976 (blue) were determined by 1D CLEANEX-PM experiments and plotted as bar graphs; X=NHC.

At position 3 and the corresponding position G10, R976 showed the highest exchange rates of the proton, which was in line with the lower melting temperature of the wobble-containing RNA duplex (Table 7). The G-U wobble base pair of U3 with G10 was weaker, consequently the proton exchanges faster, compared to NHC-G and the canonical C-G base pair. Interestingly, the amino proton of NHC at position 3 of R961 was exchanging slower than the imino of the wobble G-U. Taking into account that the amino signal in the $^{1}$H-NMR (Figure 80) was much sharper compared to normal amino signals of unmodified nucleotides, it seemed that this hydrogen bond was stabilized. No comparison for a cytidine amino exchange rate could be found in the literature, as this signal normally occurs further upfield-shifted, overlayed with other amino signals. This highlights the exceptional determination of the amino exchange rate for the NHC modification. Moreover, the modification seemed to influence the neighboring base pairs. The exchange rate of the G2 in the NHC-modified RNA is markedly higher than for the corresponding base in the unmodified sequence R922 and the G-U containing R976, with an exchange rate around 16-fold higher than R922. The value at position G2 was almost as high as the amino exchange rate at position 3. The same trend was also observed towards the 3'-end for G4, U7, and U8, but distinctly less pronounced. The higher exchange rates of R961 compared to R922 are consistent with the lower melting temperature of the NHC-containing RNA duplex. This was a result of a destabilizing effect of NHC also on adjacent base pairs. However, NHC seemed to form a more stable base pair than G-U, as the latter showed a considerably higher imino exchange rate of U3.

### 3.2.9 NMR analysis of the hairpin-forming Dickerson Drew Sequences

For the imino-NMR of the hairpin-forming sequences, fewer signals were expected compared to the duplex spectra (Figure 80). The stem was supposed to comprise three or four G-C base pairs, depending on the position of the modification (Figure 79). For R962 (top line, Figure 88), 3-4 main signals were observed between 12 and 13.2 ppm. Additionally, a minor conformation showed up in the region for A-U base pairs. Due to the high concentration of the NMR sample (200 µM), the

duplex structure was formed as well, already indicated in a faint band in the gel shift assay (Figure 79). The spectra of R910 looked analogous to R962, but with a higher ratio of the second conformation (duplex). As for R961 and R963 (Figure 80), an additional signal at 11.2 ppm, which was identified to be the N4-amino proton of NHC, was detected as well. Similar to R962, the second conformation was the RNA duplex. For the hairpin structure of R910, the additional signal was not expected as the NHC-G base pair should not be involved in base pairing, but rather is localized inside the loop region. Unfortunately, the spectrum of R911 showed more than one additional signal besides the hairpin structure signals (Figure *88*). Some of the detected signals were even more downfield shifted than U7 and U8 of the unmodified RNA duplex. A similar spectrum as for R962 was expected, but more species were formed, resulting in lower signal intensity, as the overall concentration of the formed species is reduced. The results of the $^{1}$H-NMR experiments corroborated with the results obtained from the gel shift assay (Figure 79). Placing the NHC modification into more internal positions excluding the modification from base pairing favored hairpin formation.

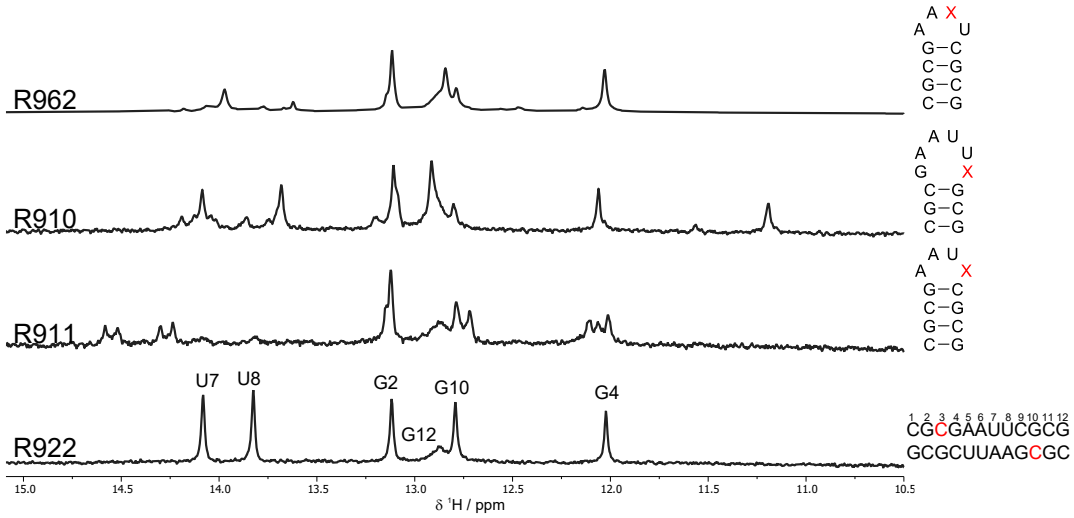



Figure 88. Overview of the Dickerson Drew sequences that form hairpin structures (R910, R911 and R962), measured at 283 K with 600 MHz, compared to the unmodified R922 (bottom spectrum). Sample concentration was 200 µM in $H_2O/D_2O$ (9:1) with 100 mM NaCl in 10 mM sodium phosphate buffer; X=NHC.

### 3.2.10 UV-Vis and CD-spectroscopy of the Dickerson Drew Sequences

UV-Vis absorption revealed no change in absorption maxima (wavelength) upon incorporation of the NHC modification (Figure 89). Only the relative absorption changed for the different dodecamers reflecting minor structural changes of the different sequences.

Furthermore, CD spectroscopy was used to investigate changes in the secondary structure of the dodecamer sequences. As mentioned earlier, it is a very sensitive method, not only to differentiate between helical structures, but also for conformational changes of the structure caused by changes in the orientation of adjacent base-pairs.[425, 467]

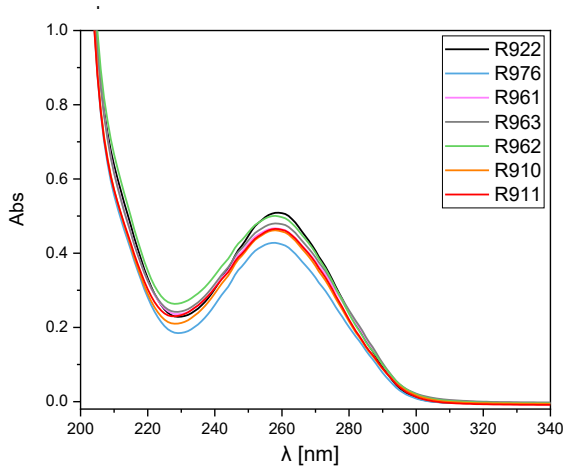



Figure 89. UV-vis spectra of all dodecamers (2 µM) in 10 mM sodium phosphate buffer containing 100 mM NaCl.

The structure of GU-wobble sequence R976 appeared to be considerably different from the unmodified as well as the NHC-modified ones (Figure 90).The shape of the CD spectrum between 230 and 280 nm deviated significantly from the canonical duplex structure, and was similar to CD spectra that have earlier been attributed to A-Z helix transition in RNA.[468-471] Thus, the wobble containing RNA seems to exhibit different base pairing and base stacking characteristics than the unmodified and NHC-modified RNAs, leading to a distinct CD-Signal between 230 and 280 nm.

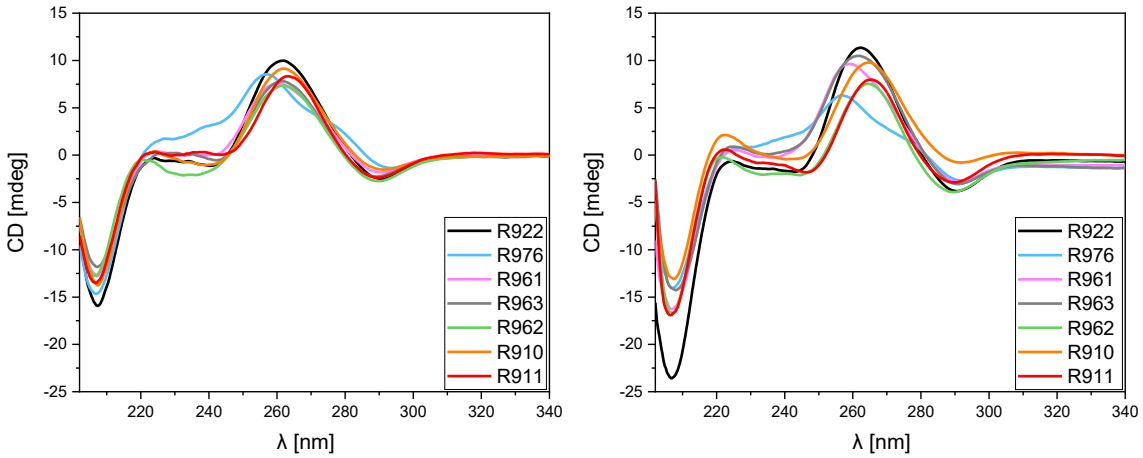



Figure 90. CD-spectra for dodecamers with 20 µM (left) and 2 µM (right) concentration. R922, R976, R961 and R963 were annealed for 2 min at 90 °C and subsequent cooling to room temperature.R962, R910 and R911 were annealed for 2 min at 95 °C and rapid cooling on ice. The samples were in 10 mM sodium phosphate buffer with 100 mM NaCl.

The overall shape of the CD spectra for the unmodified and the NHC-modified sequences were mainly in accordance with each other, but minimal changes were observed. The wavelength for the absorption maxima of the different modified dodecamers were slightly shifted compared to the unmodified one (Table 9). The duplex forming sequences R961, R963 and R976 were minimally blue shifted compared to the unmodified sequence R922. In contrast the wavelength maxima of the hairpin-forming sequences R962, R910, R911 were red shifted. This trend was enhanced for the 2 µM concentration compared to 20 µM samples (Figure 90). Additionally, R910, R911 and R962 displayed a more pronounced, readily blue shifted maximum between 215 and 230 nm. R911

and R962 were very similar with respect to the curve shape and ellipticity at 2 µM concentration, whereas R910 showed a higher ellipticity in consequence of the different length of the hairpin-stem (3 vs. 4 bp stem). The overall shape of R961 and R963 was comparable to the unmodified sequence R922 except for the shift of the wavelength maximum.

Table 9. Overview of the absorption maxima for the 20 µM and 2 µM samples in the CD-spectra. The shift of the wavelength maximum for each strand compared to R922 is given on the right.

| CD | 20 µM | | 2 µM | |
|---|---|---|---|---|
| | λ | Δλ | λ | Δλ |
| **R922** | 261.9 | 0 | 262.3 | 0 |
| **R976** | 257.1 | -4.8 | 256.8 | -5.5 |
| **R961** | 260.3 | -1.6 | 259.0 | -3.3 |
| **R963** | 261.3 | -0.6 | 261.7 | -0.6 |
| **R962** | 262.0 | +0.1 | 264.6 | +2.3 |
| **R910** | 262.1 | +0.2 | 264.7 | +2.4 |
| **R911** | 263.1 | +1.2 | 265.2 | +2.9 |

CD-spectroscopy confirmed the findings observed before in the NMR-spectra and melting curve analysis. Depending on the position of the NHC modification, either duplex or hairpin structures were formed, reflected in small changes in the shapes and ellipticities of the CD-spectra. The two tautomeric forms of NHC were base pairing with their respective counterpart but were less stable than a canonical base-pair.

### 3.2.11 $^{15}$N labeling of a mutated Dickerson Drew Sequence

To be able to analyze the NHC modification within an NHC-A base pair, the Dickerson Drew Sequence was further mutated. Similar to the wobble-containing RNA sequence, uridine was placed at position 3 and the respective guanosine at position 11 was replaced by adenosine resulting in a U-A base pair for the unmodified (R1006) and an NHC-A base pair for the modified sequence (R1007, Table 10). Due to the more terminal position of the modification, the duplex- should be favored compared to the hairpin-formation. The modified sequence was directly synthesized with the $^{15}$N labeled NHC phosphoramidite building block for the NMR studies.

Table 10. Further mutated Dickerson Drew sequence with U-A and NHC-A base pairs; X=NHC.

| Name | 5'-sequence-3' | nt | formula | Mass calc. | Mass found |
|---|---|---|---|---|---|
| **R1006** | CGUGAAUUCACG | 12 | $C_{114}H_{142}N_{45}O_{82}P_{11}$ | 3793.54386 | 3793.53462 |
| **R1007** | CGXGAAXUCACG | 12 | $C_{114}H_{143}N_{45}{}^{15}NO_{82}P_{11}$ | 3809.55180 | 3809.54283 |

Thermal denaturation experiments were performed to compare it to the melting temperatures of the duplex forming Dickerson Drew sequences. A summary of the obtained melting temperatures of these sequences is shown in Table 11 for 260 and 280 nm.

Table 11. Summary of the melting temperatures of the duplex forming sequences for comparison (X=NHC)

| $T_m$ | 260 nm | | | | | 280 nm | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **conc [µM]** | **20** | **10** | **5** | **2** | **1** | **20** | **10** | **5** | **2** | **1** |
| R922 (CGCGAAUUCGCG) | 67.2 | 65.3 | 63.4 | 60.7 | 59.0 | 68.4 | 67.0 | 65.9 | 64.5 | 64.1 |
| R976 (CGUGAAUUCGCG) | 49.0 | 47.9 | 46.3 | 44.4 | 43.1 | 49.0 | 47.9 | 46.5 | 44.7 | 43.5 |
| R961 (CGXGAAUUCGCG) | 51.1 | 49.4 | 48.1 | 46.6 | 45.5 | 51.3 | 49.8 | 48.5 | 47.2 | 46.2 |
| R1006 (CGUGAAUUCACG) | 59.1 | 57.6 | 56.5 | 54.1 | 52.7 | 59.2 | 57.9 | 56.8 | 55.0 | 53.9 |
| R1007 (CGXGAAUUCACG) | 43.8 | 42.0 | 40.6 | 38.7 | 37.0 | 43.5 | 41.8 | 40.5 | 38.6 | 37.0 |

As expected, both R1006 and R1007 sequences exhibited concentration dependent melting temperatures at 260 and 280 nm, reflecting the formation of duplex structures. R1006, the unmodified reference with the U-A base pair, revealed a melting temperature around 8 °C lower than the respective unmodified C-G duplex R922, but a 10 °C higher melting temperature than the G-U wobble sequence. Introduction of the modification (R1007, 5 µM) led to a similar decrease of the melting temperature compared to R1006, as it was the case for R961 compared to R922 with around 15.9 °C decrease at 260 nm and 16.3 °C decrease at 280 nm (5 µM). The melting curves of R1006 and R1007 are shown in Figure 91 in comparison with the other duplex forming Dickerson Drew sequences R922, R976 and R961.

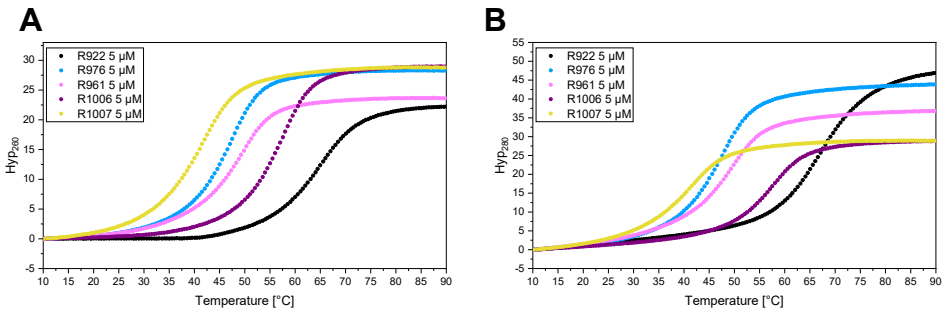



Figure 91. Melting curves of the duplex forming Dickerson Drew sequences R922 (black), R961 (magenta), R976 (blue), R1006 (purple), and R1007 (yellow) at A: 260 nm, and B: 280 nm.

In contrast to R922, the hyperchromicity stayed almost unchanged for R1006 and R1007 between 260 and 280 nm. Clearly, there was a difference between the two unmodified sequences, as no significant change was observed in the thermodynamic data (Table 12) for the two wavelengths.

Table 12. Summary of the thermodynamic data of the duplex forming Dickerson Drew sequences with X3 or U3 at 260 and 280 nm; X=NHC.

| **Thermodynamic data 260/280 nm** | $\Delta H_{260}$ **kcal/mol** | $\Delta S_{260}$ **cal/mol·K** | $\Delta G_{260}$ **kcal/mol** | $\Delta H_{280}$ **kcal/mol** | $\Delta S_{280}$ **cal/mol·K** | $\Delta G_{280}$ **kcal/mol** |
|---|---|---|---|---|---|---|
| R922 (CGCGAAUUCGCG) | -81.1 | -215.5 | -16.9 | -155.7 | -433.2 | -26.6 |
| R976 (CGUGAAUUCGCG) | -101.9 | -293.8 | -14.4 | -110.6 | -320.2 | -15.2 |
| R961 (CGXGAAUUCGCG) | -112.5 | -324.5 | -15.8 | -124.1 | -359.7 | -16.9 |
| R1006 (CGUGAAUUCACG) | -102.2 | -285.6 | -17.1 | -123.9 | -350.3 | -19.5 |
| R1007 (CGXGAAUUCACG) | -89.1 | -258.4 | -12.1 | -92.7 | -269.8 | -12.3 |

At 260 nm, R1006 showed comparable enthalpic and entropic values as the NHC-G and U-G sequences but had an almost equal free energy compared to the unmodified G-C sequence (Table 12). Interestingly, R1006 does not behave like R922 by changing from 260 to 280 nm, since the thermodynamic values were not decreasing as significantly as for the C-G RNA. The modified R1007, which revealed the lowest melting temperature of all the duplex forming Dickerson Drew sequences had the lowest free energy.

### 3.2.12 NMR analysis of the NHC-A containing Dickerson Drew sequences

Since both R1006 and R1007 were shown to form duplex structures, $^{1}$H NMR spectra were recorded to get more information about the NHC-A base pair. The imino regions are presented in Figure 92 in comparison to the unmodified R922, the modified R961, and the U-G containing R976.

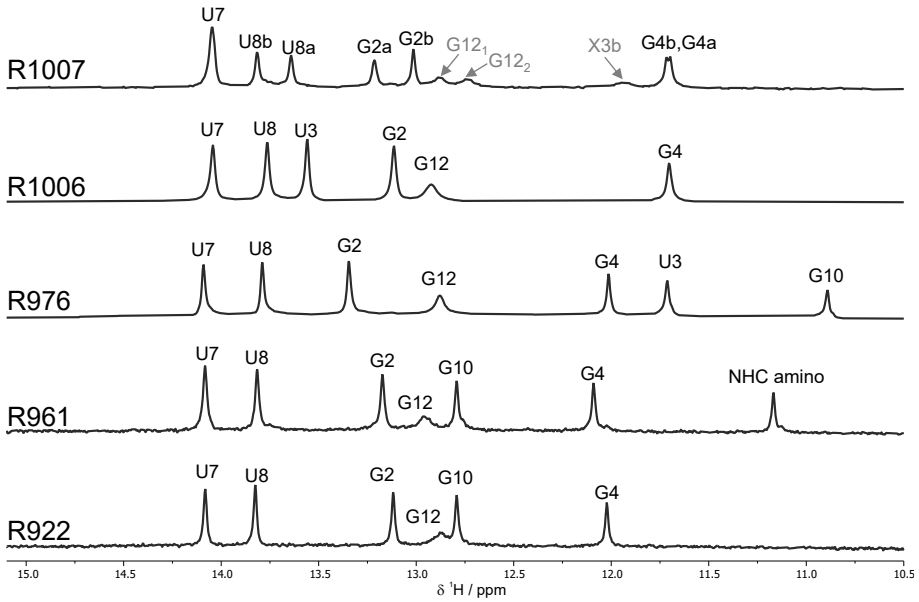



Figure 92. $^{1}$H NMR. Imino region of the duplex forming Dickerson Drew Sequences R922 (C-G, bottom line), R961 (NHC-G), R976 (U-G), and the new data of R1006 (U-A) and R1007 (NHC-A, top line). Three broad signals with low intensity were highlighted with arrows (grey); X=NHC.

The obtained signals of R1006 and R1007 were assigned with the help of 2D spectra. Six signals were detected for the unmodified R1006 RNA duplex with U-A base pair, which were in accordance with the unmodified sequence containing a C-G base pair at position 3. U7, U8, G2, and G12 of R1006 were found in almost the same spectral range as for R922. G4 was slightly shifted from around 12 ppm to 11.7 ppm for the U-A base pair. Interestingly, U3, which was found at 11.7 ppm for the U-G wobble of R976, was detected at around 13.6 ppm in the U-A base pair of R1006. These two base pairs could readily be distinguished.

For the NHC modified sequence R1007, more than the six expected signals were observed in the imino region; some of them appeared as sharp peaks, whereas others only exhibited a broad signal with low intensity (marked with arrows, Figure 92). The broad signals were the first to disappear when heated from 283 to 298 K (Figure 93).

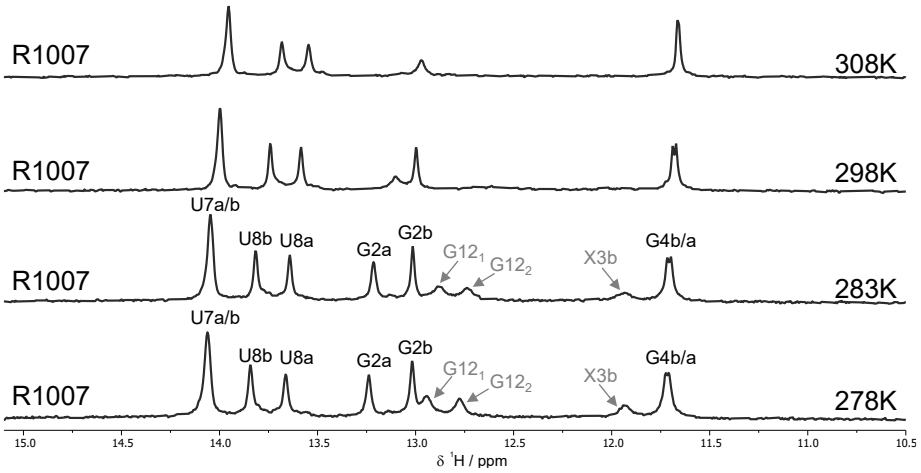



Figure 93. Imino region of the $^1$H NMR of R1007 at four different temperatures, 278 K, 283 K, 298 K, and 308 K. The broad exchangeable imino protons at 12.8 and 13.0 ppm were assigned to two different conformations of G12 H1 of the terminal base pair and the signal at 11.9 ppm to the NHC (X) modification.

A conformational equilibrium was detected for the NHC-A dodecamer, which was proven by the evaluation of the 2D NOESY spectrum. The two conformations with almost equal distribution were named ‘a’ and ‘b’ hereafter. Considering the NMR time scale, the two conformations exhibited a slow exchange, causing distinct exchange peaks for U8 and G2 and some aromatic protons (G2, X3, G4, and A10). For the imino proton of G4, it was only slightly visible by a splitting of the signal. Whereas for U7, the signal for the two conformations coincided. This slow exchange made it possible to assign the sharp peaks of the two conformations (Figure 92) by two distinguishable walks in the NOESY of R1007 at 283K (see 6.2.5). The assignment of the broad signals with low intensity were proposed by Dr. Irene Bessi, based on the obtained NOESY cross peaks in the imino-aromatic/amino region (278K). The signals at 12.8 and 13.0 ppm appeared to belong to H1 of two different conformations of the terminal G12-C1 base pair. Based on a cross peak with A10b, the signal at 11.9 ppm was assigned to the NHC modification (X3b, H3) at position 3. As mentioned earlier, Fazakerley *et al.* analyzed the base pairing behavior of N$^4$OMe opposite to A and G.[353] They reported a similar downfield shift of the N4 amino signal (see chapter 3.2.7) as it appeared for the NHC amino signal in the NHC-G base pair (Figure 80). For the N$^4$OMe-A base pair in the DNA heptamer, they detected the imino proton of the modification at 12.05 ppm, which was in the same chemical shift region as the imino proton proposed to be X3b in the NHC-A base pair (R1007) of the Dickerson Drew sequence. According to the chemical shift perturbation analysis (CSP, see 6.2.7) of the NHC-A protons in relation to the unmodified U-A duplex conformation ‘a’ seemed to be more perturbed, especially in the region of C9-A10-C11 than conformation ‘b’. For both conformations, the G2-C3-G4 region was perturbed as well, due to the substitution of the uridine with the NHC modification. After evaluation of all the collected data, we presumed, that NHC might be present in the anti imino form in conformation ‘b’ forming a WC base pair with A. The conformation ‘a’ still needed to be elucidated.

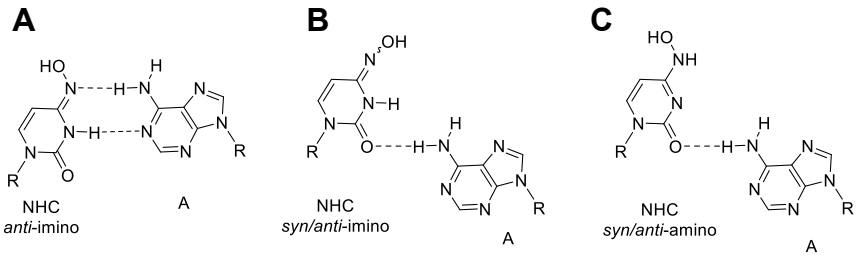



Figure 94. Possible base pairs between the NHC modification and adenosine. A: WC geometry. B: anti or syn NHC:A wobble base pair. C: amino NHC:A wobble base pair.

Stone *et al.* performed NMR analysis with a similar sequence, a DNA octamer 5'-CGAATTCG-3'.[355] By replacing T at position 6 with the $N^4$OMe modification they obtained the $N^4$OMe-A base pair. More than one set of signals for the expected octamer duplex was detected with different intensities caused by a fast equilibrium between the Watson-Crick and the wobble conformation of the $N^4$OMe-A base pair. The predominant structure was found to be the Watson-Crick conformation, with the modification being present in the *anti* imino form. Temporarily, the A3 nucleobase opposite to the modification moved towards the minor grove, facilitating wobble base pair formation. Accordingly, these structures are theoretically also possible for the NHC-A base pair. A selection of possible conformations is shown in Figure 94. The determination of the structure of conformation 'a' was difficult and could only be hypothesized with the available data by excluding some of the proposed structures (Figure 94). The imino signal for the NHC-A of conformation 'a' could not be detected, presumably due to even broader signals than conformation 'b' indicated with arrows (Figure 93), because no imine is involved in base pairing. In general, the presence of NHC in the amino form could be excluded as the experiments shown in Figure 85 were repeated with the $^{15}$N labeled R1007. No splitting of the signal could be observed by $^{15}$N decoupling, as well as no signal could be detected in the imino region using the $^{15}$N filter. Thus, no amino proton directly coupled to the $^{15}$N was forming a stable hydrogen bond. The occasional movement of the adenosine opposite to the $N^4$OMe modification in the DNA octamer described earlier, resulting in wobble base pair formation, would be in line with the observed perturbance of the region C9-A10-C11 for conformation 'a' found in the CSP. A10 of the Dickerson Drew sequence could be shifted to the minor groove as well, maybe because NHC adopted the *syn* conformation leading to steric repulsion. Only a weak base pair between the exocyclic amino group $N^6$-H61 (A10) and the O2 of NHC could be formed. Thus, structure B and C in Figure 94 would be possible. Even though the second conformation could not be unambiguously assigned, we could provide evidence that NHC is indeed able to form a Watson-Crick base pair with A: NHC was adopting the imino form with the hydroxy group *anti* to N3. This conformation was in slow exchange with another almost equally populated species, whose conformation could not be precisely determined. The imino exchange rates for the new Dickerson Drew sequences with the U-A and NHC-A ('b') base pairs were determined at 298K as well. They were plotted in comparison with the G-U wobble containing RNA R976 (Figure 95).

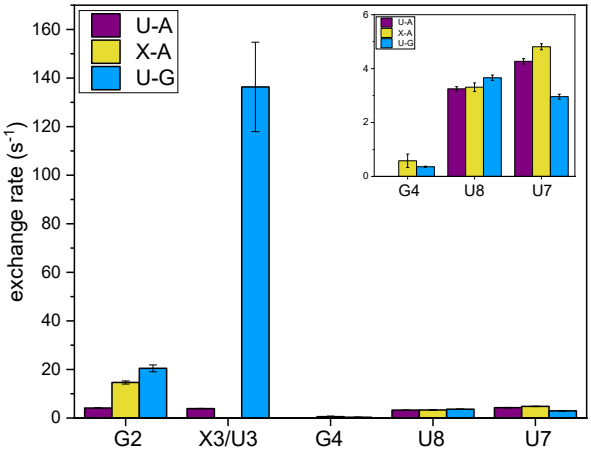



Figure 95. Imino-water exchange rates of the signals obtain in the $^{1}$H imino region at 298K for R1006 (purple) and R1007 (yellow, conformation b) in comparison with the wobble-containing sequence R976 (blue); X=NHC.

The imino signal of X3b was not detectable anymore at 298 K and the imino exchange rate for this signal could not be determined (see 6.2.6). This base pair is, in line with the melting temperatures (Table 11), destabilized compared to the unmodified reference (R1006). However, compared to R961 with the NHC-G base pair, the adjacent position G2 is less affected ($k_{EX}$: 43.7 vs. 14.6).

### 3.2.13 Crystallography

Besides NMR, it was also tried to solve the structure of the different Dickerson Drew sequences by crystallography. R922 (C-G), R961 (NHC-G) and R962 (NHC-A) were chosen for setting up the crystallization experiments. The initial screening experiments were performed with the HELIX™ screen using the sitting drop vapor-diffusion method (Figure 96A) on a 96-well plate to find conditions for successful crystal growth (“hits”). The “hits” found in the sitting drop method can be further fine-tuned by changing different parameters, like the pH value or the composition of the crystallization buffer (precipitant). The actual experiments were set up using the hanging-drop vapor-diffusion method. A drop containing a 1:1 mixture of the sample and the reagent (crystallization buffer) was applied on a siliconized coverslip and placed on top of a reservoir of the reagent (Figure 96B).[472] By forming an equilibrium between the drop, which contains a lower reagent concentration, and the reservoir, water vapor passes over from the drop to the reservoir. In the equilibrium, both drop, and reservoir have approximately the same concentration, and the sample inside the drop gets relatively supersaturated. Multiple droplets containing different samples can be applied to one reservoir. After the formation of crystals, they are easily accessible by taking off the coverslip (Figure 96C).

The obtained crystals were picked from the selective drops by special loops, cryoprotected (reduce the amount of water) using perfluoropolyether (PFPE) oil and stored in liquid nitrogen. The X-ray diffraction data were collected at 100 K on P11 (DESY) beamlines. The evaluation of the high-resolution diffraction data is currently ongoing. For R961 and R962 this procedure was performed

as well, using the crystallization buffers from the first round of hanging drop approach. Less crystals were growing and the ones picked for DESY measurement did not provide high resolution during X-ray diffraction measurements.

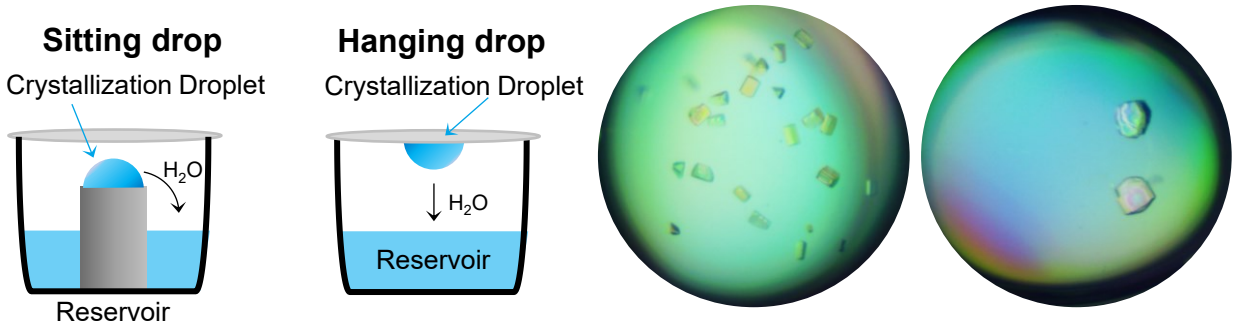



Figure 96. Common techniques to obtain crystals by vapor diffusion. A: sitting drop B: hanging drop. Water vapor is leaving to the reservoir until an equilibrium between the drop and the reservoir is achieved. C: obtained crystals for R911 and R922, respectively, with the hanging drop method

Besides the Dickerson Drew sequence, the NHC modification was also analyzed in another sequence context that reflects different stages of viral replication.

### 3.2.14 NHC in another sequence context

To determine the mechanism of action, a selection of RNA oligonucleotides was synthesized containing the NHC modification with different unmodified counterparts (Table 13) to create the NHC-G and NHC-A base pairs at different positions within an RNA strand (see 3.2.3).[439] They were further used to investigate the NHC-G and NHC-A base pairs in a different sequence context than Dickerson Drew.

Table 13. Synthesized sequences with the corresponding high-resolution ESI-MS-Data; X=NHC.

| Name | 5'-sequence-3' | nt | formula | Mass calc. | Mass found |
|---|---|---|---|---|---|
| **R840** | CACUGCGUAGGCUCA | 15 | $C_{142}H_{178}N_{56}O_{103}P_{14}$ | 4748.67387 | 4748.67166 |
| **R841** | UGAGCCUACGCAGUG | 15 | $C_{143}H_{178}N_{58}O_{103}P_{14}$ | 4788.68002 | 4788.66430 |
| **R880** | UGAGCCUACGCAAUG | 15 | $C_{143}H_{178}N_{58}O_{102}P_{14}$ | 4772.68510 | 4772.66377 |
| **R892** | CAXUGCGUAGGCUCA | 15 | $C_{142}H_{178}N_{56}O_{104}P_{14}$ | 4764.66878 | 4764.67966 |
| **R894** | UGAGXCUACGCAGUG | 15 | $C_{143}H_{178}N_{58}O_{104}P_{14}$ | 4804.67493 | 4804.66837 |
| **R917** | CACUGCGUAGG | 11 | $C_{105}H_{131}N_{43}O_{75}P_{10}$ | 3503.51347 | 3503.51600 |
| **R918** | CACUGCGUAGA | 11 | $C_{105}H_{131}N_{43}O_{74}P_{10}$ | 3487.51856 | 3487.51528 |
| **R921** | CACUGCGUAGACUCA | 15 | $C_{142}H_{178}N_{56}O_{102}P_{14}$ | 4732.67895 | 4732.67155 |
| **R945** | UGAGUCUACGCAGUG | 15 | $C_{143}H_{177}N_{57}O_{104}P_{14}$ | 4789.66403 | 4789.65330 |
| **R947** | CAUUGCGUAGGCUCA | 15 | $C_{142}H_{177}N_{55}O_{104}P_{14}$ | 4749.65843 | 4749.65855 |

NHC was either incorporated at position 3 or 5 of a 15mer to resemble different stages of viral replication and to investigate duplex stability (Table 14). For the 15mer with the modification at position 5, a full-length and a shorter complementary strand was tested that resembled the RNA strand right after the incorporation of one nucleotide opposite to position five.

Table 14. Summary of the different sequence contexts, always with CG, XG, UA, XA and UG version, X=NHC.

| RNA | Sequence | RNA | Sequence | RNA | Sequence |
|---|---|---|---|---|---|
| R841<br>R917 | 5'-UGAG**C**CUACGCAGUG-3'<br>3'-**G**GAUGCGUCAC-5' | R841<br>R840 | 5'-UGAG**C**CUACGCAGUG-3'<br>3'-ACUC**G**GAUGCGUCAC-5' | R841<br>R840 | 5'-UGAGCCUACGCA**G**UG-3'<br>3'-ACUCGGAUGCGU**C**AC-5' |
| R894<br>R917 | 5'-UGAG**X**CUACGCAGUG-3'<br>3'-**G**GAUGCGUCAC-5' | R894<br>R840 | 5'-UGAG**X**CUACGCAGUG-3'<br>3'-ACUC**G**GAUGCGUCAC-5' | R841<br>R892 | 5'-UGAGCCUACGCA**G**UG-3'<br>3'-ACUCGGAUGCGU**X**AC-5' |
| R945<br>R918 | 5'-UGAG**U**CUACGCAGUG-3'<br>3'-**A**GAUGCGUCAC-5' | R945<br>R921 | 5'-UGAG**U**CUACGCAGUG-3'<br>3'-ACUC**A**GAUGCGUCAC-5' | R880<br>R947 | 5'-UGAGCCUACGCA**A**UG-3'<br>3'-ACUCGGAUGCGU**U**AC-5' |
| R894<br>R918 | 5'-UGAG**X**CUACGCAGUG-3'<br>3'-**A**GAUGCGUCAC-5' | R894<br>R921 | 5'-UGAG**X**CUACGCAGUG-3'<br>3'-ACUC**A**GAUGCGUCAC-5' | R880<br>R892 | 5'-UGAGCCUACGCA**A**UG -3'<br>3'-ACUCGGAUGCGU**X**AC-5' |
| R945<br>R917 | 5'-UGAG**U**CUACGCAGUG-3'<br>3'-**G**GAUGCGUCAC-5' | R945<br>R840 | 5'-UGAG**U**CUACGCAGUG-3'<br>3'-ACUC**G**GAUGCGUCAC-5' | R841<br>R947 | 5'-UGAGCCUACGCA**G**UG-3'<br>3'-ACUCGGAUGCGU**U**AC-5' |

For the 15 sequences shown in Table 14 thermal denaturation experiments were performed using five concentrations and measuring two heating and two cooling ramps (Figure 97). A sigmoidal curve shape was obtained for all of them.

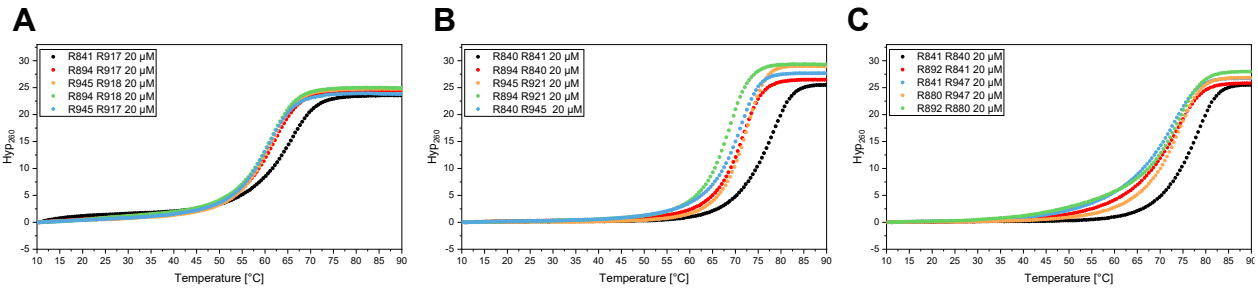



Figure 97. Overview of the different sequences, displaying different stages of viral replication with NHC inside the template and the corresponding melting curves, with 20 µM concentration each. A: mimics the growing strand during viral replication, where opposite to the NHC containing template ATP or GTP was just incorporated. B: Fully elongated strand. C: Template strand, with a more terminal NHC position. All of them are shown in comparison with the unmodified and G-U-containing sequence.

One part of this selection was a 15mer base pairing with a shorter 11mer RNA oligonucleotide (Table 13). This should mimic virus replication with the modification inside the template strand (Figure 97). More precisely, it showed the stage directly after incorporation opposite NHC. For comparison, the two unmodified-, and G-U wobble-sequences were also synthesized. The modification opposite to G showed a -3.3 °C lower melting temperature compared to the unmodified C-G sequence (Table 15A, 20 µM), whereas NHC-A showed only a 0.1 °C lower melting temperature then U-A. Both sequences, NHC-G and NHC-A showed the same value for the free energy, being slightly higher than C-G ($\Delta G$= +1.9 kcal/mol). On the other hand, their free energies were slightly lower than the unmodified U-A ($\Delta G$= -1.6 kcal/mol). Thus, destabilization of the duplex structure inside the active site of the RdRp should be rather small upon incorporation of either the correct or incorrect triphosphate (GTP or ATP) opposite to NHC. Apparently, the destabilization of the NHC-ATP base pair seemed to be lower compared to the NHC-GTP base pair. Since the imino form, needed to form NHC-ATP, is energetically more favorable,[454] this might result in energetically bet-

ter base pairing abilities in the active center. Still, the melting temperatures of both modified sequences are comparable, resulting in accumulations of mutations by forming stable base-pairs with either ATP or GTP. The wobble-sequence showed the lowest melting temperature and free energy of this group. The next sequence corresponded to the fully elongated duplex structure after incorporation of ATP and GTP opposite NHC at position 5 (Figure 97B). As soon as the modification was located more towards the center of the full-length sequence, it had a higher impact onto duplex destabilization. The melting temperature of the NHC-G sequence decreased by 6.4 °C compared to C-G. The melting temperature of the NHC-A sequence was reduced too, but only by 4.1 °C compared to the unmodified U-A-sequence (Table 15B, 20 µM). The NHC-G and NHC-A modified sequences were less enthalpically favored with Δ*H*= + 22.2 and 9.4 kcal/mol, respectively compared to C-G and resulted in a lower free energy value by around 5.5 kcal/mol. In contrast, if the modification was on a more terminal position (Figure 97C), the decrease was less pronounced with $\Delta T_m$ = -4.3 °C for NHC-G and $\Delta T_m$ = -0.1 °C for NHC-A compared to the respective unmodified reference (Table 15, 20 µM). The latter revealed similar values than the example with the shorter complementary strand (Figure 97A).

Table 15. Melting temperatures at five concentrations and thermodynamic data of the sequences with NHC (X) in the template-like strand in comparison with the respective unmodified reference and the G-U containing sequence. A: 15mer with 11mer complementary strand. B: substitutions at position 5. C: substitutions at position 3.

| | | **20** | **10** | **5** | **2** | **1** | **$\Delta T_m$** | **Δ*H*** **kcal/mol** | **Δ*S*** **cal/mol·K** | **Δ*G*** **kcal/mol** |
|---|---|---|---|---|---|---|---|---|---|---|
| A | R841 R917 (C/G) | 64.7 | 63.3 | 62.1 | 60.3 | 59.1 | 5.6 | -121.3 | -335.7 | -21.3 |
| | R894 R917 (X/G) | 61.4 | 59.9 | 58.6 | 57.0 | 55.5 | 5.9 | -114.5 | -319.1 | -19.4 |
| | R945 R918 (U/A) | 60.9 | 59.3 | 57.8 | 56.0 | 54.4 | 6.5 | -102.7 | -284.8 | -17.8 |
| | R894 R918 (X/A) | 60.8 | 59.1 | 58.1 | 56.4 | 55.0 | 5.8 | -116.3 | -325.1 | -19.4 |
| | R945 R917 (U/G) | 60.3 | 58.5 | 57.1 | 55.2 | 53.5 | 6.8 | -99.8 | -276.6 | -17.3 |
| B | R841 R840 (C/G) | 77.6 | 75.9 | 75.0 | 74.1 | 73.5 | 4.1 | -189.1 | -517.4 | -35.0 |
| | R894 R840 (X/G) | 71.2 | 70.1 | 69.0 | 68.0 | 66.9 | 4.3 | -166.9 | -461.1 | -29.4 |
| | R945 R921 (U/A) | 72.3 | 71.0 | - | 69.2 | 67.7 | 4.6 | -163.8 | -450.9 | -29.4 |
| | R894 R921 (X/A) | 68.2 | 67.3 | 66.5 | 65.2 | 64.4 | 3.8 | -179.7 | -504.1 | -29.5 |
| | R945 R840 (U/G) | 70.5 | 69.3 | 68.0 | 66.5 | 65.8 | 4.7 | -145.5 | -400.4 | -26.2 |
| C | R840 R841 (C/G) | 77.6 | 75.9 | 75.0 | 74.1 | 73.5 | 4.1 | -189.1 | -517.4 | -35.0 |
| | R841 R892 (X/G) | 73.3 | 71.8 | 70.4 | 69.2 | 68.5 | 4.8 | -150.5 | -412.7 | -27.5 |
| | R880 R947 (U/A) | 73.9 | 72.4 | 71.2 | 69.5 | 68.2 | 5.7 | -126.9 | -342.3 | -24.9 |
| | R880 R892 (X/A) | 73.8 | 72.4 | 70.7 | 69.4 | 67.4 | 6.4 | -115.0 | -308.8 | -22.9 |
| | R841 R947 (U/G) | 72.4 | 70.1 | 69.0 | 67.4 | 66.2 | 6.2 | -118.4 | -320.6 | -22.9 |

Consequently, changing from the incorporation of GTP or ATP opposite to NHC towards a sequence with the modification being at a more central position (= full elongation), the destabilizing effect of the modification increased. Furthermore, the destabilization of the NHC-G base pairs was

in all experimental cases larger than for NHC-A, indicating that the imino form is more favorable compared to the amino form.

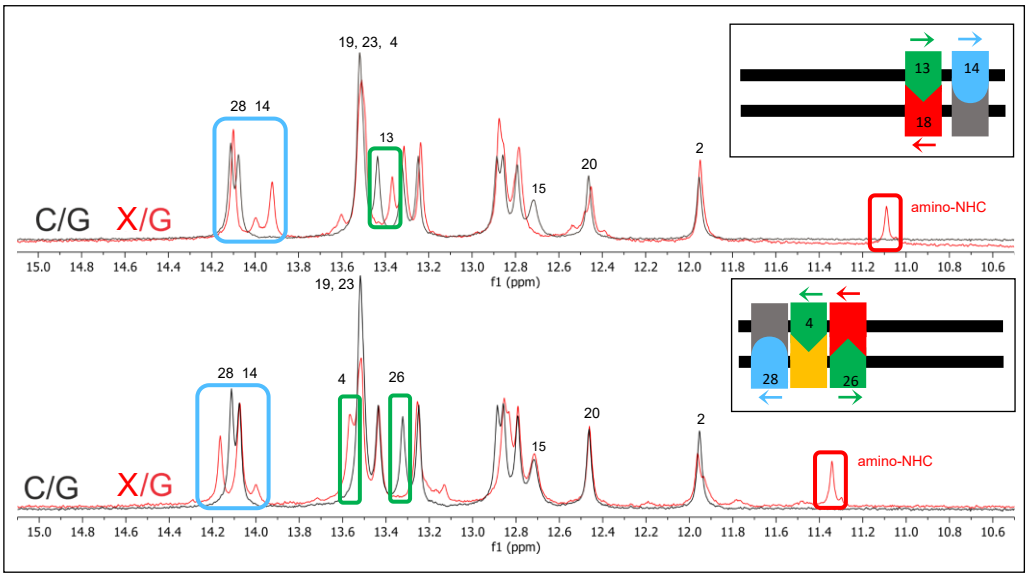



Figure 98. $^{1}$H NMR spectra for comparison of C-G and NHC-G 15mers with the modification at position 3 (top line, red, 18) or at position 5 (bottom line, red). The neighboring nucleotides are partially shifted (direction of the shift indicated by arrows), whereas most signals are in line with the unmodified references (black). Green: G, yellow: C, blue: U, grey: A, red: NHC; X=NHC.

For the two 15mer full-length duplex strands (Figure 97B/C), $^{1}$H-NMR spectra were recorded using around 200 µM duplex concentration in 10 mM sodium phosphate buffer (pH 7) containing 100 mM NaCl and DSS as internal standard. The assignment of the unmodified sequences was only based on predictions and was not confirmed by 2D assignment.[446] It was then tried to adapt this predicted assignment to the modified sequences. The sequences with the substitutions at position 3 are shown in the top spectrum and at position 5 in the bottom spectrum (Figure 98). By comparing the $^{1}$H-NMR spectra of the two NHC-G-containing sequences, a clear shift of three signals was observed relative to the unmodified references. In the top spectrum, the signal of the guanosine opposite, as well as one uridine signal adjacent to the modification were shifted (Figure 98, indicated with arrows). Furthermore, as for the Dickerson Drew Sequence an additional signal at 11.1 ppm appeared, which should correspond to the amino proton of NHC as well. The signal for the terminal C-G (G15) base pair disappeared, due to partial destabilization of the duplex structure by the NHC modification in proximity to the 3'-terminal end. For the 15mer with the modification at position 5, the guanosine opposite to the modification, the adjacent G4 (only slightly) and U28 were shifted. The amino signal of NHC exhibited a downfield shift as expected to 11.35 ppm. In contrast to the first example, the signal for the terminal C-G base pair was still visible due to the longer distance to NHC. The other signals were in accordance with the unmodified sequence. Thus, the modification in the G-NHC base pair, which decreased the stability of the duplex as shown with the obtained melting temperatures (Table 15), had a mild effect on the adjacent nucleotides. The signals of the

neighboring nucleotides are slightly shifted, while the overall structure was in accordance with the unmodified references.

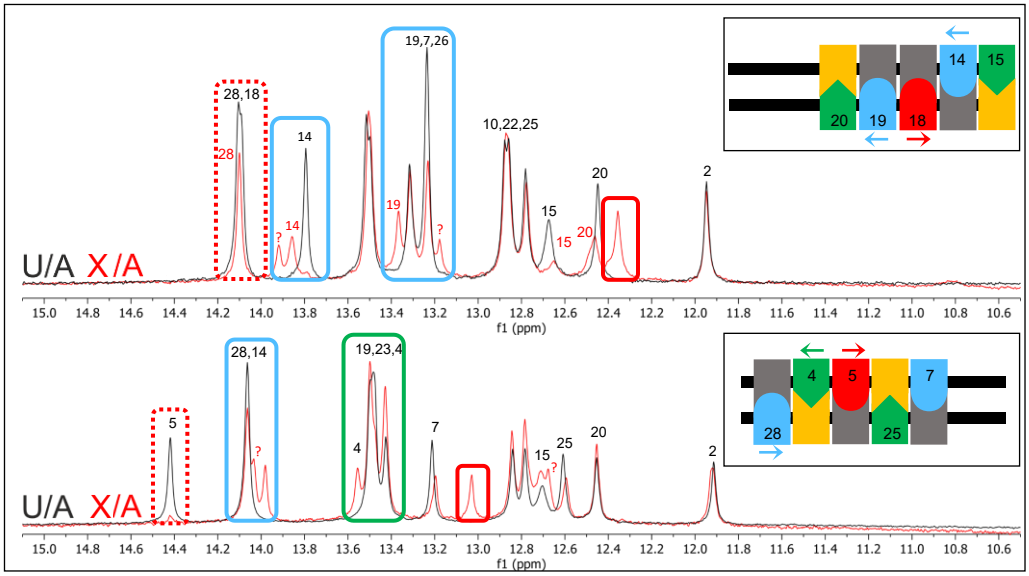



Figure 99. Comparison of U-A and NHC-A 15mers with the modification at position 3 (top line, red, 18) or at position 5 (bottom line, red) in comparison with the respective unmodified reference. The signal for the unmodified uridine which was replaced by NHC (box with dashed lines), disappeared. Instead, multiple new signals were present in the modified NMRs; X=NHC.

For the NHC-A spectra, the changes for the replaced uridine (U18 and U5) were more obvious since NHC is a cytidine derivative, but the base pairing ability is equal to uridine (Figure 99). The respective signals at 14.1 and 14.4 disappeared compared to the unmodified one, and a new major signal at 12.35 ppm and 13.0 ppm appeared, respectively. As these signals were offset to the other signals obtained for the unmodified sequence, they might belong to the imino-NHC. However, besides this major peak, the signals for a second conformation were detected as well (question marks in Figure 99), making the predicted assignment of the NHC-containing sequence challenging. The trend of the adjacent base pairs to be affected as shown for the NHC-G structures in Figure 98 seemed to remain. The signals of U19 and U14 adjacent and diagonally opposite to NHC (position 18, top line) and U28 and G4 (position 5, bottom line) were affected (Figure 99). The other signals mostly coincide with the unmodified references. In case the signals at 12.35 ppm and 13.0 ppm indeed belonged to NHC, it appears that the shift of the NHC signal was dependent on its sequence context. In both cases, NHC-G and NHC-A, the new signals were more upfield shifted to 11.35 and 13.0 ppm in a U-A context (Figure 98 and Figure 99 bottom line), whereas more downfield-shifted to 11.1 and 12.35 ppm in a C-G context (Figure 98 and Figure 99 top line). However, for the Dickerson Drew Sequence R1007, the imino proton of the NHC-A base pair was assigned to 11.9 ppm. There, the NHC-A pair was present in a G-C context. Consistently, a second conformation was detected for the Dickerson Drew sequence as well, presumably caused by the *syn*-imino form of NHC.

The collected data after the RNA elongation assays and cryo-EM structure analyses could give some more insights into the base pairing behavior of the NHC modification. By thermal denaturation experiments and NMR spectroscopy, the formation of the G-NHC base pair was demonstrated. The additional hydroxy group created a unique environment compared to a canonical C-G base pair, leading to an exceptional downfield shift of the amino signal to 11.2 ppm. Even though the incorporation of NHC into a duplex strand led to a decrease in stability, NHC-G was forming a stable base pair with NHC in the *anti*-amino form. While base pairing with adenosine, NHC presumably adopts the imino form with the hydroxy group *anti* to N3. This conformation was in slow exchange with another almost equally populated species. This might arise due to the *syn*-conformation of the imino form, forcing the A10 opposite to NHC to move to the minor groove.

## 3.3 deaza Guanosine Crosslinker for DNA-Protein interactions

The laboratory of Prof. Dr. Christian Haering and his coworker Julia Locherer investigate the protein complex Condensin, which is part of the structural maintenance of chromosomes (SMC) protein family.[473] These ATPases play a central role in the structural maintenance of the chromosomes (segregation during mitosis and meiosis) as well as DNA repair.[473, 474] In eukaryotes, they are forming three pairs of heterodimers, each of them for a respective function:[475] SMC1 and SMC3 are the basis of the cohesion complex required for cohesion during mitosis (sister chromatid cohesion);[476 477] SMC2 and SMC4 are the basis of the condensin complex, needed for chromosome condensation, to form compact structures of chromatin fibers.[478, 479] The heterodimer of SMC5 and SMC6 was found to be involved in DNA repair.[480] The general architecture of a SMC protein is shown in Figure 100A. These proteins are large polypeptides, each consisting of around 1000 – 1300 amino acids, which form different domains.[474] The overall architecture resembles a V-shaped molecule with two long coiled-coil arms (rod-like) emerging by an antiparallel self-folding of the SMC sequence. Each arm has a length of around 50 nm. At each end of the SMC molecule, a N-terminal and a C-terminal domain are located, which are brought into contact by the coiled-coil interactions. Walker A and Walker B motifs, two nucleotide-binding motifs, are placed at the N- and C-terminal domain. This end is called the "head" and represents the ATP-binding domain, whereas the other end is called "hinge". The C motif is not involved in ATP binding, but is needed for the ATP hydrolysis.[474] The heterodimer is formed by association of the two subunits at the hinge domain (hinge-hinge interaction).[474]

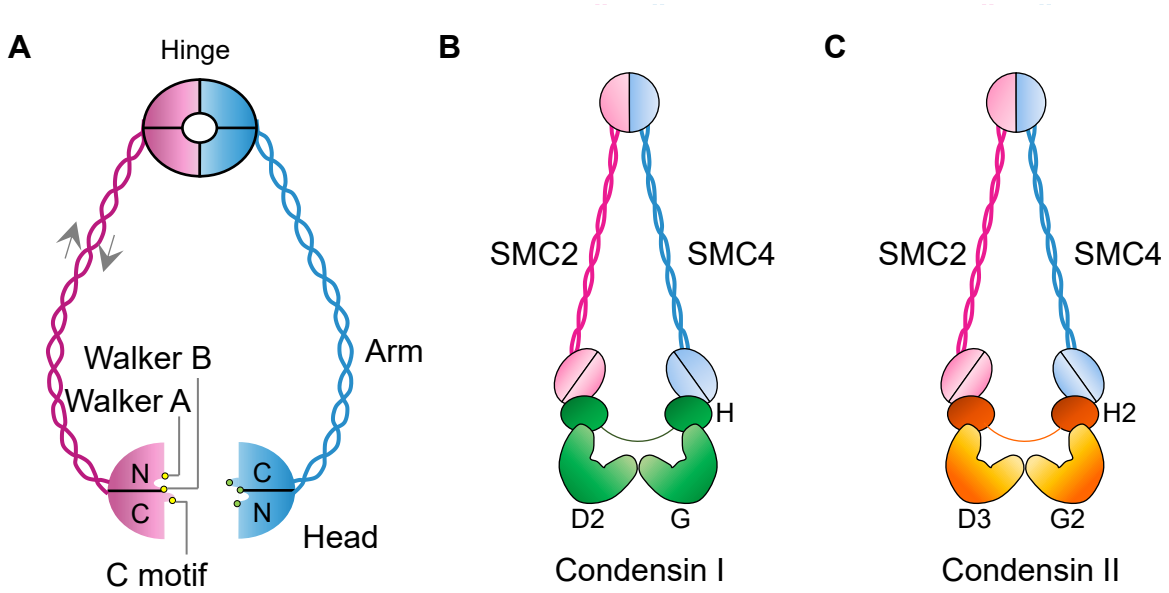



Figure 100. A: Schematic structure of a SMC heterodimer. The subunits (magenta, blue) form antiparallel coiled-coil interactions by self-folding.[474] The antiparallel composition is indicated with arrows. Through this assembly a hinge domain, as well as an ATP binding head domain is formed at both ends of the coils. The hinge interfaces of the two subunits create a V-shaped complex.[474] The figure was reproduced from Ref.[474] B: Schematic structure of condensin I with the Heat repeat subunits D2 and G (green), and the kleisin subunit H (dark green). C: Schematic structure of condensin II with the heat repeat subunit D3 and G2 (orange), and the kleisin subunit H2 (dark orange). Figure B,C based on Ref.[481]

Condensin is involved in mitotic chromosome assembly and segregation.[481] In eukaryotes, two types of condensin complexes are present – condensin I and condensin II, built up by 5 subunits (Figure 100B,C). Both possess SMC2 and SMC4 as core subunits, but they are accompanied by a different set of regulatory (non-SMC) subunits. Condensin I contains CAP-D2 and CAP-G, and condensin II CAP-D3 and CASP-2, which are HEAT repeat subunits.[481] The third regulatory subunit is part of the kleisin family. The member contained in condensin I is called CAP-H, and CAP-H2 for condensin II.[481] These regulatory subunit are binding to the HEAD domain of condensin I and II as a subcomplex.

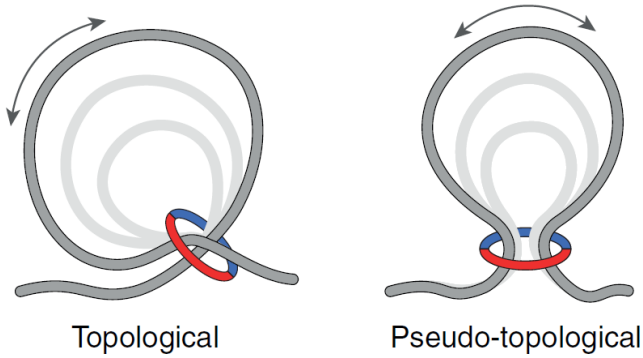



Figure 101. The nucleobases could be captures by the ring-shaped SMC complex in a topological or pseudo-topological manner to create and further enlarge DNA loops (loop extrusion). For the topological process, the ring would need to open for the interaction with the DNA strand, in contrast to the pseudo-topological example. Figure taken with permission from Ref.[482]

The proposed mechanism for the chromosome assembly by the SMC protein complexes is the performance of loop extrusion: They create and progressively enlarge loops of DNA.[482] This would bring different regions of a DNA strand in proximity with each other to either activate a specific function or to fold the chromatin fibers into the special rod-shaped chromosomes.[482] The SMC heterodimers could act as a ring-like structure and capture the nucleobases to form DNA loops in two distinct ways (Figure 101): In the topological mechanism, the SMC ring would need to open for the DNA to enter and exit, which would not be the case for the pseudo-topological process.[482] Different hypotheses about the precise process of the loop extrusion have been made, like the extrusion by external motors, or by osmotic pressure;[482] The continuous addition of SMC complexes at one specific site would move the loaded rings along the DNA. SMC complexes were also assumed to act as DNA motors themselves.[482] Real-time imaging of DNA loop-extrusion by condensin by Ganji et al. provided evidence for the creation and extension of DNA loops by condensin in 2018.[483] A model for the loop extrusion by condensin was predicted as follows: The DNA is anchored by the regulatory subunits kleisin, and heat repeat at a DNA binding site and one condensin complex reels it in from one side reflecting an asymmetric process (Figure 102). The energy for this process is obtained by ATP hydrolysis.[483] To get more detailed information about the exact DNA binding site of condensin and how the DNA runs through the complex, an artificially induced covalent DNA-protein crosslink is a useful tool.

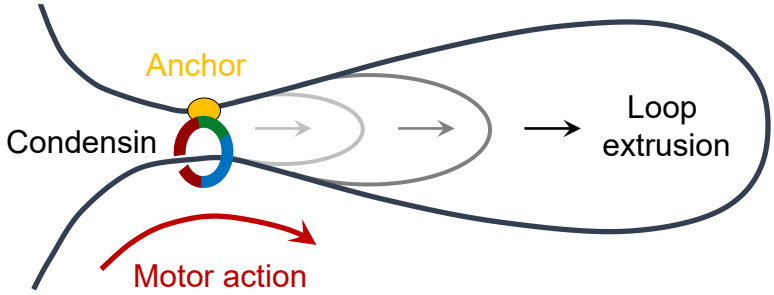
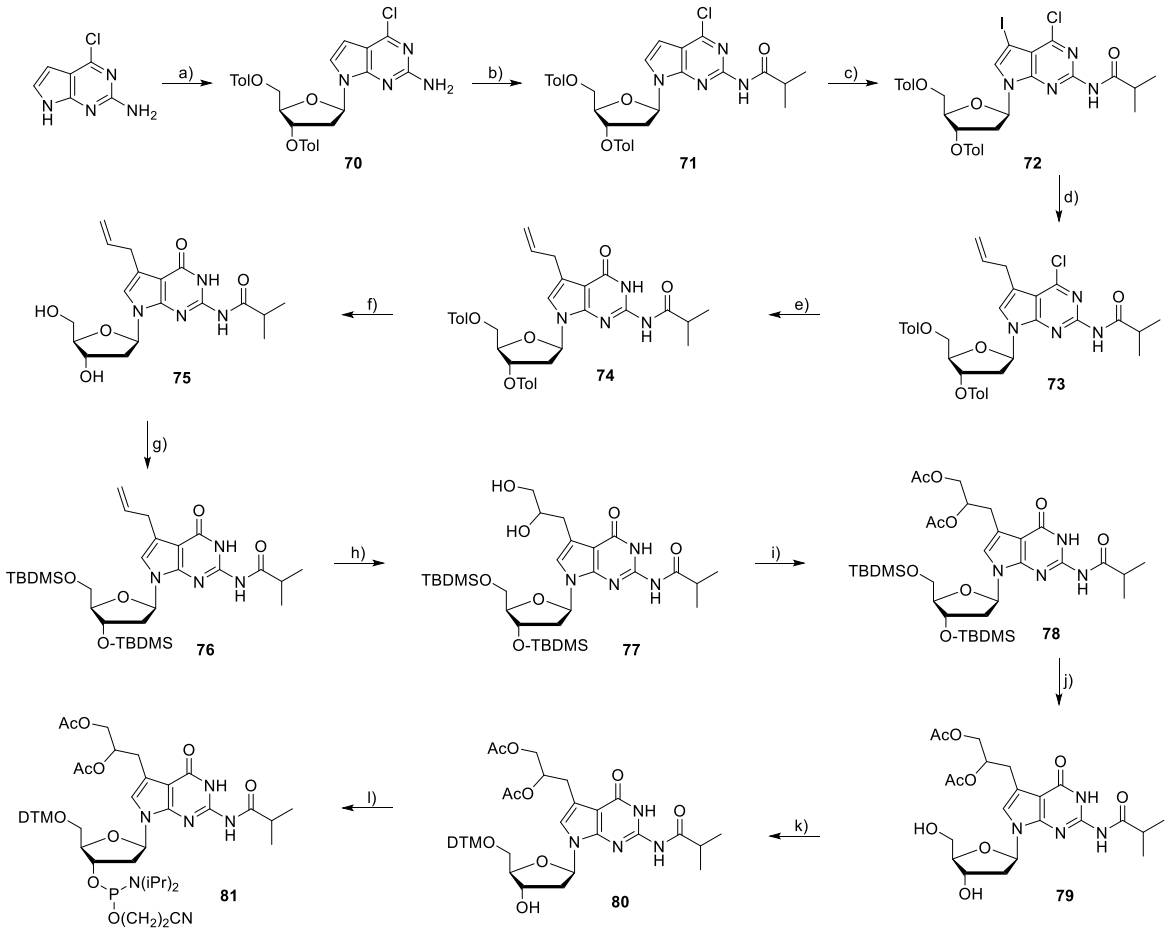



Figure 102. Schematic structure of the asymmetric DNA loop-extrusion by condensin. The figure was reproduced from Ref.[483]

For this reason, the 7-deaza-guanosine crosslinker was synthesized as a phosphoramidite as well as a triphosphate building block. The diol function of the crosslinker can by oxidized resulting in an aldehyde residue, which can react with Lys and Arg residues of condensin (see Figure 19A). With this covalent bond, it is expected that the exact interaction sites of condensin with DNA can be investigated. The synthesis of the two building blocks will be described in the next chapter.

### 3.3.1 Synthesis of the c7dG-crosslinker building blocks

The synthesis of the phosphoramidite **81** (Scheme 22) has been published earlier by Angelov *et al.*[484, 485] The synthetic pathway from their study was mainly followed, however, some of the steps had to be optimized. The synthesis started with the glycosylation of 2-amino-6-chloropurine with the Hoffer's chlorosugar in MeCN. Tris[2-(2-methoxyethoxy)ethyl]amine (TDA-1) was used as a phase-transfer catalyst under basic conditions. The addition of KOH had to be handled carefully to prevent hydrolysis of the chlorosugar or the chloropurine due to the hygroscopic properties of KOH. Addition of the Hoffer's chlorosugar all at once resulted predominantly in the formation of a side product with a higher mass. The sugar moiety seemed to be added twice and the product **70** could only be isolated with a yield of 38%. Treating the reaction with the sugar in three portions within 15 min increased the yield of product formation to 78%. The second step was the protection of the exocyclic amino group with isobutyrylchloride in pyridine, which could be obtained in high yields (**71**, 87%). To introduce the allyl function at position 7, iodination had to be performed with *N*-iodosuccinimide (NIS) in DMF at 80°C. Angelov *et al.* used an excess of NIS (7.00 eq),[484] which led to the formation of a side product, running slightly less polar on the TLC compared to the product. A pure fraction of the product could only be isolated with a yield of 44%. Reducing the amount of NIS to 2.00 eq, as well as reducing the reaction time from 8 h to 4.5 h helped to improve the synthesis of the desired product **72** with 96% yield. The iodinated compound **72** enabled the introduction of the allyl group by Stille cross-coupling reaction using $Pd(PPh_3)_4$ and allyl-$SnBu_3$ in toluene at 95 °C. This reaction resulted in a comparable yield (**73**, 86%) as the earlier report. The reaction conditions for the conversion of the chloropurine-riboside **73** to the 7-deaza guanosine analogue had to be slightly modified compared to the published procedure.[484]

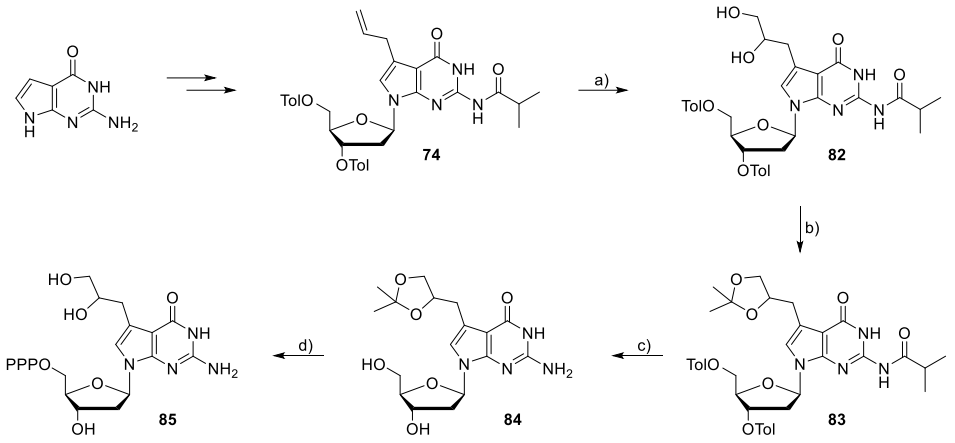


Scheme 22. Synthesis of the c7dG-crosslinker over 12 steps starting from 2-amino-6-chloropurine. a) MeCN, KOH, Tris[2-(2-methoxyethoxy)ethyl]amine (TDA-1), hoffer's chlorosugar, r. t., 30 min, (78%) b) pyridine, isobutyryl chloride, 0 °C – r. t., 1 h, (87%) c) DMF, *N*-iodosuccinimide (NIS), 80 °C, 4.5 h, (96%) d) toluene, $Pd(PPh_3)_4$, allyl tributyl stannane, 95 °C, 10.5 h, (86%), e) dioxane/DMF (1:1.07), pyridine-2-carboxaldoxime, 1,1,3,3-tetramethylguanidine, 40 °C, 72 h, (80%) f) THF, MeOH, 1M NaOMe in MeOH, 0 °C – r. t., 4 h, (79%) g) DMF, $^t$butyl-chlorodimethylsilane, imidazole, r. t., 16 h, (88%) h) THF, $H_2O$, 2% $OsO_4$ in tBuOH, 4-methylmorpholine-4-oxide monohydrate, 0 °C, 3 h, (68%) i) pyridine, acetic anhydride, 0°C – r. t., 3 h, (98%) j) THF, TBAF·$3H_2O$, acetic acid, r. t., 15 h, (99%.) k) pyridine, DMT-Cl, r. t., 2.5 h, (75%) l) DCM, $^i$Pr$_2$NEt, CEP-Cl, r. t., 1 h, (92%).

Treating the 6-chloropurine compound **73** with pyridine-2-carboxaldoxime and 1,1,3,3-tetramethylguanidine for 48 h at room temperature resulted in around 59% yield of the deaza guanosine **74**. Letting it stir for 72 h at 40 °C increased the reaction yield to 80%. As an alternative for the commercially available 1 M NaOMe-solution in MeOH, which was used in the earlier report to deprotect the tolyl-protection from the hydroxy groups,[484] NaOMe powder dissolved in MeOH was used. However, this solution led to the formation of multiple side products, which were identified as partially deprotected hydroxy groups as well as a fully deprotected compound lacking the isobutyryl group. The basic deprotection solution must be really anhydrous to prevent the deprotection of the isobutyryl group, which could not be achieved by dissolving the NaOMe solid in MeOH. Performing this reaction with commercially available 0.5 M NaOMe in MeOH predominantly resulted in the deprotection of the hydroxy groups, obtaining compound **75** with 79% yield. Subsequently, the hydroxy groups had to be protected with the silylation agent TBDMS-Cl. The protected compound **76** was obtained in a yield of 88%. The functionalization of the ally group was achieved by Upjohn

dihydroxylation creating a vicinal diol **77** with a catalytic amount of osmium tetroxide ($OsO_4$). *N*-methylmorpholine *N*-oxide (NMO) is added to the reaction as a stoichiometric co-oxidant to reoxidize the reduced Os-species to Os(V). The hydroxy groups were subsequently masked with acetyl protection groups. Less equivalents of acetic anhydride have been used compared to the earlier report (10 instead of 44 eq). The acetyl protected compound **78** was obtained with 98% yield. The TBDMS groups were selectively deprotected with the fluoride agent TBAF in quantitative yield resulting in the free c7dG derivative **79**. The second to last step was the 5'-DMT protection of compound **79** using DMT-Cl in pyridine for 1.5 h. The protected compound **80** was obtained in 75% yield, which was further converted to the phosphoramidite building block **81** by phosphitylation.

For the synthesis of the corresponding triphosphate building block, the first five steps remained unchanged. Starting with compound **74**, the trisphosphate was obtained after four additional steps (Scheme 23).




Scheme 23. Synthesis pathway to synthesize the c7dG-crosslinker triphosphate **85** for enzymatic incorporation. a) $OsO_4$, NMMO, THF/$H_2O$, 0 °C, 4 h, (59%) b) dimethoxypropane, *p*-TsOH, DCM, r. t., 2 h, (89%), c) $MeNH_2$ in $H_2O$ (40%), reflux, 2 h, (86%) d) proton sponge, $POCl_3$, $NBu_3$, pyrophosphate, $PO_4Me_3$, DMF.

The alkyl group of compound **74** was oxidized with $OsO_4$ before deprotecting the sugar moiety, in contrast to the synthetic pathway in Scheme 22. This facilitated the selective acetal protection of compound **82.** This acetonide is stable under the basic deprotection conditions for the tolyl- and isobutyryl-protection groups of compound **83**. This deprotection resulted in the free nucleoside **84**, which was further converted to the triphosphate **85**. The latter was achieved by reacting the 5'-OH with $POCl_3$, followed by the treatment with the bis(tetra-n-butylammonium) dihydrogen pyrophosphate.[486] The purification of the triphosphate was performed by HPLC.

To separate triphosphate from monophosphate and remaining starting material, the reaction mixture was first purified by anion-exchange-chromatography. Diethyl-aminoethyl (DEAE) was used as the positively charged ion-exchange resin. The starting material and the monophosphate were

eluting faster from the HPLC column, compared to the highly negatively charged triphosphate (Figure 103A). As a buffer system, triethylammonium bicarbonate buffer (pH 7.5) was used, which had to be prepared freshly beforehand from an aqueous $NEt_3$ solution and dry ice.

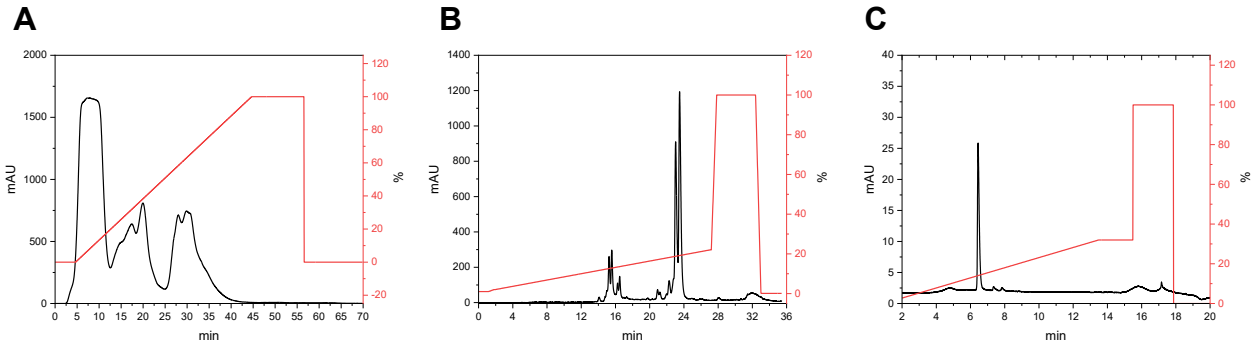



Figure 103. Purification of the c7dG triphosphate by A: anion-exchange chromatography (DEAE) and B: reversed phase (C18) HPLC. A: The product was eluting between 25 – 40 min. B: the product was eluting between 23 – 24 min. C: purity evaluation by anion-exchange chromatography. The gradient is shown in red for all graphs.

The fraction between 25 – 40 min from anion exchange HPLC was collected (Figure 103A), lyophilized, and further purified by reversed-phase HPLC (Nucleosil, C18, Figure 103B). Some impurities after the first round of purification could be separated off by reversed phase. These impurities were eluting faster from the column than the product, which was collected from 23 – 24 min. As buffer system, 100 mM triethylammonium acetate (TEAA, pH 7) in $H_2O$ and MeCN was used. After lyophilization, the collected fraction was validated by HR-ESI/MS and NMR. The triphosphate contained the deprotected dihydroxy functional group at position c7. The purity was further shown by anion exchange chromatography (DNAPac PA200) using Dionex buffer system (Figure 103C).

Both building blocks **81** and **85** were then incorporated into DNA by solid phase synthesis and enzymatically, respectively.

### 3.3.2 Incorporation of the c7dG-crosslinker triphosphate into DNA

For the incorporation of the phosphoramidite building block **81** into an DNA oligonucleotide by solid phase synthesis, a sequence from an earlier report from Hearing and coworkers was used.[487] They crystallized a complex of *S. cerevisiae* Ycg1–Brn1$_{\text{short kleisin loop}}$ with an 18mer DNA sequence (5'-GATGTGTAGCTACACATC-3') and solved its crystal structure (PDB:5oqn).[487] Since some structural information were available for this DNA sequence, it was chosen to replace one nucleotide with the c7dG-crosslinker. With the help of the molecular visualization tool PyMOL, a nucleotide with proximity to a Lys residue of the protein Condensin was selected to substitute it with the crosslinker (Figure 104). The Lys residue should be located close to the major groove, as the dihydroxy residue at position c7 points towards this direction. Adenosine at position 12 was substituted with c7dG as there was a Lys located close by. Accordingly, the thymidine residue at position 5 of the palindromic sequence was changed to cytosine, to enable base pairing in the duplex.

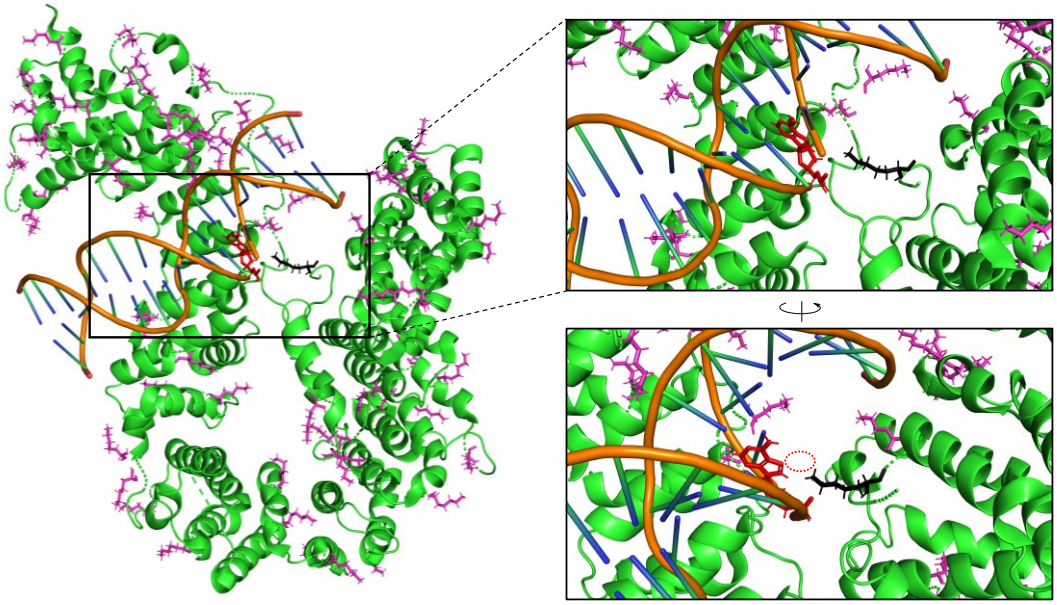


Figure 104. Crystal structure of the *S. cerevisiae* Ycg1-Brn1$_{\text{short kleisin loop}}$ (green) – DNA (orange) complex (PDB:5oqn[487]). Lysin residues are shown in magenta. One lysine residue pointing towards the major groove of the 18mer is highlighted in black. The potential area to form the covalent crosslinker between DNA and the enzyme is indicated with a red sphere.

The oligonucleotide was synthesized with a 5'-alkyne end to enable fluoresceine labeling: Alk-GATGTGCAGCTXCACATC (X=c7dG). This was achieved by Copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) of 5' alkynes with 6-FAM-azide (see 5.4.2). This fluorescein-labeled product was validated by anion-exchange chromatography and HR-ESI/MS. The collaboration partner, Julia Locherer, wants to form a complex with the unmasked DNA (oxidized) and *Ct* Ycg1:Brn1, in which the crosslinker can form a covalent bond with a lysine (or arginine) residue. These experiments will be investigated by mass spectroscopy. With these studies, they want to find out how the DNA runs through the complex.

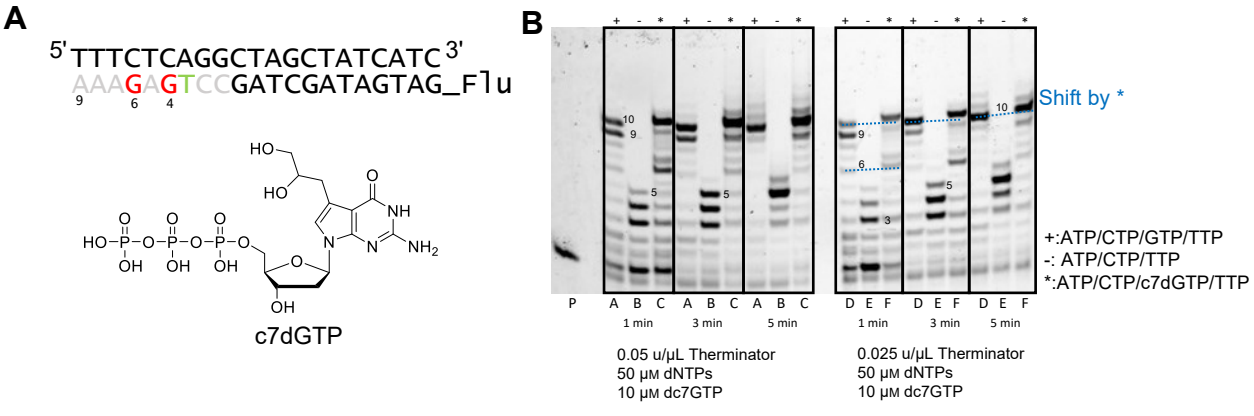



Figure 105. A: Primer extension with a 21mer template DNA strand and a 12mer DNA primer, which was 5'-fluroesently labeled. The primer could get elongated by 9 more nucleotides, of which position 4 and 6 are guanosines (red). On these two positions the c7dGTP should be incorporated. B: Results of the primer extension experiments using two different concentrations (0.05 (A,B,C) and 0.025 u/µL (D,E,F)) of the Therminator DNA polymerase and an incubation time of 1, 3, and 5 min. The modification is resulting in a slight shift of the elongated band, marked with blue dashed lines. +: positive control with the four canonical dNTPS; -: without dGTP; *: c7dGTP instead of dGTP.

To incorporate the c7dG crosslinker frequently within an oligonucleotide, the corresponding triphosphate should be used. After the successful synthesis, the incorporation efficiency of the modified triphosphate by a polymerase had to be tested. Thus, primer elongation experiments were performed first using Terminator, KOD, Taq, Dream-Taq, and PFU polymerases. The primer should be elongated by nine more nucleotides, of which two would be guanosines. The positive control contained the four canonical dNTPs, the negative control was lacking dGTP, and the last one contained the c7dG-triphosphate instead of dGTP (Figure 105A). Terminator polymerase was found to be the most efficient. Two concentrations of the enzymes were used for the experiments (0.05 and 0.025 u/µL), and an incubation time was varied between 1, 3, and 5 min. It was shown, that the crosslinker triphosphate could be installed as the positive control and the reaction containing the c7dGTP showed fully elongated product bands. It appeared that the modification caused a small shift in the migration of the product band (Figure 105B) compared to the positive control, but both reactions were equally efficient. However, all three samples overextended a few nucleotides. The negative control should not be able to add more than three nucleotides, as dGTP was left out from the reaction. However, 5-6 nucleotides were incorporated before termination. The two other reactions were elongated by 10 nucleotides. Therminator has a higher error rate (70 errors per 1000),[488] than for example Taq polymerase (~$2x10^{-4}$).[489] To verify if the crosslinker triphosphate was actually incorporated, the primer extension experiment was performed on a larger scale. The reaction was separated on PAGE and the product band was analyzed by MALDI. For mass measurement, the samples were desalted using the Merck ZipTip™ 0.6 µL-C18 resin. The found mass of 7577.05894 g/mol corresponded to the product strand, containing the c7dG-crosslinker with the dihydroxy function group twice (as the template contained two cytidines) and one additional adenosine. Thus, the primer was elongated by 10 instead of 9 nucleotides and the incorporation of the modification occurred twice.


Figure 106. PCR experiment with a 50mer DNA template strand (black), a 34mer reverse primer (blue) and a 17mer forward primer (green). The complementary regions are underlined.

Since these initial experiments were successful, the fidelity of crosslinker triphosphate incorporation into a longer DNA sequence by PCR was tested. A 50mer DNA template was utilized together with a 34mer reverse primer and a 17mer forward primer (Figure 106).

The PCR was performed according to the PCR cycle shown in Figure 107A with a 2 pmol template and 10 pmol of the two primers. Taq- and Dream-Taq polymerase (0.025 u/µL) were tested for the PCR with two different concentrations of c7dGTP (10 µM and 50 µM) in place of dGTP (Figure 107B). Two positive controls with the four canonical dNTPs at 200 µM concentration each were performed as a reference. The four reactions containing c7dGTP showed the same product band as the two positive control samples but were slightly less efficient and a higher amount of the primers remained. The reaction with c7dGTP appeared to be successful. Taq and Dream-Taq were equally efficient, but for Taq, a home-made enzyme preparation was used, which produced byproducts that got stuck in the well of the gel. Hence, commercially available Dream-Taq was chosen for further experiments.

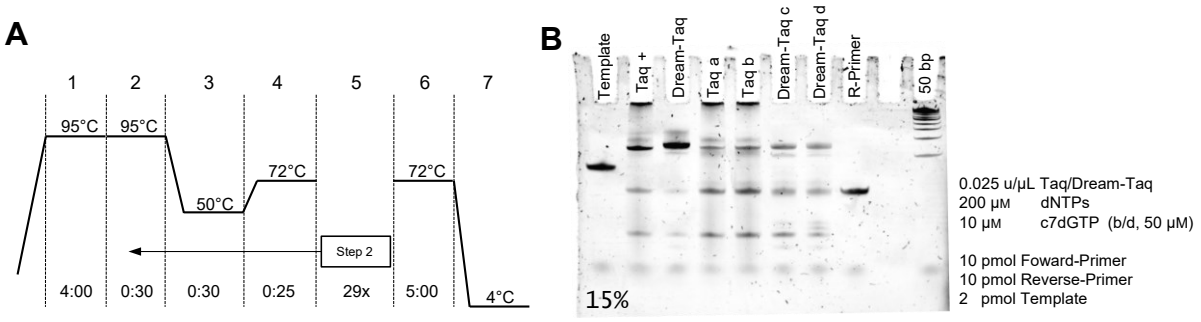



Figure 107. A: PCR cycle B: 15% Denaturing PAGE (mini gel), stained with SYBR Gold. Taq+ and Dream-Taq+ are the positive control samples. For Taq a) and Dream-Taq c) a c7dGTP concentration of 10 µM, and for Taq b) and Dream-Taq d) a c7dGTP concentration of 50 µM were used. The template and the reverse primer were loaded to the gel in addition. A 50 bp DNA ladder is shown on the right as a size marker. Using c7dGTP instead of dGTP (a,b,c,d) led to a product band on the same level as the positive control samples (+).

Then, the positive control was performed with a lower dGTP concentration, 10 instead of 200 µM, to see if the c7dGTP concentration (10 µM) used in the PCR experiments was sufficient to create a full-length product or if higher concentrations should be added. This reaction resulted in a band at the same level as the positive control (Dream-Taq+) with a similar intensity (Figure 108), indicating that the concentrations of c7dGTP should essentially be enough. Next, a mixture of dGTP and c7dGTP was used in the reactions to evaluate the incorporation efficiency and whether c7dGTP can be incorporated in the presence of canonical dGTP. The intensity of the band decreased with increasing concentration of the modification, suggesting that the c7dGTP gets incorporated less efficiently into the DNA oligonucleotide. However, it appeared that there is no preference for only dGTP if both GTP derivatives are present. Otherwise, a band with the same intensity as the positive control should be obtained. The reaction with c7dGTP instead of dGTP showed the lowest intensity of the product band, but at the same level as the other reactions.

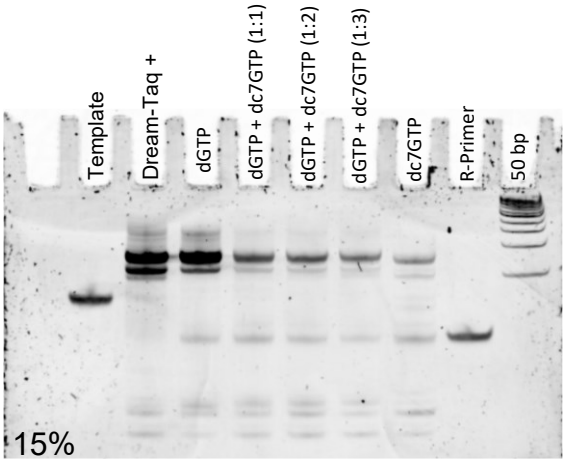



Figure 108. 15% Denaturing PAGE of the PCR rections. The template and the reverse primer (R-primer) were loaded as a reference. A positive control (Dream-Taq+) with 200 µM dNTP and 0.025 u/µL enzyme concentration was performed as a reference. In the next line, a lower dGTP concentration (10 instead of 200 µM) was chosen, to see if the low c7dGTP concentration is sufficient. Next, different ratios of dGTP and c7dGTP were used together, followed by c7dGTP instead of dGTP. The 50 bp DNA ladder is given as a size marker.

The c7dGTP triphosphate seemed to be incorporated, even though with a lower yield. To prove that the modification was present in the PCR product, and that the respective reactions to form the covalent crosslink with Lys or Arg residues are working, the reaction conditions for the oxidation of the diol were tested. Therefore, the conditions reported from Angelov et al. were applied onto the DNA oligonucleotide Alk-GATGTGCAGCTXCACATC, which was synthesized by solid phase synthesis.[484] A solution of the oligonucleotide in $H_2O$ was mixed with sodium phosphate buffer (pH 5.3) and a freshly prepared sodium periodate ($NaIO_4$) solution. The resulting mixture was kept for 6 h at 4 °C in the dark.[484] The reaction was stopped by the addition of a sodium sulfite ($Na_2SO_3$) solution. After ethanol precipitation, the reaction product was analyzed by MALDI-MS (see mass spectra in 6.3). The unmasked oligonucleotide with the aldehyde function on position 7 of the modification was detected in the spectra, while the starting material was not visible anymore. Reducing the reaction time to 1 h but increasing the temperature to 37 °C was also sufficient. Keeping the ratio between the reactants, but decreasing the $NaIO_4$ concentration around 10-fold significantly decreased the yield of the oxidized product and the starting material remained. Thus, the reaction conditions of the oxidation have to be carefully chosen for an efficient unmasking of the crosslinker.

These reaction conditions should be applied to the PCR product obtained with c7dGTP instead of dGTP. The reaction shown in Figure 107 was repeated, but with a negative control lacking dGTP. The negative control was elongated as well and led to a band which was migrating slightly faster on the gel. This band was also visible in the positive control to a minor extent. The PCR cycle was reduced from 29 to 20 and 10 repetitions to see if the elongation of the negative control can be reduced (Figure 109). However, the band in the negative control was observed for all the three conditions, which might either occur due to contaminations, or due to misincorporations opposite to cytidine. Besides, a band migrating slightly slower than the reverse primer was observed. The reverse primer got partially elongated, but also a longer band with almost the same length as the

positive control appeared. The reduction of the dNTP concentration from 200 µM to 100 µM or 20 µM resulted in the same behavior. Even though the negative control showed some unusual band migration, it could be distinguished from the positive control and the one containing c7dG.

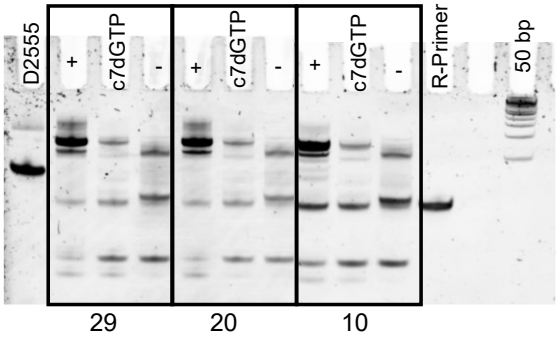



Figure 109. PCR with different repetition cycles: 29, 20 and 10. The positive control (+) was performed with Dream-Taq polymerase and 200 µM dNTPs, while in the negative control (-) the dGTP was left out. In the middle lane, dGTP was replaced by c7dGTP. As shown before, the modified TP got less efficiently incorporated but the obtained band was on the same level as the positive control. The negative control got elongated as well, and resulted in a band, which was slightly migrating faster on the gel. A minor band on the same level was also observed in the positive control.

After ethanol precipitation of the three reactions (positive, c7dG, and negative), they were incubated with the oxidation conditions described earlier (see 5.4.11 for details) to unmask the diol and create the reactive aldehyde moiety (Figure 110A). The reaction was worked up by ethanol precipitation. The pellet was dissolved in water, mixed with sodium cacodylate (pH 5) and fluorescein-5-thiosemicarbazide (FAM-TSC), and incubated for 3 h at 37 °C in the dark (Figure 110A). The reaction was ethanol precipitated and separated on denaturing PAGE. An image in the fluorescein channel was recorded first and then the gel was stained with SYBR gold to visualize the bands without fluorescein label (Figure 110B-D).

In the fluorescein channel, remaining Fluorescein-TSC can be detected for the three reactions after PCR, oxidation, and fluorescent labeling (Figure 110B). Otherwise, only the reaction containing c7dGTP showed two fluorescent bands. Staining with SYBR gold visualized all the other bands and revealed that two fluorescent bands were on the same level as the bands obtained from the positive control (Figure 110C). Again, the negative control contained the band slightly eluting faster as the main band of the positive control. This band present in the positive control and the sample containing c7dG seemed to be an incompletely elongated strand. Taking an image in the fluorescein channel after SYBR gold staining highlighted the bands, which contained a fluorescent label (Figure 110D). This was only the case for the c7dG-containing sample. This experiment could clearly demonstrate that the oxidation (unmasking) of the diol and the further functionalization of the aldehyde was possible after incorporation of the triphosphate via PCR. The functionalization of the crosslinker with fluorescein was chosen to get a direct read out of the reaction by a fluorescent signal on the gel image. At this step, the collaboration partners would form a covalent bond with Lys or Arg residues of Condensin.

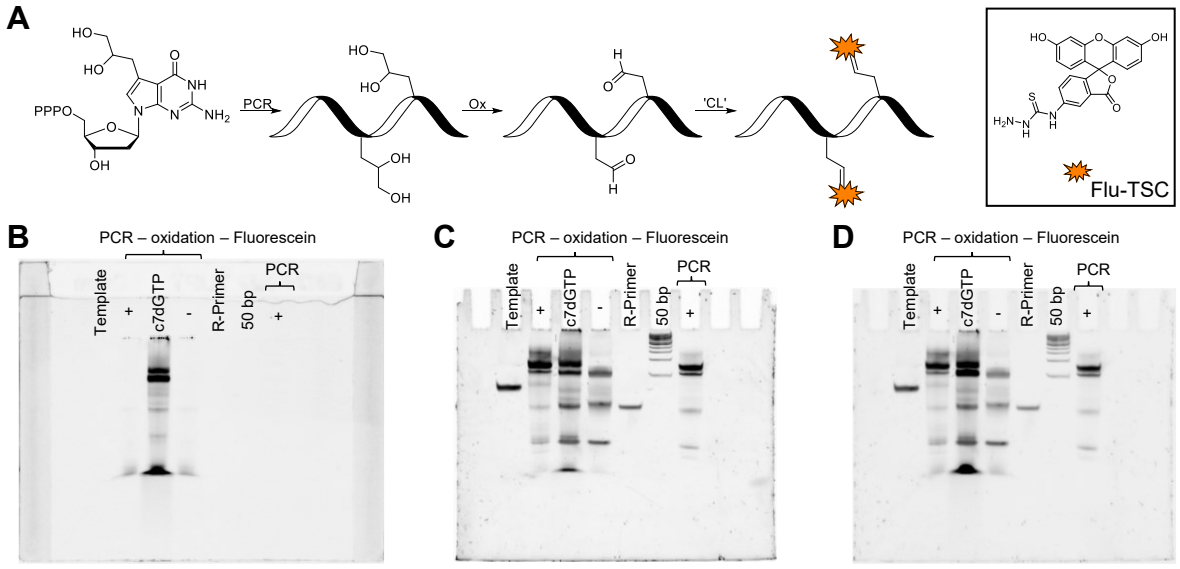



Figure 110. A: mechanism of the incorporation, oxidation, and fluorescent labeling with FAM-TSC. Denaturing gel images. B: Fluorescein channel. C: SYBR gold staining. D: Fluorescein channel after SYBR gold staining. Template, reverse primer, and the positive control after PCR are given as reference, a 50 bp ladder as size marker. Positive control, c7dG-containing sample, and the negative control were oxidized after PCR. The resulting aldehyde was further functionalized with FAM-TSC. B: The Fluorescein label, which was attached to the crosslinker, as well as remaining Fluorescein-TSC (the three bands on the same level) can be visualized in the fluorescein channel. Only the lane with c7dG showed fluorescent bands, which might belong to the product-strand. C: To uncover the other bands, the gel was stained with SYBR gold. D: The stained gel was again detected in the fluorescein channel to highlight the bands which are containing a fluorescent label.

The synthesis of both, the phosphoramidite as well as the triphosphate building block of the crosslinker was successfully performed. A selected DNA oligonucleotide was synthesized by solid phase synthesis containing the modification at position 12. This position was chosen based on the crystal structures of a Ycg1–Brn1$_{\text{short kleisin loop}}$ – DNA complex (Figure 104). The crosslinker was selected to be in proximity to a Lys residue of Condensin. Furthermore, the successful incorporation of the trisphosphate was demonstrated by Primer extension experiments using Therminator polymerase, and by PCR using Taq and Dream-Taq polymerase. The diol was unmasked by oxidation ($NaIO_4$) and the resulting aldehyde was further reacted with FAM-TSC. The successful incorporation and functionalization was observed by a fluorescent signal on PAGE.

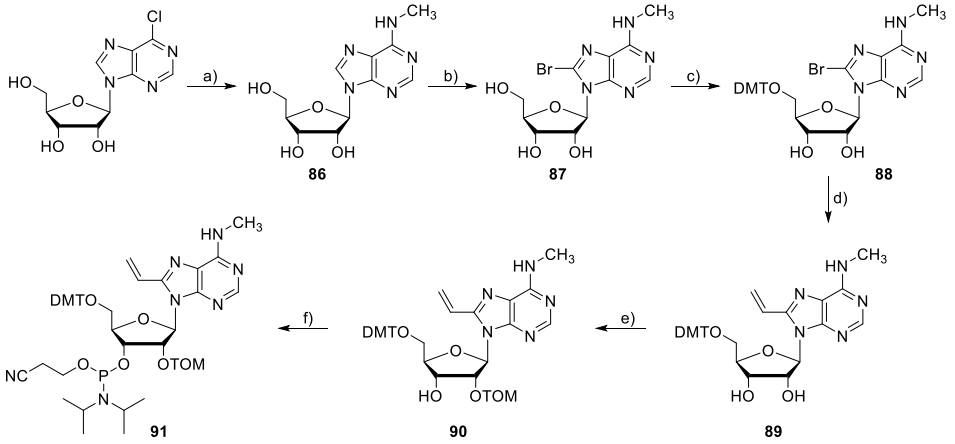
## 3.4 Fluorescent nucleosides analogues

$N^6$-methyladenosine ($m^6A$) is the most abundant post-transcriptional RNA modification in eukaryotic cells. In concert with the methylation and demethylation enzymes shown for $m^3C$, the methylation level of $m^6A$ is regulated by the interplay of writer and eraser enzymes.[490] Assays that detect changes in this methylation pattern could provide new insights into important biological processes. Fluorescent nucleosides like 8-vinyladenosine modified analogs are used as molecular tools in such assays. The fluorescent nucleoside analogue $N^6$-methyl-8-vinyladenosine ($m^6v^8A$) was synthesized as the phosphoramidite building block **91**. It was supposed to be incorporated into RNA oligonucleotides as an analytical tool to facilitate the investigation of the demethylation and methylation activity of the respective eraser and writer enzymes described earlier (see chapter 1.3.3 and 1.3.4) by fluorescence changes upon structural changes of the RNA. For this important research topic, there is the imperative need for highly sensitive detection methods of the enzyme activity *in vitro* and *in vivo*. The synthesis of the deoxy-analogue 8vdA was published by Gaied *et al.* in 2005.[182] In this report, they compared the thermodynamic and spectroscopic properties of 8vdA with the frequently used 2AP. The 15mer DNA duplex, with 8vdA at position 8 revealed a higher thermal stability with flanking A or T residues than the corresponding duplex containing 2-AP. 8vdA was proposed to be present in the *anti*-conformation and less perturbing than 2AP. The absorption (8vdA: 290 nm, 2AP: 309 nm) and emission (8vdA: 382 nm, 2AP: 370 nm) of the free nucleoside 8vdA was in the same spectral range than 2AP.[182] Thus, the Stokes shift of 8vdA was larger by 25 nm compared to 2AP (8vdA: 92 nm, 2AP: 67 nm). Both 8vdA and 2AP had comparable quantum yields as free nucleosides with 0.65 and 0.68, respectively. However, 8vdA was more sensitive than 2AP, as its absorption coefficient was almost twice that of 2AP ($\phi$; 8vdA: 12600 $M^{-1}cm^{-1}$, 2AP: 7200 $M^{-1}cm^{-1}$).[182] In single and double stranded DNAs, 8vdA had a higher quantum yield than 2AP. The absorption maximum of 8vdA in a single strand was 265 nm, with a shoulder at around 310 nm attributed to the fluorescent nucleoside, while the emission maximum was similar to the free nucleoside (387 nm).[182] 8vdA appeared to behave as a non-perturbing analog of 2AP with higher sensitivity and therefore could be used for the same research purposes. The synthesis of the $m^6v^8A$ RNA phosphoramidite was based on the synthetic pathway described in this report.

### 3.4.1 Synthesis of the $v^8m^6A$ phosphoramidite and free nucleoside

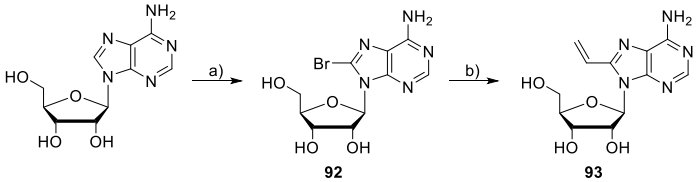
The phosphoramidite building block of $m^6v^8A$ **91** was obtained over six steps starting from 6-chloropurin riboside (Scheme 24). The starting material was stirred with a 33% $MeNH_2$ solution in EtOH for 4 h at room temperature. After evaporation of the solvent, the solid product was used without further purification. The 6-chloro compound was converted in quantitative yields. Bromination at position 8 was achieved by reacting **86** with elementary bromine in sodium acetate buffer (pH 4) for 72 h at room temperature. After neutralization with a NaOH solution, the product precipitated

and was obtained after filtration. The product **87** was evaporated with toluene to facilitate removal of remaining water. To selectively protect the 5'-OH group with DMT-Cl, the 2'-and 3'-hydroxy groups were protected with DMF-DMA first, thereby forming a cyclic acetal. The free 5'-OH group was protected, and the acetal was cleaved off under workup conditions resulting in compound **88** with 78% yield. The vinyl moiety was introduced via Still-cross coupling. Gaied et al. used tetrakis(triphenylphosphine)palladium ($Pd(PPh_3)_4$) as a Pd(0) catalyst and $Sn(CH{=}CH_2)_4$ in NMP at 110 °C.[182] However, these reaction conditions used for the 8vdA compound resulted in only 28% yield. Such low efficiency for the functionalization of a 8-bromoadenosine derivative with $Pd(PPh_3)_4$ has also been observed by Tu et al.[491] Besides, they stated that protection of the 5'-OH is a pre-requisite for the success of the reaction. Thus, the reaction conditions based on Tu *et al.* resulted in an increase of the reaction yield to 75%.[491] Only the equivalents of the organostannane and the Pd-catalyst were slightly increased. To the starting material $Bu_3Sn(CH{=}CH_2)$, $Pd^{II}(OAc)_2$, and the phosphine ligand $PPh_3$ were added to generate the Pd-catalyst *in situ*.[491] Pd(II) is getting reduced intramolecularly to a $Pd^0$-complex.[492, 493] The anionic compound $[Pd(OAc)(PPh_3)_2]^-$ is the active complex, which performs the oxidative addition.[492] Due to its fluorescent properties, product **89** was spotted under the UV lamp using 365 nm. The fluorescent product was isolated by column chromatography with 75% yield. The introduction of the vinyl group could, unambiguously, be shown by NMR spectroscopy in $CDCl_3$. The diastereotopic protons of the vinyl-$CH_2$-group gave two multiplets at distinct chemical shifts, 5.57 ppm (J = 11.1, 1.7 Hz) and 5.95 ppm (J = 17.2, 1.7 Hz). The proton at 5.57 ppm was orientated *cis* to the CH proton of the vinyl group as it had the smaller vicinal coupling constant (11.1 vs. 17.2 Hz). The geminal coupling constant was 1.7 Hz. The last two steps were protection of the 2'-OH group with TOM-Cl and conversion of compound **90** into the phosphoramidite **91** with 76% yield by phosphitylation using CEP-Cl.




Scheme 24. Synthesis of the $m^6v^8A$ phosphoramidite building block **91**. a) $MeNH_2$ (33% in EtOH), $NEt_3$, r. t., 4 h, (99%) b) $Br{\cdot}H_2O$, NaOAc-buffer (pH 4), r. t., 72 h, (82%) c) DMT-Cl, pyridine, DMAP, r. t., 4 h (78%) d) $Bu_3Sn(CH{=}CH_2)$, $Pd(OAc)_2$, $PPh_3$, CuI, THF, 60 °C, 24 h, (75%) e) 1. $Bu_2SnCl_2$, $^iPr_2NEt$, 1,2-DCE, r. t., 1 h; 2. TOM-Cl, 80 °C, 20 min, (26%) f) CEP-Cl, DCM, $Me_2NEt$, r. t., 2.5 h, (76%).

Additionally, the free $m^6v^8A$ and $v^8A$ nucleosides were synthesized for fluorescent measurements. For $m^6v^8A$, compound **89** (Scheme 24) was used. With 2% formic acid in DCM the 5'-*O*-DMT group was cleaved off, resulting in the free nucleoside. $v^8A$ was synthesized over three steps starting with adenosine (Scheme 25). The free 5'-OH group of **92** led to low yields (35%) of the product.




Scheme 25. Synthesis of the free $v^8A$ nucleoside over two steps. a) $Br{\cdot}H_2O$, NaOAc buffer (pH 4), r. t., 72 h, (81%) b) $Pd(PPh_3)_4$, DMF, $Bu_3Sn(CH{=}CH_2)$, 100 °C, 6 h, (35%).

### 3.4.2 Incorporation of the $m^6v^8A$ phosphoramidite into a palindromic sequence

The phosphoramidite building block **91** was used for the incorporation into an RNA oligonucleotide. The 12mer palindromic sequence (Table 16) was chosen from a report by Cheong *et al.* (2018), where all nucleotides were carrying 2'-OMe groups.[494]

Table 16. Sequence, formula, and mass-data of the unmodified reference R691 and R692, which is containing the fluorescent $m^6v^8$-adenosine modification (X).

| Name | 5'-sequence-3' | nt | formula | Mass calc. | Mass found |
|---|---|---|---|---|---|
| **R691** | GGCAGAUCUGCC | 12 | $C_{114}H_{143}N_{46}O_{82}P_{11}$ | 3808.55421 | 3808.57225 |
| **R692** | GGCXGAUCUGCC | 12 | $C_{117}H_{147}N_{46}O_{82}P_{11}$ | 3848.58661 | 3848.62532 |

In the report of Cheong et al. they were able to monitor the FTO activity *in vitro* and in living cells depending on the $m^6A$ methylation dynamics.[494] If $m^6A$ at position 4 was present, the palindromic sequence formed a hairpin structure with the modification in the loop region (Figure 111).

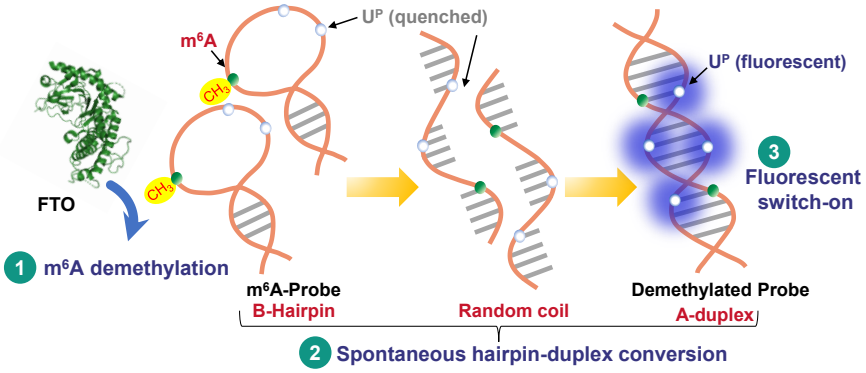



Figure 111. Highly sensitive, methylation-switchable probe to analyze FTO induced demethylation of $m^6A$. The fluorescent nucleoside $U^P$ was quenched if located in the loop but showed fluorescent signal in the duplex structure. The switch from hairpin to duplex occurred upon demethylation of $m^6A$. The figure (FTO pdb:3LFM) was reproduced from Ref.[494]

The fluorescent nucleotide 2'-*O*-(1-pyrenylmethyl)uridine ($U^P$), which was either placed at position U7 or U9 was also located inside the loop. In a single-stranded or hairpin loop environment, the fluorescent nucleotide $U^p$ is quenched by photo-induced charge transfer (PET, pyrene to adjacent nucleobases) due to stacking with adjacent nucleobases. Inside a duplex, the pyrene moiety is

shifted to the minor groove and the nucleotide can emit fluorescence. Accordingly, as soon as $m^6A$ was demethylated to A, the fluorescent intensity increased at least 4-fold by forming a stable duplex structure.[494]

Using $m^6v^8A$ inside this system (Figure 111), could make the additional fluorescent nucleotide U redundant. The dynamic changes of $m^6A$ could directly be monitored at the position of interest depending on the methylation level. The unmodified reference was deprotected under standard conditions using $MeNH_2$ in $H_2O$/EtOH (1:1). In the chromatogram of the anion-exchange HPLC, a broad peak with a second peak in front of the main peak was observed (Figure 112A). A temperature increased from 60 °C to 80 °C, resulted in one sharp peak (Figure 112B). Since it is a palindromic sequence, 60 °C was not enough for denaturation of secondary structures.

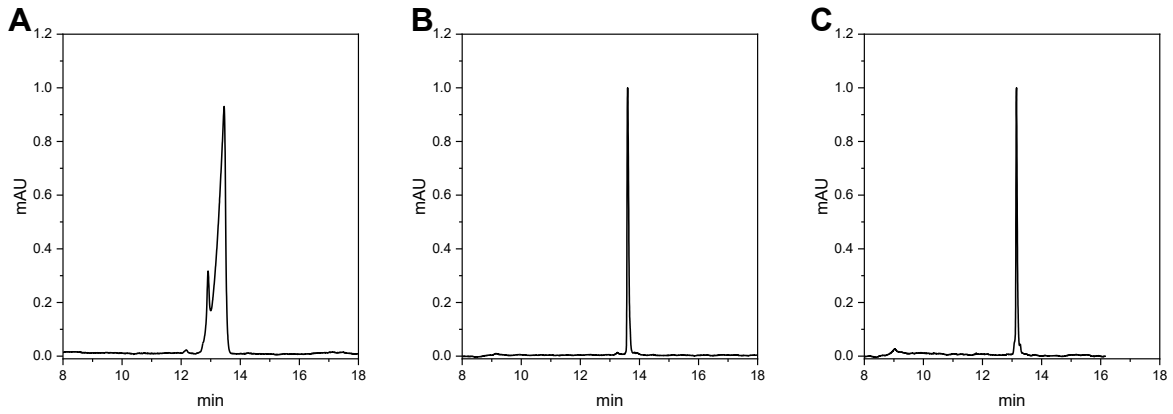



Figure 112. Anion-exchange chromatograms of R691 at A: 60 °C and B: 80 °C. At 60 °C the self-complementary structure is not completely denatured giving a broad signal with an additional small peak in front of the main peak. Analysis of the same sample at 80 °C resulted in a single, sharp peak due to complete denaturation. C: R692 at 60 °C revealed on sharp signal. The chromatograms were plotted after normalization as mAU vs. min.

The deprotection conditions used for the unmodified reference could not be adapted for the modified oligonucleotide. Gaied et al. reported the appearance of a side product after the treatment with ammonia (28%) for 4 h, which is eluting faster on HPLC.[182] The ratio of the side product further increased with prolong deprotection time. MALDI-TOF analysis revealed a mass difference of 17, compared to the correct product peak (eluting slower). Ammonia seemed to be added to the vinyl moiety. Thus, milder reactions conditions were tested for the deprotection of $m^6v^8A$. The oligonucleotide was incubated with a mixture of $H_2O$, MeOH and $NEt_3$ in the ratio 5:4:1 for 48 h at 25 °C. After TBAF deprotection of the 2'-*O*-TOM group and PAGE purification, the oligonucleotide was analyzed by anion-exchange chromatography. The chromatogram at 60 °C showed one sharp product peak (Figure 112C). Both products, the unmodified and the $m^6v^8A$-containing oligonucleotide, were confirmed by HR-ESI-MS spectroscopy (Table 16).

### 3.4.3 UV melting analysis

Thermal denaturation experiments were performed with the unmodified (R691) and the modified (R692) sequence, with two heating and two cooling ramps between 10 and 90 °C for five different concentrations (1, 2, 5, 10 and 20 µM single strand concentration) of the oligonucleotides. The oligonucleotides were incubated in 10 mM sodium phosphate buffer (pH 7) containing 100 mM NaCl. As described earlier, the self-complementary sequence enabled the formation of duplex and hairpin structures.

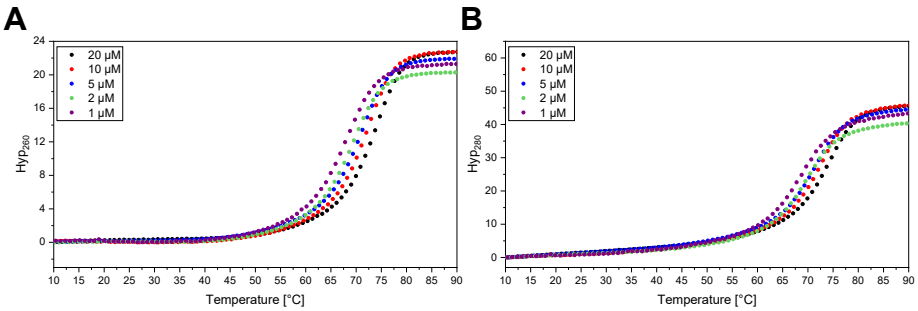



Figure 113. UV melting curve analysis of the unmodified reference R691 measured in a range of 10 to 90 °C for five concentrations at A: 260 nm and B: 280 nm in 10 mM sodium phosphate buffer (pH 7), containing 100 mM NaCl. Hyperchromicity was plotted vs. temperature. A sigmoidal curve shape with a concentration dependency of the melting temperature was obtained, revealing the formation of a duplex structure.

For the unmodified reference, a sigmoidal curve shape was obtained (Figure 113). The two heating and two cooling ramps are mostly coinciding, representing a reversible melting process. The obtained melting temperatures showed a concentration dependency with $\Delta T_m$ = 4.8 °C at 260 nm and $\Delta T_m$ = 4.0 °C at 280 nm. As expected, R691 adopted a duplex structure, with a melting temperature of around 69.7 °C at 5 µM total strand concentration (Table 17).

Table 17. Obtained melting temperature at 260 and 280 nm for R691 and R692.

| $T_m$ | 260 nm | | | | | 280 nm | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Conc [µM]** | **10** | **5** | **2.5** | **1** | **0.5** | **10** | **5** | **2.5** | **1** | **0.5** |
| R691 (GGCAGAUCUGCC) | 72.1 | 70.2 | 69.7 | 68.3 | 67.3 | 72.0 | 70.2 | 70.0 | 68.8 | 68.0 |
| R692 (GGCXGAUCUGCC) | 53.4 | 49.4 | 47.8 | 44.6 | 42.4 | 57.3 | 55.3 | 54.9 | 54.3 | 54.2 |

The melting temperatures are almost similar at 260 and 280 nm. The inverse temperature was plotted against ln c to calculate enthalpy, entropy, and free energy. The thermodynamic data obtained by the Van't Hoff Plot is summarized in Table 18.

Table 18.Thermodynamic data of R691.

| **Thermodynamic data** | $\Delta H_{260}$ **kcal/mol** | $\Delta S_{260}$ **cal/mol·K** | $\Delta G_{260}$ **kcal/mol** | $\Delta H_{280}$ **kcal/mol** | $\Delta S_{280}$ **cal/mol·K** | $\Delta G_{280}$ **kcal/mol** |
|---|---|---|---|---|---|---|
| R691 (GGCAGAUCUGCC) | -155 | -423 | -28.7 | -190 | -525 | -33.4 |

The slight changes of the values upon changing from 260 to 280 nm was already discussed in chapter 3.2.5, caused by the higher Δε of G/C compared to A/U base pairs. Cheong *et al.* performed their melting curve analysis with the 2'-OMe(GGCAGAUCUGCC) oligonucleotide in 10 mM

sodium phosphate buffer (pH 7.4) containing 150 mM NaCl, too. They obtained a melting temperature of 67.3 °C (5 µM total strand concentration), which is similar to the obtained value without 2'-OMe-modification.

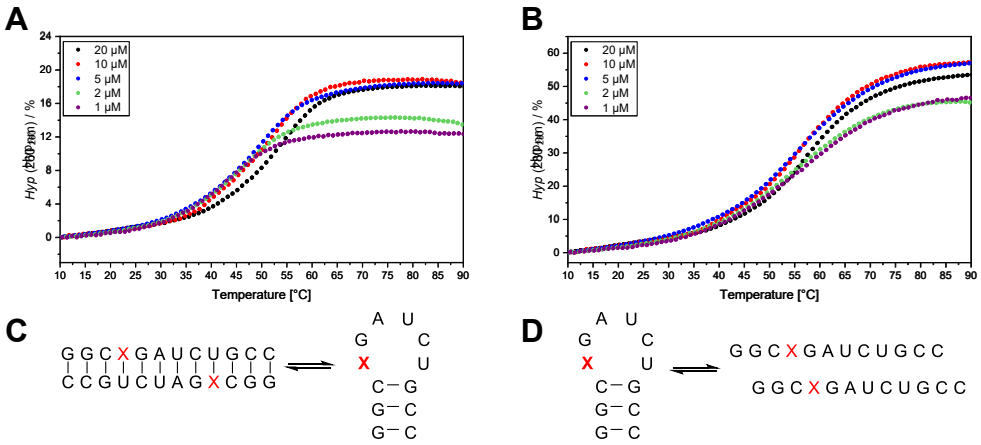



Figure 114. UV melting curve analysis of R692 at A: 260 nm and B: 280 nm in 10 mM sodium phosphate buffer (pH 7), containing 100 mM NaCl. Hyperchromicity was plotted vs. temperature. C: duplex-hairpin equilibrium D: hairpin-single strand equilibrium. Concentration is given as single strand concentration; the second cooling ramp is shown.

As for the unmodified reference, thermal melting experiments were performed for the modified sequence R692 at 260 and 280 nm (Figure 114A,B). A distinct curve shape and lower melting temperatures were observed for the modified sequence. At the two highest concentrations a minor second transition at around 40 °C could be surmised. This biphasic curve shape might result from a duplex-hairpin equilibrium. The lower melting temperature reflected a concentration dependent conversion of the duplex to hairpin (Figure 114C), whereas the higher melting temperature reflect the hairpin to single strand conversion (Figure 114D), a concentration independent process. At lower concentrations, an almost monophasic curve shape was obtained. Under these conditions, predominantly the hairpin structure was formed. This was in accordance with the lower hyperchromicity values at 1 and 2 µM single stand concentration compared to 5, 10 and 20 µM, reflecting a lower number of base pairs, as it is the case for the hairpin structure (Figure 114A). The melting temperatures obtained at 260 nm were just a rough approximation, especially for the 20- and 10-mM samples, as the second transition was not taken into account for the calculations. At 280 nm, the melting temperatures for the lower concentrations were almost concentration independent, as the resulting hairpin structures only contain C/G base pairs. Thus, the hairpin structures had a higher impact on the absorption. In contrast to the report of Cheong *et al.*, both duplex and hairpin structures are present, which rendered the thermodynamic evaluation by the Van't Hoff method impossible. The main difference between these two studies is, besides the 2'-OMe residues of Cheong *et al.*, the presence of the vinyl moiety of $m^6v^8A$. The vinyl group seemed to influence the duplex-hairpin conversion due to hydrophobic interactions with neighboring base pairs. For an un-

ambiguous comparison between the methylated and demethylated oligonucleotides for an evaluation of enzyme activity, conditions would need to be found where only one conformation is present for the modified oligonucleotide. Additionally, the 8-vinyladenosine ($v^8A$), the demethylated analogue of $m^6v^8A$, would be needed as unmodified reference to determine the differences in the fluorescent signals upon demethylation.

### 3.4.4 Fluorescence spectroscopy

The fluorescent properties of the free nucleosides $v^8m^6A$ and $v^8A$ were compared. For this, fluorescent spectra were measured in triplicates for 10 µM samples in $H_2O$ (60 µL).

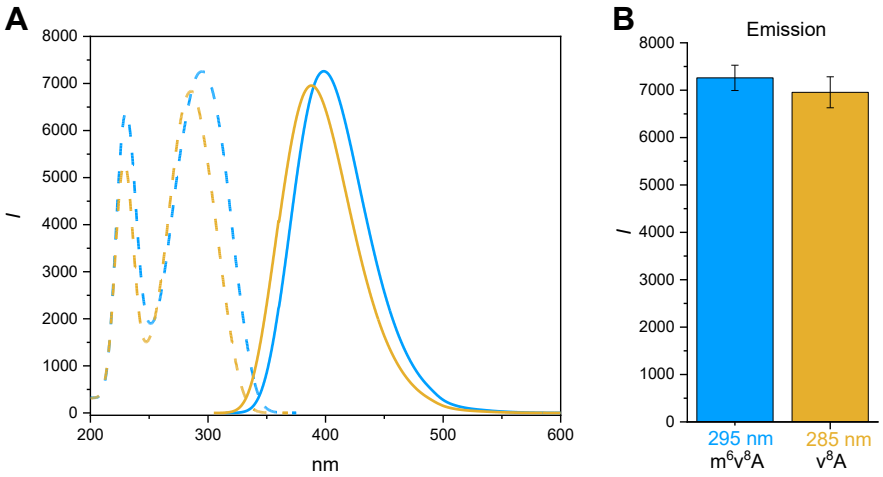



Figure 115. A: Absorption and emission spectra of the free nucleosides $m^6v^8A$ (blue) and $v^8A$ (yellow). B: Fluorescence intensity of the emission after excitation at 295 ($m^6v^8A$) and 285 nm ($v^8A$) as bar graphs with standard deviation (±).

Both nucleosides exhibited fluorescence in an almost identical spectral range. The absorption and emission wavelength of both nucleosides were shifted by 10 nm (Figure 115A). $v^8m^6A$ and $v^8A$ revealed the same Stokes shift of around 100 nm. Both showed comparable emission intensity after excitation at 295 nm and 285 nm, respectively. The mean values of the fluorescence intensity differed by around 300 with $m^6v^8A$ being slightly brighter (Table 19).

Table 19. Absorption and emission values obtained from the fluorescent spectra.

| Name | Absorption max | Emission max | Fluorescence intensity |
|---|---|---|---|
| **$m^6v^8A$** | 295 nm | 399 nm | 7260 ± 266 |
| **$v^8A$** | 285 nm | 388 nm | 6955 ± 326 |

For a good discrimination between the methylated and the demethylated sample, a larger change in fluorescence intensity or absorption and emission wavelength would have been desirable. Still, compared to the canonical nucleobases that are in principle non-fluorescent, a significant enhancement of fluorescence could be achieved be the incorporation of the vinyl group at position 8.

After incorporation of the $m^6v^8A$ phosphoramidite in the self-complementary sequence, the fluorescence spectrum of this R692 was measured and compared to the unmodified reference R961.

The latter did not reveal any fluorescence. The absorption maximum of the 12mer was slightly red shifted to 301 nm, compared to 295 nm of the free nucleotide $m^6v^8A$, while the emission maximum remained (398 nm, Figure 116A). Even though there was still a fluorescent signal available for R692, it was significantly reduced compared to the free nucleoside $m^6v^8A$ (Figure 116B). This was already shown for the deoxy analog 8vA due to a reduced quantum yield by Gaied et al.[182] For complete comparison, the oligonucleotide containing demethylated $v^8A$ would be needed.

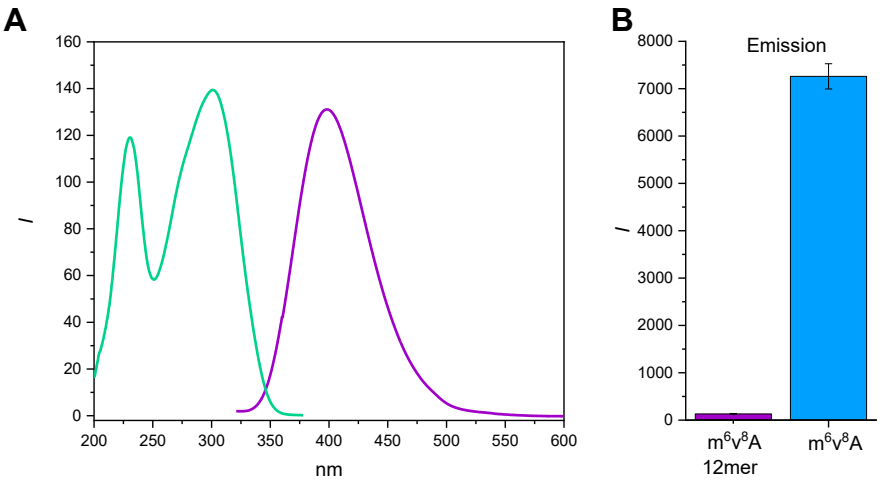



Figure 116. A: Absorption (green) and emission (purple) spectra of the 12mer containing $m^6v^8A$. B: Fluorescence intensity of the emission after excitation at 301 nm of R692 ($m^6v^8A$) compared to the fluorescence intensity of the free $m^6v^8A$ nucleoside (blue) after excitation at 295 nm as bar graphs with standard deviation (±).

Due to missing reference oligonucleotides, the obtained results are only preliminary and not much can be predicted about the potential use of this fluorescent nucleoside analogue yet. However, the successful incorporation of the $m^6v^8A$ phosphoramidite into an oligonucleotide and their required deprotection conditions were shown. Furthermore, fluorescent measurements were recorded with the oligonucleotide, as well as the $m^6v^8A$ and $v^8A$ free nucleosides for the first time. It needs to be tested if the $m^6v^8A$ modification is still accepted as a substrate by FTO or ALKBH5. The resulting product would then contain the missing demethylated fluorescent nucleoside analog reference. More experiments and comparisons must be performed to see if this modification can be used as an environment sensitive fluorescent probe to investigate the influence of methylation states on the fluorescent properties and structural effects of $m^6A$.

# 4.Conclusion and outlook

This thesis includes the synthesis of several naturally occurring as well as artificially modified nucleoside analogs as phosphoramidite or triphosphate building blocks. Some of the synthetic pathways were newly established, others were reproduced from already published reports and had to be optimized. All the obtained building blocks were successfully incorporated into RNA or DNA oligonucleotides, mainly by solid phase synthesis, and used for diverse research topics, working together with Prof. Dr. Markus Bohnsack, Dr. Katherine Bohnsack, Prof. Dr. Patrick Cramer, Prof. Dr. Christian Haering, and their coworkers.

The naturally occurring modifications $m^3C$, $m^4C$, $m^4{}_2C$, $i^6A$, $ms^2i^6A$, and $t^6A$ were synthesized as phosphoramidite building blocks. After their successful incorporation into RNA oligonucleotides, specific deprotection conditions were determined for each of them. Especially for the $m^3C$ deprotection, methylamine had to be avoided to prevent side product formation. By placing $m^3C$, $i^6A$, $ms^2i^6A$, and $t^6A$ into their respective positions in the tRNA-ASL, these oligonucleotides were used for spectroscopic and biochemical analysis of writer and eraser enzymes of $m^3C$. The RNA methyltransferase, METTL8, was found to be a mitochondrial matrix protein. Interestingly, the human mt-$tRNA^{Thr}$ and mt-$tRNA^{Ser(UCN)}$ were identified as selective substrates of this protein. This was the first report of METTL8 as tRNA methyltransferase, which was earlier thought to be responsible for the installation of $m^3C$ in mRNAs. Recognition elements within the ACL like U34 and G35 present in both tRNAs serve as recognition elements for METTL8. The efficient installation of the methyl group at position C32 was permitted by the respective modification on position A37. Thereby, $(ms^2)i^6A$ is a prerequisite for the methylation of C32 (mt-$tRNA^{Ser(UCN)}$) by METTL8, while $t^6A$ (mt-$tRNA^{Thr}$) enhances the reaction. The resulting changes in the methylation level of $m^3C$ indicated consequences for cellular processes: NMR analysis, CD spectroscopy, thermal denaturation experiments, and native page separation revealed a structural impact of $m^3C_{32}$ alone and in combination with the respective modifications $t^6A$ and $(ms^2)i^6A$ at position 37, which has not been reported before. The presence of the methyl group of $m^3C_{32}$ led to a fine-tuning of the ASL structure, which was supposed to relate to cellular functions as the level of some of the *de novo* synthesized mitochondrial proteins was decreased from cell lines lacking $m^3C$ on position 32 compared to WT. Thus, the introduction of the methyl group to form $m^3C_{32}$ by METTL8 seems to promote efficient mitochondrial translation by fine-tuning the mt-$tRNA^{Thr/Ser(UCN)}$ structure. Future crystallography experiments or structural probing assays[495] might clarify in greater detail the exact changes between the structures of methylated and demethylated ASLs. Even though the observed structural changes might have been comparatively small, the influence on an important cellular process with respect to protein levels was shown. Besides the remarkable biochemical assays performed by the collaboration partners, the chemical synthesis of the naturally occurring modifications enabled

the clarification of important interdependencies of the modifications within the ACL, which in turn explained the selectivity of METTL8 to the two tRNA substrates, as well as their structural impact by spectroscopic analysis.

The precise regulation of the methylation level of $m^3C_{32}$ to dynamically fine-tune cellular functions like mitochondrial translation remains elusive. If the METTL8 expression level varies or if the opposing mechanism, the removal of the methyl group by demethylases, is involved in the regulation still needs to be elucidated. As an example, Schöller *et al.* observed that the METTL8 activity on mt-tRNA$^{Thr/Ser(UCN)}$ was upregulated upon tumorigenesis in a pancreatic cancer cell line and cell growth can be reduced by knockout of METTL8.[429] To determine if the eraser enzymes might be involved in the dynamic modification pathway on tRNAs, the synthesized mitochondrial tRNA-ASLs with the $m^3C$, $t^6A$ and $(ms^2)i^6A$ modifications were tested in demethylation assays using ALKBH1 and ALKBH3. Interestingly, ALKBH3, whose subcellular localization was shown to be in the nucleus and cytoplasm, revealed activity for the demethylation of the $m^3C_{32}$ containing mt-tRNA ASLs. The removal of the methyl group could be further enhanced by leaving out $Mg^{2+}$ in the oxidation buffer. However, if this activity is physiologically relevant can only be stated by detailed examination of the cellular localizations of ALKBH3. Thus, these results might be an indication for an additional localization of ALKBH3 in the mitochondria. Earlier, the $m^3C$ demethylation activity of ALKBH3 was shown only for cyt-tRNAs. In contrast, ALKBH1 with a mitochondrial subcellular localization was not able to demethylate $m^3C_{32}$ of the mt-tRNA$^{Thr/Ser(UCN)}$ ASLs. However, no $m^3C$ demethylation activity for tRNAs has been observed before, but the methyl group could only be removed from $m^3C$ in ssDNA and mRNA. By investigating the demethylation activity of ALKBH1, indication for other potential activities of this protein besides being a demethylase was observed: By HR-ESI-MS experiments it appeared that ALKBH1 was able to remove the isopentenyl group of $i^6A$ from mt-tRNA$^{Thr/Phe}$ by oxidation. However, for both tRNAs, ALKBH1 had contrary response in the presence of Mg inside the oxidation buffer. Furthermore, an exonuclease activity for ALKBH1 was proposed. A higher reaction yield for the removal of the isopentenyl group must be achieved to analyze the proceeding reactions *in vitro*. Direct readout by LC-MS would facilitate the evaluation and assignment of the newly emerging peak in HPLC. To further examine the exonuclease activity, the incubation could be performed without cofactors, which would be needed for the demethylation reaction. So far, the obtained results of the different demethylation assays are just preliminary data, which needs to be further validated. They will help to learn more about the dynamic network of methylases and demethylases to regulate cellular functions.

My colleague Anam Liaqat used a selection of the natural modifications for gel-based *in vitro* selection experiments to find DNA enzymes that can cleave RNA substrates in a modification dependent manner. She was able to identify DNAzymes that selectively cleaved RNA strands if the

$i^6A$ modification was present, while the enzyme was inhibited by a canonical adenosine. Furthermore, a set of DNAzymes was reported, each of which was selective for one of the cytidine isomers $m^3C$, $m^4C$, and $m^5C$, as well as for the canonical cytidine. These DNAzymes are a useful tool for the detection of modifications as well as their modification sites.

In addition to the natural occurring nucleotides, nucleoside analogs can be modified artificially to fulfill a special purpose. Especially during the Covid-19 pandemic, antiviral nucleoside analogues became a highly investigated research topic for the treatment of the viral infection. One promising drug candidate was molnupiravir, the isopropyl ester prodrug of $N^4$-hydroxycytidine. It was synthesized as the $N^4$-hydroxycytidine phosphoramidite building block and incorporated into multiple RNA oligonucleotides. Workup conditions were found to separate the product from a side product, which was emerging during the deprotection conditions. These oligonucleotides were used for biochemical and structural analysis to elucidate the mechanism of action of the drug candidate molnupiravir, as at the beginning of these investigations, molnupiravir was being tested in clinical trial (phase 2/3) by MERCK and Ridgeback Biotherapeutics. A two-step model for the NHC-induced mutagenesis of SARS-CoV-2 was proposed based on RNA elongation, thermal denaturation, and cryo-EM experiments: the viral polymerase can install MTP opposite to G or A during viral replication. This incorporation does not lead to a stalling of the RNA extension. When the NHC containing RNA is used as a template, G and A can be installed opposite to the modification, leading to an accumulation of mutations, as molnupiravir seems to evade the ExoN proof-reading activity of RdRp. This so-called error catastrophe mechanism can be explained by the formation of two distance base pairs of NHC. It mimics cytidine and uridine and is either base pairing with guanine (NHC-G) in the amino form or with adenosine (NHC-A) in the imino form. Following this, UV melting analysis and NMR studies were performed using the Dickerson Drew sequence to learn more about the base pairing abilities of NHC and its influence on the duplex or hairpin stability. The NHC phosphoramidite was incorporated into different positions of this self-complementary sequence, which should result in NHC-G and NHC-A at a more terminal or central position. When NHC was incorporated at the more external position, the Dickerson Drew sequence was present as duplex structures, but revealed a distinctly destabilized melting temperature. This is the reason why the Dickerson Drew sequences with the modification being located at central positions are predominantly forming hairpin structures; the modification is placed in the loop region and excluded from base pairing. In the $^1$H-NMR spectra of the duplex structures with the NHC-G base pair an additional signal at 11.2 ppm was detected compared to the unmodified reference. The introduction of a $^{15}$N label at position N4 of NHC facilitated the assignment of this additional peak to the $N^4$-proton, which was significantly downfield shifted compared to the amino signal of a canonical C-G base pair. This also applied to the nitrogen chemical shift of the N4, which was downfield shifted to 156 ppm compared to a canonical cytidine at around 98 ppm. With this experiment, the additional

signal could unambiguously be assigned to the N4-amino proton, proving the formation of a NHC-G base pair with NHC in the *anti*-amino form, which has not been reported before for the NHC compound. To obtain a NHC-A base pair inside a duplex structure, the Dickerson Drew sequence was mutated by placing a U-A base pair at position 3. By NMR analysis, a conformational equilibrium of two species with an almost equal population was detected for the NHC-A 12mer. We proposed that one species is the NHC-A base pair with NHC in the *anti*-imino form, while the other could be the NHC-A base pair with NHC in the *syn*-imino form, inducing the movement of the opposing adenosine to the minor groove. However, presence of the two species in the NMR spectra, as well as the low NMR-concentration makes it impossible to conclusively determine the two structures. Besides higher sample concentrations, the synthesis of a doubly $^{15}$N-labeled NHC, at position N4 and N3,[496] might help to distinguish and assign the signals of the two species. Despite the encouraging results and the FDA-emergency-use authorization, it was hypothesized that NHC could get converted to the dNHC by RNR, which could get incorporated into the host DNA. To impede this conversion of NHC, the mutation of the 2'-position might prevent this mutagenic reaction. The NHC building block could be synthesized starting from gemcitabine, a 2',2'-diflurodesoxycytidine derivative with known broad spectrum antiviral activity. The 2'-position would be blocked for the RNR. Interestingly, antiviral nucleoside analogues can also occur naturally *in vivo*. Viperin is a member of the radical *S*-adenosyl-L-methionine (SAM) superfamily and catalyzes the formation of 3'-deoxy-3′,4'-didehydro-CTP (ddhCTP) from CTP.[497] This naturally occurring cytidine analog was shown to be a chain terminator for the RdRp of several viruses. NHC could be combined with this chain terminating sugar moiety to provide a twofold functional antiviral nucleoside analog.

The 7-deaza-guanosine crosslinker was synthesized as phosphoramidite and triphosphate building block over 12 and 9 steps, respectively. This nucleoside bears a propyl moiety on position 7 with a diol functional group. This can be unmasked by oxidation and the resulting aldehyde can potentially react with the amine of Lys or Arg residues within proteins. These reactions were successfully tested after the incorporation of the triphosphate into an DNA oligonucleotide by PCR. The oxidation with $NaIO_4$ enabled the linkage to FAM-TSC to get a direct readout on PAGE by a fluorescent signal. This building block can be used for the investigation of the protein complex condensin and its precise mechanism for structural maintenance of the chromosomes by loop extrusion. Therefore, one DNA oligonucleotide with the crosslinker on a specific position was synthesized by solid phase synthesis, based on a crystal structure of the Ycg1-Brn1$_{\text{short kleisin loop}}$ – DNA complex. Covalent crosslinking to the protein with subsequent mass measurements are currently ongoing with this DNA 18mer by Julia Locherer.

In the last part of this thesis, the fluorescent nucleoside analog $m^6v^8A$ was synthesized over six steps as a phosphoramidite building block. In addition, the free nucleoside of $m^6v^8A$ and $v^8A$ were

synthesized for fluorescent measurements. The absorption end emission maxima of both molecules only differed from each other by around 10 nm and the fluorescence intensity was in the same range. Incorporated into an oligonucleotide, the fluorescence intensity of the $m^6v^8A$ dropped compared to the free nucleoside, but it was still enhanced compared to the non-fluorescent canonical oligonucleotide. Next, this oligonucleotide should be subjected to a demethylation assay to see if the demethylation enzyme would still accept this modification as a substrate and to monitor the fluorescent changes upon demethylation of $m^6v^8A$. Furthermore, for comparison, the $v^8A$ phosphoramidite building block would also be needed to test if demethylation or methylation by the respective writer and eraser enzymes of $m^6A$ could be followed by fluorescent changes. These modified oligonucleotides could then be used for investigations of the influence of methylation states on the fluorescent properties and structural effects of $m^6A$, that may be revealed by an environment sensitive fluorescent probe.

# 5. Experimental part

## 5.1 Chemical Syntheses

### 5.1.1 General Remarks

#### 5.1.1.1 Materials

All synthetic procedures were performed under nitrogen atmosphere. The chemicals were purchased 'Pro analysis'- or 'For synthesis' grade and used without additional purification from commercial suppliers. Anhydrous solvents (DCM, MeCN, DMF, and THF) were obtained from a solvent purification system (SPS). 1,2-DCE was dried over molecular sieves overnight. Solvents for column chromatography (technical quality), recrystallization or extraction were distilled prior to use. The obtained compounds were stored at 4 °C (-20 °C in case of TOM-protected compounds and the phosphoramidite building blocks). Column chromatography was performed on silica gel (Kieselgel 60, Merck, 0.040-0.063 mm). For molecules containing the 5'-*O*-DMT group 1% triethylamine was added to the eluent system for packing to column to neutralize the silica gel. During the elution of the product, $NEt_3$ was left out. TLC was performed on Alugram DIL G/UV254 (Machery-Nagel) and visualized with UV light (254 nm and 365 nm) or staining reagents (e.g., Ninhydrin for primary amino group). 5% $H_2SO_4$ in EtOH was used for visualizing the trityl group resulting in an orange color (trityl cation).

#### 5.1.1.2 NMR and mass spectroscopy

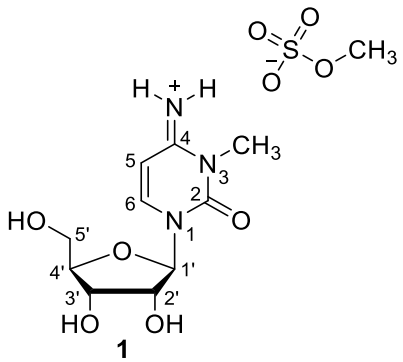
NMR spectra were recorded on a Bruker Avance III HD 400 spectrometer. Chemical shifts ($\delta$) are reported relative to the residual solvent signals: $CDCl_3$: 7.26 ppm ($^1$H), 77.16 ppm ($^{13}$C); DMSO-$d_6$: 2.50 ppm ($^1$H), 39.52 ppm ($^{13}$C); $CD_3OD$: 3.31 ($^1$H), 49.0 ($^{13}$C); $D_2O$: 4.79 ($^1$H), or on the unified scale for $^{31}$P. The data is given as s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad with the coupling constants (J) in Hz. To neutralize the acidic $CDCl_3$ solvent it was filtered through basic aluminum oxide prior to the usage for 5'-*O*-DMT protected molecules and phosphoramidites. For the precise assignment, the following 2D-NMR spectra were recorded in addition to the $^1$H and $^{13}$C spectra: COSY, HSQC (multiplicity-edited: CH and $CH_3$ give the same color, whereas $CH_2$ gives another color) and HMBC. The $^{15}$N-NMRs of the NHC building blocks were references to nitromethane. HR-ESI-MS spectra were recorded on a Bruker micrOTOF-Q III spectrometer. The detected mass-to-charge ratio (m/z) and the calculated monoisotopic mass values are indicated. The small molecules were dissolved in acetonitrile or methanol in HPLC grade.

## 5.1.2 Synthetic procedure

Parts of these syntheses have been published and will be adopted from the cited publication.

### 5.1.2.1 Preparation of $N^3$-methylcytidine phosphoramidite **5**[437]

**$N^3$-Methylcytidine methosulfate (1)**

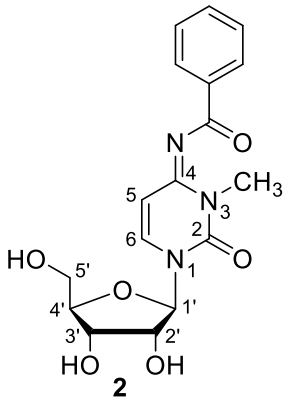


**1**

As described earlier,[498] cytidine (500 mg, 2.06 mmol, 1.00 eq.) was suspended in anhydrous dimethylformamide (5 mL) and warmed to 37 °C under nitrogen atmosphere. Afterwards, dimethyl sulfate (2.00 mL, 21.1 mmol, 10.2 eq.) was added. After stirring for 30 min, the resulting clear solution was diluted to 20 mL with methanol, and ethyl acetate was added until a precipitate formed. The title compound was obtained by filtration as a colorless solid (510 mg, 1.38 mmol, 67%).

**$^1$H-NMR** ($CD_3OD$, 400 MHz): $\delta$ (ppm) = 8.50 (d, J = 7.9 Hz; 1H, H-6), 6.16 (d, J =7.9 Hz; 1H, H-5), 5.87 (d, J = 2.7 Hz; 1H, H-1'), 4.18 (dd, J = 4.9, 2.7 Hz; 1H, H-2'), 4.15 (dd, J = 6.1, 4.9 Hz; 1H, H-3'), 4.08 (dt, J = 6.1, 2.5 Hz; 1H, H-4'), 3.94 (dd, J = 12.4, 2.5 Hz; 1H, $H^2$-5'), 3.78 (dd, J = 12.4, 2.5 Hz; 1H, $H^1$-5'), 3.68 (s, 3H, $^-OS(O)_2OCH_3$), 3.50 (s, 3H, $NCH_3$).
**$^{13}$C-NMR** (100 MHz, $CD_3OD$): $\delta$ (ppm) = 161.29 ($C=NH_2^+$),149.13 (C=O), 143.57 (C6), 94.75 (C5), 93.05 (C1'), 86.22 (C4'), 76.28 (C2'), 70.09 (C3'), 61.15 (C5'), 55.11 ($^-OS(O)_2OCH_3$), 30.93 ($N-CH_3$).
**HR-MS** ($ESI^+$): Exact mass calculated for $C_{10}H_{16}N_3O_5$ $[M+H]^+$: 258.1084, found: 258.1086.

***N*⁴-Benzoyl-*N*³-methylcytidine (2)**

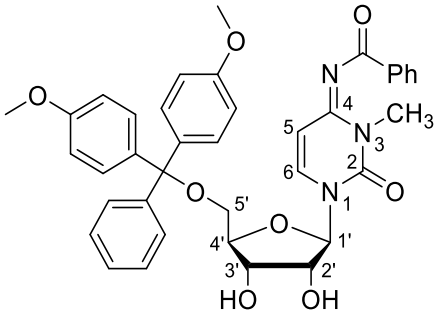


**2**

*N*³-Methylcytidinium methyl sulfate **1** (395 mg, 1.07 mmol, 1.00 eq.) was treated with trimethylsilyl chloride (TMSCl, 528 µL, 4.17 mmol, 3.90 eq.) in anhydrous pyridine (7 mL) under nitrogen atmosphere. After stirring for 1 h, triethylamine (297 µL, 2.14 mmol, 2.00 eq.) was added, and the reaction mixture was stirred for further 5 min at room temperature. Afterwards, benzoyl chloride (149 µL, 1.28 mmol, 1.20 eq.) was added dropwise. After stirring for 4 h at room temperature, the reaction mixture was quenched with water (1.5 mL). The resulting mixture was stirred for 5 min at room temperature followed by addition of aqueous ammonia (2.6 mL). After further 15 min stirring, the volatiles were removed under reduced pressure. The crude residue was purified by column chromatography (DCM/MeOH, 20:1) to give the title compound **2** as a colorless foam (344 mg, 952 µmol, 89%).

**$^1$H-NMR** (MeOD, 400 MHz): *δ* (ppm) = 8.09-8.13 (m, 2H, bz), 7.93 (d, J = 8.2 Hz; 1H, H-6), 7.53–7.58 (m, 1H, bz), 7.42–7.48 (m, 2H, bz), 6.41 (d, J = 8.2 Hz; 1H, H-5), 5.94 (d, J = 4.0 Hz; 1H, H-1'), 4.19 (dd, J = 5.2, 4.0 Hz; 1H, H-2'), 4.15 (t, J = 5.2 Hz; 1H, H-3'), 4.02 (dt, J = 5.2, 2.9 Hz; 1H, H-4'), 3.86 (dd, J = 12.3, 2.9 Hz; H$^2$-5'), 3.74 (dd, J = 12.3, 2.9 Hz; 1H, H$^1$-5'), 3.54 (s, 3H, N-$CH_3$).
**$^{13}$C-NMR** (MeOD, 100 MHz): *δ* (ppm) = 179.35 (C=O), 158.18 (C4), 151.94 (C2), 138.80 (C6), 137.22 ($C_q$-bz), 133.71 (bz), 130.72 (bz), 129.34 (bz), 98.54 (C5), 91.79 (C1'), 86.19 (C4'), 75.94 (C2'), 70.96 (C3'), 61.99 (C5'), 30.48 (N-$CH_3$).
**HR-MS** (ESI$^+$): Exact mass calculated for $C_{17}H_{19}N_3NaO_6$ [M+Na]$^+$: 384.1166, found: 384.1171.

***N*⁴-Benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-*N*³-methylcytidine (3)**



**3**

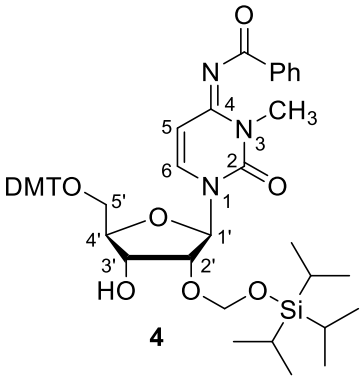
Compound **2** (265 mg, 734 µmol, 1.00 eq.) was co-evaporated with pyridine (2 mL) and re-dissolved in anhydrous pyridine (2.5 mL) under nitrogen atmosphere. *N*,*N*-dimethylformamide dimethyl acetal (DMF-DMA, 175 mg, 196 µL, 1.47 mmol, 2.00 eq.) was added. The resulting mixture was stirred for 18 h at room temperature and then evaporated to dryness. The residue was re-dissolved in dry pyridine (2.5 mL) and 4,4'-dimethoxytrityl chloride (298 mg, 881 µmol, 1.20 eq.) was added in one portion. After stirring for 2 h at room temperature, the reaction was stopped by the addition of methanol (3 mL). The reaction mixture was evaporated to dryness and then co-evaporated with toluene (3 mL). The residue was diluted with dichloromethane (30 mL) and then washed with aqueous citric acid (5%), water and saturated aqueous $NaHCO_3$. The organic layer was dried with $Na_2SO_4$ and concentrated under reduced pressure. The crude product was purified by column chromatography (DCM/MeOH, 80:1 + 2% $NEt_3$) to give the title compound as a colorless foam (433 mg, 653 µmol, 89%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 8.11-8.14 (m, 2H, bz), 7.66 (d, J = 8.2 Hz; H-6), 7.50-7.55 (m, 1H, bz), 7.41-7.46 (m, 2H, bz), 7.32-7.35 (m, 2H, trityl), 7.21-7.28 (m, 6H, trityl), 7.16-7.21 (m, 1H, trityl), 6.78-6.83 (m, 4H, trityl), 6.21 (d, J = 8.2 Hz; H-5), 5.85 (d, J = 3.6 Hz; 1H, H-1'), 4.35 (t, J = 5.0 Hz; 1H, H-3'), 4.28 (dd, J = 5.0, 3.6 Hz; 1H, H-2'), 4.23 (dt, J = 5.0, 3.0 Hz; 1H, H-4'), 3.77 (d, J = 2.2 Hz; 6H, $OCH_3$), 3.56 (s, 3H, N-$CH_3$), 3.47 (dd, J = 11.0, 3.0 Hz; 1H, $H^2$-5'), 3.38 (dd, J = 11.0, 3.0 Hz; 1H, $H^1$-5').

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 177.43 (*C*=O, bz), 158.77 ($C_q$-$OCH_3$, trityl), 155.86 (C4), 151.27 (C2), 144.18 ($C_q$-trityl), 136.02 ($C_q$-bz), 135.74 (C6), 135.30 ($C_q$-trityl), 132.55 (bz), 130.16 (trityl), 130.12 (trityl), 129.83 (bz), 128.32 (bz), 128.22 (trityl), 128.11 (trityl), 127.24 (trityl), 113.53 (trityl), 113.51 (trityl), 113.40 (trityl), 98.24 (C5), 91.85 (C1'), 87.14 ($C_q$-trityl), 84.73 (C4'), 76.41 (C2'), 70.86 (C3'), 62.48 (C5'), 55.37 ($OCH_3$), 30.17 (N-$CH_3$).

**HR-MS** (ESI$^+$): Exact mass calculated for $C_{38}H_{37}N_3NaO_8$ $[M+Na]^+$: 686.2473, found: 686.2467.

### *N*$^4$-Benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-*N*$^3$-methyl-2'-*O*-[[(triisopropylsilyl)oxy]methyl]cytidine (4)

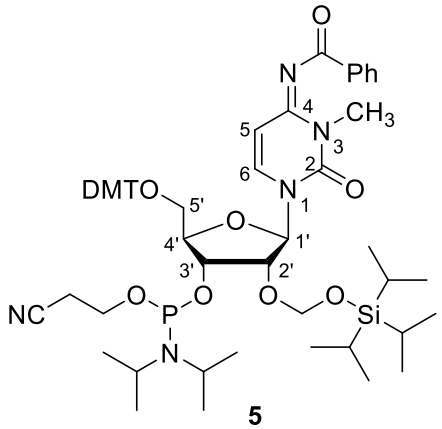



**4**

To a solution of compound **3** (345 mg, 520 µmol, 1.00 eq.) in anhydrous 1,2-dichloroethane (4 mL), *N*,*N*-diisopropylethylamine (311 µL, 1.82 mmol, 3.50 eq.) and di-*tert*-butyltin dichloride (174 mg, 572 µmol, 1.1 eq.) were added under nitrogen atmosphere. After stirring at room temperature for 1 h, the reaction mixture was heated to 80 °C, and TOM-Cl (150 mg, 676 µmol, 1.30 eq.) was added. After 20 min, the reaction mixture was cooled down to room temperature and diluted with dichloromethane (40 mL). The organic layer was separated, washed with saturated aqueous $NaHCO_3$ (3 x 20 mL) and dried over $Na_2SO_4$. The crude compound was purified by column chromatography (hexane/EtOAc, 5:1+ 2% $NEt_3$), to give the title compound as a colorless foam (207 mg, 243 µmol, 47%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 8.11-8.14 (m, 1H, bz), 7.79 (d, J = 8.2 Hz; H-6), 7.50-7.55 (m, 1H, bz), 7.41-7.46 (bz), 7.34-7.38 (trityl), 7.23-7.29 (m, 6H, trityl, bz), 7.17-7.22 (m, 1H, trityl), 6.80-6.84 (m, 4H, trityl), 6.06 (d, J = 8.2 Hz; H-5), 6.04 (d, J = 2.9 Hz; 1 H, H-1'), 5.25 (d, J = 4.7 Hz; 1H, $OCH_2O$), 5.06 (d, J = 4.7 Hz; 1H, $OCH_2O$), 4.39-4.48 (m, 1H, H-3'), 4.21 (dd, J = 5.0, 2.9 Hz; 1H, H-2'), 4.09 (dt, J = 6.6, 2.5 Hz; 1H, H-4'), 3.78 (d, J = 1.0 Hz; 6H, $OCH_3$), 3.54 (s, 3H, N-$CH_3$), 3.48-3.50 (m; 2H, H-5'), 3.20 (d, J = 6.4 Hz; 1H, C3'-O*H*), 1.05-1.13 (m, 21H, $Si(CH(CH_3)_2)_3$).
**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): *δ* (ppm) = 177.44 (C=O, bz), 158.77 (trityl, $C_q$-$OCH_3$), 155.83 (C4), 150.40 (C2), 144.18 (trityl, $C_q$), 136.07 (bz, $C_q$), 135.88 (C6), 135.64 (trityl, $C_q$), 135.34 (trityl, $C_q$), 132.50 (bz), 130.24 (trityl), 130.15 (trityl), 129.82 (bz), 128.35 (trityl, bz), 128.30 (trityl, bz), 128.13 (trityl), 127.24 (trityl), 113.40 (trityl), 98.29 (C5), 90.80 ($OCH_2O$), 88.72 (C1'), 87.15 ($C_q$-$OCH_2$), 83.66 (C4'), 83.27 (C2'), 69.10 (C3'), 61.98 (C5'), 55.36 ($OCH_3$), 30.02 (N-$CH_3$), 17.94 ($Si(CH(CH_3)_2)_3$), 11.99 ($Si(CH(CH_3)_2)_3$).
**HR-MS** ($ESI^+$): Exact mass calculated for $C_{48}H_{59}N_3NaO_9Si$ $[M+Na]^+$: 872.3913, found: 872.3895.

**$N^4$-Benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-$N^3$-methyl-2'-*O*-[[(triisopropylsilyl)oxy]methyl]cytidine 3'-*O*-(2-cyanoethyl diisopropylphosphoramidite) (5)**




**5**

To a solution of compound **4** (188 mg, 221 µmol, 1.00 eq.) in anhydrous dichloromethane (4 mL), *N*,*N*-dimethylethylamine (240 µL, 2.21 mmol, 10.0 eq.) and 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (78.6 mg, 332 µmol, 1.50 eq.) were added, and the reaction mixture was stirred for 2.5 h at room temperature under nitrogen atmosphere. Afterwards, the reaction mixture was evaporated to dryness under reduced pressure. The crude product was purified by column chromatography (hexane/EtOAc, 3:1 + 2% $NEt_3$) to yield the title compound **5** as a colorless foam (211 mg, 201 µmol, 91%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 8.11-8.15 (m, 4H, bz, diast.), 7.74 (d, 1H, J = 8.2 Hz; H-6), 7.66 (d, 1H, diast., J = 8.2 Hz; H-6), 7.50-7.55 (m, 2H, bz, diast.), 7.41-7.46 (m, 4H, bz, diast.), 7.33-7.40 (m, 4H, trityl, diast.), 7.24-7-30 (m, 7H, trityl), 7.17-7.23 (m, 2H, trityl, diast.), 6.79-6.85 (m, 8H, trityl, diast.), 6.18 (d, 1H, diast., H-1'), 6.16 (d, 1H, J = 4.2 Hz; H-1'), 6.13 (d, 1H, J = 8.2 Hz, H-5), 6.08 (d, 1H, diast., H-5), 5.09 (d, 2H, J = 4.8 Hz, $H^1$, $OCH_2O$, diast.), 5.05 (d, 1H, diast. $H^2$, $OCH_2O$), 5.03 (d, 1H, J = 4.8 Hz; $H^2$, $OCH_2O$), 4.45 (dt, 1H, J = 10.3, 5.4 Hz; H-3'), 4.35-4.38 (m, 1H, H-2'), 4.39-4.42 (m, diast. H-3', H-2'), 4.27 (dd, 1H, J = 5.3, 2.5 Hz; H-4'), 4.20 (dt, 1H, J = 4.7, 2.7 Hz; diast., H-4'), 3.94 (ddt, 1H, J = 10.5, 7.5, 6.1 Hz; diast., $H^1$, $POCH_2$), 3.81-3.88 (m, 1H, $H^1$, $POCH_2$), 3.79 (d, 6H, J = 0.8 Hz; $OCH_3$), 3.78 (d, 5H, diast., $OCH_3$), 3.61-3.68 (m, 1H, $H^2$, $POCH_2$), 3.53-3.60 (m, 4H, diast., H-5', $N(CH(CH_3)_2)_2$ and $H^2$, $POCH_2$), 3.53 (s, 3H, $N\text{-}CH_3$), 3.53 (s, 3H, diast., $N\text{-}CH_3$), 3.38 (td, 2H, J = 10.6, 2.90 Hz; H-5', diast.), 2.64 (td, 2H, J = 6.1, 4.0 Hz; $POCH_2CH_2$, diast.), 2.37 (td, 2H, J = 6.5, 2.0 Hz; $POCH_2CH_2$, diast.), 1.14-1.19 (m, 18H, $CH_3$ (isopropyl-N), diast.), 0.99-1.08 (m, 42H, $CH_3$ (isopropyl-N); $Si(CH(CH_3)_2)_3$, diast.).

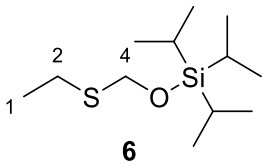
**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): *δ* (ppm) = 177.4 (C=O, bz), 158.8 ($C_q$-$OCH_3$), 155.9 (C4), 155.8 (C4), 150.4 (C2), 144.2 (trityl), 144.1 (trityl), 136.2 (bz, $C_q$), 136.1 (bz, $C_q$), 135.6 (trityl), 135.4 (trityl), 135.3 (trityl), 132.4 (bz), 130.3 (trityl), 130.2 (trityl), 129.8 (bz), 128.5 (trityl), 128.4 (bz, trityl), 128.3 (bz, trityl), 128.1 (bz, trityl), 127.3 (trityl), 117.8 (CN), 117.5 (CN), 113.4 (trityl), 89.5 ($OCH_2O$), 89.2 ($OCH_2O$), 88.3 (C1'), 88.1 (C1'), 87.3, 87.1 (trityl, $C_qOCH_2$), 83.3 (C4'), 83.2 (C4'), 78.5 (C2'), 77.9 (C2'), 70.7 (C3'), 70.2 (C3'), 61.9 (C5'), 59.1, 58.9, 58.1, 57.4 ($POCH_2$), 55.4 ($OCH_3$), 43.4 ($N(CH(CH_3)_2)_2$), 43.2 ($N(CH(CH_3)_2)_2$), 30.1 ($N\text{-}CH_3$), 24.8, 24.7 ($CH_3$, isopropyl-N), 20.5, 20.3, ($POCH_2CH_2$), 18.0 ($Si(CH(CH_3)_2)_3$), 12.1 ($Si(CH(CH_3)_2)_3$).

**$^{31}$P-NMR** ($CDCl_3$, 162 MHz): 150.12, 150.50.

**HR-MS** ($ESI^+$): Exact mass calculated for $C_{57}H_{76}N_5NaO_{10}PSi$ $[M+Na]^+$: 1072.4991, found: 1072.4968.

5.1.2.2 Preparation of (Triisopropylsiloxy)-methyl chloride **7**[369]

**Triisopropylsilyl (Ethylthio)methyl ether (6)**

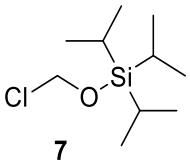



**6**

Paraformaldehyde (3.00 g, 99.9 mmol, 1.0 eq) was suspended in ethanethiol (6.00 mL, 99.9 mmol, 1.0 eq) and treated with aqueous NaOH (2 M, 2 drops) at 0 °C. The reaction was brought back to room temperature before being heated to 40 °C and the mixture was stirred for one hour. The reaction was allowed to cool to room temperature and DCM (98 mL), imidazole (13.6 g, 200 mmol, 2.0 eq) and($^{i}$Pr)$_3$SiCl (21.4 mL, 99.9 mmol, 1.0 eq) were added. The mixture was stirred overnight at room temperature. Then n-hexane (196 mL) was added, and the organic phase was washed with a $NaH_2PO_4$-solution (10% in $H_2O$, 3x100mL). The organic phase was dried over $Na_2SO_4$. The solvent was evaporated under reduced pressure. The residue was distilled in vacuum (0.2 mBar). The product (**6**, 13.5 g, 54.3 mmol, 54%) was obtained as a colorless liquid with a boiling temperature of 88-94 °C at 1.7 mbar.[369]

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 4.87 (s, 2H, $CH_2$, H-4), 2.71 (q, $J$ = 7.4 Hz, 2H, $CH_2$, H-2), 1.31 (t, $J$ = 7.4 Hz, 3H, $CH_3$, H-1), 1.10 – 1.06 (m, 21H, $^{i}$Pr$_3$Si).

**[(Triisopropylsilyl)oxy]methyl chloride (7)**




**7**

A solution of triisopropylsilyl(ethylthiomethyl)ether (**6**, 13.5 g, 54.3 mmol, 1.0 eq) in anhydrous DCM (50 mL) was cooled to 0 °C. $SO_2Cl_2$ (4.33 mL, 53.5 mmol, 0.99 eq) was added and the mixture was stirred at room temperature for 1 h. The reaction was evaporated under reduced pressure. The residue was distilled in vacuum (0.33 mBar) to obtain the product (**7**, 3.23 g, 13.0 mmol, 24%) as a colorless liquid with a boiling temperature of 48-54 °C at 0.33 mbar.[369]

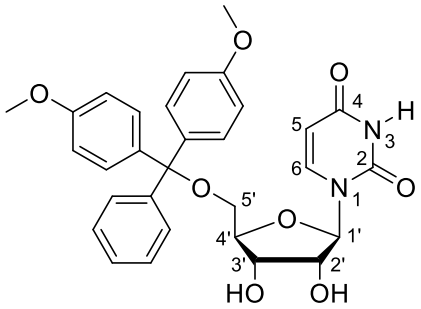
**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 5.66 (s, 2H, $CH_2$), 1.12 – 1.07 (m, 21H, $^{i}$Pr$_3$Si).

5.1.2.3 Preparation of $N^4$-methylcytidine phosphoramidite and the free $N^4$-methyl cytidine nucleoside

**$N^4$-methyl cytidine phosphoramidite (15)**

The phosphoramidite building block was synthesized according to published procedures of Pitsch *et al.* (tritylation and TOM-protection),[369] and Höbartner *et al.*[372] The NMR data were in agreement with the reports. The $^1$H-NMR of the phosphoramidite building block **15** contained four species, two diastereomers and presumably two rotamers, as assumed from Höbartner *et al.*[372] Some signals showed an additional splitting (H-6, $N^4CH_3$), which might arise from rotamers around the C4-NH($CH_3$) bond or the glycosidic bond.[372]

**5'-*O*-(4,4'-dimethoxytrityl)-uridine (12)**




**12**

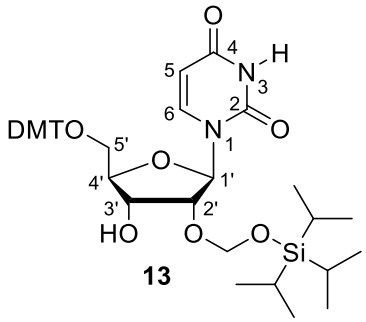
Uridine (1.00 g, 4.10 mmol, 1.00 eq) was dissolved in anhydrous pyridine (15 mL) under nitrogen atmosphere. DMT-Cl (1.67 g, 4.91 mmol, 1.20 eq) was added in 3 portions within 2 h. After stirring for one more hour at room temperature, MeOH (5 mL) was added to stop the reaction. The mixture was diluted with DCM and consecutively washed with a 5% citric acid solution, water and then $NaHCO_3$. The combined organic phases were dried over $Na_2SO_4$, and the solvent was removed under reduced pressure. The crude compound was purified by column chromatography (DCM:MeOH, 98:2, with 1% $NEt_3$ to 64:4) to yield **12** as a colorless foam (1.98 g, 3.63 mmol, 88%).[369]

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 7.99 (d, 1H, $J$ = 8.1 Hz, H-6), 7.39 – 7.35 (m, 2H, DMT), 7.31 – 7.25 (m, 4H, DMT), 7.24 – 7.20 (m, 1H, DMT), 6.85 – 6.81 (m, 4H, DMT), 5.88 (d, 1H, $J$ = 2.8 Hz, H-1'), 5.36 (d, 1H, $J$ = 8.1 Hz, H-5), 4.41 (dd, 1H, $J$ = 6.1, 5.2 Hz, H-3'), 4.31 (dd, 1H, $J$ = 5.2, 2.8 Hz, H-2'), 4.17 (dt, 1H, $J$ = 5.8, 2.6 Hz, H-4'), 3.78 (d, 6H, $J$ = 0.7 Hz, $OCH_3$), 3.52 (dd, 1H, $J$ = 11.0, 2.5 Hz, H-5'), 3.48 (d, 1H, $J$ = 2.9 Hz, H-5').

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 163.89 (C4), 158.77 ($C_q$-$OCH_3$), 151.25 (C2), 144.46 ($C_q$, DMT), 149.49 (C6) 135.43 ($C_q$, DMT), 135.24 ($C_q$, DMT), 130.28 (DMT), 130.21 (DMT), 128.24 (DMT), 128.13 (DMT), 127.25 (DMT), 113.42 (DMT), 113.40 (DMT), 102.35 (C5), 90.65 (C1'), 87.14 ($C_q$, DMT), 84.07 (C4'), 75.64 (C2'), 69.98 (C3'), 62.14 (C5'), 55.39 ($CH_3$).

**HR-MS** (ESI$^+$): Exact mass calculated for $C_{30}H_{30}N_2NaO_8$ [M+Na]$^+$: 569.1894, found: 569.1908.

**5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-[[(triisopropylsilyl)oxy]methyl]uridine (13)**




**13**

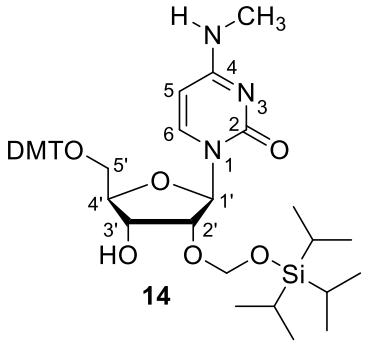
5'-*O*-(4,4'-dimethoxytrityl)-uridine (915 mg, 1.68 mmol, 1.00 eq, **12**) was dissolved in anhydrous 1,2-DCE under nitrogen atmosphere. $^i$Pr$_2$NEt (1.00 mL, 5.86 mmol, 3.50 eq) and $tBu_2SnCl_2$ (560 mg, 1.84 mmol, 1.10 eq) were added and the mixture was stirred for 1 h at room temperature. The reaction was heated to 80 °C before adding TOM-Cl (447 mg, 2.01 mmol, 1.20 eq). After stirring for 30 min at 80 °C, the reaction mixture was allowed to cool to room temperature and diluted with DCM. The organic solution was washed with a saturated $NaHCO_3$-solution. The organic phase was dried over $Na_2SO_4$, and the solvent was evaporated under reduced pressure. The crude product was purified by column chromatography (hexane/EtOAc, 3:1 with 1% $NEt_3$, stepwise to 1:1) to yield **13** as a colorless foam (392 mg, 535 µmol, 32%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 7.94 (d, 1H, *J* = 8.2 Hz, H-6), 7.39 – 7.36 (m, 2H, DMT), 7.32 – 7.29 (m, 2H, DMT), 7.29 – 7.26 (m, 4H, DMT), 7.26 – 7.21 (m, 1H, DMT), 6.86 – 6.81 (m, 4H, DMT), 6.03 (d, 1H, *J* = 3.4 Hz, H-1'), 5.28 (d, 1H, *J* = 8.2 Hz, H-5), 5.23 (d, 1H, *J* = 4.8 Hz, $OCH_2O$), 5.03 (d, 1H, *J* = 4.8 Hz,$OCH_2O$), 4.47 (q, 1H, *J* = 5.4 Hz, H-3'), 4.26 (dd, 1H, *J* = 5.0, 3.5 Hz, H-2'), 4.12 – 4.09 (m, 1H, H-4'), 3.80 (d, 6H, *J* = 0.5 Hz, $OCH_3$), 3.51 (d, 2H, *J* = 2.4 Hz, H-5'), 3.18 (d, 1H, *J* = 5.7 Hz, 3'-OH), 1.15 – 1.04 (m, 21H, $Si(CH(CH_3)_2)_3$).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 162.89 (C4), 158.84 ($C_qOCH_3$), 158.81 ($C_qOCH_3$), 150.12 (C2), 144.43 ($C_q$, DMT), 140.28 (C6), 135.42 ($C_q$, DMT), 135.15 ($C_q$, DMT), 130.30 (DMT), 130.25 (DMT), 128.27 (DMT), 128.14 (DMT), 127.30 (DMT), 113.42 (DMT), 113.41 (DMT), 102.34 (C5), 90.80 ($OCH_2O$), 87.93 (C1'), 87.26 ($C_q$, DMT), 83.81 (C4'), 83.06 (C2'), 69.54 (C3'), 62.25 (C4'), 55.39 ($OCH_3$), 17.94 ($CH(CH_3)_2$), 11.98 ($CH(CH_3)_2$).

**HR-MS** (ESI$^+$): Exact mass calculated for $C_{40}H_{52}N_2NaO_9Si$ [M+Na]$^+$: 755.3334, found: 755.3324.

**5'-*O*-(4,4'-dimethoxytrityl)-*N*$^4$-methyl-2'-*O*-[[(triisopropylsilyl)oxy]methyl]cytidine (14)**




**14**

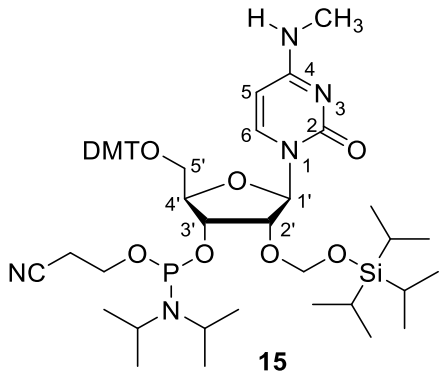
5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-[[(triisopropylsilyl)oxy]methyl]uridine (155 mg, 211 µmol, 1.00 eq) was dissolved in anhydrous 1,2-DCE (2 mL) under nitrogen atmosphere. DMAP (2.58 mg, 21.1 mmol µmol, 0.10 eq) and $NEt_3$ (295 µL, 2.11 mmol, 10.0 eq) were added. The reaction mixture was treated with Trisyl-Cl (96.1 mg, 317 µmol, 1.50 eq), stirred for 2 h at room temperature, and diluted with DCM. The mixture was washed with a 50% $NaHCO_3$-solution and the organic phase was dried with $Na_2SO_4$. After evaporation of the solvent, $MeNH_2$ in EtOH (33%, 2.5 mL) was added and the reaction was stirred for 15.5 h. The solution was evaporated under reduced pressure and the crude product was purified by column chromatography (EtOAc + 1% $NEt_3$) to yield **14** (125 mg, 168 µmol, 79%) as a colorless foam.

**$^1$H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 8.02 (d, 1H, *J* = 7.3 Hz, H-6) 7.43 – 7.40 (m, 2H, DMT), 7.33 – 7.28 (m, 4H, DMT), 7.27 – 7.19 (m, 1H, DMT), 6.84 – 6.82 (m, 4H, DMT), 5.98 (d, 1H, *J* = 9.8 Hz, H-1'), 5.27 (d, 1H, *J* = 4.7 Hz, $OCH_2O$), 5.17 (d, 1H, *J* = 7.5 Hz, H-5), 5.14 (d, 1H, *J* = 4.7 Hz, $OCH_2O$), 4.92 (d, 1H, *J* = 5.3 Hz, NH), 4.36 (d, 1H, *J* = 5.5 Hz, H-3'), 4.21 (d, 1H, *J* = 4.5 Hz, H-2'), 4.04 (d, 1H, *J* = 8.6 Hz, H-4'), 3.82 – 3.76 (m, 6H, $OCH_3$), 3.58 (d, 1H, *J* = 10.5 Hz, H-5'), 3.47 (d, 1H, *J* = 10.7 Hz, H-5'), 3.34 (d, 1H, *J* = 8.3 Hz, 3'-OH), 2.99 (d, 3H, *J* = 4.6 Hz, $CH_3$), 1.12 – 1.02 (m, 21H, $Si(CH(CH_3)_2)_3$).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): *δ* (ppm) =164.45 (C4), 158.69 ($C_qOCH_3$), 156.11 (C2), 144.71 ($C_q$, DMT), 139.94 (C6), 135.80 ($C_q$, DMT), 135.61 ($C_q$, DMT), 130.39 (DMT), 130.35 (DMT), 128.45 (DMT), 128.03 (DMT), 127.06 (DMT), 113.32 (DMT), 94.51 (C5), 90.96 ($OCH_2O$), 89.73 (C1'), 86.85 ($C_q$, DMT), 83.77 (C2'), 83.01 (C4'), 68.08 (C3'), 61.56 (C5'), 55.38 ($OCH_3$), 28.11 ($CH_3$), 17.94 ($CH(CH_3)_2$), 11.98 ($CH(CH_3)_2$).

**HR-MS** ($ESI^+$): Exact mass calculated for $C_{41}H_{56}N_3O_8Si$ $[M+H]^+$: 746.38312, found: 746.38394.

**5'-*O*-(4,4'-dimethoxytrityl)-$N^4$-methyl-2'-*O*-[[(triisopropylsilyl)oxy]methyl]cytidine 3'-*O*-(2-cyanoethyl diisopropylphosphoramidite) (15)**




**15**

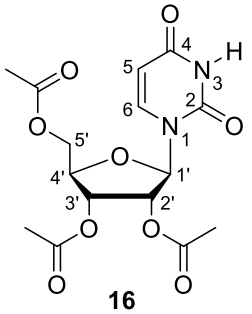
5'-*O*-(4,4'-dimethoxytrityl)-$N^4$-methyl-2'-*O*-[[(triisopropylsilyl)oxy]methyl]cytidine **14** (110 mg, 147 µmol, 1.00 eq) was dissolved in anhydrous DCM (2 mL) under nitrogen atmosphere. The mixture was treated with $Me_2NEt$ (160 µL, 1.47 mmol, 10.0 eq) and CEP-Cl (52.4 mg, 221 µmol, 1.50 eq). After 3 h, the solution was evaporated under reduced pressure. The crude product was purified by column chromatography (hexane:EtOAc, 4:1 + 1% $NEt_3$) to yield the product **15** (119 mg, 126 µmol, 86%) as a colorless foam.[372]

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 8.22, 8.13 (2 br, H-6, rotamers), 7.98, 7.89 (2d, 2H, H-6), 7.45 – 7.37 (m, 4H, DMT), 7.33 – 7.22 (m, 9H, DMT), 6.86 – 6.79 (m, 8H, DMT), 6.18 (d, 1H, *J* = 3.1 Hz, H-1'), 6.12 (d, 1H *J* = 2.3 Hz, diast, H-1'), 5.21 – 5.07 (m, 6H, H-5, $OCH_2O$), 4.84 (br, 2H, NH), 4.54 – 4.37 (2m, 2H, H-3'), 4.36 – 4.24 (2m, 2H, H-2'), 4.23 – 4.10 (2m, 4H, H-4'), 3.98 – 3.81 (m, 2H, $POCH_2$), 3.80, 3.79 (2s, 12H, 2x$OCH_3$), 3.69 – 3.60 (m, 2H, H-5'), 3.60 – 3.44 (m, 6H, $POCH_2$, $P(NCH)_2$), 3.42 – 3.31 (m, 2H, H-5'), 2.98/2.80 (d, *J* = 4.9 Hz / br s, 6H, N4$CH_3$, rotamers), 2.62 (dt, 2H, *J* = 8.2, 6.4 Hz, diast. $CH_2CN$), 2.36 (t, *J* = 6.5 Hz, 1H, $CH_2CN$), 1.27 (dd, 16H, *J* = 6.8, 5.6 Hz, $NCH(CH_3)_2$), 1.18 – 0.91 (m, 66H, $Si(CH(CH_3)_2)_3$).

**$^{31}$P-NMR** (162 MHz, $CDCl_3$) $\delta$ (ppm) = 149.99, 149.55.

**HR-MS** ($ESI^+$): Exact mass calculated for $C_{50}H_{73}N_5O_9PSi$ $[M+H]^+$: 946.49097, found: 946.49019.

**2',3',5'-tri-*O*-acetyl-uridine (16)**

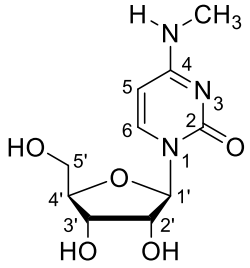



**16**

Uridine (500 mg, 2.05 mmol, 1.00 eq) was dissolved in anhydrous pyridine (3 mL) under nitrogen atmosphere. Acetic acid anhydride (1.35 mL, 14.3 mmol, 7.00 eq) were slowly added to the solution and stirred for 18.5 h. The solvent was evaporated under reduced pressure. The crude product was purified by column chromatography (EtOAc) to yield **16** in quantitative yields.[373]

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 8.90 (s, 1H, NH), 7.39 (d, $J$ = 8.2 Hz, H-6), 6.04 (d, 1H, $J$ = 5.0 Hz, H-1'), 5.79 (dd, 1H, $J$ = 8.2, 2.1 Hz, H-5), 5.34-5.33 (m, 1H, H-3'), 5.33 – 5.32 (m, 1H, H-2'), 4.36-4.33 (m, 3H, H-4', H-5'), 2.14 (s, 3H, Ac), 2.13 (s, 3H, Ac), 2.10 (s, 3H, Ac).
**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 170.26, 169.79, 169.78 (3x $C_q$-Ac), 162.68 (C4), 150.19 (C2), 139.41 (C6), 103.55 (C5), 87.58 (C1'), 80.07 (C4'), 72.85 (C3'), 70.31 (C2'), 63.27 (C5'), 20.95, 20.67, 20.58 (3x $CH_3$, Ac).
**HR-MS** ($ESI^+$): Exact mass calculated for $C_{15}H_{18}N_2NaO_9$ $[M+Na]^+$: 393.09045, found: 393.09124.

***N*$^4$-methyl cytidine (94)**



**94**

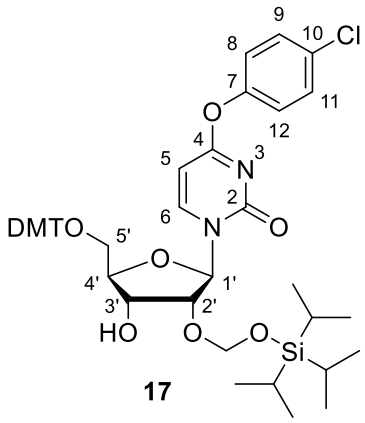
2',3',5'-tri-*O*-acetyl-*N*$^4$-methyl cytidine (400 mg, 1.08 mmol, 1.00 eq) was dissolved in anhydrous 1,2-DCE (5 mL) under nitrogen atmosphere. DMAP (13.2 mg, 108 µmol, 0.10 eq) and $NEt_3$ (1.51 mL, 10.8 mmol, 10.0 eq) were added and the mixture was stirred for 10 min at room temperature. After the addition of 2,4,6-triisopropyl-benzolsulfonylchlorid (Trisyl-Cl, 491 mg, 1.62 mmol, 1.50 eq), the reaction mixture was stirred for 2 h at room temperature. The mixture was diluted with DCM and washed with a 50% $NaHCO_3$-solution. The organic fractions were combined and dried over $Na_2SO_4$. The organic solvent was removed under reduced pressure. The residue was dissolved in $MeNH_2$ in EtOH (33%, 5 mL) and stirred for 16 h at room temperature. The solvent was evaporated under reduced pressure and the crude product was purified by column chromatography (DCM:MeOH, 5:1) to yield **94** as a colorless foam (224 mg, 871 µmol, 81%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 7.78 (d, 1H, $J$ = 7.5 Hz, H-6), 7.69 (q, 1H, $J$ = 4.7 Hz, NH), 5.76 (d, 1H, $J$ = 3.5 Hz, H-1'), 5.71 (d, 1H, $J$ = 7.5 Hz, H-5), 5.31 – 5.27 (m, 1H, 2'-OH), 5.05 (t, 1H, $J$ = 5.2 Hz, 5'-OH), 5.00 – 4.96 (m, 1H, 3'-OH), 3.94 – 3.89 (m, 2H, H-2', H-3'), 3.80 (q, 1H, $J$ = 3.8 Hz, H-4'), 3.64 (ddd, 1H, $J$ = 12.0, 5.2, 3.1 Hz, H-5'), 3.53 (ddd, 1H, $J$ = 12.1, 5.3, 3.5 Hz, H-5'), 2.74 (d, 3H $J$ = 4.7 Hz, $CH_3$).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 163.78 (C4), 155.39 (C2), 140.07 (C6), 94.47 (C5), 89.09 (C1'), 83.99 (C4'), 74.00 (C3'), 69.41 (C2'), 60.62 (C5'), 26.90 ($CH_3$).
**HR-MS** ($ESI^+$): Exact mass calculated for $C_{10}H_{15}N_3NaO_5$ $[M+Na]^+$: 280.09039, found: 280.09047.

5.1.2.4 Preparation of $N^4$,$N^4$-dimethylcytidine phosphoramidite (**19**)

**4-chlorophenyl-5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-[[(triisopropylsilyl)oxy]methyl]cytidine (17)**



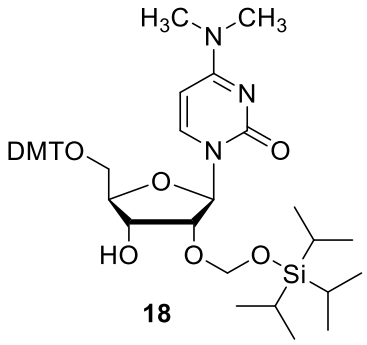
**17**

5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-[[(triisopropylsilyl)oxy]methyl]uridine **13** (400 mg, 546 µmol, 1.0 eq) was dissolved in anhydrous DCM (6 mL) and cooled to 0 °C after the addition of 4-Dimethylaminopyridine (DMAP, 10.0 mg, 81.9 µmol, 0.15 eq) and $NEt_3$ (761 µL, 5.46 mmol, 10.0 eq). Trisyl-Cl (218 mg, 720 µmol, 1.32 eq) was added and the reaction was stirred for 10 min at 0°C, followed by 2 h at room temperature. The reaction was diluted with DCM and washed with a saturated $NaHCO_3$-solution. The organic phase was dried over $Na_2SO_4$, and the solvent was removed under reduced pressure. The residue was re-dissolved in anhydrous DCM (12 mL) and mixed with 4-chlorophenol (351 mg, 2.73 mmol, 5.0 eq) and *N*,*N*-dimethylethylamine (887 µL, 8.19 mmol, 15.0 eq). DBU (81.5 µL, 546 µmol, 1.0 eq) was added in four portions within 5 min. The reaction was stirred for 2 h at room temperature and then washed with a saturated $NaHCO_3$-solution. The organic phase was dried over $Na_2SO_4$, and the solvent removed under reduced pressure. The residue was purified by column chromatography (DCM to DCM/MeOH = 98:2 + 1% $NEt_3$) to obtain the product (**17**, 396 mg, 469 µmol, 86%) as a colorless foam.

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 8.44 (d, *J*=7.4, 1H, H-6), 7.37 – 7.33 (m, 2H,O-$CH_3$), 7.27 (d, *J*=6.8, 2H, H-9, H-11), 7.26 – 7.22 (m, 7H, DMT), 7.06 – 6.97 (m, 2H, H8, H-12), 6.83 – 6.74 (m, 4H, DMT), 5.87 (s, 1H, H-1'), 5.58 (d, *J*=7.3, 1H, H-5), 5.20 (d, *J*=4.7, 1H, $OCH_2O$), 5.05 (d, *J*=4.7, 1H, $OCH_2O$), 4.35 (td, *J*=8.6, 4.8, 1H, H-3'), 4.14 (d, *J*=4.9, 1H, H-2'), 4.01 (dt, *J*=8.8, 2.4, 1H, H-4'), 3.74 (d, *J*=1.3, 6H, $OCH_3$), 3.59 – 3.44 (m, 2H, H-5'), 3.29 (d, *J*=8.4, 1H, C3'-OH), 0.98 – 1.02 (m, 21H, $Si(CH(CH_3)_2)_3$).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 171.37 (C4), 158.81 ($C_q$-$OCH_3$), 155.15 (C2), 150.19 ($C_q$, C7), 144.73 (C6), 135.60 ($C_q$-DMT), 135.35 ($C_q$-DMT), 131.36 ($C_q$, C10), 130.40 (DMT), 130.33 (DMT), 129.74 (C9, C11), 128.40 (DMT), 128.13 (DMT), 127.23 ($C_q$-DMT), 123.30 (C8, C12), 113.41 (DMT), 113.40 (DMT), 94.98 (C5), 90.96 ($OCH_2O$) 90.27 (C1'), 87.11 ($C_q$-DMT), 83.59 (C2'), 83.38 (C4'), 67.68 (C3'), 61.10 (C5'), 55.40 ($OCH_3$), 17.91, 11.94 (Si(CH(CH3)2)3).
**HR-MS** (ESI$^+$): Exact mass calculated for $C_{46}H_{55}N_2NaClO_9Si$ $[M+Na]^+$: 865.3258, found: 865.3271.

**5'-*O*-(4,4'-dimethoxytrityl)-*N*$^4$,*N*$^4$-dimethyl-2'-*O*-[[(triisopropylsilyl)oxy]methyl]cytidine (18)**

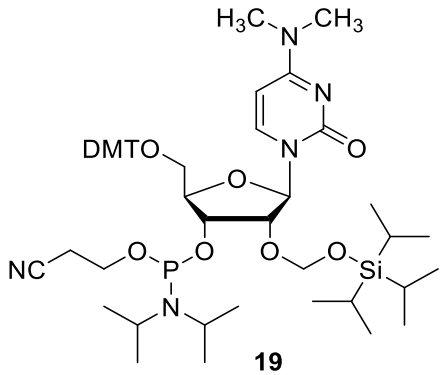



**18**

The convertible nucleoside 4-chlorophenyluridine **17** (105 mg, 124 µmol, 1.0 eq) was dissolved in dioxane (1.4 mL) and mixed with dimethylamine (40% in $H_2O$, 697 µL), The reaction mixture was stirred for 45 min at room temperature. After evaporation of the solvent in vacuum, the crude product was purified by column chromatography (DCM + 1% $NEt_3$ → DCM:MeOH = 99:1) to yield the product **18** as a colorless foam (92.8 mg, 122 µmol, 98%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 8.14 (d, *J* = 7.8 Hz, 1H, H-6), 7.44 – 7.40 (m, 2H, DMT), 7.34 – 7.27 (m, 6H, DMT), 7.26 – 7.21 (m, 1H, DMT), 6.87 – 6.82 (m, 4H, DMT), 5.98 (d, *J* = 0.9 Hz, 1H, H-1'), 5.31 (d, *J* = 7.8 Hz, 1H, H-5), 5.27 (d, *J* = 4.7 Hz, 1H, $OCH_2O$), 5.13 (d, *J* = 4.7 Hz, 1H, $OCH_2O$), 4.37 (td, *J* = 8.5, 4.9 Hz, 1H, H-3'), 4.21 (d, *J* = 4.9 Hz, 1H, H-2'), 4.04 (dt, *J* = 8.8, 2.5 Hz, 1H, H-4'), 3.79 (s, 6H, 2x$OCH_3$), 3.61 (dd, *J* = 11.0, 2.2 Hz, 1H, H-5'), 3.50 (dd, *J* = 11.1, 2.7 Hz, 1H, H-5'), 3.37 (d, *J* = 8.4 Hz, 1H, OH-3'), 3.19 (s, 3H, $N^4CH_3$), 2.95 (s, 3H, $N^4CH_3$), 1.12 – 1.00 (m, 21H, $Si(CH(CH_3)_2)_3$).
**HR-MS** (ESI$^+$): Exact mass calculated for $C_{42}H_{58}N_3O_8Si$ $[M+H]^+$: 760.3988, found: 760.4052.

**5'-*O*-(4,4'-dimethoxytrityl)-$N^4$,$N^4$-dimethyl-2'-*O*-[[(triisopropylsilyl)oxy]methyl]cytidine 3'-*O*-(2-cyanoethyl diisopropylphosphoramidite) (19)**



**19**

5'-*O*-(4,4'-dimethoxytrityl)-$N^4$,$N^4$-dimethyl-2'-*O*-[[(triisopropylsilyl)oxy]methyl]cytidine **18** (90.0 mg, 118 µmol, 1.0 eq) was dissolved in anhydrous DCM. $Me_2Net$ (128 µL, 1.18 mmol, 10.0 eq) and CEP-Cl (42.0 mg, 39.6 mL, 178 µmol, 1.5 eq) were added and the reaction mixture was stirred for 2.5 h at room temperature. The crude compound was purified by column chromatography (EtOAc:hexane = 3:1 + 1% $NEt_3$) and the product **19** was obtained as a colorless foam (92.0 mg, 95.8 µmol, 81%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 8.11 (d, $J$ = 7.8 Hz, 1H, H-6, diast.), 8.02 (d, $J$ = 7.8 Hz, 1H, H-6), 7.45 – 7.41 (m, 1H, DMT, diast.), 7.40 – 7.37 (m, 2H, DMT), 7.35 – 7.26 (m, 9H, DMT), 7.26 – 7.21 (m, 2H, DMT), 6.87 – 6.78 (m, 7H, DMT), 6.16 (d, $J$ = 2.8 Hz, 1H, H-1'), 6.11 (d, $J$ = 1.9 Hz, 1H, H-1', diast.), 5.31 (d, $J$ = 7.8 Hz, 1H, H-5, diast), 5.25 (d, $J$ = 7.8 Hz, 1H, H-5), 5.22 (d, $J$ = 4.1 Hz, 1H, $OCH_2O$, diast.), 5.21 (d, $J$ = 4.2 Hz, 1H, $OCH_2O$), 5.12 (d, $J$ = 4.2 Hz, 1H, $OCH_2O$, diast.), 5.11 (d, $J$ = 4.2 Hz, 1H, $OCH_2O$), 4.50 (ddd, $J$ = 10.4, 7.9, 5.0 Hz, 1H, H-3', diast.), 4.39 (ddd, $J$ = 9.7, 7.1, 5.0 Hz, 1H, H-3'), 4.31 (dd, $J$ = 5.0, 2.7 Hz, 1H, H-2'), 4.28 – 4.24 (m, 2H, H-2', H-4'), 4.23 – 4.20 (m, 1H, H-4'), 4.20 – 4.07 (m, 1H, $POCH_2$), 3.94 (ddt, $J$ = 10.5, 7.7, 6.1 Hz, 1H, $POCH_2$), 3.88 – 3.80 (m, 1H, $POCH_2$), 3.79 (d, $J$ = 3.1 Hz, 9H, $OCH_3$), 3.67 (ddd, $J$ = 11.0, 5.0, 2.4 Hz, 2H, H-5'), 3.62 – 3.45 (m, 5H, $POCH_2$, $N(CH(CH_3)_2)_2$), 3.38 (ddd, $J$ = 10.9, 5.8, 2.7 Hz, 2H, H-5'), 3.17 (s, 5H, $N^4CH_3$), 2.93 (s, 5H, $N^4CH_3$), 2.78 – 2.73 (m, 1H, $CH_2CN$), 2.69 – 2.53 (m, 2H, $CH_2CN$), 2.36 (t, $J$ = 6.5 Hz, 1H, $CH_2CN$), 1.27 (dd, $J$ = 6.8, 5.6 Hz, 10H, $N(CH(CH_3)_2)_2$ ), 1.17 – 1.10 (m, 14H, $N(CH(CH_3)_2)_2$), 1.09 – 1.01 (m, 34H, $Si(CH(CH_3)_2)_3$), 0.95 (d, $J$ = 6.7 Hz, 6H, $N(CH(CH_3)_2)_2$).
**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 163.71 (C4), 163.62 (C4), 158.74 ($OCH_3$), 155.48 (C=O, C2), 155.44 (C=O, C2), 144.63 ($C_q$, DMT), 144.50 ($C_q$, DMT), 141.13 (C6), 141.10 (C6), 135.70 ($C_q$, DMT), 135.64 ($C_q$, DMT), 135.57 ($C_q$, DMT), 130.55 (DMT), 128.68 DMT), 128.01 (DMT), 127.12 (DMT), 118.02 (CN), 117.57 (CN), 113.25 (DMT), 91.34 (C5), 91.21 (C5), 89.74 ($OCH_2O$), 89.17 (C1'), 87.07 ($C_q$, DMT), 86.88 ($C_q$, DMT), 82.23 (C4'), 82.18 (C4'), 78.65 (C2'), 78.61 (C2'), 69.83 (C3'), 69.66 (C3'), 61.93 (C5'), 61.27 (C5'), 58.29 ($POCH_2$), 58.23 ($POCH_2$), 55.38 ($OCH_3$), 55.35 ($OCH_3$), 43.33 ($N(CH(CH_3)_2)_2$), 43.20 ($N(CH(CH_3)_2)_2$), 37.90 ($N^4CH_3$), 37.48 ($N^4CH_3$), 24.80

($N(CH(CH_3))_2$), 24.73 ($N(CH(CH_3))_2$), 18.03 ($Si(CH(CH_3)_2)_3$), 12.09 $Si(CH(CH_3)_2)_3$), 12.08 ($Si(CH(CH_3)_2)_3$)).

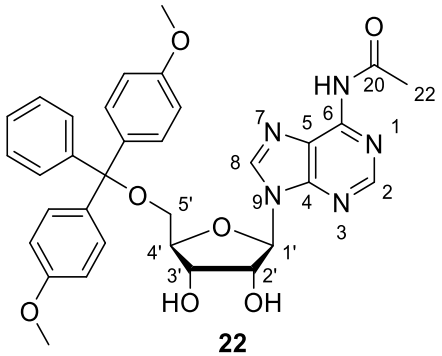
**$^{31}$P-NMR** (162 MHz, $CDCl_3$) $\delta$ (ppm) = 150.0, 149.53.

**HR-MS** ($ESI^+$): Exact mass calculated for $C_{51}H_{74}N_5NaO_9PSi$ $[M+H]^+$: 982.4891, found: 982.5196.

### 5.1.2.5 Preparation of $N^6$-isopentenyl adenosine phosphoramidite

The synthesis was based on the report of Tararov et. al.[377]

**$N^6$-Acetyl-5'-*O*-(4,4'-dimethoxytrityl)adenosine (22)**




**22**

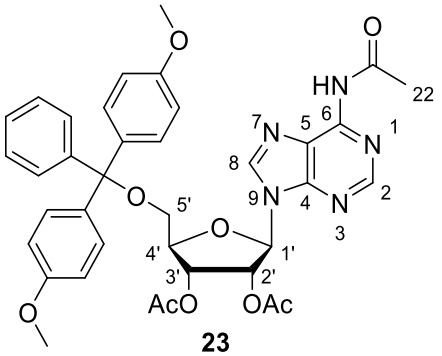
Starting from adenosine, $N^6$-acetyl-5'-*O*-DMT-adenosine (**22**) was produced following the published procedure.[499] The analytical parameters coincided with the ones reported by Pitsch *et al.*, but the assignments of H2 and H8 need to be revised, as described here.

**$^1$H-NMR** (400 MHz, $CDCl_3$) $\delta$ (ppm) = 8.63 (s, 1H, **2-H**), 8.22 (s, 1H, **8-H**), 7.30 – 7.21 (m, 2H, DMT), 7.20 – 7.14 (m, 7H, DMT), 6.77 – 6.69 (m, 4H, DMT), 6.03 (d, *J* = 5.7 Hz, 1H, 1'-H), 4.86 (dd, *J* = 5.7, 5.1 Hz, 1H, 2'-H), 4.46 (dd, *J* = 5.1, 2.3 Hz, 1H, 3'-H), 4.44 – 4.40 (m, 1H, 4'-H), 3.76 (d, *J* = 0.8 Hz, 6H, 2x $CH_3$), 3.44 (dd, *J* = 10.6, 3.5 Hz, 1H, 5'-H), 3.30 (dd, *J* = 10.6, 3.4 Hz, 1H, 5'-H), 2.61 (s, 3H, Acetyl-$CH_3$).

**$^{13}$C-NMR** (100 MHz, $CDCl_3$) $\delta$ (ppm) = 170.38 ($C_q$, Acetyl), 158.68, 158.65 ($C_q$, C-$OCH_3$, DMT), 152.11 (C-2), 150.60 (C-4), 149.45 (C-6), 144.41 ($C_q$, DMT), 141.40 (C-8), 135.54, 135.53 ($C_q$, DMT), 130.11 (DMT), 130.06 (DMT), 128.10 (DMT), 127.99 (DMT), 127.08 (DMT), 122.14 (C-5), 113.27 (DMT), 90.73 (C-1'), 86.76 ($C_q$, C-O, DMT), 86.01 (C-4'), 75.88 (C-2'), 72.41 (C-3'), 63.55 (C-5'), 55.35 (2x$CH_3$, DMT), 25.83 (C-22, Acetyl-$CH_3$).

**HR-MS ($ESI^+$):** Exact mass calculated for $C_{33}H_{33}N_5NaO_7$ $[M+Na]^+$: 634.22722, found: 634.22888.

**$N^6$-Acetyl-2',3'-*O*-diacetyl-5'-*O*-(4,4'-dimethoxytrityl)adenosin (23)**




**23**

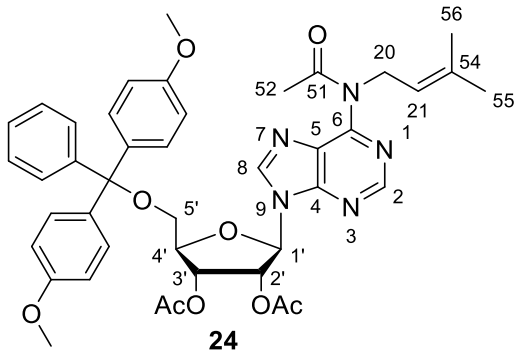
$N^6$-acetyl-5'-*O*-(4,4'-dimethoxytrityl)adenosine (**22**, 230 mg, 376 µmol, 1 eq) was dissolved in anhydrous pyridine (1.5 mL), treated with $Ac_2O$ (190 µL, 1.90 mmol, 5.05 eq) and stirred for 1.5 h at room temperature. After addition of EtOH (2.5 mL) the solvent was removed under reduced pressure. The reaction mixture was diluted with ethyl acetate (20 mL) and washed with brine and the organic phase was dried over $Na_2SO_4$. After evaporation of the solvent under reduced pressure, the product (**23**, 208 mg, 299 µmol, 79%) was obtained as a colorless foam.

**$^1$H-NMR** (400 MHz, $CDCl_3$) $\delta$ (ppm) = 8.64 (s, 1H, 2-H), 8.13 (s, 1H, 8-H), 7.43 – 7.39 (m, 2H, DMT), 7.34 – 7.23 (m, 6H, DMT), 7.24 – 7.19 (m, 1H, DMT), 6.84 – 6.75 (m, 4H, DMT), 6.32 (d, *J* = 6.6 Hz, 1H, 1'-H), 6.11 (dd, *J* = 6.6, 5.3 Hz, 1H, 2'-H), 5.69 (dd, *J* = 5.3, 3.0 Hz, 1H, 3'-H), 4.36 (q, *J* = 3.2 Hz, 1H, 4'-H), 3.78 (t, *J* = 0.7 Hz, 6H, 2 x $CH_3$), 3.51 (dd, *J* = 10.7, 3.1 Hz, 1H, 5'-H), 3.46 (dd, *J* = 10.7, 3.5 Hz, 1H, 5'-H), 2.59 (d, *J* = 1.5 Hz, 3H, Acetyl-$CH_3$), 2.13 (s, 3H, Acetly-$CH_3$), 2.05 (s, 3H, Acetyl-$CH_3$).

**$^{13}$C-NMR** (100 MHz, $CDCl_3$) $\delta$ (ppm) = 170.28 ($C_q$, Acetyl), 169.87 ($C_q$, Acetyl), 169.54 ($C_q$, Acetyl), 158.81 ($C_q$, C-$OCH_3$, DMT), 152.83 (C-2), 151.43 (C-4), 149.36 (C-6), 144.19 ($C_q$, DMT), 141.08 (C-8), 135.34, 135.30 ($C_q$, DMT), 130.29, 130.27 (DMT), 128.34 (DMT), 128.16 (DMT), 127.27 (DMT), 121.97 (C-5), 113.42 (DMT), 87.27 ($C_q$, C-O, DMT), 85.16 (C-1'), 82.79 (C-4'), 73.41 (C-2'), 71.83 (C-3'), 63.12 (C-5'), 55.39 (2x $CH_3$, DMT), 25.81 (C-22, $CH_3$, Acetyl), 20.82 ($CH_3$, Acetyl), 20.56 ($CH_3$, Acetyl).

**HR-MS (ESI$^+$):** Exact mass calculated for $C_{37}H_{37}N_5NaO_9$ $[M+Na]^+$: 718.24835, found: 718.24876.

**$N^6$-Acetyl-2',3'-*O*-diacetyl-5'-*O*-(4,4'-dimethoxytrityl)-$N^6$-isopentenyladenosine (24)**

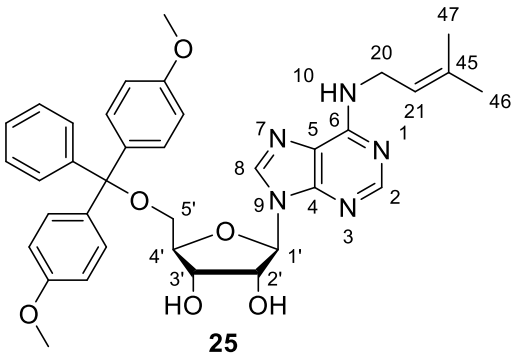



**24**

$N^6$-Acetyl-2',3'-*O*-diacetyl-5'-*O*-(4,4'-dimethoxytrityl)adenosine (**23**, 1.0 g, 1.44 mmol, 1 eq) in dry DMF (15 mL) was treated with $K_2CO_3$ (1.39 g, 10.1 mmol, 7 eq) and isopentenyl bromide (1.29 g, 996 µL, 8.62 mmol, 6 eq) and stirred for 20 h at 50 °C. The solution was allowed to cool down to room temperature, diluted with ethyl acetate (42 mL) and washed with brine. The organic phase was dried over $Na_2SO_4$, and the solvent was removed under reduced pressure. The residual was purified by column chromatography (DCM + 1% $NEt_3$ → DCM:EtOH = 99:1 + 1% $NEt_3$) to yield the product (**24**, 590 mg, 54%) as a slightly yellow foam. A significant fraction of unreacted starting material was recovered (**23**, 225 mg, 22%).

**$^1$H-NMR** (400 MHz, $CDCl_3$) *δ* (ppm) = 8.72 (s, 1H, 2-H), 8.20 (s, 1H, 8-H), 7.43 – 7.39 (m, 2H, DMT), 7.33 – 7.29 (m, 4H, DMT), 7.26 (s, 2H, DMT), 7.24 – 7.18 (m, 1H, DMT), 6.84 – 6.78 (m, 4H, DMT), 6.35 (d, *J* = 6.5 Hz, 1H, 1'-H), 6.11 (dd, *J* = 6.4, 5.3 Hz, 1H, 2'-H), 5.71 (dd, *J* = 5.3, 3.2 Hz, 1H, 3'-H), 5.23 (ddt, *J* = 6.6, 5.3, 1.4 Hz, 1H, 21-H), 4.84 (d, *J* = 6.7 Hz, 2H, $CH_2$, 20-H), 4.37 (q, *J* = 3.3 Hz, 1H, 4'-H), 3.78 (d, *J* = 1.6 Hz, 6H, 2x$CH_3$, DMT), 3.52 (dd, *J* = 10.7, 3.1 Hz, 1H, 5'-H), 3.46 (dd, *J* = 10.8, 3.6 Hz, 1H, 5'-H), 2.27 (s, 3H, $CH_3$, H-52), 2.14 (s, 3H, $CH_3$, Acetyl), 2.07 (s, 3H, $CH_3$, Acetyl), 1.60 (d, *J* = 1.3 Hz, 3H), 1.59 (d, *J* = 1.3 Hz, 3H, H-55, H-56).
**$^{13}$C-NMR** (100 MHz, $CDCl_3$) *δ* (ppm) = 171.41 ($C_q$ C-51), 169.84 ($C_q$ Acetyl), 169.52 ($C_q$ Acetyl), 158.80 ($C_q$, C-$OCH_3$, DMT), 153.85 (C-6), 153.08 (C-4), 152.35 (C-2), 144.20 ($C_q$, DMT), 141.89 (C-8), 135.96 ($C_q$, C-54), 135.37, 135.33 ($C_q$, DMT), 130.26 (DMT), 128.30 (DMT), 128.13 (DMT), 127.40 (C-5), 127.23 (DMT), 120.28 (C-21), 113.41 (DMT), 87.23 (C-22), 85.37 (C-1'), 82.65 (C-4'), 73.30 (C-2'), 71.68 (C-3'), 63.03 (C-5'), 55.37 (2x$CH_3$, DMT), 45.79 (C-20), 25.78 (C-55, $CH_3$), 24.46 (C-52, $CH_3$), 20.81, 20.59 (2x$CH_3$, Acetyl), 18.07 (C-56, $CH_3$).
**HR-MS (ESI$^+$):** Exact mass calculated for $C_{42}H_{45}N_5NaO_9$ $[M+Na]^+$: 786.31095, found: 786.31438.

### *N*$^6$-Isopentenyl-5'-*O*-(4,4'-dimethoxytrityl)adenosine (25)




**25**

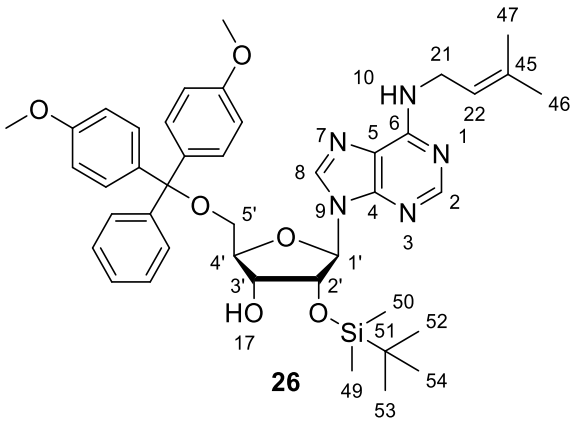
*N*$^6$-Acetyl-2',3'-*O*-diacetyl-5'-*O*-(4,4'-dimethoxytrityl)-*N*$^6$-isopentenyladenosine (**24**, 500 mg, 655 µmol, 1 eq) in MeOH (6 mL) was treated with methylamine in EtOH (33%, 5.7 mL, 45.8 mmol, 70 eq) and stirred at room temperature overnight. After evaporation of the solvent, the residue was purified by column chromatography (DCM + 1% $NEt_3$ → DCM:EtOH = 99:1 + 1% $NEt_3$ → DCM:EtOH = 97:3 + 1% $NEt_3$) to yield the product (**25**, 293 mg, 459 µmol, 70%) as a colorless foam.

**$^1$H-NMR** (400 MHz, $CDCl_3$) *δ* (ppm) = 8.34 (s, 1H, 2-H), 8.02 (s, 1H, 8-H), 7.28 – 7.24 (m, 2H, DMT), 7.20 – 7.14 (m, 7H, DMT), 6.76 – 6.70 (m, 4H, DMT), 5.94 (d, *J* = 6.0 Hz, 1H, 1'-H), 5.79 (s, 1H, NH), 5.41 – 5.34 (m, 1H, 21-H), 4.76 (dd, *J* = 6.1, 5.1 Hz, 1H, 2'-H), 4.42 (td, *J* = 3.3, 1.8 Hz, 1H, 4'-H), 4.37 (dd, *J* = 5.1, 1.9 Hz, 1H, 3'-H), 4.22 (s, 2H, $CH_2$, 20-H), 3.76 (d, *J* = 1.4 Hz, 6H, 2x$CH_3$, DMT), 3.43 (dd, *J* = 10.5, 3.5 Hz, 1H, 5'-H), 3.22 (dd, *J* = 10.5, 3.4 Hz, 1H, 5'-H), 1.74 (dd, *J* = 4.0, 1.3 Hz, 6H, 2xCH3,46-H, 47-H).

**$^{13}$C-NMR** (100 MHz, $CDCl_3$) *δ* (ppm) = 158.62 ($C_q$, C-$OCH_3$), 154.87 (C-6), 152.70 (C-2), 147.89 (C-4), 144.40 ($C_q$, DMT), 137.92 (C-8), 137.36 (C-45), 135.70, 135.50 ($C_q$, DMT), 130.09 (DMT), 130.04 (DMT), 128.11 (DMT), 127.96 (DMT), 127.00 (DMT), 120.29 (C-5), 119.96 (C-22), 113.24 (DMT), 91.23 (C-1'), 86.71, 86.61 (C-4', $C_q$, C-O), 76.48 (C-2'), 73.15 (C-3'), 63.73 (C-5'), 55.32 ($CH_3$, DMT), 38.77 (C-21), 25.83 ($CH_3$, C-46), 18.17 ($CH_3$, C-47).

**HR-MS (ESI$^+$):** Exact mass calculated for $C_{36}H_{40}N_5O_6$ $[M+H]^+$: 638.29731, found: 638.29719

**$N^6$-Isopentenyl-5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-(tert-butyldimethylsilyl)adenosine (26)**



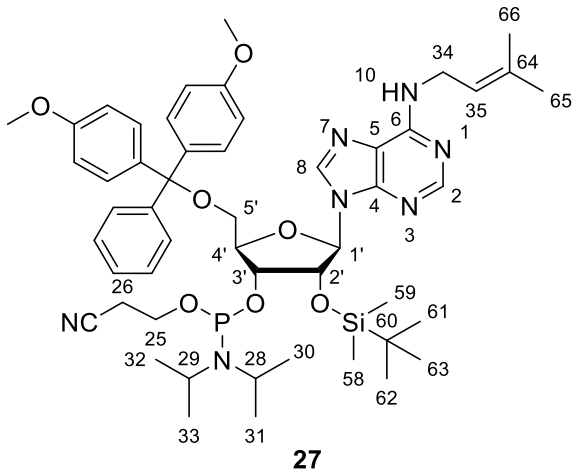

To $N^6$-isopentenyl-5'-*O*-(4,4'-dimethoxytrityl)adenosine (**25**, 335 mg, 525 µmol, 1 eq) in dry THF/pyridine (1:1, 4 mL) $AgNO_3$ (134 mg, 788 µmol, 1.5 eq) and TBDMS-Cl (87.1 mg, 578 µmol, 1.1 eq) were added and stirred for 19 h at room temperature. The reaction was stopped with MeOH (4 mL) and filtered over celite. The filtrate was evaporated to dryness under reduced pressure. The mixture of 2' and 3' silylated isomers was repeatedly separated by column chromatography (hexane:ethylacetate = 4:1 + 1% $NEt_3$ → 3:1 → 2:1) to yield the pure isomer **26** (182 mg, 242 µmol, 46%) as a colorless foam.

**$^1$H-NMR** (400 MHz, $CDCl_3$) $\delta$ (ppm) = 8.31 (s, 1H, 2-H), 7.94 (s, 1H, 8-H), 7.47 – 7.43 (m, 2H, DMT), 7.36 – 7.31 (m, 4H, DMT), 7.27 (s, 2H, DMT), 7.23 – 7.18 (m, 1H, DMT), 6.84 – 6.77 (m, 4H, DMT), 6.01 (d, *J* = 5.5 Hz, 1H, 1'-H), 5.61 (s, 1H, NH), 5.38 (tdt, *J* = 7.1, 2.9, 1.4 Hz, 1H, 22-H), 4.99 (t, *J* = 5.3 Hz, 1H, 2'-H), 4.33 (dt, *J* = 5.2, 3.7 Hz, 1H, 3'-H), 4.24 (q, *J* = 3.5 Hz, 1H, 4'-H), 4.22 (s, 2H, $CH_2$, 21-H), 3.78 (d, *J* = 1.2 Hz, 6H, 2x$CH_3$, DMT), 3.51 (dd, *J* = 10.6, 3.3 Hz, 1H, 5'-H), 3.37 (dd, *J* = 10.6, 4.0 Hz, 1H, 5'-H), 2.72 (d, *J* = 4.0 Hz, 1H, 3'-OH), 1.78 – 1.73 (m, 6H, 2xCH3, 46-H, 47-H), 1.26 (t, *J* = 7.1 Hz, 1H), 0.84 (s, 9H, 3x$CH_3$, 52-H, 53-H, 54-H), -0.02 (s, 3H, $CH_3$, H-49), -0.14 (s, 3H, $CH_3$, 50-H).

**$^{13}$C-NMR** (100 MHz, $CDCl_3$) $\delta$ (ppm) = 158.66 ($C_q$, C-$OCH_3$), 154.73 (C-6), 153.43 (C-2), 149.28 (C-4), 144.75 ($C_q$, DMT), 138.56 (C-8), 137.13 ($C_q$, C-45), 135.83 ($C_q$, DMT), 130.23 (DMT), 128.30 (DMT), 128.03 (DMT), 127.05 (DMT), 120.32 (C-5), 120.21 (C-22), 113.32 (DMT), 88.23 (C-1'), 86.71 ($C_q$, DMT), 84.18 (C-4'), 75.73 (C-2'), 71.72 (C-3'), 63.58 (C-5'), 55.36 (2x$CH_3$, DMT), 38.85 (C-21), 25.85 (C-46), 25.73 (C-52, C-53, C-54), 18.17 (C-47), -4.82, -5.11 (C-49, C-50).

**HR-MS (ESI$^+$):** Exact mass calculated for $C_{42}H_{54}N_5O_6Si$ $[M+H]^+$: 752.38379, found: 752.38394

**$N^6$-Isopentenyl-5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-(*tert*-butyldimethylsilyl)adenosine 3' cyano-ethyl *N*,*N*-diisopropyl phosphoramidite (27)**



**27**

$N^6$-Isopentenyl-5'-*O*-DMT-2'-*O*-TBDMS-adenosine (**26**, 70 mg, 93.1 µmol, 1 eq) was dissolved in dry DCM (1 mL) and $Me_2NEt$ (101 µL, 931 µmol, 10 eq) and CEP-Cl (33.0 mg, 140 µmol, 1.5 eq) were added. After stirring for 2.5 h at room temperature, the solvent was removed under reduced pressure. The residue was purified by column chromatography (ethylacetate:hexane = 2:1 + 1% $NEt_3$) to yield compound **27** (79 mg, 83.0 µmol, 89%).

**$^1$H-NMR** (400 MHz, $CDCl_3$) $\delta$ (ppm) = 8.29 (s, 1H, 2-H), 8.27 (s, 1H, 2-H diast.), 7.97 (s, 1H, 8-H), 7.94 (s, 1H, 8-H diast.), 7.49 – 7.42 (m, 4H, DMT), 7.38 – 7.31 (m, 7H, DMT), 7.31 – 7.21 (m, 6H, DMT), 7.24 – 7.17 (m, 2H, DMT), 6.84 – 6.77 (m, 8H, DMT), 6.00 (d, *J* = 6.5 Hz, 1H, 1'-H), 5.96 (d, *J* = 6.1 Hz, 1H, 1'-H diast.), 5.58 (s, 2H, NH), 5.38 (dddd, *J* = 6.8, 5.4, 3.2, 1.5 Hz, 2H, 35-H), 5.07 (ddd, *J* = 8.6, 6.3, 4.6 Hz, 2H, 2'-H), 4.39 (m, 2H, 3'-H, 4'-H), 4.32 (q, *J* = 4.0 Hz, 1H, 4'-H), 4.21 (s, 4H, $CH_2$, 34-H), 4.01 – 3.92 (m, 1H, $CH_2$, 25-H), 3.88 (ddt, *J* = 10.3, 8.3, 6.7 Hz, 1H, $CH_2$, 25-H), 3.78 (d, *J* = 1.3 Hz, 7H, 2x$CH_3$, DMT), 3.78 (d, *J* = 0.6 Hz, 6H, 2x$CH_3$, DMT diast.), 3.71 – 3.62 (m, 1H, $CH_2$, 25-H diast.), 3.59 (dq, *J* = 10.2, 3.2 Hz, 4H, 25-H, 5'-H, 28-H, 29-H), 3.54 (dd, *J* = 10.7, 4.0 Hz, 1H, 5'-H diast.), 3.32 (dd, *J* = 9.8, 3.4 Hz, 1H, 5'-H), 3.28 (dd, *J* = 9.9, 3.4 Hz, 1H, 5'-H), 2.65 (td, *J* = 6.5, 4.4 Hz, 2H, $CH_2$, 26-H), 2.30 (td, *J* = 6.7, 1.9 Hz, 1H, $CH_2$, 26-H diast.), 1.76 (d, *J* = 1.3 Hz, 6H, 2x$CH_3$, 65-H, 66-H), 1.74 (d, *J* = 1.3 Hz, 5H, 2x$CH_3$, 65-H, 66-H diast.), 1.20 – 1.02 (m, 21H, 4x$CH_3$, 30-H, 31-H, 32-H, 33-H), 0.77 (s, 8H, 3x$CH_3$, 61-H, 62-H, 63-H), 0.76 (s, 8H, 3x$CH_3$, 61-H, 62-H, 63-H diast.), -0.03 (s, 2H, $CH_3$, 58-H), -0.06 (s, 3H, $CH_3$, 58-H diast.), -0.19 (s, 2H, $CH_3$, 59-H), -0.21 (s, 3H, $CH_3$, 59-H diast.).

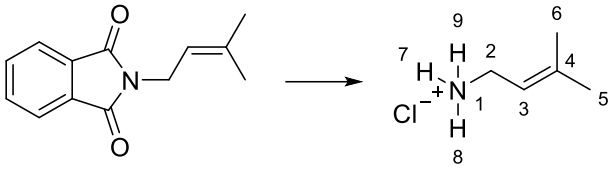
**$^{13}$C-NMR** (100 MHz, $CDCl_3$) $\delta$ (ppm) = 158.64 ($C_q$-OMe, DMT), 154.72 (C-6), 153.30 (C-2), 149.20 (C-4), 144.79, 144.69 ($C_q$, DMT), 139.02 (C-8), 137.10 (C.64), 135.97, 135.93, 135.80, 135.76 ($C_q$, DMT), 130.32, 130.27, 130.23 (C-DMT), 128.41, 128.29 (C-DMT), 128.02, 128.00 (C-DMT), 127.03 (C-DMT), 120.47 (C-5), 120.22 (C-35), 117.80, 117.46 (CN), 113.30, 113.26 (C-DMT), 88.40, 88.13 (C-1'), 86.77, 86.63 ($C_q$-O, DMT), 84.06, 83.81 (C-4'), 75.13, 75.11, 74.59, 74.54 (C-

2'), 73.56, 73.46, 73.01, 72.86 (C-3'), 63.55, 63.38 (C-5'), 59.08, 58.91, 57.93, 57.72 (C-25), 55.37, 55.35 (O-$CH_3$, DMT), 43.58, 43.45, 43.12, 43.00 (C-28, C-29), 38.76 (C-34), 25.84, 25.80, 25.76 (C-61, C-62, C-63), 24.92, 24.85, 24.78, 24.67 (C-30, C-31, C-32, C-33), 20.62, 20.56, 20.22, 20.15 (C-26), 18.16, 18.10, 18.04 (C-65, C-66), -4.58, -4.60, -5.00 (C58, C-59).

**$^{31}$P-NMR** (162 MHz, $CDCl_3$) $\delta$ (ppm) = 150.81, 148.84.

**HR-MS (ESI$^+$):** Exact mass calculated for $C_{51}H_{71}N_7O_7PSi$ $[M+H]^+$: 952.49164, found: 952.49298.

**Isopentenyl amine hydrochloride (29)**




**28** **29**

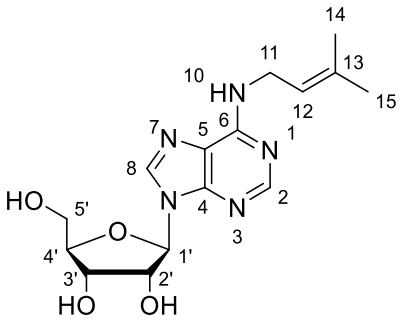
Potassium phtalimide (2.50 g, 13.5 mmol, 1.0 eq.) was dissolved in anhydrous DMF under nitrogen atmosphere. After addition of 1-bromo-3-methylbut-2-ene (2.01 g, 1.56 mL, 13.5 mmol, 1.0 eq) the reaction mixture was stirred for 16 h at 80°C. The reaction mixture was allowed to cool to ambient temperature and poured into ice water (~ 150 mL). The resulting precipitate was filtered off. The colorless intermediate **28** was suspended in EtOH (150 mL), heated to 50°C and treated with hydrazine monohydrate (1.35 g, 1.31 mL, 27.0 mmol, 2.0 eq.). After 2 h, the reaction mixture was again allowed to cool to ambient temperature and 6M HCl was added. The reaction mixture was stirred for 18 h at ambient temperature and filtered off over celite. The filtrate was dried over $Na_2SO_4$, and the solvent was removed under reduced pressure, to yield **29** as a colorless solid, slightly contaminated with phthalhydrazide.[390]

**$^1$H-NMR** (DMSO-d6, 400 MHz): $\delta$ (ppm) = 8.06 (s, br, 3H, $CH_3$), 5.22 (tdq, $J$ = 1.5, 3.1, 7.4 Hz, 1H, CH), 3.31 – 3.40 (m, 2H, $CH_2$), 1.71 – 1.73 (m, 3H, $CH_3$), 1.64 – 1.67 (m, 3H, $CH_3$).

**$^{13}$C-NMR** (DMSO-d6$_3$, 100 MHz): $\delta$ (ppm) = 138.40 ($C_q$, C4), 117.09 (CH), 36.44 ($CH_2$), 25.49 (C-5, $CH_3$), 17.95, (C-6, $CH_3$).

**HR-MS (ESI$^+$):** Exact mass calculated for $C_5H_{12}N$ $[M+H]^+$: 86.09697, found: 86.09635.

***N*$^6$-Isopentenyladenosine (30)**




**30**

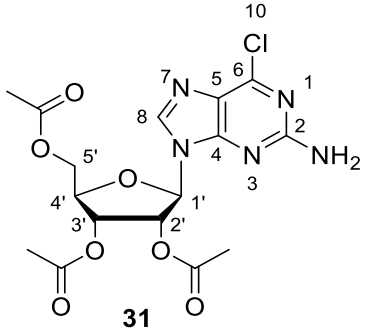
6-chloropurinriboside (1.5 g, 5.23 mmol, 1.0 eq) was suspended in ethanol (14 mL) under nitrogen atmosphere. Isopentenyl amine hydrochloride (1.06 g, 8.71 mmol, 1.7 eq) and triethylamine (7.00 mL, 50.5 mmol, 9.7 eq) were added and the reaction was stirred at 80°C for 4.5 h. The solvent was removed under reduced pressure and the crude product was purified by column chromatography (DCM:MeOH = 94:6 + 1% $NEt_3$) to obtain the product (**30**, 1.10 g, 3.28 mmol, 62%) as a colorless foam.

**$^1$H-NMR** (400 MHz, DMSO-d6) *δ* (ppm) = 8.35 (d, *J* = 5.2 Hz, 1H, H-8), 8.21 (s, 1H, H-2), 7.94 (s, 1H, NH), 5.88 (d, *J* = 6.3 Hz, 1H, H-1'), 5.48 – 5.37 (m, 2H, 2'-OH, 5'-OH), 5.34 – 5.26 (m, 1H, $CH_2CH$, H-12), 5.19 (dd, *J* = 4.7, 1.5 Hz, 1H, 3'-OH), 4.60 (td, *J* = 6.3, 5.0 Hz, 1H, H-2'), 4.14 (td, *J* = 4.8, 3.0 Hz, 1H, H-3'), 4.06 (br s, 2H, $CH_2CH$, H-11), 3.96 (q, *J* = 3.4 Hz, 1H, H-4'), 3.72 – 3.62 (m, 1H, H-5'), 3.55 (ddd, *J* = 12.1, 7.3, 3.6 Hz, 1H, H-5'), 1.70 (d, *J* = 1.3 Hz, 3H, $CH_3$), 1.66 (t, *J* = 1.3 Hz, 3H, $CH_3$).

**$^{13}$C-NMR** (100 MHz, DMSO-d6) *δ* (ppm) = 152.37 (C2), 148.26 (C4), 139.72 (C8), 133.24($C_q$, C13), 122.1 ($CH_2CH$, C12), 119.83 (C5), 87.94 (C1'), 85.91 (C4'), 73.48 (C2'), 70.68 (C3'), 61.69 (C5'), 37.69 ($CH_2CH$, C11), 25.44, 17.89 (2x$CH_3$).

5.1.2.6 Preparation of *N*$^6$-isopentenyl-2-thiomethyl adenosine phosphoramidite

**2',3',5'-*O*-Triacetyl-6-chloro-2-aminopurine riboside (31)**

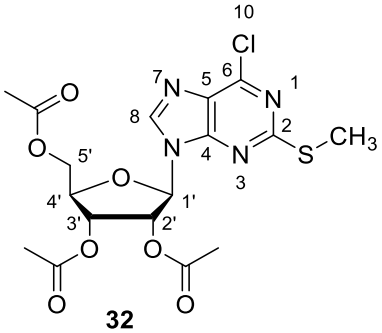



**31**

2-Amino-6-chloropurine riboside (2.00 g, 6.63 mmol, 1.0 eq) was suspended in anhydrous acetonitrile (25 mL) under nitrogen atmosphere. After adding DMAP (81.0 mg, 663 µmol, 0.1 eq) and $NEt_3$ (5.51 mL, 39.8 mmol, 6.0 eq), the mixture was cooled to 0°C. Then acetic anhydride (1.87 mL, 19.8 mmol, 2.99 eq) was added dropwise and the mixture was stirred at room temperature for 3 h. The solvent was evaporated under reduced pressure, followed by recrystallization of the crude product from isopropanol (~ 35 mL). The product **31** was obtained as a colorless solid (2.33 g, 5.45 mmol, 82%).[500]

$^1$**H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 7.87 (s, 1H, 8-H), 6.01 (d, *J* = 4.9 Hz, 1H, 1'-H), 5.95 (t, *J* = 5.1 Hz, 1H, 2'-H), 5.74 (t, *J* = 5.0 Hz, 1H, 3'-H), 5.22 (s, 2H, $NH_2$), 4.47 – 4.42 (m, 2H, 4'-H, 5'-H), 4.41 – 4.34 (m, 1H, 5'-H), 2.14 (s, 3H, acetyl), 2.10 (s, 3H, acetyl), 2.08 (s, 3H, acetyl).
**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): *δ* (ppm) = 170.62 (C=O, acetyl), 169.71 (C=O, acetyl), 169.49 (C=O, acetyl), 159.21 (C-6), 153.19 (C-4), 151.97 (C-2), 140.83 (C-8), 125.89 ($C_q$-5), 86.71 (C1'), 80.12 (C4'), 72.84 (C2'), 70.58 (C3'), 63.04 (C5'), 20.84 ($CH_3$, acetyl), 20.67 ($CH_3$, acetyl), 20.56 ($CH_3$, acetyl).
**HR-MS** ($ESI^+$): Exact mass calculated for $C_{16}H_{18}ClN_5NaO_7$ $[M+Na]^+$: 450.0788, found: 450.0798.

**2',3',5'-*O*-Triacetyl-2-methylthio-6-chloropurine riboside (32)**



**32**

2',3',5'-*O*-Triacetyl-6-methylthiopurine riboside **31** (1.30 g, 3.04 mmol,1.0 eq) was dissolved in 20 mL anhydrous acetonitrile under nitrogen atmosphere. Then dimethyl disulfide (2.70 mL, 30.4 mmol, 10.0 eq) and isopentyl nitrite (817 µL, 6.08 mmol, 2.0 eq) were added and the mixture was heated to 60 °C. After 2 h the reaction mixture was concentrated to ~ 5 mL. The residue was diluted with DCM and washed with a saturated aqueous $NaHCO_3$ solution. The organic phase was dried over $Na_2SO_4$ and the solvent was removed under reduced pressure. The residual was purified by column chromatography (DCM → DCM:MeOH = 99:1 → 98:2) to yield **32** as a colorless foam (922 mg, 2.01 mmol, 66%).[387]

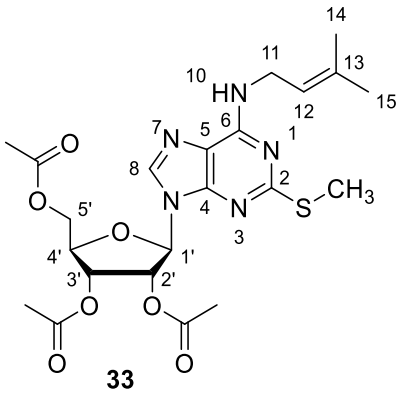
$^1$**H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 8.09 (s, 1H, H-8), 6.11 (d, *J* = 4.6 Hz, 1H, H-1'), 5.99 (dd, *J* = 4.6, 5.5 Hz, 1H, H-2'), 5.65 (t, *J* = 5.5 Hz, 1H, H-3'), 4.42 – 4.47 (m, 2H, H-4', H-5'a), 4.32 (dd, *J*

= 5.1, 13.0 Hz, 1H, H-5'b), 2.64 (s, 3H, S-$CH_3$), 2.15 (s, 3H, acetyl), 2.11 (s, 3H, acetyl), 2.09 (s, 3H, acetyl).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): *δ* (ppm) = 170.43 (C=O, acetyl), 169.64 (C=O, acetyl), 169.46 (C=O, acetyl), 167.54 (C-2), 152.08 (C-4), 151.46 (C-6), 142.29 (C-8), 129.26 (C-5), 87.19 (C-1'), 80.17 (C-4'), 73.07 (C-2'), 70.27 (C-3'), 62.85 (C-5'), 20.87 ($CH_3$, acetyl), 20.65 ($CH_3$, acetyl), 20.55 ($CH_3$, acetyl), 15.01 (S-$CH_3$).

**HR-MS** ($ESI^+$): Exact mass calculated for $C_{17}H_{19}ClN_4NaO_7S$ $[M+Na]^+$: 481.0561, found: 481.0574.

**2',3',5'-*O*-triacetyl-*N*$_6$-isopentenyl-2-thiomethyl-adenosine (33)**




**33**

2',3',5'-*O*-triacetyl-2-methylthio-6-chloropurine riboside (**32**, 200 mg, 436 µmol, 1.0 eq) was dissolved in anhydrous pyridine (5 mL) under nitrogen atmosphere. After addition of $NEt_3$ (846 µL, 6.10 mmol, 14.0 eq) and isopentenyl amine hydrochloride (159 mg, 1.31 mmol, 3.0 eq) the reaction mixture was stirred at room temperature for 17 h. The solvent was evaporated under reduced pressure. The residue was re-dissolved in DCM and washed with 5% citric acid, brine and saturated $NaHCO_3$, respectively. The organic phase was dried over $Na_2SO_4$. The crude product was purified by column chromatography (DCM → DCM:MeOH = 99:1) to yield **33** as a colorless foam (179 mg, 353 µmol, 81%).[387]

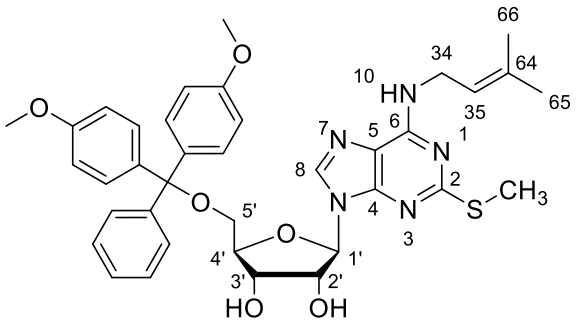
**$^1$H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 7.71 (s, 1H, 8-H), 6.04 (d, *J* = 4.6 Hz, 1H, H-1'), 6.01 (dd, *J* = 4.6, 5.5 Hz, 1H, H-2'), 5.72 (t, *J* = 5.5 Hz, 1H, H-3'), 5.61 (s, 1H, NH), 5.33 (ddq, *J* = 1.4, 5.8, 8.6 Hz, 1H, CH=$C_q$), 4.45 (dd, *J* = 3.4, 12.0 Hz, 1H, H-5'), 4.39 (ddd, *J* = 3.4, 4.6, 5.5 Hz, 1H, H-4'), 4.29 (dd, *J* = 4.6, 12.0 Hz, 1H, H-5'), 4.17 (s, br, 2H, NH-$CH_2$), 2.57 (s, 3H, $SCH_3$), 2.13 (s, 3H, acetyl), 2.09 (s, 3H, acetyl), 2.08 (s, 3H, acetyl), 1.74 (d, *J* = 1.3 Hz, 3H, $C_q$-$CH_3$), 1.73 (d, *J* = 1.3 Hz, 3H, $C_q$-$CH_3$).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): *δ* (ppm) = 170.59 (C=O, acetyl), 169.64 (C=O, acetyl), 169.51 (C=O, acetyl), 166.82 (C-2), 153.92 (C-6), 149.46 (C-4), 137.30 (C-8), 120.22 (C-32, *C*H=$C_q$), 118.02 (C-5), 86.85 (C-1'), 79.84 (C-4'), 73.15 (C-2'), 70.45 (C-3'), 63.11 (C-5'), 38.60 (CH2), 25.83 (C-15,

$CH_3$), 20.88 ($CH_3$, acetyl), 20.68 ($CH_3$, acetyl), 20.59 ($CH_3$, acetyl), 18.15 (C-14, $CH_3$), 14.76 (S-$CH_3$),

**HR-MS** (ESI$^+$): Exact mass calculated for $C_{22}H_{29}N_5NaO_7S$ [M+Na]+: 530.1686, found: 530.1675.

### *N*$_6$-Isopentenyl-5'-*O*-(4,4'-dimethoxytrityl)-2-thiomethyl-adenosine (34)

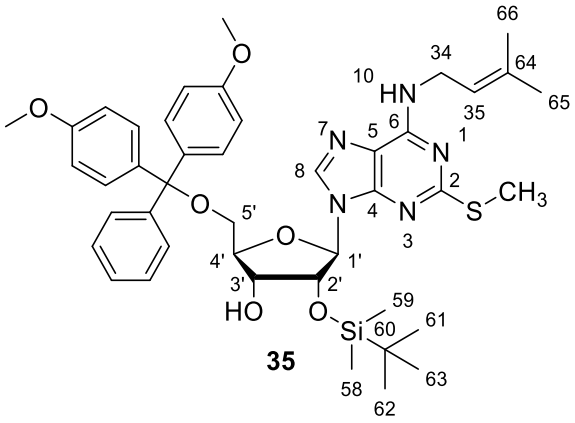



**34**

Compound **33** (170 mg, 335 µmol, 1.0 eq) was dissolved in MeOH (4.5 mL) and triethyl amine (500 µL) and stirred for 48 h at room temperature. After evaporation of the solvent under reduced pressure the residue was co-evaporated with pyridine. The residue was re-dissolved in anhydrous pyridine (3 mL) under nitrogen atmosphere and 4,4'-dimethoxytriyl chloride (DMT-Cl, 136 mg, 402 µmol, 1.2 eq) was added in three portions within 1 h. After stirring for an additional hour at room temperature MeOH (3 mL) was added to stop the reaction. The solvent was removed under reduced pressure. Then the residue was re-dissolved with DCM and washed with 5% citric acid, water and saturated $NaHCO_3$. The crude product was purified by column chromatography (DCM + 1% NEt3 → DCM:MeOH = 99:1) to yield **34** as a colorless foam (186 mg, 272 µmol, 81%).[387]

***H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 7.88 (s, 1H, H-8), 7.24 – 7.29 (m, 2H, DMT), 7.13 – 7.22 (m, 7H, DMT), 6.71 – 6.78 (m, 4H, DMT), 6.53 (s, br, 1H, C2'-OH), 5.89 (d, *J* = 6.2 Hz, 1H, H-1'), 5.84 (s, br, 1H, NH), 5.35 (dddd, *J* = 1.4, 2.9, 5.7, 8.6 Hz, 1H, CH=$C_q$), 4.71 – 4.77 (m, 1H, H-2'), 4.42 (td, *J* = 1.6, 3.5 Hz, 1H, H-4'), 4.36 (dt, *J* = 1.6, 5.2 Hz, 1H, H-3'), 4.19 (s, br, 2H, NH-$CH_2$), 3.77 (d, *J* = 1.6 Hz, 6H, $OCH_3$), 3.41 (dd, *J* = 3.5, 10.5 Hz, 1H, H-5'a), 3.22 (dd, *J* = 3.5, 10.5 Hz, 1H, H-5'b), 3.16 (d, *J* = 1.6 Hz, 1H, C3'-OH), 2.54 (s, 3H, S-$CH_3$), 1.73 – 1.76 (m, 6H, 2x $C_q$-$CH_3$).
**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): *δ* (ppm) = 165.97 (C-2), 158.64 ($C_q$-O, DMT), 153.86 (C-6), 148.86 (C-4), 144.37 ($C_q$, DMT), 137.15 (CH=*C*$_q$), 136.78 (C-8), 135.74 ($C_q$, DMT), 135.53 ($C_q$, DMT), 130.11 (DMT), 130.03 (DMT), 128.15 (DMT), 127.99 (DMT), 127.02 (DMT), 120.08 (*C*H=$C_q$), 117.81 (C-5), 113.26 (DMT), 90.96 (C-1'), 86.63, 86.56 (C-4'; $C_q$, DMT), 76.40 (C-2'), 73.12 (C-3'), 63.78 (C-5'), 55.33 ($OCH_3$, DMT), 38.63 (NH-$CH_2$), 25.84 (C-65, $CH_3$), 18.16 (C-66, $CH_3$), 14.65 (S-$CH_3$).
**HR-MS** (ESI$^+$): Exact mass calculated for $C_{37}H_{42}N_5O_6S$ [M+H]$^+$: 684.2856, found: 684.2862.

***N*$_6$-Isopentenyl-5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-(tert-butyldimethylsilyl)-2-thiomethyladenosine (35)**




**35**

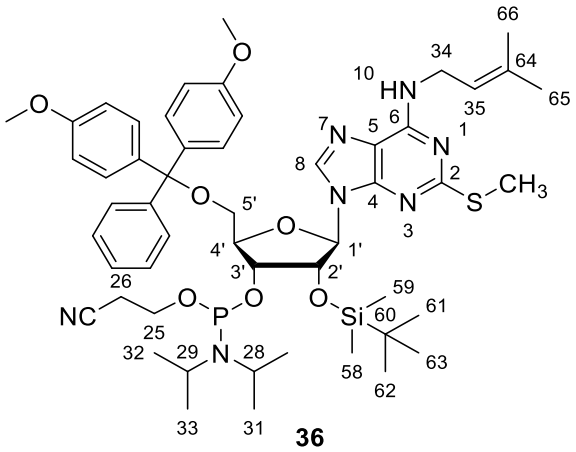
Compound **34** (170 mg, 249 µmol, 1.0 eq) was dissolved in anhydrous THF/pyridine (1:1, 3 mL) under nitrogen atmosphere followed by the addition of $AgNO_3$ (63.3 mg, 373 µmol, 1.5 eq.) and TBDMS-Cl (41.2 mg, 273 µmol, 1.1 eq). The reaction was stirred at room temperature for 20 h. After addition of MeOH (1 mL), the suspension was filtered over celite. The solvent of the filtrate was removed under reduced pressure. The residue was purified by column chromatography (hexane:EtOAc = 5:2 + 1% $NEt_3$ → 2:1) to obtain **35** as a colorless foam (63.5 mg, 79.6 µmol, 32%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 7.81 (s, 1H, H-8), 7.39 – 7.45 (m, 2H, DMT), 7.29 – 7.34 (m, 4H, DMT), 7.22 – 7.31 (m, 2H, DMT), 7.17 – 7.23 (m, 1H, DMT), 6.77 – 6.83 (m, 4H, DMT), 5.94 – 5.99 (d, *J* = 4.9 Hz, 1H, H-1'), 5.62 (s, br, 1H, NH), 5.35 (tdt, *J* = 1.4, 2.9, 7.2 Hz, 1H, CH=$C_q$), 4.89 (t, *J* = 4.9 Hz, 1H, H-2'), 4.34 (q, *J* = 4.8 Hz, 1H, H-3'), 4.12 – 4.24 (m, 3H, H-4', $CH_2$), 3.78 (s, 6H, $OCH_3$), 3.46 (dd, *J* = 3.4, 10.5 Hz, 1H, H-5'a), 3.40 (dd, *J* = 4.3, 10.5 Hz, 1H, H-5'b), 2.45 (d, *J* = 4.8 Hz, 1H, C3'-OH), 2.45 (s, 3H, S-$CH_3$), 1.72 – 1.76 (m, 6H, $C_q(CH_3)_2$), 0.86 (s, 9H, $Si(C(CH_3)_3)$), -0.00 (s, 3H, Si-$CH_3$), -0.10 (s, 3H, Si-$CH_3$).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): *δ* (ppm) = 166.20 (C-2), 158.66 ($C_q$-O, DMT), 149.98 (C-4), 144.71 ($C_q$, DMT), 137.33 (C-8), 136.91 (CH=*C*$_q$), 135.79 ($C_q$, DMT), 135.73 ($C_q$, DMT), 130.22 (DMT), 128.26 (DMT), 128.04 (DMT), 127.06 (DMT), 120.33 (*C*H=$C_q$), 117.76 (C-5), 113.33 (DMT), 88.27 (C-1'), 86.69 ($C_q$, DMT), 83.91 (C-4'), 75.75 (C-2'), 71.63 (C-3'), 63.71 (C-5'), 55.35 ($OCH_3$), 38.61 (NH-$CH_2$), 25.85 (C-65, $CH_3$), 25.74, ($Si(C(CH_3)_3)$), 18.15 (C-66, $CH_3$), 14.58 (S-$CH_3$), -4.70 (Si-$CH_3$), -5.11 (Si-$CH_3$).

**HR-MS** ($ESI^+$): Exact mass calculated for $C_{43}H_{55}N_5NaO_6SSi$ $[M+Na]^+$: 820.3540, found: 820.3537.

***N*6-Isopentenyl-5'-*O*-(4,4'-dimethoxytrityl)-2'-O-(tert-butyldimethylsilyl)-2-thiomethyl adenosine 3' cyanoethyl *N*,*N*-diisopropyl phosphoramidite (36)**



**36**

Compound **35** (105 mg, 132 µmol, 1.0 eq) was dissolved in anhydrous DCM (1 mL) under nitrogen atmosphere. $Me_2NEt$ (142 µL, 1.32 mmol, 10.0 eq) and 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite (CEP-Cl) (46.7 mg, 197 µmol, 1.5 eq) were added and the resulting mixture was stirred at room temperature for 2 h. After evaporation of the solvent under reduced pressure, the crude residue was purified by column chromatography (hexane:EtOAc = 3:1 + 1% $NEt_3$) to obtain **36** as a colorless foam (114 mg, 114 µmol, 86%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 7.87 (s, 1H, H-8), 7.84 (s, 1H, H-8), 7.40 – 7.46 (m, 4H, DMT), 7.29 – 7.37 (m, 7H, DMT), 7.21– 7.29 (m, 2H, DMT), 7.17 – 7.23 (m, 1H, DMT), 6.77 – 6.84 (m, 7H, DMT), 6.00 – 6.03 (d, *J* = 6.2 Hz, 1H, H-1'), 5.98 (d, *J* = 5.8 Hz, 1H, H-1'), 5.56 (s, br, 2H, NH), 5.32 – 5.37 (m, 2H, $CH=C_q$), 4.89 – 4.96 (m, 2H, H-2'), 4.36 – 4.42 (m, 1H, H-3'), 4.32 – 4.36 (m, 1H, H-4'), 4.28 – 4.32 (m, 1H, H-3', H-4'), 4.19 (s, 3H, $NH-CH_2$), 3.91 – 4.00 (m, 1H, $CH_2CN$), 3.82 – 3.91 (m, 1H, $CH_2CN$), 3.78 (s, 3H, $OCH_3$), 3.77 – 3.78 (m, 7H, $OCH_3$), 3.56 – 3.73 (m, 5H, $N(CH(CH_3)_2)$, 3.50 – 3.55 (m, H-5'a, H-5'b), 3.46 (dd, *J* = 3.7, 10.4 Hz, 1H, H-5'b), 3.33 (dt, *J* = 4.4, 10.4 Hz, 2H, H-5'a), 2.56 – 2.72 (m, 2H, $POCH_2$), 2.44 (s, 3H, $SCH_3$), 2.42 (s, 2H, $SCH_3$), 2.27 – 2.33 (m, 1H, C25, $POCH_2$), 1.74 – 1.75 (m, 5H, $C_q(CH_3)_2$), 1.73 – 1.74 (m, 6H, $C_q(CH_3)_2$), 1.02 – 1.19 (m, 20H, 4x $CH_3$, $N(CH(CH_3)_2)$, 0.76 – 0.82 (m, 14H, $Si(C(CH_3)_3)$), -0.03 (s, 2H, $Si-CH_3$), -0.05 (s, 3H, $Si-CH_3$), -0.17 (s, 5H, $Si-CH_3$).

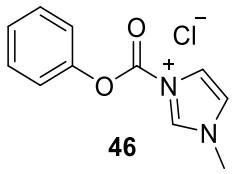
**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 165.94 (C-2), 158.53 ($C_q$-O, DMT), 149.97, 149.77 (C-4), 144.61, 144.50 ($C_q$, DMT), 137.27 (C-8), 136.74 ($CH=C_q$), 135.51, 135.48 ($C_q$, DMT), 130.12, 130.07 (DMT), 128.09 (DMT), 127.90 (DMT), 126.93 (DMT), 120.22 ($CH=C_q$), 117.58, 117.51 (C-5, CN), 113.20, 113.18, 113.15 (DMT), 87.55 (C-1'), 86.64, 86.51 ($C_q$, DMT), 83.43 (C-3'), 75.48 (C-2'), 72.99 (C-4'), 63.59, 63.44 (C-5'), 58.83 ($CH_2CN$), 55.20 ($OCH_3$), 43.02, 42.89 (($N(CH(CH_3)_2)$), 25.70 (C-65, $CH_3$), 25.64, 25.63, 25.61 ($Si(C(CH_3)_3)$), 24.68, 24.60 ($N(CH(CH_3)_2)$), 20.45, 20.39 ($POCH_2$), 18.00, 17.98 (C-66, $CH_3$), 14.35, 14.28 ($SCH_3$), 1.02, -5.04 ($Si-CH_3$).

**$^{31}$P-NMR ($CDCl_3$, 162 MHz):** δ (ppm) = 149.08, 150.64

**HR-MS ($ESI^+$):** Exact mass calculated for $C_{52}H_{72}N_7NaO_7PSSi$ [M+Na]+: 1020.4619, found: 1020.4613.

5.1.2.7 Preparation of $N^6$-threonylcarbamoyl adenosine phosphoramidite[401]

**1-*N*-methyl-3-phenoxycarbonyl-imidazolium chloride (46)**



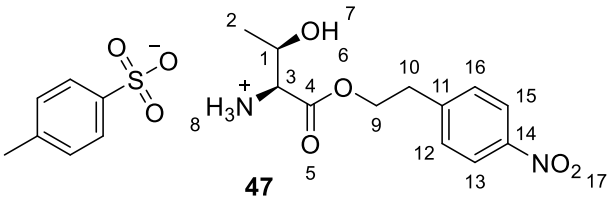
**46**

Phenyl chloroformate (1.50 mL, 11.9 mmol, 1.0 eq.) was dissolved in dry DCM. *N*-Methyl-imidazole (947 µL, 11.9 mmol, 1.0 eq.) was added at 0 °C and the reaction stirred for 2 h at room temperature. The precipitate was collected by filtration, washed with DCM, and dried to give the product **46** as a colorless solid (2.78 g, 11.7 mmol, 98%).

**$^1$H-NMR** (DMSO-d6, 400 MHz): *δ* (ppm) = 10.23 (td, *J* = 1.6, 0.9 Hz, 1H, C*H*$_{imidazole}$, NCHN), 8.42 – 8.34 (m, 1H, C*H*$_{imidazole}$), 7.99 (dd, *J* = 2.2, 1.5 Hz, 1H, C*H*$_{imidazole}$), 7.61 – 7.54 (m, 2H, C*H*$_{npe}$), 7.49 – 7.43 (m, 3H, C*H*$_{npe}$, C*H*$_{npe}$), 4.01 (d, *J* = 0.8 Hz, 3H, N-$CH_3$).

**HR-MS** ($ESI^+$): Exact mass calculated for $C_{11}H_{11}N_2O_2$ $[M+H]^+$: 203.0815, found: 203.0822.

**L-Threonine-2-(4-Nitrophenyl)ethyl ester (47)**



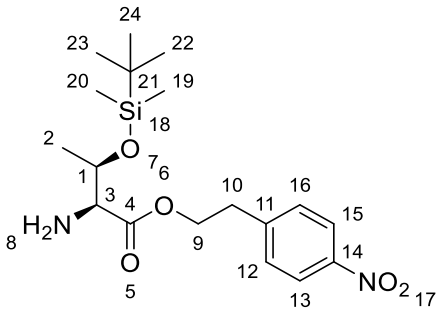
**47**

L-Threonine (500 mg, 4.20 mmol, 1.0 eq.), 2-(4-nitrophenyl) ethanol (2.11 g, 12.6 mmol, 3.0 eq) and TsOH (2.40 g, 12.6 mmol, 3.0 eq.) were refluxed in toluene (20 mL) for four days using a *Dean-Stark* apparatus. After cooling down to room temperature, $Et_2O$ was added, and the upper layer was removed to get the oily residue. Further purification with column chromatography (DCM/MeOH = 11:1) gave the product **47** as a yellow oil (1.28 g, 2.92 mmol, 70%).

**$^1$H-NMR** (DMSO-d6, 400 MHz): *δ* (ppm) = 8.25 – 8.13 (m, 5H, $NH_3^+$, bz), 7.63 – 7.57 (m, 2H, bz), 7.49 – 7.42 (m, 2H, OTs), 7.08-7.14 (m, 2H, OTs), 4.50 – 4.38 (m, 2H, O-$CH_2$), 4.05 (qd, *J* = 6.5, 3.8 Hz, 1H, O-CH), 3.92 (q, *J* = 5.0 Hz, 1H, $NH_2$-C*H*), 3.10 (t, *J* = 6.3 Hz, 2H, O-$CH_2$-C*H*$_2$), 2.29 (s, 3H, $CH_3$, OTs), 1.13 (d, *J* = 6.5 Hz, 3H, $CH_3$).

**$^{13}$C-NMR** (DMSO-d6, 100 MHz): *δ* (ppm) = 174.09 (C=O, C4), 146.71 ($C_q$, C11), 146.26 ($C_q$, C14), 130.29 (C12, C16), 123.42 (C13, C15), 67.77 (C1-OH), 63.84 ($CH_2$, C9), 59.29 (CH, C3), 34.10 ($CH_2$, C10), 20.10 ($CH_3$, C2).

**L-Threonine-*O*-(tert-butyldimethylsilyl)-2-(4-Nitrophenyl)ethyl ester (48)**




**48**

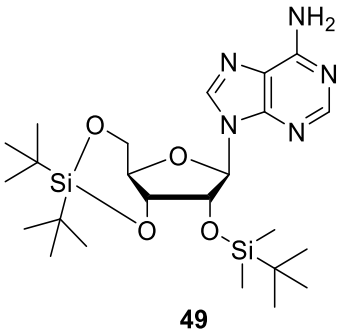
L-Threonine-2-(4-Nitrophenyl)ethyl ester (**47**, 785 mg, 1.79 mmol, 1.0 eq.) was dissolved in anhydrous pyridine (15 mL) and imidazole (366 mg, 5.37 mmol, 3.0 eq) and TBDMS-Cl (809 mg, 5.37 mmol, 3.0 eq) were added. The reaction mixture was stirred overnight at room temperature and then diluted with DCM (80 mL). The organic phase was washed with sat. $NaHCO_3$ solution (2 x 80 mL), water (80 mL) and brine (80 mL) and then dried over $Na_2SO_4$. The solvents were removed *in vacuo* and the product purified with column chromatography (1.5% MeOH in DCM) to give the protected amino acid **48** as a yellow oil (505 mg, 1.32 mmol, 74%).

**$^{1}$H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 8.21 – 8.14 (m, 2H, C$H_{npe}$ C13, C15), 7.44 – 7.36 (m, 2H, C$H_{npe}$,, C12, C16), 4.43 (dt, *J* = 11.0, 6.8 Hz, 1H, O-$CH_2$ C9), 4.31 – 4.24 (m, 1H, O-$CH_2$), 4.21 (qd, *J* = 6.3, 2.3 Hz, 1H, $CH_3$-C*H*OR, C2), 3.28 (d, *J* = 2.8 Hz, 1H, C*H*-$NH_2$), 3.08 (t, *J* = 6.8 Hz, 2H, $CH_2$, C10), 1.23 (d, *J* = 6.3 Hz, 3H, $CH_3$, C1), 0.82 (s, 9H, Si-C($CH_3$)$_3$), 0.01 (s, 3H, Si-$CH_3$), -0.08 (s, 3H, Si-$CH_3$).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): *δ* (ppm) = 174.16 (C=O), 147.10 ($C_{npe}$-$NO_2$, C14), 145.57 ($C_{npe}$, C11), 129.92 ($C_{npe}$, C12, C16), 123.96 ($C_{npe}$, C13, C15), 69.63 ($CH_3$-*C*HOR, C2), 64.63 (O-*C*$H_2$,C9), 60.86 (*C*H-$NH_2$, C4), 35.02 (*C*$H_2$, C10), 25.75 (Si-C(*C*$H_3$)$_3$, C22, 23, 24), 21.02 (*C*$H_3$ C1), 18.00 (Si-*C*($CH_3$)$_3$, C21), -4.15 (Si-*C*$H_3$), -5.11 (Si-*C*$H_3$).

**HR-MS** (ESI$^+$): Exact mass calculated for $C_{18}H_{31}N_2O_5Si$ $[M+H]^+$: 383.1997, found: 383.2007.

**2'-*O*-(tert-butyldimethylsilyl)-3'-*O*,5'-*O*-di-tert-butylsilylene adenosine (49)**



**49**

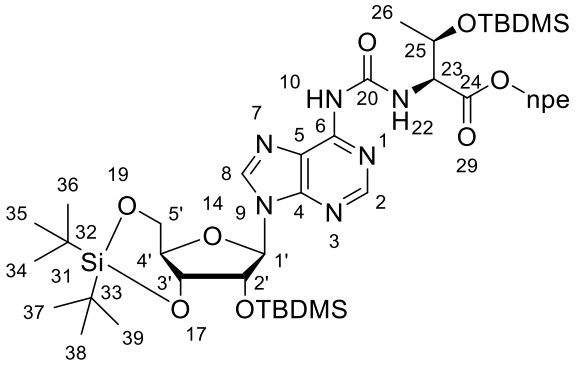
To a suspension of adenosine (1.00 g, 3.74 mmol, 1.0 eq.) in DMF (10 mL) di-*tert*-butylsilyl ditriflate (1.46 mL, 4.49 mmol, 1.2 eq) was added dropwise at 0 °C. The solution was stirred at 0 °C for 45 min. After addition of imidazole (1.27 g, 18.7 mmol, 5.0 eq) the reaction mixture was allowed to reach room temperature over 30 min. TBDMS-Cl (680 mg, 4.49 mmol, 1.20 eq) was added and then the mixture was heated to 60 °C overnight. After cooling to room temperature, the reaction mixture was diluted with EtOAc (40 mL) and washed with water (2 x 50 mL) and brine (60 mL). The organic layer was dried over $Na_2SO_4$ and evaporated. Column chromatography (EtOAc/Hexan = 1:1) gave the protected adenosine **49** as a colorless solid (1.40 g, 2.68 mmol, 72%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 8.32 (s, 1H, H-2), 7.83 (s, 1H, H-8), 5.91 (s, 1H, H-1'), 5.89 – 5.82 (br s, 2H, N$H_2$), 4.62 (d, $J$ = 4.7 Hz, 1H, H-2'), 4.54 (dd, $J$ = 9.5, 4.7 Hz, 1H, H-3'), 4.51 – 4.46 (m, 1H, H-5'), 4.21 (td, $J$ = 10.1, 5.1 Hz, 1H, H-4'), 4.03 (dd, $J$ = 10.5, 9.1 Hz, 1H, H-5'), 1.08 (s, 9H, Si-C(C$H_3$)$_3$), 1.04 (s, 9H, Si-C(C$H_3$)$_3$), 0.93 (s, 9H, TBDMS-Si-C(C$H_3$)$_3$), 0.16 (s, 3H, Si-$CH_3$), 0.14 (s, 3H, Si-$CH_3$).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 155.44 (C6), 152.98 (C2), 149.41 (C4), 139.05 (C8), 120.45 (C5), 92.59 (C1'), 75.95 (C3'), 75.59 (C2'), 74.82 (C4'), 67.95 (C5'), 27.63 (Si-C(*C*$H_3$)$_3$), 27.16 (Si-C(*C*$H_3$)$_3$), 26.03 (TBDMS-Si-C(*C*$H_3$)$_3$), 22.89 ($C_q$, Si-*C*(CH$_3$)$_3$), 20.49 ($C_q$, Si-*C*(CH$_3$)$_3$), 18.45 ($C_q$, TBDMS-Si-*C*(CH$_3$)$_3$), -4.16 (TBDMS-Si-*C*H$_3$), -4.85 (TBDMS-Si-*C*H$_3$).

**HR-MS** (ESI$^+$): Exact mass calculated for $C_{24}H_{44}N_5O_4Si_2$ $[M+H]^+$: 522.2932, found: 522.2928.

**$N^6$-(*O*-tert-butyldimethylsilyl-2-(4-nitrophenylethyl)-*N*-L-threoninyl)carbamoyl-2'-*O*-(tert-butyldimethylsilyl)-3'-*O*,5'-*O*-di-tert-butylsilylene adenosine (50)**




**50**

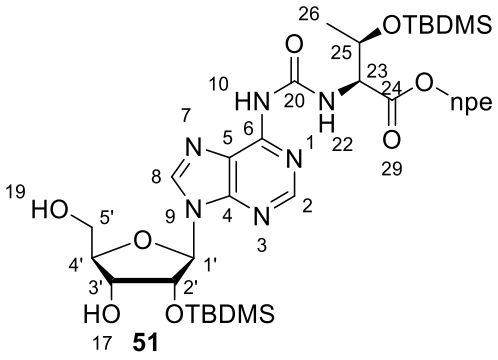
The silyl-protected adenosine derivative **49** (334 mg, 640 µmol, 1.0 eq.) was dissolved in dry DCM (8 mL). After addition of 1-*N*-methyl-3-phenoxycarbonyl-imidazolium chloride (**46**, 306 mg, 1.28 mmol, 2.00 eq) the reaction was stirred for 2 h. Then, a solution of TBDMS-protected threonine 4-nitrophenylethyl ester (**48**, 490 mg, 1.28 mmol, 2.00 eq) and $NEt_3$ (130 mg, 1.28 mmol, 2.00 eq) in dry DCM (8 mL) were added and the mixture was stirred at room temperature overnight. After addition of saturated $NaHCO_3$ solution (25 mL) the mixture was extracted with DCM (3 x 20 mL). The combined organic phases were dried over $Na_2SO_4$ and the solvents removed *in vacuo*. Further purification with column chromatography (EtOAc/Hexan = 3:4) gave **50** as a colorless foam (479 mg, 520 µmol, 81%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 9.93 (d, $J$ = 9.0 Hz, 1H, thr-NH, 22), 8.42 (s, 1H, H-2), 8.06 (s, 1H, H-8), 7.98 – 7.91 (m, 2H, C$H_{npe}$), 7.37 – 7.30 (m, 2H, C$H_{npe}$), 5.98 (s, 1H, H-1'), 4.66 (d, $J$ = 4.6 Hz, 1H, H-2'), 4.58 – 4.47 (m, 4H, H-3', H-5', thr-NH-C*H* (C23), $CH_3$-C*H* (C25)), 4.38 (td, $J$ = 6.6, 2.5 Hz, 2H, O-C$H_2$), 4.25 (d, $J$ = 5.0 Hz, 1H, H-4'), 4.06 (d, $J$ = 9.2 Hz, 1H, H-5'), 3.03 (t, $J$ = 6.5 Hz, 2H, O-$CH_2$-C$H_2$), 1.25 (d, $J$ = 6.2 Hz, 3H, thr-C$H_3$), 1.09 (s, 9H, Si-C(C$H_3$)$_3$), 1.05 (s, 9H, Si-C(C$H_3$)$_3$), 0.95 (s, 9H, Si-C(C$H_3$)$_3$), 0.90 (s, 9H, Si-C(C$H_3$)$_3$), 0.19 (s, 3H, Si-C$H_3$), 0.17 (s, 3H, Si-C$H_3$), 0.07 (s, 3H, Si-C$H_3$), -0.03 (s, 3H, Si-C$H_3$).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 170.94 (C=O, C24), 154.33 (C=O, C20), 150.14 (C-6), 149.68 (C-4), 146.78 ($C_q$, $C_{npe}$), 145.64 ($C_q$, $C_{npe}$), 141.11 (C-8), 129.87 ($CH_{npe}$), 123.64 ($CH_{npe}$), 120.85 (C-5), 92.67 (C1'), 75.93 (C3'), 75.64 (C2'), 74.96 (C4'), 68.72 (CH, C25), 67.91 (C5'), 64.67 (O-$CH_2$), 59.76 (thr-NH-CH, C23), 34.92 (O-$CH_2$-$CH_2$), 27.61 (Si-C($CH_3$)$_3$), 27.16 (Si-C($CH_3$)$_3$), 26.04 (Si-C($CH_3$)$_3$), 25.66 (Si-C($CH_3$)$_3$), 22.89 ($C_q$, Si-C($CH_3$)$_3$), 21.24 (thr-$CH_3$), 20.50 ($C_q$, Si-C($CH_3$)$_3$), 18.45 ($C_q$, Si-C($CH_3$)$_3$), 17.95 ($C_q$, Si-C($CH_3$)$_3$), -4.15 (Si-$CH_3$), -4.86 (Si-$CH_3$), -5.27 (Si-$CH_3$).

**HR-MS** ($ESI^+$): Exact mass calculated for $C_{43}H_{71}N_7NaO_{10}Si_3$ $[M+Na]^+$: 952.4468, found: 952.4458.

**$N^6$-(*O*-tert-butyldimethylsilyl-2-(4-nitrophenylethyl)-*N*-L-threoninyl)carbamoyl-2'-*O*-(tert-butyldimethylsilyl) adenosine (51)**



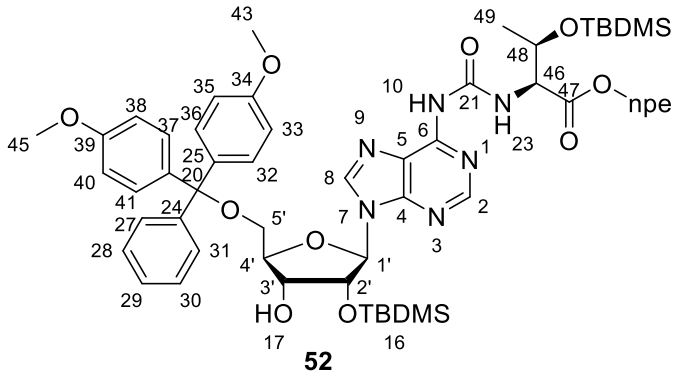

Compound **50** (250 mg, 270 µmol, 1.0 eq.) was dissolved in anhydrous DCM (5 mL). After addition of pyridine (300 µL) the solution was cooled to 0 °C and HF·Pyridine (70%, 44.0 µL) was added. The reaction mixture was stirred for 2 h at 0 °C and then quenched with sat. $NaHCO_3$ solution (5 mL). The aqueous phase was extracted with DCM (3 x 10 mL), then the combined organic phases were washed with water (30 mL), dried over $Na_2SO_4$ and evaporated. Column chromatography (DCM/MeOH = 11:1) gave the product **51** as a colorless foam (180 mg, 230 µmol, 85%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 9.88 (d, *J* = 9.1 Hz, 1H, thr-N*H*, 22), 8.50 (s, 1H, H-2), 8.40 (s, 1H, N*H*, 10), 8.14 (s, 1H, H-8), 8.12 – 8.06 (m, 2H, C$H_{npe}$), 7.41 – 7.33 (m, 2H, C$H_{npe}$), 5.86 (d, *J* = 6.9 Hz, 1H, H-1'), 5.08 (dd, *J* = 7.0, 4.7 Hz, 1H, H-2'), 4.58 (dd, *J* = 9.1, 1.7 Hz, 1H, thr-NH-C*H,* H23), 4.53 – 4.42 (m, 2H, O-C*H₂, H25*), 4.41 – 4.36 (m, 2H, H-3', H-4'), 4.30 (dt, *J* = 11.0, 6.8 Hz, 1H, O-C$H_2$), 3.98 (dd, *J* = 12.9, 1.7 Hz, 1H, H-5'), 3.78 (dd, *J* = 12.9, 1.6 Hz, 1H, H-5'), 3.07 (t, *J* = 6.7 Hz, 2H, O-$CH_2$-C$H_2$), 1.24 (d, *J* = 6.2 Hz, 3H, thr-C$H_3$), 0.89 (s, 9H, Si-C(C$H_3$)$_3$), 0.82 (s, 9H, Si-C(C$H_3$)$_3$), 0.05 (s, 3H, Si-C$H_3$), -0.06 (s, 3H, Si-C$H_3$), -0.14 (s, 3H, Si-C$H_3$), -0.34 (s, 3H, Si-C$H_3$).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): *δ* (ppm) = 170.84 (C=O, C24), 153.94 (C=O, C20), 151.24 (C-6), 150.87 (C-2), 149.25 (C-4), 147.02 ($C_q$, $C_{npe}$), 145.53 ($C_q$, $C_{npe}$), 142.85 (C-8), 129.97 ($CH_{npe}$), 123.87 ($CH_{npe}$), 121.69 (C-5), 91.48 (C1'), 87.61 (C3'), 74.87 (C2'), 72.69 (C4'), 68.74 (C25), 65.06 (O-$CH_2$), 63.24 (C5'), 59.69 (thr-NH-*C*H, C23), 34.98 (O-$CH_2$-$CH_2$), 25.65 (Si-C($CH_3$)$_3$), 21.25 (thr-$CH_3$), 17.97 ($C_q$, Si-*C*($CH_3$)$_3$), 17.95 ($C_q$, Si-*C*($CH_3$)$_3$), -4.10 (Si-*C*$H_3$), -5.15 (Si-*C*$H_3$), -5.20 (Si-*C*$H_3$), -5.29 (Si-*C*$H_3$).

**HR-MS** (ESI$^+$): Exact mass calculated for $C_{35}H_{55}N_7NaO_{10}Si_2$ $[M+Na]^+$: 812.3441, found: 812.3447.

**5'-*O*-(4,4'-Dimethoxyrtriyl)-$N^6$-(*O*-tert-butyldimethylsilyl-2-(4-nitrophenylethyl)-*N*-L-threoninyl)carbamoyl-2'-*O*-(tert-butyldimethylsilyl) adenosine (52)**




**52**

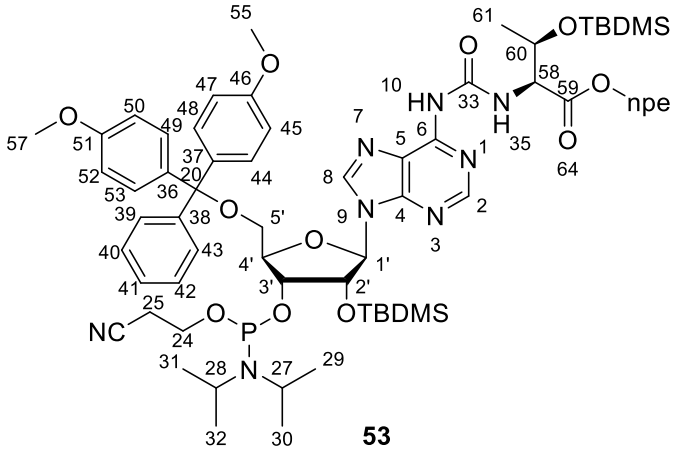
To a solution of **51** (325 mg, 411 µmol, 1.0 eq) in anhydrous pyridine (6 mL) DMT-Cl (227 mg, 617 µmol, 1.5 eq) was added and the resulting mixture was stirred overnight at room temperature. After removing the volatiles in vacuum, the compound was purified with column chromatography (1. DCM + 1 % $NEt_3$, 2. Toluene/EtOAc = 3:2 + 1 % $NEt_3$) to give **52** (290 mg, 265 µmol, 64%) as a colorless foam.

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 9.97 (d, $J$ = 9.1 Hz, 1H, thr-N*H,* 23), 8.41 (s, 1H, H-2), 8.18 (s, 1H, H-8), 8.03 – 7.99 (m, 2H, C$H_{npe}$), 7.98 (s, 1H, N*H,* 10), 7.46 – 7.42 (m, 2H, DMT), 7.36 – 7.31 (m, 6H, C$H_{npe}$, DMT), 7.31 – 7.18 (m, 3H, DMT), 6.84 – 6.78 (m, 4H, DMT), 6.08 (d, $J$ = 4.9 Hz, 1H, H-1'), 4.97 (t, $J$ = 5.0 Hz, 1H, H-2'), 4.57 (dd, $J$ = 9.1, 1.6 Hz, 1H, thr-NH-C*H,* C46), 4.51 (qd, $J$ = 6.2, 1.6 Hz, 1H, CH, C48), 4.44 – 4.30 (m, 3H, H-3', O-C$H_2$, npe), 4.26 (q, $J$ = 3.6 Hz, 1H, H-4'), 3.78 (d, $J$ = 1.4 Hz, 6H, DMT-OC$H_3$), 3.54 (dd, $J$ = 10.7, 3.0 Hz, 1H, H-5'), 3.41 (dd, $J$ = 10.7, 3.9 Hz, 1H, H-5'), 3.04 (t, $J$ = 6.6 Hz, 2H, O-$CH_2$-C$H_2$, npe), 2.67 (d, $J$ = 4.7 Hz, 1H, 3'-O*H*), 1.27 (d, $J$ = 6.3 Hz, 3H, thr-C$H_3$, C49), 0.88 (s, 9H, Si-C(C$H_3$)$_3$), 0.86 (s, 9H, Si-C(C$H_3$)$_3$), 0.06 (s, 3H, Si-C$H_3$), 0.03 (s, 3H, Si-C$H_3$), -0.05 (s, 3H, Si-C$H_3$), -0.08 (s, 3H, Si-C$H_3$).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 171.00 (C=O, C47), 158.72 (DMT-$C_q$-$OCH_3$), 154.25 (C=O, C21), 151.36 (C2), 150.38 (C4), 150.21 (C6), 146.89 ($C_q$, $C_{npe}$), 145.58 ($C_q$, $C_{npe}$), 144.69 (DMT-$C_q$), 141.54 (C8), 135.70 (DMT-$C_q$), 135.66 (DMT-$C_q$), 130.24 (DMT), 130.22 (DMT), 129.88 ($CH_{npe}$), 128.26 (DMT), 128.05 (DMT), 127.12 (DMT), 123.74 ($CH_{npe}$), 120.96 (C5), 113.34 (DMT), 88.71 (C1'), 86.85 (DMT-O*C*), 84.32 (C4'), 75.98 (C2'), 71.53 (C3'), 68.74 (C48), 64.83 (O-$CH_2$, npe), 63.43 (C5'), 59.70 (C46), 55.36 (DMT-O$CH_3$), 34.94 (O-$CH_2$-$CH_2$, npe), 25.72 (Si-C($CH_3$)$_3$), 25.62 (Si-C($CH_3$)$_3$), 21.28 (thr-$CH_3$), 18.05 ($C_q$, Si-$C$($CH_3$)$_3$), 17.92 ($C_q$, Si-$C$($CH_3$)$_3$), -4.18 (Si-$CH_3$), -4.74 (Si-$CH_3$), -5.03 (Si-$CH_3$), -5.29 (Si-$CH_3$).

**HR-MS** ($ESI^+$): Exact mass calculated for $C_{56}H_{73}N_7NaO_{12}Si_2$ $[M+Na]^+$: 1114.4753, found: 1114.4745.

**5'-*O*-(4,4'-Dimethoxyrtrityl)-*N*$^6$-(*O*-tert-butyldimethylsilyl-2-(4-nitrophenylethyl)-N-L-threonyl)carbamoyl-2'-*O*-(tert-butyldimethylsilyl) adenosine 3' cyanoethyl *N*,*N*-diisopropyl phosphoramidite (53)**




The 5'-*O*-DMT-2'-*O*-TBDMS-protected nucleoside **52** (90.0 mg, 82.0 µmol, 1.0 eq) was dissolved in anhydrous DCM (1.5 mL) and Hünig's Base (56.0 µL, 328 µmol, 4.0 eq) was added dropwise. After addition of 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite (49.0 mg, 205 µmol, 2.5 eq) the reaction was stirred at room temperature for 4 h. The solvent was removed, and the product purified by column chromatography (hexane/EtOAc = 2:1 + 1 % $NEt_3$) to obtain **53** as a colorless foam (62.0 mg, 48.0 µmol, 59%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 9.99 (d, J =7.5 Hz, 1H, thr-NH, 35) 9.97 (d, J =7.3 Hz, 1H, thr-NH, 35, diast.), 8.40, 8.39 (2s, 2H, H-2, diast.), 8.22, 8.18 (2s, 2H, H-8, diast.), 8.04 – 7.99 (m, 6H, $CH_{npe}$), 7.46 (ddd, J = 7.7, 6.4, 1.4 Hz, 4H, DMT), 7.38 – 7.30 (m, 12H, $CH_{npe}$, DMT), 7.30 – 7.19 (m, 6H, DMT), 6.85 – 6.78 (m, 8H, DMT), 6.10 (d, J = 6.0 Hz, 1H, H-1'), 6.05 (d, J = 5.7 Hz, 1H, H-1', diast.), 5.00 (dd, J = 5.9, 4.6 Hz, 2H, H-2', diast.), 4.56 (dd, J = 9.0, 1.6 Hz, 2H, thr-NH-CH, C58, diast.), 4.55 – 4.46 (m, 2H, CH, C60, diast.), 4.47 – 4.37 (m, 3H, H-3', O-$CH_2$, npe), 4.37 – 4.27 (m, 2H, H-4', O-$CH_2C\mathit{H}_2$, npe), 4.00 – 3.83 (m, 2H, $POCH_2$, C24), 3.78 (dd, J = 2.1, 1.4 Hz, 12H, DMT-$OCH_3$), 3.69 – 3.51 (m, 7H, $POCH_2$, H-5', NC$\mathit{H}(CH_3)_2$), 3.33 (td, J = 10.5, 3.8 Hz, 2H, H-5'), 3.04 (t, J = 6.6 Hz, 4H, O-$CH_2$, npe), 2.65 (tt, J = 6.3, 2.9 Hz, 2H, $CH_2CN$), 2.31 (td, J = 6.5, 2.5 Hz, 2H, $CH_2CN$), 1.28 – 1.24 (m, 12H, thr-$CH_3$, C61, NCH(C$\mathit{H}_3)_2$), 1.20 – 0.98 (m, 24H, NCH(C$\mathit{H}_3)_2$), 0.88 (d, J = 3.5 Hz, 15H, Si-C$(CH_3)_3$), 0.78 (d, J = 1.1 Hz, 15H, Si-C$(CH_3)_3$), 0.05 (d, J = 2.5 Hz, 6H, Si-$CH_3$), -0.01 (d, J = 10.8 Hz, 3H, Si-$CH_3$), -0.06 (d, J = 1.7 Hz, 3H, Si-$CH_3$), -0.16 (d, J = 8.4 Hz, 3H, Si-$CH_3$).

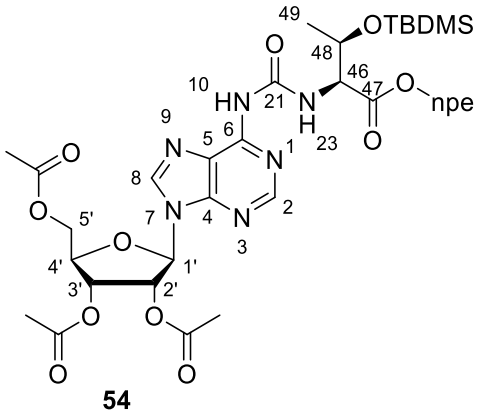
**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 170.90 (C=O, C59), 158.57 (DMT-$C_q$-$OCH_3$), 154.38 (C=O, C33), 151.16 (C2), 150.40 (C6), 150.02 (C4), 146.78 ($C_q$, $C_{npe}$), 145.43 ($C_q$, $C_{npe}$), 144.41 ($C_q$-DMT), 141.58 (C8), 135.47 ($C_q$, DMT), 130.19, 130.10 (DMT), 129.74 ($CH_{npe}$), 128.23 (DMT), 128.12 (DMT), 127.92 (DMT), 127.88 (DMT), 126.99 (DTM), 126.98 (DMT), 123.62 ($CH_{npe}$), 120.77

(C5), 117.30 (CN), 113.20 (DMT), 113.14 (DMT), 88.21 (C1'), 86.80 ($C_q$, DMT-OC), 83.68 (C4'), 74.89 (C2'), 72.94 (C3'), 68.62 (C60), 64.72 (O*C*$H_2$$CH_2$, npe), 63.07 (C5'), 59.56 (thr-NH-CH, C58), 58.67 (PO-$CH_2$), 57.53 (PO-$CH_2$), 55.22 (DMT-$OCH_3$), 42.88 ((N*C*H($CH_3$)$_2$, C27, C28), 34.80 (OCH$_2$*C*H$_2$), 25.61 (Si-C($CH_3$)$_3$), 25.58 (Si-C($CH_3$)$_3$), 25.47 (Si-C($CH_3$)$_3$), 24.70 (NCH($CH_3$)$_2$), 24.61 (NCH($CH_3$)$_2$), 21.13 (thr-$CH_3$, C61), 20.02 (*C*$H_2$CN), 17.92 ($C_q$, Si-C($CH_3$)$_3$), 17.77 ($C_q$, Si-C($CH_3$)$_3$), -4.33 (Si-$CH_3$), -4.69 (Si-$CH_3$), -5.08 (Si-$CH_3$), -5.43 (Si-$CH_3$).

**$^{31}$P-NMR** ($CDCl_3$, 162 MHz): *δ* (ppm) = 149.09, 150.77.

**HR-MS** (ESI$^+$): Exact mass calculated for $C_{65}H_{90}N_9NaO_{13}PSi_2$ [M+Na]$^+$: 1314.5826, found: 1314.5848.

**2',3',5'-Tri-*O*-acetyl-2-(4-nitrophenylethyl)-*N*-L-threoninyl)carbamoyl-2'-*O*-(tert-butyldimethylsilyl)adenosine (54)**




**54**

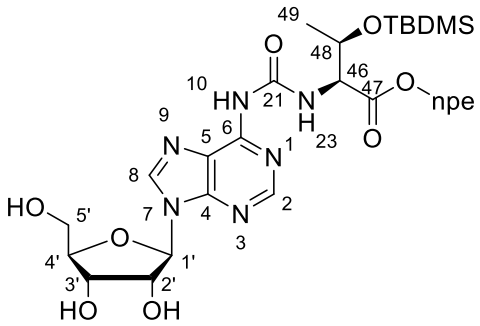
To 2',3',5'-Tri-*O*-acetyl-adenosine (500 mg, 1.27 mmol, 1.0 eq) in anhydrous DCM (20 mL) was added 1-methyl-3-(phenoxycarbonyl)-1H-imidazol-3-ium chloride (1.00 g, 4.19 mmol, 3.7 eq) and stirred at room temperature for 64 hours. After that, the protected threonine **48** (729 mg, 1.91 mmol, 1.5 eq) and triethylamine (350 µL, 2.54 mmol, 2.0 eq) were added to the reaction mixture and stirred for further 64 hours. The solvent was removed in vacuum and the crude product was purified by column chromatography (hexane/EtOAc = 1/1 to EtOAc) to obtain the product **54** (637 mg, 794 µmol, 63%) as a colorless foam. The product coelutes on a TLC plate (dichloromethane/ methanol 15/1) with the protected threonine. To differentiate them TLC plates were stained after elution and UV detection with ninhydrin which stains the protected threonine and not the product.

**$^1$H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 9.94 (d, *J* = 9.0 Hz, 1H, NH, H-23 ), 8.48 (s, 1H, H-2), 8.18 (s, 1H, H-8), 8.12 (s, 1H, NH, H-10), 7.90 – 7.82 (m, 2H, npe), 7.26 – 7.21 (m, 2H, npe), 6.22 (d, *J* = 5.4 Hz, 1H, H-1'), 6.05 (t, *J* = 5.4 Hz, 1H, H-2'), 5.67 (dd, *J* = 5.6, 4.4 Hz, 1H, H-3'), 4.53 – 4.27 (m, 8H, H-4', H-5', 3x $CH_2$, H-46, H-48, C(=O)$CH_2$), 3.03 (t, *J* = 6.5 Hz, 2H, $CH_2$-npe), 2.12 – 2.02

(m, 9H, 3x$CH_3$, acetyl), 1.19 (d, *J* = 6.2 Hz, 3H, $CH_3$, H-49), 0.90 (s, 9H, $CH_3$, Si(CH(C*H*$_3$)$_3$), 0.07 (s, 3H, Si((C$H_3$)$_2$), -0.10 (s, 3H, Si((C$H_3$)$_2$).

**HR-MS** (ESI$^+$): Exact mass calculated for $C_{35}H_{47}N_7NaO_{13}Si$ $[M+Na]^+$: 824.2893, found: 824.2894.

**2-(4-nitrophenylethyl)-*N*-L-threoninyl)carbamoyl-2'-*O*-(tert-butyldimethylsilyl)adenosine (55)**

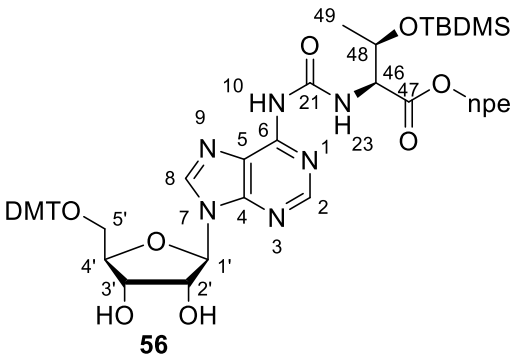



**55**

Compound **54** (619 mg, 770 µmol, 1.0 eq) was dissolved in methanol (3 mL) and ammonia in methanol (7 M, 1.22 mL 8.47 mmol, 11.0 eq) was added. The reaction mixture was stirred for 14.5 h and then concentrated in vacuum giving a light-yellow solid. The crude product was purified by column chromatography (DCM:MeOH = 94:6) yielding the product (**55**, 574 mg, 795 µmol, 82%) as a colorless solid.

**$^1$H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 9.66 (d, *J* = 8.6 Hz, 1H, NH, H-23), 8.31 (s, 1H, NH, H-10), 8.27 (s, 1H, H-2), 8.16 (s, 1H, H-8), 7.72 (d, *J* = 8.3 Hz, 2H, npe), 7.33 – 7.28 (m, 2H, npe), 5.95 (d, *J* = 7.5 Hz, 1H, H-1'), 5.12 (t, *J* = 5.6 Hz, 1H, H-2'), 4.64 – 4.36 (4m, 5H, H-3', H-4', H-46, H-48, C(=O)C$H_2$), 4.36 – 4.29 (m, 1H, C(=O)C$H_2$), 3.96 (d, *J* = 11.7 Hz, 1H, H-5'), 3.79 (d, *J* = 12.6 Hz, 1H, H-5'), 3.02 (t, *J* = 5.9 Hz, 2H, $CH_2$-npe), 1.22 (d, *J* = 6.3 Hz, 3H, $CH_3$, H-49), 0.87 (s, 9H, 3x$CH_3$, Si(CH(C*H*$_3$)$_3$), 0.06 (s, 3H, Si((C$H_3$)$_2$), -0.01 (s, 3H, Si((C$H_3$)$_2$).

**5'-*O*-(4,4'-Dimethoxyrtriyl)-*N*$^6$-(O-tert-butyldimethylsilyl-2-(4-nitrophenylethyl)-*N*-L-threoninyl)carbamoyl-adenosine (56)**



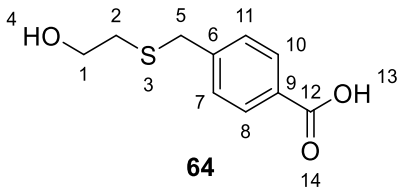

**56**

Compound **55** (559 mg, 775 µmol, 1.0 eq) was dissolved in anhydrous pyridine (2.5 mL). *N,N*-dimethylformamide dimethyl acetal (287 mL, 2.1 mmol, 3.0 eq) was added and the reaction was stirred for 17 h at ambient temperature. The solvent was removed under reduced pressure and and the residue was redissolved in pyridine (2.5 mL). DMT-Cl (285 mg, 841 µmol, 1.2 eq) was added. After two hours, additional DMT-Cl (71 mg, 0.3 eq) was added. After 23 h of stirring at room ambient temperature, methanol (2 mL) was added and the reaction mixture was stirred for 45 minutes. The solvents were removed under reduced pressure. The residue was dissolved in DCM and washed with 5% citric acid, water and a saturated $NaHCO_3$-solution, respectively. The organic phase was dried over $Na_2SO_4$ and the solvent removed under reduced pressure. This residue was purified by column chromatography (DCM + 1% $NEt_3$ to DCM/MeOH = 97:3) yielding the product (**56**, 335 mg, 327 µmol, 49%) as a colorless foam.

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 9.87 (d, $J$ = 9.1 Hz, 1H, NH, 23), 8.36 (s, 1H, H-2), 8.20 (s, 1H, H-8), 8.05 (s, 1H, NH, H-10), 7.93 – 7.85 (m, 2H, npe), 7.35 – 7.29 (m, 2H, npe), 7.29 – 7.24 (m, 3H, DMT), 7.20 – 7.13 (m, 6H, DMT), 6.77 – 6.69 (m, 4H, DMT), 6.04 (d, $J$ = 6.0 Hz, 1H, H-1'), 4.92 (dd, $J$ = 5.9, 4.8 Hz, 1H, H-2'), 4.59 – 4.32 (m, 6H, H-3', H-4', H-46, H-48, C(=O)C$H_2$), 3.76 (s, 6H, $OCH_3$), 3.44 (dd, $J$ = 10.7, 3.6 Hz, 1H, H-5'), 3.30 (dd, $J$ = 10.5, 3.4 Hz, 1H, H-5'), 3.03 (t, $J$ = 6.4 Hz, 1H, $CH_2$-npe), 1.34 (t, $J$ = 7.3 Hz, 3H, $CH_3$, H-49), 0.89 (s, 9H, Si(CH(C$H_3$)$_3$), 0.07 (s, 3H, Si((C$H_3$)$_2$), -0.03 (s, 3H, Si((C$H_3$)$_2$).
**HR-MS** (ESI$^+$): Exact mass calculated for $C_{50}H_{59}N_7NaO_{12}Si$ $[M+Na]^+$: 1000.3883, found: 1000.3862

5.1.2.8 Preparation of 3'-Phosphate CPG solid support[501]

**4-(((2-Hydroxyethyl)thio)methyl)benzoic acid (64)**

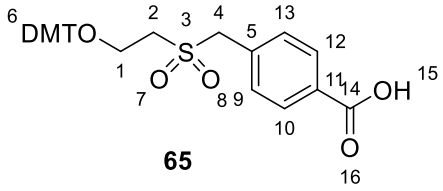


**64**

4-Bromomethylbenzoic acid (300 mg, 1.48 mmol, 1.0 eq) was suspended in anhydrous methanol (6 mL). 2-mercaptoethanol (126 mg, 114 µL, 1.62 mmol, 1.16 eq) and $NEt_3$ (386 µL, 2.76 mmol, 1.98 eq) were added and the reaction was stirred at ambient temperature for 17 h. The solvent was removed under reduced pressure and the residue was re-dissolved in water (12 mL). By acidifying with HCl (1 M), the product was precipitating as a white solid. The resulting suspension was extracted with ethyl acetate and the organic phase was dried over $Na_2SO_4$. After removing the solvent under reduced pressure, the product (**64**, 278 mg, 1.31 mmol, 88%) was obtained as a colorless solid and was used for the next reaction without further purification.[501]

**$^1$H-NMR** ($CD_3CN$, 400 MHz): $\delta$ (ppm) = 7.96 – 7.92 (m, 2H, bz), 7.45 – 7.41 (m, 2H, bz), 3.80 (s, 2H, $CH_2$, H-5), 3.58 (t, $J$ = 6.6 Hz, 2H, $CH_2$, H-1), 2.53 (t, $J$ = 6.6 Hz, 2H, $CH_2$, H-2).

**4-(((2-(4,4'-Dimethoxytrityl)oxyethyl)thio)methyl)benzoic acid (65)**

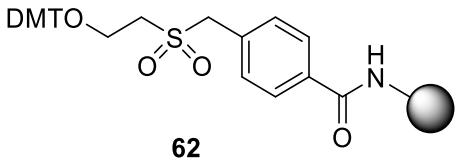



**65**

Compound **64** (270 mg, 1.27 mmol, 1.0 eq) was co-evaporated with anhydrous pyridine and dissolved in anhydrous pyridine (6 mL). DMT-Cl (647 mg, 1.91 mmol, 1.5 eq) was added in three portions within 1 h. The resulting reaction mixture was stirred for 19 h at ambient temperature. The solvent was removed under reduced pressure and the residue was co-evaporated with toluene (2x). The residue was re-dissolved in DCM and washed with 5% citric acid, water and a saturated $NaHCO_3$-solution, respectively. The organic phase was dried over $Na_2SO_4$, and the solvent was removed under reduces pressure. The crude residue was purified by column chromatography (DCM:MeOH, = 10:1 to 9:1 + 1% $NEt_3$) to obtain the product (**65**, 484 mg, 885 µmol, 69%) as a colorless solid.[501]

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 8.01 – 7.96 (m, 2H, bz), 7.45 – 7.21 (3m, 10H, DMT, bz), 7.23 – 7.17 (m, 1H, DMT), 6.87 – 6.79 (m, 4H, DMT), 3.79 (s, 6H, $OCH_3$), 3.68 (s, 2H, H-4), 3.22 (t, $J$ = 6.8 Hz, 2H, $CH_2$, H-1), 2.58 (t, $J$ = 6.8 Hz, 2H, $CH_2$, H-2).

**3'-Phosphate CPG solid support (62)**




**62**

Native amino CPG support (200 mg, 14.8 µmol, 1.00 eq) was transferred into a syringe and washed with anhydrous DMF (1.5 mL). The syringe was loaded with a mixture of compound **65** (20.0 mg, 39.0 µmol, 2.63 eq), BOP reagent (17.6 mg, 39.8 µmol, 2.69 eq) and $^i$Pr$_2$NEt (9.60 µL, 54.2 µmol, 3.66 eq) in anhydrous DMF (1.2 mL). The reaction was shaken for 2 h at 40 °C in the dark. The solution was removed, and the CPG was washed with anhydrous DMF (5 mL) and anhydrous MeCN (5 mL), respectively. Cap A (pyridine/acetic anhydride/THF = 10/10/80, 2 mL) and Cap B (NMI, 16% in THF, 2 mL) were added to the CPG within the syringe and the reaction was shaken for 10 min at ambient temperature. The solution was removed, and the CPG was washed with anhydrous MeCN (5 mL), MeOH and DCM (5 mL), respectively. A fresh solution of *m*CPBA (138 mg, 600 µmol, 40.5 eq) in DCM (9.60 mL) was prepared and added to the CPG. The reaction

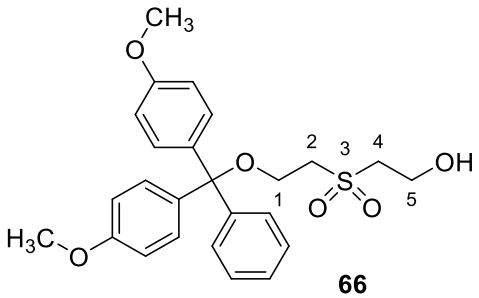
was shaken for 30 min at ambient temperature, the solvent was removed, and the CPG was washed with anhydrous MeCN (5 mL) and DCM (5 mL), respectively. The CPG support was dried in vacuum. To determine the loading density, around 3.0 mg of **62** were suspended in 10 mL of a dichloroacetic acid solution (3% in DCM). 1 mL of the orange solution was transferred into a cuvette (1 cm path length) and the absorption was measured.[502]

$$\frac{n}{m} = \frac{A}{\varepsilon \cdot 1cm} \cdot \frac{10\ mL}{m} \qquad (3)$$

With ε = 70000 L/mol·cm for the trityl cation at 498 nm, the exact mass "m" and the measured absorption at 498 nm a loading density of 52 µmol/g was obtained.

5.1.2.9 Preparation of 1-*O*-(2-cyanoethyl-*N*,*N*-diisopropylphosphoramidite)-1'-*O*-(4,4'-dimethoxytrityl)-2,2'- sulfonyldiethanol

**1-*O*-(4,4'-dimethoxytrityl)-2,2'-sulfonyldiethanol (66)**




**66**

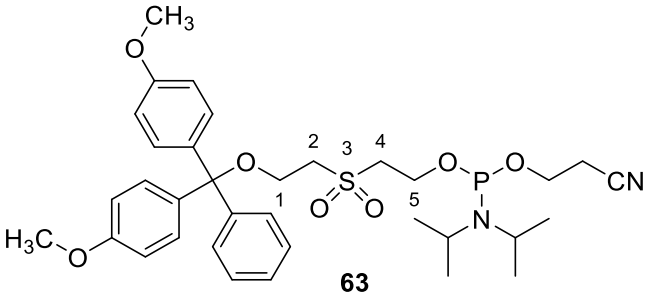
2,2'-Sulfonyldiethanol (65% in ethanol, 1.70 g, 6.49 mmol, 1.0 eq) was co-evaporated with pyridine (3x) and dried in high vacuum. The residue was redissolved in anhydrous pyridine (15 mL) and DMT-Cl (1.54 g, 4.54 mmol, 0.7 eq) were added in three portions within 1 h. After stirring the reaction for 20 h at ambient temperature, the solvent was removed under reduced pressure. The residue was taken up in ethyl acetate and washed with a saturated $NaHCO_3$- and a saturated NaCl-solution, respectively. The crude compound was purified my column chromatography (DCM + 1% $NEt_3$) to obtain the product (**66**, 1.31 g, 2.87 mmol, 44%) as a colorless solid.[409]

**$^1$H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 7.42 – 7.37 (m, 2H, DMT), 7.34 – 7.27 (m, 6H, DMT), 7.25 – 7.20 (m, 1H, DMT), 6.87 – 6.81 (m, 4H, DMT), 4.13 – 4.07 (m, 2H, $CH_2$, H-5), 3.79 (d, *J* = 0.5 Hz, 6H, $OCH_3$), 3.67 (t, *J* = 5.5 Hz, 2H, $CH_2$, H-1), 3.42 – 3.34 (m, 2H, $CH_2$, H-4), 3.19 (t, *J* = 5.5 Hz, 2H, $CH_2$, H-2).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): *δ* (ppm) = 158.87 ($C_q$, *C*-$OCH_3$, 144.24 ($C_q$, DMT), 135.30 ($C_q$, DMT), 130.07 (DMT), 128.21 (DMT), 128.07 (DMT), 127.28 (DMT), 113.50 (DMT), 87.77 ($C_q$, DMT-OC), 57.92 (C1), 56.76 (C4), 56.56 (C5), 55.39 ($CH_3$), 55.23 (C2).

**HR-MS** ($ESI^+$): Exact mass calculated for $C_{25}H_{28}NaO_6S$ $[M+Na]^+$: 479.15043, found: 479.14917.

**1-*O*-(2-cyanoethyl-*N*,*N*-diisopropylphosphoramidite)-1'-*O*-(4,4'-dimethoxytrityl)-2,2'-sulfonyldiethanol**



**63**

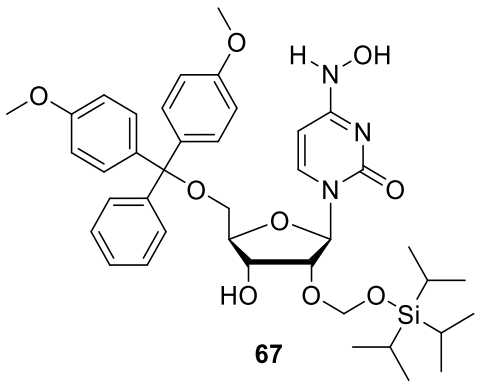
Compound **66** (500 mg, 1.10 mmol, 1.0 eq) in anhydrous DCM (5 mL) was mixed with diisopropylethylamine (DIPEA, 206 µL, 1.1 eq) and CEP-Cl (259 mg, 244 µL, 1.10 mmol, 1.0 eq). The reaction was stirred for 1.5 h at ambient temperature. After removing the solvent under reduced pressure, the crude product was purified by column chromatography (hexane:EtOAc = 3:1 + 1% $NEt_3$) to obtain the product (**63**, 570 mg, 868 µmol, 79%) as a colorless oil.[409]

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 7.44 – 7.40 (m, 2H, DMT), 7.33 – 7.27 (m, 6H, DMT), 7.25 – 7.20 (m, 1H, DMT), 6.86 – 6.82 (m, 4H, DMT), 4.16 – 4.01 (m, 2H, $CH_2$, H-5), 3.89 – 3.73 (m, 2H, $CH_2$, $POCH_2$), 3.79 (s, 6H, $OCH_3$), 3.67 – 3.56 (m, 4H, H-1, NC*H*$(CH_3)_2$, H-1), 3.52 – 3.36 (m, 2H, $CH_2$, H-4), 3.36 – 3.22 (m, 2H $CH_2$, H-2), 2.56 (tdd, *J* = 6.3, 1.4, 0.6 Hz, 2H, $CH_2$, $CH_2CN$), 1.19 (dd, *J* = 9.5, 6.8 Hz, 12H, NCH($CH_3$)$_2$).
**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 158.80 ($C_q$-$CH_3$), 144.01 ($C_q$, DMT), 135.44 ($C_q$, DMT), 130.17 (DMT), 128.70 (DMT), 128.12 (DMT), 127.58 (DMT), 117.70 (CN), 112.81 (DMT), 87.48 ($C_q$, DMT-C-O), 58.83 ($POCH_2$), 58.63 ($POCH_2$), 57.76, 57.55 (C1, C5), 55.84, 55.76 (C2, C4), 55.38 ($OCH_3$), 43.47 (N*C*H$(CH_3)_2$), 43.35 (N*C*H$(CH_3)_2$), 24.83 (NCH($CH_3$)$_2$), 24.76 (NCH($CH_3$)$_2$), 20.55 ($CH_2$CN), 20.48 ($CH_2$CN).
**$^{31}$P-NMR** ($CDCl_3$, 162 MHz): $\delta$ (ppm) = 148.84
**HR-MS** ($ESI^+$): Exact mass calculated for $C_{34}H_{46}N_2O_7PS$ $[M+H]^+$: 657.2758, found: 657.2743.

5.1.2.10 Preparation of $N^4$-hdroxycytidine phosphoramidite (**69**)[439]

5'-*O*-DMT-2'-*O*-TOM-$O^4$-chlorophenyluridine was prepared as described earlier by Büttner *et al.*[374] The convertible nucleoside was further reacted to the corresponding $N^4$-hydroxy-$N^4$-benzoylcytidine phosphoramidite in three steps.[439]

**5'-*O*-(4,4'-Dimethoxytrityl)-$N^4$-hydroxy-2'-*O*-(triisopropylsilyloxy)methylcytidine (67)**



**67**

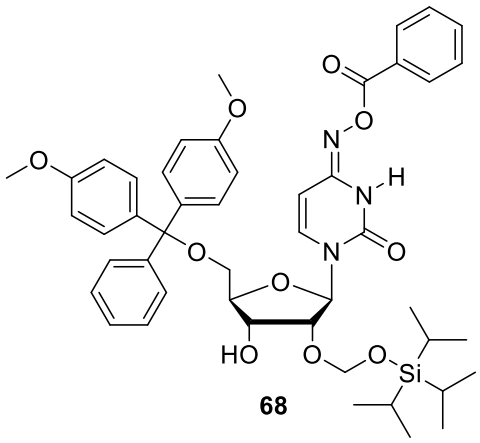
To synthesize compound **67** (5'-*O*-(4,4'-Dimethoxytrityl)-$N^4$-hydroxy-2'-*O*-(triisopropyl-silyloxymethylcytidine), 5'-*O*-DMT-2'-*O*-TOM-$O^4$-chlorophenyluridine (**17**, 400 mg, 474 µmol, 1.0 eq) was dissolved in anhydrous $CH_3CN$ (4 mL) under nitrogen atmosphere. DMAP (174 mg,1.42 mmol, 3.0 eq.) and $NEt_3$ (331 µL, 2.37 mmol, 5.0 eq) were added, followed by hydroxyl amine hydrochloride (165 mg, 2.37 mmol, 5.0 eq). After stirring for 21 h at ambient temperature, the reaction mixture was diluted with $CH_2Cl_2$ and washed with saturated aq. $NaHCO_3$ (2x). The organic phase was dried over $Na_2SO_4$ and the solvent was removed under reduced pressure. The crude product was purified by column chromatography (*n*-hexane:EtOAc + 1% $NEt_3$ = 1:2 to 2:1) to yield the product (compound **67**, 230 mg, 308 µmol, 64%) as a colorless foam.

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 8.53 (s, NH), 7.35 – 7.42 (m, 2H, DMT), 7.24 – 7.33 (m, 6H, DMT), 7.18 – 7.27 (m, 1H, DMT), 7.17 (d, *J* = 8.3 Hz, 1H, H-6), 6.79 – 6.88 (m, 4H, DMT), 6.03 (d, *J* = 4.4 Hz, 1H, H-1'), 5.25 (d, *J* = 8.3 Hz, 1H, H-5), 5.20 (d, *J* = 4.8 Hz, 1H, $OCH_2O$), 5.00 (d, *J* = 4.8 Hz, 1H, $OCH_2O$), 4.38 – 4.46 (q, *J* = 4.8 Hz, 1H, H-3'), 4.24 (t, *J* = 4.8 Hz, 1H, H-2'), 4.07 – 4.11 (dt, *J* = 2.6, 5.1 Hz, 1H, H-4'), 3.80 (s, 6H, $OCH_3$), 3.45 – 3.49 (dd, *J* = 2.4, 10.8 Hz, 1H, H-5'a), 3.40 – 3.45 (dd, *J* = 2.8, 10.8 Hz, 1H, H-5'b), 3.13 (d, *J* = 4.7 Hz, 1H, C3'-O*H*), 1.04 – 1.15 (m, 21H, $Si(CH(CH_3)_2)_3$).

**$^{13}$C-NMR** (100 MHz, $CDCl_3$): $\delta$ (ppm) = 158.70, 158.68 ($C_q$-$OCH_3$), 149.57 (C2), 145.43 (C4), 144.42 ($C_q$-DMT), 135.75 ($C_q$-DMT), 135.47 ($C_q$-DMT), 130.32 (DMT), 130.29 (DMT), 129.64 (C6), 128.36 (DMT), 128.10 (DMT), 127.19 (DMT), 113.39 (DMT), 98.44 (C5), 90.80 ($OCH_2O$), 87.08, 87.04, (C1', $C_q$-DMT), 83.48 (C4'), 82.37 (C2'), 70.24 (C3'), 62.91 (C5'), 55.44 ($OCH_3$), 55.43 ($OCH_3$), 17.95, 11.99 ($Si(CH(CH_3)_2)_3$).

**HR-MS (ESI$^+$)**: Exact mass calculated for $C_{40}H_{53}NaN_3O_9Si$ $[M+Na]^+$: 770.3443, found: 770.3459.

***N*⁴-*O*-Benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-*N*⁴-hydroxy-2'-*O*-(triisopropylsilyloxy)methyl cytidine (68)**

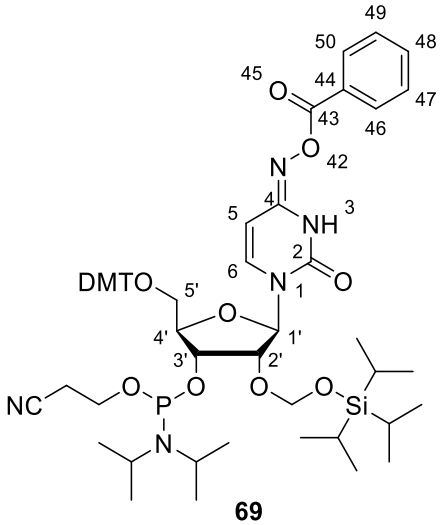


**68**

A solution of compound **67** (200 mg, 267 µmol, 1.0 eq) in anhydrous $CH_2Cl_2$ (4 mL) was treated with DMAP (65.3 mg, 535 µmol, 2.0 eq) and $NEt_3$ (149 µL, 1.07 mmol, 4.0 eq) under nitrogen atmosphere. Benzoic anhydride (59.9 mg, 265 µmol, 0.99 eq) was added in three portions within 3 h and the resulting reaction mixture was stirred for one more hour at ambient temperature. Volatiles were removed under reduced pressure. The crude residue was purified by column chromatography (*n*-hexane:EtOAc + 1% $NEt_3$ = 1:2) to yield compound **68** (183 mg, 215 µmol, 80%) as a colorless foam.

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 8.10 (d, *J* = 2.2 Hz, 1H, NH), 8.02 – 8.05 (m, 2H, bz), 7.60 – 7.64 (m, 1H, bz), 7.48 – 7.53 (m, 2H, bz), 7.45 (d, *J* = 8.3 Hz, 1H, H-6), 7.36 – 7.40 (m, 2H,DMT), 7.22 – 7.34 (m, 6H, DMT), 6.81 – 6.86 (m, 4H, DMT), 6.06 (d, *J* = 4.7 Hz, 1H, H-1'), 5.49 (dd, *J* = 2.3, 8.3 Hz, 1H, H-5), 5.22 (d, *J* = 4.8 Hz, 1H, $OCH_2O$), 5.01 (d, *J* = 4.8 Hz, 1H, $OCH_2O$), 4.48 (q, *J* = 4.6 Hz, 1H, H-3'), 4.28 (t, *J* = 4.9 Hz, 1H, H-2'), 4.11 – 4.16 (m, 1H, H-4'), 3.80 (2s, 6H, $OCH_3$), 3.44 – 3.50 (m, 2H, H-5'), 3.10 (d, *J* = 4.4 Hz, 1H, C3'-O*H*), 1.05 – 1.10 (m, 21H, $Si(CH(CH_3)_2)_3$).
**$^{13}$C-NMR** (100 MHz, $CDCl_3$): $\delta$ (ppm) = 163.50 (C=O), 158.82 ($C_q$-$OCH_3$), 149.58 (C4), 148.44 (C2), 144.47 ($C_q$-DMT), 135.38 ($C_q$-DMT), 135.10 ($C_q$-DMT), 133.89 (bz), 130.66 (C6), 130.31 (DMT), 130.24 (DMT), 129.72 (bz), 128.83 (bz), 128.23 (DMT), 128.15 (DMT), 127.34 (DMT), 113.41 (DMT), 97.46 (C5), 90.89 ($OCH_2O$), 87.29 ($C_q$-DMT), 87.08 (1'), 83.88 (4'), 82.58 (2'), 70.48 (C3'), 62.94 (C5'), 55.39 ($OCH_3$), 17.95, 11.99 ($Si(CH(CH_3)_2)_3$).
**HR-MS (ESI$^+$)**: Exact mass calculated for $C_{47}H_{57}N_3NaO_{10}Si$ $[M+Na]^+$: 874.3705, found: 874.3727.

**$N^4$-*O*-Benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-$N^4$-hydroxy-2'-*O*-(triisopropylsilyloxy)methyl cytidine 3'-cyanoethyl-*N*,*N*-diisopropylphosphoramidite (69)**




**69**

Compound **68** (150 mg, 176 µmol, 1.0 eq) was dissolved in anhydrous $CH_2Cl_2$ (2 mL) and cooled to 4 °C. 2-Cyanoethyl *N,N,N',N'*-tetraisopropyl phosphoramidite (67.3 µL, 212 µmol, 1.2 eq) and 4,5-dicyanoimidazol (23.0 mg, 194 µmol, 1.1 eq) were added in two portions within 1 hour. After one additional hour at 4°C, the reaction mixture was allowed to warm up to ambient temperature and stirred for one more hour. Then, the solvent was evaporated, and the crude residue was purified by column chromatography (*n*-hexane:EtOAc + 1% $NEt_3$ = 2:1 to 1:1) to yield compound **69** (75.0 mg, 71.3 µmol, 40%) as a colorless foam.

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 8.03 (m, 5H, bz), 7.59 – 7.65 (m, 2H, bz), 7.47 – 7.53 (m, 3H, bz), 7.39 – 7.44 (m, 3H, DMT), 7.35 – 7.39 (m, 1H, H-6), 7.21 – 7.35 (m, 8H, DMT), 6.79 – 6.88 (m, 6H, DMT), 6.08 – 6.14 (2d, *J* = 8.2 Hz, 1H, H-1'), 5.58 (d, *J* = 8.2 Hz, 1H, C-5, diast.), 5.54 (d, *J* = 8.2 Hz, 1H, C-5, diast.), 5.00 – 5.05 (m, 2H, $OCH_2O$), 4.96 (d, *J* = 5.0 Hz, 1H, $OCH_2O$, diast.), 4.45 – 4.51 (q, *J* = 5.6 Hz, 1H, H-2'), 4.39 – 4.44 (m, 2H, H-2', H-3'), 4.24 – 4.28 (m, 1H, H-4'), 4.15 – 4.18 (m, 1H, H-4', diast.), 3.81 – 3.99 (m, 1H, $CH_2$-CN), 3.79 – 3.81 (s, 6H, $OCH_3$), 3.63 – 3.71 (m, 1H, $CH_2$-CN, diast.), 3.51 – 3.63 (m, 3H, H-5', NC*H*$(CH_3)_2$, $CH_2$-CN), 3.45 – 3.51 (m, 1H, H-5', diast.), 3.33 – 3.41 (m, 1H, H-5', diast.), 2.63 – 2.68 (td, *J* = 2.9, 6.4 Hz, 1H, $POCH_2$, diast.), 2.36 – 2.41 (t, *J* = 6.4 Hz, 2H, $POCH_2$), 1.13 – 1.19 (m, 13H, NCH(C*H*$_3)_2$), 1.00 – 1.10 (m, 33H, Si(CH(C*H*$_3)_2)_3$, diast.).

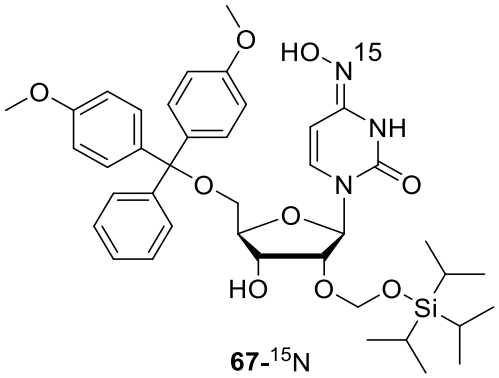
**$^{13}$C-NMR** (100 MHz, $CDCl_3$): $\delta$ (ppm) = 163.51 (C=O, C43), 158.82 ($C_q$-$OCH_3$),149.62, 149.55 (C4, 148.41 (C2), 144.49, 144.33 ($C_q$-DMT), 135.45, 135.35 ($C_q$-DMT), 134.12, 134.08 (C6), 133.63, 133.61 (Bz, C48), 130.36, 130.27 (DMT), 129.73 (Bz), 128.81 (bz), 128.36, 128.27 (DMT), 128.14, 128.11 (DMT), 127.31 (DMT), 117.50 (CN), 113.40, 113.36 (DMT), 97.55, 97.43 (C5), 89.05, 89.02 ($OCH_2O$), 87.36, 87.23 (O-$C_q$, DMT), 87.01, 86.78 (C1'), 83.71 (C4'), 77.36 (C2'), 71.51, 71.37 (C3'), 62.03, 62.63 (C5'), 59.21, 58.00 ($CH_2$-CN), 55.39 ($OCH_3$), 43.52,43.31 (N(*C*H$(CH_3)_2)_2$),

24.75, 24.70 (N(CH(*C*H$_3$)$_2$)$_2$), 20.55, 20.28 (POCH$_2$), 17.95, 17.91 (Si(CH(*C*H$_3$)$_2$)$_3$), 12.06, 12.03 (Si(*C*H(CH$_3$)$_2$)$_3$).

**$^{31}$P-NMR** (162 MHz, CDCl$_3$): *δ* (ppm) = 150.12, 150.61.

**HR-MS (ESI$^+$)**: Exact mass calculated for $C_{56}H_{75}N_5O_{11}PSi$ $[M+H]^+$: 1052.4964, found: 1052.4933 and for $C_{56}H_{74}N_5NaO_{11}PSi$ $[M+Na]^+$: 1074.4784, found: 1074.4749.

**5'-*O*-(4,4'-Dimethoxytrityl)-$^{15}$*N*$^4$-hydroxy-2'-*O*-(triisopropylsilyloxy)methylcytidine (67-$^{15}$N)**




**67-$^{15}$N**

To 5′-*O*-DMT-2′-*O*-TOM-*O*4-chlorophenyluridine (**17**, 115 mg, 136 µmol, 1.0 eq) in anhydrous CH$_3$CN, 4-Dimethylaminopyridine (DMAP, 50.0 mg, 409 µmol, 3.0 eq) and NEt$_3$ (190 µL, 1.36 mmol, 10.0 eq) were added under nitrogen atmosphere. After addition of $^{15}$N-hyroxylamine hydrochloride ($^{15}$NH$_2$OH, 38.4 mg, 545 µmol, 4.0 eq) the reaction mixture was stirred for 24h at ambient temperature, diluted with CH$_2$Cl$_2$ and washed with saturated aqueous NaHCO$_3$ (2x). The organic phase was dried over Na$_2$SO$_4$, and the solvent was removed under reduced pressure. The crude product was purified by column chromatography (*n*-hexane:EtOAc + 1% NEt$_3$ = 1:2 to 2:1) to yield compound **67-$^{15}$N** (68.0 mg, 90.8 µmol, 67%) as a colorless foam.

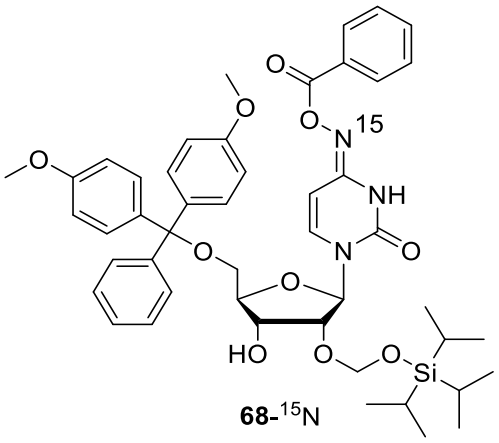
**$^1$H-NMR** (CDCl$_3$, 400 MHz): *δ* (ppm) = 8.55 (s, 1H, N3-H), 7.41 – 7.36 (m, 2H, DMT), 7.32 – 7.26 (m, 6H, DMT), 7.25 – 7.20 (m, 1H, DMT), 7.16 (d, *J* = 8.3 Hz, 1H, H-6), 6.86 – 6.81 (m, 4H, DMT), 6.03 (d, *J* = 4.4 Hz, 1H, H-1'), 5.25 (d, *J* = 8.2 Hz, 1H, H-5), 5.21 (d, *J* = 4.7 Hz, 1H, OCH$_2$O), 4.42 (s, 1H, H-3'), 4.24 (t, *J* = 4.8 Hz, 1H, H-2'), 4.09 (dt, *J* = 5.1, 2.6 Hz, 1H, H-4'), 3.80 (s, 6H, OCH$_3$), 3.49 – 3.39 (m, 2H, H-5'), 3.12 (s, 1H, C2'-O*H*), 1.18 – 1.01 (m, 21H, Si(CH(CH$_3$)$_2$)$_3$)

**$^{13}$C-NMR** (100 MHz, CDCl$_3$): *δ* (ppm) = 158.69 (C$_q$-OCH$_3$), 149.58 (C2), 145.34 (C4), 144.42 (C$_q$-DMT), 135.77 (C$_q$-DMT), 135.48 (C$_q$-DMT), 130.68 (C6), 130.31 (DMT), 128.36 (DMT), 128.09 (DMT), 127.19 (DMT), 113.39 (DMT) 98.42 (C5), 90.79 (OCH$_2$O), 87.07, 87.03 (C1', C$_q$-DMT), 83.47 (C4'), 82.33 (C2'), 70.24 (C3'), 62.92 (C5'), 55.44, 55.43 (OCH$_3$), 17.95, 11.99 (Si(CH(CH$_3$)$_2$)$_3$).

**HR-MS (ESI$^+$)**: Exact mass calculated for $C_{40}H_{53}NaN_2{}^{15}NO_9Si$ $[M+Na]^+$: 771.34136, found: 771.34164

**$^{1}$H-NMR** (DMSO-d6, 400 MHz): *δ* (ppm) = 9.99 (d, *J* = 2.2 Hz, 1H, $^{15}$N-O*H*), 9.53 (d, *J* = 2.2 Hz, 1H, N3-H), 7.40 – 7.36 (m, 2H, DMT), 7.33 – 7.28 (m, 2H, DMT), 7.28 – 7.24 (m, 4H, DMT), 7.23 – 7.20 (m, 1H, DMT), 6.91 – 6.87 (m, 4H, DMT), 6.86 (d, *J* = 8.4 Hz, 1H, H-6), 5.87 (d, *J* = 6.1 Hz, 1H, H-1'), 5.39 – 5.35 (m, 1H, H-5), 5.13 (d, *J* = 5.7 Hz, 1H, C2'-O*H*), 4.95 (d, *J* = 5.3 Hz, 1H, $OCH_2O$), 4.90 (d, *J* = 5.3 Hz, 1H, $OCH_2O$), 4.24 (t, *J* = 5.8 Hz, 1H, H-2'), 4.10 – 4.05 (m, 1H, H-3'), 3.91 (q, *J* = 3.8 Hz, 1H, H-4'), 3.73 (s, 6H, $OCH_3$), 3.19 (qd, *J* = 10.6, 3.8 Hz, 2H, H-5'), 1.05 – 0.92 (m, 21H, $Si(CH(CH_3)_2)_3$).
**$^{13}$C-NMR** (100 MHz, DMSO-d6): *δ* (ppm) = 158.12 ($C_q$-$OCH_3$), 149.27 (C2), 144.68 (Cq-DMT), 143.17 (C4), 135.43, 135.33 ($C_q$-DMT), 129.77 (DMT, C6), 127.91 (DMT), 127.72 (DMT), 126.79 (DMT), 113.24 (DMT), 98.71 (C5), 88.29 ($OCH_2O$), 85.93, 85.72 (C1', $C_q$-DMT), 83.07 (C4'), 76.18 (C2'), 68.91 (C3'), 63.65 (C5'), 55.05 ($OCH_3$), 17.65, 11.38 ($Si(CH(CH_3)_2)_3$).
**$^{15}$N-NMR** (40.532 MHz, DMSO-d6): *δ* (ppm) = -262.8 (N3, amide), -93.3 (N4).

***N*$^{4}$-*O*-Benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-$^{15}$*N*$^{4}$-hydroxy-2'-*O*-(triisopropylsilyloxy)methylcytidine (68-$^{15}$N)**



**68-**$^{15}$N

5'-*O*-(4,4'-Dimethoxytrityl)-*N*$^{4}$-hydroxy-2'-*O*-(triisopropylsilyloxy)methylcytidine **67-$^{15}$N** (62.0 mg, 82.8 µmol, 1.0 eq) was dissolved in anhydrous $CH_2Cl_2$ under nitrogen atmosphere and treated with 4-Dimethylaminopyridine (DMAP, 20.2 mg, 166 µmol, 2.0 eq) and $NEt_3$ (46.2 µL, 331 µmol, 4.0 eq). Benzoic anhydride (18.5 mg, 81.95 µmol, 0.99 eq) was added in three portions within 3 hours and stirred for additional three hours at ambient temperature. The solvent was removed under reduced pressure. The crude product was purified by column chromatography (*n*-hexane:EtOAc + 1% $NEt_3$ = 2:1) to yield compound **68-$^{15}$N** (57.0 mg, 66.8 µmol, 80%) as a colorless foam.

**$^{1}$H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 8.08 – 8.12 (s, 1H, N3), 8.02 – 8.06 (m, 2H, bz), 7.59 – 7.64 (m, 1H, bz), 7.47 – 7.52 (m, 2H, bz), 7.43 – 7.46 (dd, *J* = 0.7, 8.3 Hz, 1H, H-6), 7.36 – 7.41 (m, 2H, DMT), 7.26 – 7.33 (m, 6H, DMT), 7.20 – 7.26 (m, 1H, DMT), 6.79 – 6.87 (m, 4H, DMT), 6.04 – 6.07 (d, *J* = 4.7 Hz, 1H, H-1'), 5.47 – 5.51 (dd, *J* = 1.0, 8.2 Hz, 1H, H-5), 5.20 – 5.24 (d, *J* = 4.8 Hz, 1H, $OCH_2O$), 4.98 – 5.02 (d, *J* = 4.8 Hz, 1H, $OCH_2O$), 4.45 – 4.50 (q, *J* = 4.5 Hz, 1H, H-3'), 4.25 – 4.30

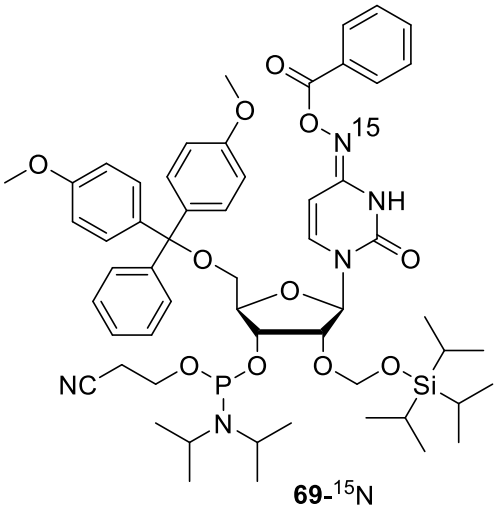
(t, *J* = 4.9 Hz, 1H, H-2'), 4.12 – 4.15 (m, 1H, H-4'), 3.78 – 3.82 (d, *J* = 0.6 Hz, 6H, $OCH_3$), 3.43 – 3.51 (m, 2H, H-5'), 3.07 – 3.12 (d, *J* = 4.3 Hz, 1H, C2'-O*H*), 1.04 – 1.17 (m, 21H, $Si(CH(CH_3)_2)_3$).

**$^{13}$C-NMR** (100 MHz, $CDCl_3$): *δ* (ppm) = 163.51 (C(O)bz), 158.86, 158.82 ($C_q$-$OCH_3$), 149.56 (C4), 148.44 (C2), 144.47 ($C_q$-DMT), 135.38, 135.11 ($C_q$-DMT), 133.91 (C6), 133.67 (bz), 130.32 (DMT), 130.24 (DMT), 129.73 (bz), 128.83 (bz), 128.24 (DMT), 128.15 (DMT), 127.34 (DMT), 113.41 (DMT), 97.41 (C5), 90.89 ($OCH_2O$), 87.30 ($C_q$-DMT), 87.09 (C1'), 83.89 (C4'), 82.58 (C2'), 70.48 (C3'), 62.94 (C5'), 55.39 ($OCH_3$), 17.95, 11.99 ($Si(CH(CH_3)_2)_3$).

**$^{15}$N-NMR** (40.532 MHz, DMSO-d6): *δ* (ppm) = -256.2 (N3, amide), -91.6 (N4).

**HR-MS (ESI$^+$)**: Exact mass calculated for $C_{47}H_{57}NaN_2{}^{15}NO_{10}Si$ $[M+Na]^+$: 875.36758, found: 875.36985

**$N^4$-*O*-Benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-$N^4$-hydroxy-2'-*O*-(triisopropylsilyloxy)methyl cytidine 3'-cyanoethyl-*N*,*N*-diisopropylphosphoramidite (69)**




**69-$^{15}$N**

To $N^4$-*O*-Benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-$N^4$-hydroxy-2'-*O*-(triisopropylsilyloxy)methyl cytidine **68-$^{15}$N** (50.0 mg, 58.6 µmol, 1.0 eq) in anhydrous $CH_2Cl_2$ (2 mL), 4,5-dicyanoimidazol (DCI, 7.62 mg, 64.5 µmol, 1.1 eq) and 2-Cyanoethyl *N*,*N*,*N*′,*N*′-tetraisopropyl phosphoramidite (22.4 µL, 70.6 µmol, 1.2 eq) were added at 4°C. The reaction mixture was allowed to warm to ambient temperature and stirred for 3 more hours. The solvent was removed under reduced pressure and the crude product was purified by column chromatography (n-hexane:EtOAc + 1% $NEt_3$ = 2:1) to yield compound **69-$^{15}$N** (25.0 mg, 23.7 µmol, 40%) as a colorless foam.

**$^1$H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 8.02 – 8.06 (m, 6H, bz, diast.), 7.59 – 7.65 (m, 2H, bz, diast.), 7.47 – 7.52 (m, 4H, bz, diast.), 7.41 – 7.43 (d, *J* = 8.2 Hz, 1H, H-6, diast.), 7.39 – 7.42 (p, *J* = 1.2 Hz, 2H, DMT, diast.), 7.35 – 7.38 (d, *J* = 8.1 Hz, 2H, H-6), 7.25 – 7.34 (m, 9H, DMT, diast.), 7.20 – 7.26 (m, 1H, DMT), 6.81 – 6.88 (m, 8H, DMT, diast.), 6.09 – 6.13 (dd, *J* = 5.3, 8.1 Hz, 2H, H-1', diast.), 5.56 – 5.59 (d, *J* = 8.2 Hz, 1H, H-5), 5.52 – 5.55 (d, *J* = 8.2 Hz, 1H, H-5, diast.), 5.00 – 5.05 (m, 3H, $OCH_2O$), 4.95 – 4.97 (d, *J* = 5.1 Hz, 1H, $OCH_2O$, diast.), 4.45 – 4.51 (q, *J* = 5.5 Hz,

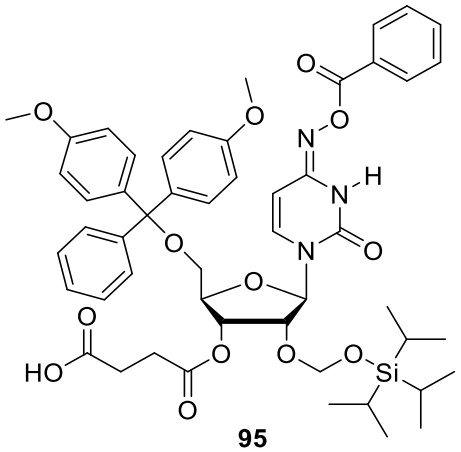
1H, H-2'), 4.40 – 4.44 (m, 2H, H-2', diast., H-3'), 4.22 – 4.29 (dd, *J* = 2.9, 5.9 Hz, 2H, H-4'), 4.14 – 4.21 (dd, *J* = 2.9, 5.9 Hz, 1H, H-4', diast.), 3.82 – 4.00 (m, 2H, $CNCH_2$, diast.), 3.78 – 3.81 (d, *J* = 3.0 Hz, 12H, $OCH_3$, diast.), 3.62 – 3.70 (m, 1H, $CNCH_2$, diast.), 3.53 – 3.61 (m, 4H, H-5', $NCH(CH_3)_2$,$CNCH_2$), 3.46 – 3.51 (dd, *J* = 2.6, 10.8 Hz, 1H, H-5', diast.), 3.33 – 3.41 (ddd, *J* = 2.8, 6.8, 10.4 Hz, 2H, H-5', diast.), 2.62 – 2.69 (td, *J* = 2.8, 6.4 Hz, 2H, $POCH_2$, diast.), 2.35 – 2.43 (t, *J* = 6.4 Hz, 2H, $POCH_2$), 1.14 – 1.22 (m, 17H, $NCH(CH_3)_2$, diast.), 0.98 – 1.11 (m, 44H, $Si(CH(CH_3)_2)_3$, diast.).

**$^{31}$P NMR** (162 MHz, $CDCl_3$): δ (ppm) = 150.12, 150.60.

**HR-MS (ESI$^+$)**: Exact mass calculated for $C_{56}H_{74}NaN_4{}^{15}NO_{11}PSi$ $[M+Na]^+$:1075.47543, found: 1075.47851

5.1.2.11 Preparation of *N*$^4$-hdroxycytidine CPG (**96**)[503] and triphosphate (**97**)

***N*$^4$-*O*-Benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-*N*$^4$-hydroxy-2'-*O*-(succinic acid)cytidine-CPG (95)**

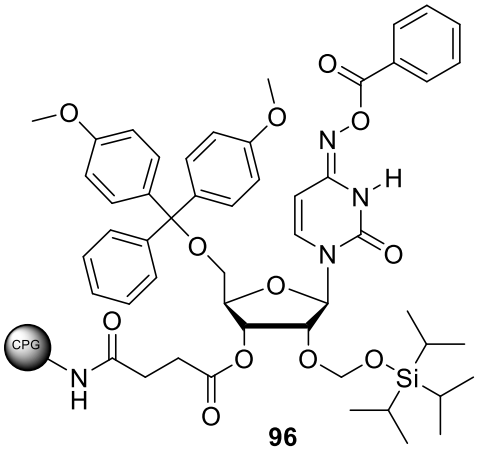


**95**

Compound **68** (31.0 mg, 36.4 µmol, 1.0 eq.) was dissolved in DCM (200 µL). Succinic anhydride (3.60 mg, 36.4 µmol, 1.0 eq) and DMAP (6.67 mg, 54.6 µmol, 1.5 eq) were added and the mixture was stirred for 4 h at room temperature. The mixture was diluted with DCM (3 mL) and washed with 10% citric acid and brine, respectively. The organic phase was dried over $Na_2SO_4$ and the solvent was removed under reduced pressure to yield compound **95** (26.0 mg, 27.4 µmol, 75%) as a colorless solid.

**$^1$H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 8.60 (s, 1H, N-OH), 8.16 – 7.98 (m, 2H, Bz), 7.68 – 7.54 (m, 1H, Bz), 7.49 (ddd, *J*= 8.1, 6.7, 1.2 Hz, 2H, Bz), 7.40 – 7.35 (m, 2H, DMT), 7.32 – 7.21 (m, 6H, H-6, DMT), 6.87 – 6.81 (m, 3H, DMT), 6.11 (d, *J*= 7.2 Hz, 1H, H-1'), 5.53 (dd, *J*= 8.2, 2.2 Hz, 1H, H-5), 5.29 (d, *J*= 4.8 Hz, 2H, H-3'), 4.92 – 4.88 (m, 2H, OCH2O), 4.68 (s, 1H, $(CH_2)_2COOH$), 4.57 (dd, *J*= 7.2, 5.1 Hz, 1H, H-2'), 4.13 (d, *J*= 2.3 Hz, 1H, H-4'), 3.79 (d, *J*= 1.2, 6H Hz, OCH3), 3.50 – 3.45 (m, 2H, H5'), 2.64 (d, *J*= 4.3 Hz, 3H, $(CH_2)_2COOH$), 1.08 – 0.98 (m, 22H, $(Si(CH(CH_3)_2)_3$).

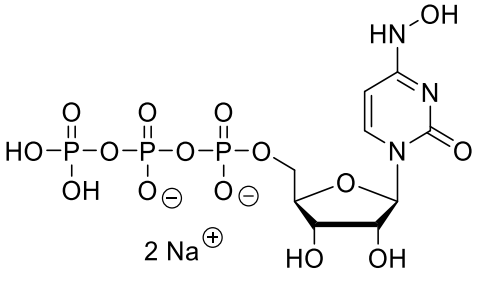
**$N^4$-O-Benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-$N^4$-hydroxy-2'-*O*-(triisopropylsilyloxy)methyl cytidine-CGP (96)**




**96**

Compound **95** (26.0 mg, 30.9 µmol, 1.0 eq) was dissolved in anhydrous acetonitrile (3 mL), BOP (27.3 mg, 61.7 µmol, 2.0 eq) and *N*-methylimidazole (9.20 mg, 117 µmol, 3.8 eq) were added to the LCAA-CPG (150 mg, 500 Å, amine loading: 82 µmol/g) in a syringe with a filter. The mixture was shaken for 20 h at room temperature. The solution was filtered off and the solid support was washed with acetonitrile and DCM. A mixture of pyridine, acetic anhydride (3:1, 3.6 mL), and DMAP (24.5 mg, 201 µmol, 6.5 eq) were added to the CPG within the syringe. The suspension was shaken for 2 h at room temperature. The solution was filtered off, and the solid support was washed with DCM and dried under reduced pressure. A loading density of 54.0 µmol/g was determined with equation (3) for compound **96**.

**N4-hydroxycytidine-5'-triphosphate (97)**




**97**

Cytidine 5'-triphosphate disodium salt (137 mg, 260 µmol, 1.0 eq) was dissolved in an aqueous $NH_2OH$ solution (2 M, 2.00 mL, 4.00 mmol, pH 5, 15.4 eq) in a sealable pressure tube. The reaction mixture was stirred for 5 h at 55 °C. After allowing it to cool to ambient temperature, 2 mL of TEAB buffer (100 mM, pH 7.5) was added, and the solvent was removed under reduced pressure. The residue was dissolved in 10 mL $H_2O$ and a fraction was purified via DEAE and RP-HPLC, respectively.[504]

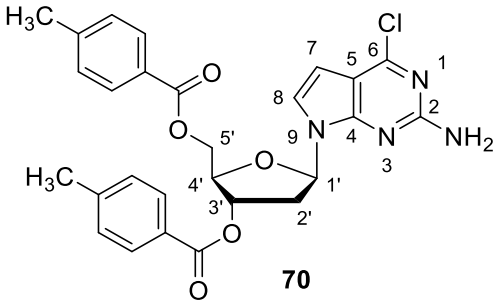
**$^{1}$H** NMR (400 MHz, $D_2O$) δ (ppm) = 7.19 (d, *J*=8.4, 1H), 5.95 (d, *J*=6.4, 1H), 5.82 (d, *J*=8.3, 1H), 4.44 – 4.30 (m, 2H), 4.24 – 4.14 (m, 4H).

**$^{31}$P** NMR (162 MHz, Deuterium Oxide) δ (ppm) = -10.65, -11.52 (d, *J*=19.8), -23.24.

**HR-MS** (ESI+): Exact mass calc. for $C_9H_{15}N_3O_{15}P_3$ $[M+H]^+$: 497.9722, found: 497.9710.

5.1.2.12 Preparation of c7dG-phosphoramidite (**81**)[484, 485] and triphosphate

**6-Chloro-7-deaza-2'-deoxy-3',5'-di-*O*-p-toluoyl-guanosine**




**70**

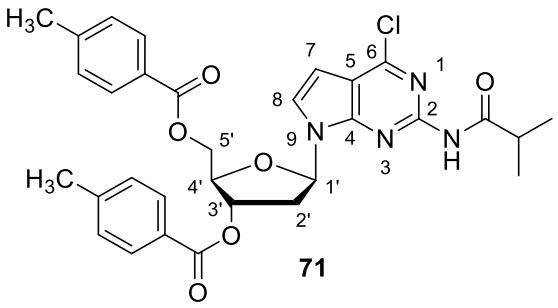
Tris[2-(2-methoxyethoxy)ethyl]amine (74.6 mg, 148 µmol, 0.05 eq) was added to a suspension of KOH (333 mg, 5.93 mmol, 2.0 eq) in dry MeCN (20 mL) and stirred for 15 min. 4-Chloro-7H-pyrrolo[2,3- d]pyrimidin-2-amine (500 mg, 2.97 mmol, 1.0 eq) was added and stirred for 10 min. Hoffer's chlorosugar (1.17 g, 3.00 mmol, 1.01 eq) was added in three portions within 15 min and then stirred for 15 min. The reaction mixture was filtrated, and the solvent was removed in vacuum. The crude product was purified by silica gel column chromatography (DCM:EtOAc, 0 to 5% EtOAc). All volatiles were removed in vacuum to give **70** as a colorless solid (1.21 g, 2.32 mmol, 78%).

**$^{1}$H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 8.01 – 7.94 (m, 2H, tol), 7.97 – 7.89 (m, 2H, tol), 7.32 – 7.24 (m, 2H, tol), 7.28 – 7.21 (m, 2H, tol), 7.01 (d, J = 3.8 Hz, 1H, H8), 6.58 (dd, J = 8.5, 5.8 Hz, 1H, H-1'), 6.39 (d, J = 3.8 Hz, 1H, H-7), 5.74 (dt, J = 6.6, 2.3 Hz, 1H, H-3'), 4.98 (s, 2H, $NH_2$), 4.76 (dd, J = 11.4, 3.9 Hz, 1H, H-5'), 4.63 – 4.59 (m, 1H, H-5'), 4.58 – 4.55 (m, 1H, H-4'), 2.98 – 2.85 (m, 1H, H-2'), 2.65 (ddd, J = 14.1, 5.8, 2.2 Hz, 1H, H-2'), 2.44 (s, 3H, $CH_3$), 2.42 (s, 3H, $CH_3$).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): *δ* (ppm) = 166.45 (C=O), 166.16 (C=O), 158.67 (C2), 153.75 (C4), 152.98 (C6), 144.60 ($C_q$-$CH_3$), 144.30 ($C_q$-$CH_3$), 129.97 (tol), 129.81 (tol), 129.42 (tol), 126.94 ($C_q$-C=O), 126.68 ($C_q$-C=O), 122.69 (C8), 111.50 (C5), 102.00 (C7), 84.29 (C1'), 82.15 (C4'), 75.35 (C3'), 64.33 (C5'), 37.44 (C2'), 21.91 ($CH_3$), 21.86 ($CH_3$).

**HR-MS** ($ESI^+$): Exact mass calculated for $C_{27}H_{25}ClN_4NaO_5$ $[M+Na]^+$: 543.1406, found: 543.1400.

**6-Chloro-7-deaza-*N*(2)-isobutyryl-3',5'-di-*O*-p-toluoyl-2'-deoxyguanosine (71)**

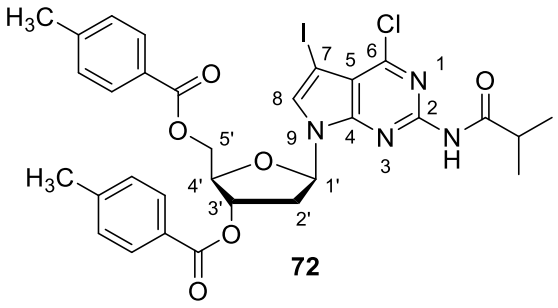


**71**

To a solution of 6-chloro-7-deaza-2'-deoxy-3',5'-di-*O*-p-toluoyl-guanosine **70** (2.82 g, 5.40 mmol, 1.0 eq) in dry pyridine (14 mL), isobutyryl chloride (623 µL, 5.94 mmol, 1.1 eq) was added at 0 °C and then stirred for 1.5 h at room temperature. After stopping the reaction with $H_2O$ (100 µL), the solvent was removed in vacuum. The solid residue was re-dissolved in ethyl acetate and washed with 5% HCl, sat. $NaHCO_3$ and brine, dried over $Na_2SO_4$ and evaporated to dryness under reduced pressure. The crude product was purified by silica gel column chromatography (DCM:EtOAc, 0 to 5% EtOAc). All volatiles were removed in vacuum to give **71** as a colorless solid (2.81 g, 4.75 mmol, 88%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 8.02 (s, 1H, NH), 8.01 – 7.94 (m, 2H, tol), 7.94 – 7.87 (m, 2H, tol), 7.31 – 7.27 (m, 2H, tol), 7.26 – 7.20 (m, 2H, tol), 6.70 (dd, J = 8.3, 5.9 Hz, 1H, H-1'), 6.51 (d, J = 3.8 Hz, 1H, H-7), 5.77 (dt, J = 6.5, 2.5 Hz, 1H, H-3'), 4.76 (dd, J = 11.7, 4.0 Hz, 1H, H-5'), 4.63 (dd, J = 11.7, 4.4 Hz, 1H, H-5'), 4.59 (td, *J* = 4.1, 2.5 Hz, 1H, H-4'), 2.96 (ddd, J = 14.5, 8.3, 6.5 Hz, 2H, H-2', C*H*$(CH_3)_2$), 2.76 (ddd, J = 14.5, 6.0, 2.3 Hz, 1H, H-2'), 2.44 (s, 3H, $CH_3$), 2.42 (s, 3H, $CH_3$), 1.29 (dd, J = 0.6 Hz, 3H, CH(C*H*$_3)_2$), 1.27 (d, *J* = 0.6 Hz, 3H, CH(C*H*$_3)_2$).
**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 175.65 (NH-C=O), 166.39 (C=O, tol), 166.20 (C=O, tol), 152.56 (C6), 152.09 (C4), 151.35 (C2) 144.62 ($C_q$-$CH_3$), 144.31 ($C_q$-$CH_3$), 129.98 (tol), 129.78 (tol), 129.43 (tol), 129.41 (tol), 126.90 (*C*$_q$-C=O), 126.60 (*C*$_q$-C=O), 125.48 (C8), 114.77 (C5), 101.46 (C7), 84.74 (C1'), 82.46 (C4'), 75.22 (C3'), 64.28 (C5'), 37.94 (C2'), 36.19 (*C*H$(CH_3)_2$), 21.91 ($CH_3$), 19.42 (CH(*C*H$_3)_2$), 19.40 (CH(*C*H$_3)_2$).
**HR-MS** (ESI$^+$): Exact mass calculated for $C_{31}H_{31}ClN_4NaO_6$ $[M+Na]^+$: 613.1830, found: 613.1823.

**6-Chloro-7-deaza-7-iodo-*N*(2)-isobutyryl-3',5'-di-*O*-p-toluoyl-2'-deoxyguanosine (72)**

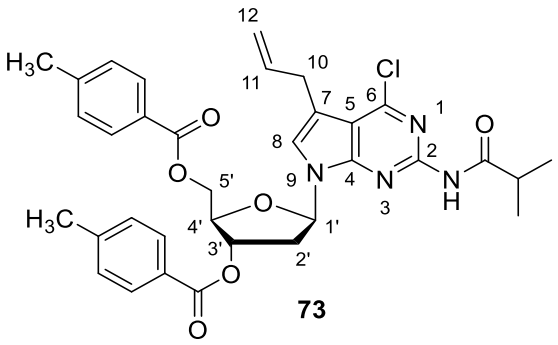


**72**

To a solution of 6-chloro-7-deaza-*N*(2)-isobutyryl-3',5'-di-*O*-p-toluoyl-2'-deoxyguanosine **71** (470 mg, 795 µmol, 1.0 eq) in dry DMF (15 mL), *N*-iodo-succinimide (358 mg, 1.59 mmol, 2.0 eq) was added and stirred for 4.5 h at 80 °C. The reaction mixture was diluted with ethyl acetate and washed with sat. $NaHCO_3$, sat. $Na_2S_2O_3$ and brine, dried over $Na_2SO_4$ and evaporated to dryness under reduced pressure. The crude product was purified by silica gel column chromatography (DCM:EtOAc, 2 to 5% EtOAc). All volatiles were removed in vacuum to give **72** as a colorless solid (547 mg, 763 µmol, 96%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 8.02 (s, 1H, NH), 8.00 – 7.93 (m, 2H, tol), 7.93 – 7.87 (m, 2H, tol), 7.41 (s, 1H, H-8), 7.28 (d, J = 8.2 Hz, 2H, tol), 7.28 – 7.19 (m, 2H, tol), 6.68 (dd, J = 8.0, 6.0 Hz, 1H, H-1'), 5.75 (dt, J = 6.3, 2.5 Hz, 1H, H-3'), 4.77 (dd, J = 12.0, 3.9 Hz, 1H, H-5'), 4.65 (dd, J = 12.0, 3.9 Hz, 1H, H-5'), 4.58 (td, J = 3.9, 2.6 Hz, 1H, H-4'), 2.90 (s, 1H, C*H*$(CH_3)_2$), 2.87 – 2.82 (m, 1H, H-2'), 2.76 (ddd, J = 14.3, 6.1, 2.5 Hz, 1H, H-2'), 2.44 (s, 3H, $CH_3$, tol), 2.42 (s, 3H, $CH_3$, tol), 1.27 (dd, J = 6.9, 0.9 Hz, 6H, CH(C*H*$_3)_2$).
**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): *δ* (ppm) = 175.61 (NH-C=O), 166.35 (C=O, tol), 166.18 (C=O, tol), 153.22 (C6), 151.67 (C4), 151.25 (C2), 144.66 (*C*$_q$-$CH_3$, tol), 144.38 (*C*$_q$-$CH_3$, tol), 130.89 (tol), 129.98 (tol), 129.76 (tol), 129.56 (tol), 129.43 (tol), 126.77 ($C_q$, tol), 126.52 ($C_q$, tol), 113.85 (C5), 84.74 (C1'), 75.13 (C4'), 63.27 (C5'), 53.31 (C7), 38.30 (C2'), 36.65 (*C*H$(CH_3)_2$), 21.90 ($CH_3$, tol), 21.89 ($CH_3$, tol), 19.37 (CH(*C*$H_3)_2$), 19.34 (CH(*C*$H_3)_2$).
**HR-MS** (ESI$^+$): Exact mass calculated for $C_{31}H_{30}IClN_4NaO_6$ $[M+Na]^+$: 739.0791, found: 739.0808.

**7-Allyl-6-chloro-7-deaza-*N*(2)-isobutyryl-3',5'-di-*O*-p-toluoyl-2'-deoxyguanosine (73)**

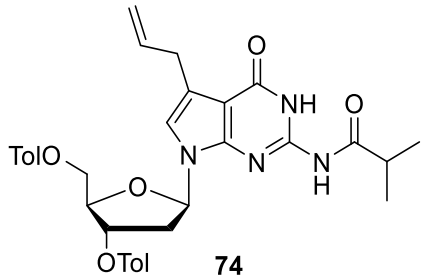



To a solution of 6-chloro-7-deaza-7-iodo-*N*(2)-isobutyryl-3',5'-di-*O*-p-toluoyl-2'- deoxyguanosine (**72**, 586 mg, 817 µmol, 1.0 eq) in anhydrous toluene (12 mL), $Pd(PPh_3)_4$ (94.5 mg, 81.7 µmol, 0.1 eq) and allyl tributyl stannane (380 µL, 1.23 mmol, 1.5 eq) was added and stirred for 16 h at 95 °C. After removing the solvent in vacuum, the crude product was purified by column chromatography (DCM:EtOAc, 2 to 6% EtOAc). All volatiles were removed *in vacuo* to give **73** as a colorless solid (443 mg, 702 µmol, 86%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 8.00 – 7.95 (m, 2H, tol), 7.94 – 7.91 (m, 2H, tol), 7.30 – 7.27 (m, 2H, tol), 7.26 – 7.22 (m, 2H, tol), 6.98 (d, J = 1.2 Hz, 1H, H-8), 6.72 (dd, J = 8.5, 5.8 Hz, 1H, H-1'), 5.86 (ddt, J = 16.6, 10.1, 6.5 Hz, 1H, C*H*=$CH_2$), 5.76 (dt, J = 6.4, 2.3 Hz, 1H, H-3'), 5.03 – 4.94 (m, 1H, CH=C$H_2$), 4.98 (p, *J* = 1.5 Hz, 1H, CH=C$H_2$), 4.78 (dd, *J* = 11.7, 3.7 Hz, 1H, H-5'), 4.60 (dd, *J* = 11.8, 4.0 Hz, 1H, H-5'), 4.56 (td, *J* = 3.8, 2.4 Hz, 1H, H-4'), 3.47 (ddq, *J* = 6.7, 2.8, 1.4 Hz, 2H, C7-$CH_2$), 2.97 (s, 1H, C*H*($CH_3$)$_2$), 2.90 (ddd, *J* = 14.7, 8.5, 6.5 Hz, 1H, H-2'), 2.71 (ddd, *J* = 14.2, 5.9, 2.2 Hz, 1H, H-2'), 2.44 (s, 3H, $CH_3$, tol), 2.42 (s, 3H, $CH_3$, tol),1.28 (d, *J* = 0.5 Hz, 3H, CH(C$H_3$)$_2$), 1.27 (d, *J* = 0.5 Hz, 2H, CH(C$H_3$)$_2$).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 175.77 (CH-C=O), 166.33 (C=O, tol), 166.23 (C=O, tol), 152.81 (C4), 152.33 (C6), 151.17 (C2), 144.61 ($C_q$-$CH_3$, tol), 144.30 ($C_q$-$CH_3$, tol), 135.73 (*C*H=$CH_2$), 129.98 (tol), 129.79 (tol), 129.44 (tol), 129.43 (tol), 126.97 ($C_q$, tol), 126.61 ($C_q$, tol), 122.68 (C8), 116.68 (CH=*C*$H_2$), 115.66 (C7), 113.24 (C5), 84.15 (C1'), 82.37 (C4'), 75.32 (C3'), 64.35 (C5'), 37.94 (C2'), 36.11 (*C*H($CH_3$)$_2$), 30.31 (C7-*C*$H_2$), 21.91 ($CH_3$ ,tol), 21.86 ($CH_3$, tol), 19.42 (CH(*C*$H_3$)$_2$), 19.39 (CH(*C*$H_3$)$_2$).

**7-Allyl-7-deaza-*N*(2)-isobutyryl-3',5'-di-*O*-p-toluoyl-2'-deoxyguanosine (74)**




**74**

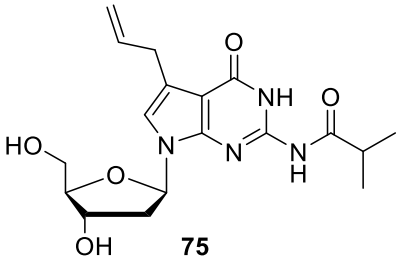
To a solution of 7-allyl-6-chloro-7-deaza-*N*(2)-isobutyryl-3',5'-di-*O*-p-toluoyl-2'- deoxyguanosine **73** (500 mg, 792 µmol, 1.0 eq) in a mixture of dry dioxane (8.5 mL) and DMF (9 mL), pyridine-2-carboxaldoxime (484 mg, 3.96 mmol, 5.0 eq) and 1,1,3,3-tetramethylguanidine (496 µL, 3.96 mmol, 5.0 eq) were added and stirred for 3 days at 40 °C. After removing the solvent *in vacuo*, the crude product was purified by column chromatography (DCM:EtOAc, 2 to 6% EtOAc). All volatiles were removed in vacuum to give **74** as a colorless solid (388 mg, 633 µmol, 80%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 11.58 (s, 1H, NH), 8.57 (s, 1H, C2-NH), 7.96 – 7.93 (m, 2H, tol), 7.92 – 7.88 (m, 2H, tol), 7.28 – 7.25 (m, 2H, tol), 7.25 – 7.22 (m, 2H, tol), 6.56 (t, *J* = 1.1 Hz, 1H, H-8), 6.30 (dd, J = 8.1, 6.2 Hz, 1H, H-1'), 5.99 (ddt, J = 16.8, 10.0, 6.8 Hz, 1H, C*H*=$CH_2$), 5.80 (dd, *J* = 5.5, 2.8 Hz, 1H, H-3'), 5.07 (dq, *J* = 17.4, 1.9 Hz, 1H, CH=C$H_2$), 5.06 – 5.00 (m, 1H, H-5'), 4.97 (ddt, *J* = 10.0, 2.2, 1.2 Hz, 1H, CH=C$H_2$), 4.60 (dd, *J* = 4.7, 2.5 Hz, 1H, H-4'), 4.56 (d, *J* = 4.7 Hz, 1H, H-5'), 3.50 (ddq, *J* = 6.4, 3.8, 1.3 Hz, 2H, C7-$CH_2$), 3.06 (ddd, *J* = 14.2, 8.1, 6.2 Hz, 1H, H-2'), 2.69 (hept, *J* = 6.9 Hz, 1H, C*H*($CH_3$)$_2$), 2.55 (ddd, *J* = 14.2, 6.2, 2.6 Hz, 1H, H-2'), 2.44 (s, 3H, $CH_3$, tol), 2.42 (s, 3H, $CH_3$, tol), 1.29 (dd, *J* = 6.9, 5.0 Hz, 6H, CH(C$H_3$)$_2$).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 178.36 (C2-NH-*C*=O), 167.00 (C=O, tol), 166.11 (C=O, tol), 157.97 (C6), 147.05 (C4), 145.87 (C2), 144.55 ($C_q$-$CH_3$, tol), 144.50 ($C_q$-$CH_3$, tol), 136.82 (*C*H=$CH_2$), 129.90 (tol), 129.84 (tol), 129.44 (tol), 129.38 (tol), 126.78 ($C_q$, tol), 126.75 ($C_q$, tol), 119.68 (C7), 117.54 (C8), 115.62 (CH=*C*$H_2$), 105.24 (C5), 85.93 (C1'), 81.94 (C4'), 75.42 (C3'), 63.92 (C5'), 36.94 (C2'), 36.67 (*C*H($CH_3$)$_2$), 30.54 (C7-$CH_2$), 21.90 ($CH_3$, tol), 21.87 ($CH_3$, tol), 19.21 (CH(*C*$H_3$)$_2$), 18.69 (CH(*C*$H_3$)$_2$).

**HR-MS** (ESI$^+$): Exact mass calculated for $C_{34}H_{36}N_4NaO_7$ [M+Na]$^+$: 635.2476 found: 635.2483.

**7-Allyl-7-deaza-*N*(2)-isobutyryl-2'-deoxyguanosine (75)**




**75**

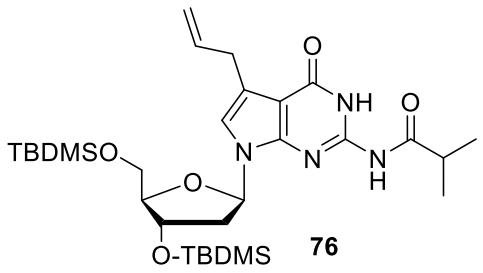
To a solution of 7-allyl-7-deaza-*N*(2)-isobutyryl-3',5'-di-*O*-p-toluoyl-2'-deoxyguanosine **74** (200 mg, 326 µmol, 1.0 eq) in dry THF (6.5 mL), 0.5 M NaOMe in MeOH (660 µL, 330 µmol, 1.01 eq) was added at 0 °C. The ice bath was removed, and the reaction was stirred for 4 h at room temperature. After neutralization of the reaction mixture with acetic acid and removing the solvent in vacuum, the crude product was purified by column chromatography (DCM:MeOH, 6 to 8% MeOH). All volatiles were removed in vacuum to give **75** as a colorless solid (96.5 mg, 256 µmol, 79%).

**$^1$H-NMR** (DMSO-d6, 400 MHz): $\delta$ (ppm) = 11.71 (s, 1H, NH), 11.47 (s, 1H, C2-NH), 6.96 (d, J = 1.1 Hz, 1H, H-8), 6.37 (dd, J = 8.6, 5.6 Hz, 1H, H-1'), 6.05 (ddt, J = 16.7, 10.0, 6.7 Hz, 1H, C*H*=$CH_2$), 5.23 (d, J = 3.4 Hz, 1H, C3'-OH ), 5.07 (ddt, J = 17.1, 2.2, 1.6 Hz, 1H, CH=C*H*$_2$), 4.98 (ddt, J = 10.0, 2.2, 1.3 Hz, 1H, CH=C*H*$_2$), 4.88 (t, J = 5.4 Hz, 1H, C5'-OH), 4.29 (dq, J = 5.8, 2.5 Hz, 1H, H-3'), 3.76 (td, J = 4.8, 2.3 Hz, 1H, H-4'), 3.55 – 3.41 (m, 2H, H-5'), 3.43 (d, J = 2.1 Hz, 1H, C7-$CH_2$), 2.81 – 2.66 (m, J = 6.9 Hz, 1H,C*H*($CH_3$)$_2$), 2.34 (ddd, J = 13.0, 8.7, 5.5 Hz, 1H, H-2'), 2.15 – 2.04 (m, 1H, H-2'), 1.11 (dd, J = 6.8, 1.1 Hz, 6H, CH(C*H*$_3$)$_2$).

**$^{13}$C-NMR** (DMSO-d6, 100 MHz): $\delta$ (ppm) = 179.89 (C=O), 157.01 (C6), 147.58 (C4), 146.70 (C2), 137.43 (*C*H=$CH_2$), 117.75 (C7), 116.06 (C8), 115.14 (CH=*C*$H_2$), 102.60 (C5), 87.05 (C4'), 82.16 (C1'), 70.93 (C3'), 61.93 (C5'), 39.73 (C2'), 34.75 (*C*H($CH_3$)$_2$), 30.25 (C7-$CH_2$), 18.95 ($CH_3$), 18.92 ($CH_3$).

**HR-MS** (ESI$^+$): Exact mass calculated for $C_{18}H_{24}N_4NaO_5$ [M+Na]$^+$: 399.16389, found: 399.16492

**7-Allyl-7-deaza-*N*(2)-isobutyryl-3',5'-di-(*O*-tert-butyl-dimethylsilyl)-2'-deoxyguanosine (76)**



**76**

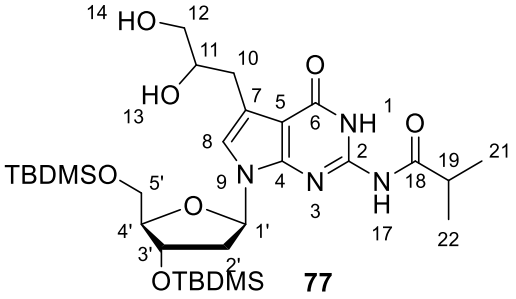
To a solution of 7-allyl-7-deaza-*N*(2)-isobutyryl-2'-deoxyguanosine **75** (500 mg, 1.33 mmol, 1.0 eq) in dry DMF (10 mL), imidazole (452 mg, 6.64 mmol, 5.0 eq) and TBDMS-Cl (501 mg, 3.32 mmol, 2.5 eq) were added and the reaction was stirred for 16 h. The reaction mixture was diluted with ethyl acetate and washed with 5% HCl, sat. $NaHCO_3$ and brine, dried over $Na_2SO_4$ and evaporated to dryness under reduced pressure. The crude product was purified by column chromatography (DCM:MeOH, 0 to 1% MeOH). All volatiles were removed in vacuum to give **76** as a colorless solid (708 mg, 1.17 mmol, 88%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 11.59 (s, 1H, NH), 8.07 (d, J = 1.4 Hz, 1H, C2-NH), 6.70 (t, J = 1.1 Hz, 1H, H-8), 6.39 (dd, J = 8.2, 5.7 Hz, 1H, H-1'), 6.07 (ddt, J = 16.8, 10.0, 6.7 Hz, 1H, C*H*=$CH_2$), 5.16 – 5.06 (m, 1H, CH=C*H*$_2$), 5.05 – 5.00 (m, 1H, CH=C*H*$_2$), 4.52 (dt, J = 5.4, 2.6 Hz, 1H, H-3'), 3.93 (dt, J = 4.3, 3.0 Hz, 1H, H-4'), 3.71 (dd, J = 5.0, 3.7 Hz, 2H, H-5'), 3.56 (ddq, J = 6.7, 5.5, 1.3 Hz, 2H, C7-$CH_2$), 2.55 (hept, J = 6.9 Hz, 1H, C*H*$(CH_3)_2$), 2.38 (ddd, *J* = 13.0, 8.3, 5.7 Hz, 1H, H-2'), 2.19 (ddd, *J* = 13.0, 5.7, 2.7 Hz, 1H, H-2'), 1-28 (d, J = 1-4 Hz, 3H, CH(C*H*$_3)_2$), 1.26 (d, J = 1.4 Hz, 3H, CH(C*H*$_3)_2$), 0.92 (d, J = 1.6 Hz, 18H, TBDMS, $CH_3$), 0.13 – 0.04 (m, 12H, TBDMS, $CH_3$).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 177.99 (C=O), 158.08 (C6), 147.60 (C4), 145.86 (C2), 137.12 (*C*H=$CH_2$), 119.84 (C7), 115.75 (C8), 115.55 (CH=*C*H$_2$), 104.19 (C5), 87.53 (C4'), 82.84 (C1'), 72.69 (C3'), 63.32 (C5'), 40.80 (C2'), 36.79 (*C*H$(CH_3)_2$), 30.77 (C7-$CH_2$), 26.12 (SiCH$(CH_3)_3$), 25.94 (SiC$(CH_3)_3$), 19.19 (C(*C*H$_3)_2$), 19.15 (CH(*C*H$_3)_2$), -4.60 (Si$(CH_3)_2$), -5.20 (Si$(CH_3)_2$)).

**HR-MS** (ESI$^+$): Exact mass calculated for $C_{30}H_{52}N_4NaO_5Si_2$ $[M+Na]^+$: 627.33684 found: 627.33691.

**7-deaza-7-(2,3-dihydroxy-propyl)-*N*(2)-isobutyryl-3',5'-di-(*O*-tert-butyl-dimethylsilyl)-2'-deoxyguanosine (77)**

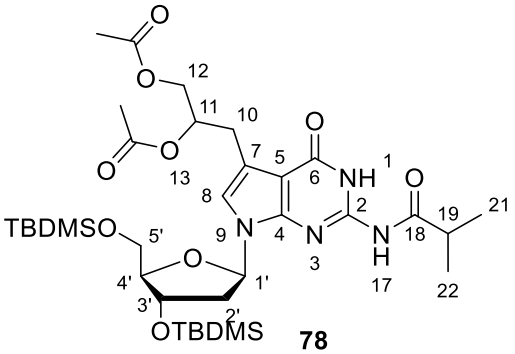


**77**

To a solution of 7-allyl-7-deaza-*N*(2)-isobutyryl-3',5'-di-(*O*-tert-butyl-dimethylsilyl)-2'-deoxyguanosine **76** (43.8 mg, 72.4 µmol, 1.0 eq) in THF (1.5 mL) and $H_2O$ (100 µL), a 2.5 % $OsO_4$-solution in $^t$BuOH (96.0 µL, 9.41 µmol, 0.13 eq) was added at 0 °C. After addition of 4-methylmorpholine-4-oxide (14.5 mg, 124 µmol, 1.71 eq), the reaction mixture was stirred for 3 h at 0 °C. The reaction mixture was diluted with ethyl acetate and washed with, 5% HCl, sat. $NaHCO_3$, sat. $Na_2SO_3$ and brine, dried over $Na_2SO_4$ and evaporated to dryness under reduced pressure. The crude product was purified by column chromatography (DCM:MeOH, 1 to 2% MeOH). All volatiles were removed in vacuum to give **77** as a colorless solid (31.4 mg, 49.2 µmol, 68%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 11.80 (s, 1H, NH), 8.19 (s, 1H, C2-NH), 6.87 (d, *J* = 6.6 Hz, 1H, H-8), 6.41 (dd, *J* = 7.9, 5.9 Hz, 1H, H-1'), 4.52 (dq, *J* = 5.8, 3.0 Hz, 1H, H-3'), 3.96 – 3.93 (m, 1H, H-4'), 3.83 – 3.68 (m, 1H, H-11), 3.75 – 3.70 (m, 2H, H-5'), 3.54 (t, *J* = 4.6 Hz, 2H, $CH_2$, H-12), 3.08 – 2.89 (m, 2H, C7-$CH_2$), 2.58 (hept, *J* = 7.1 Hz, 1H, C*H*($CH_3$)$_2$), 2.35 (ddt, *J* = 13.3, 8.0, 5.4 Hz, 1H, H-2'), 2.22 (ddt, *J* = 13.0, 5.8, 2.7 Hz, 1H, H-2'), 1.28 (d, *J* = 1.8 Hz, 3H, CH(C*H*$_3$)$_2$), 1.26 (d, *J* = 1.9 Hz, 3H, CH(C*H*$_3$)$_2$), 0.94 – 0.86 (m, 18H, SiC(C*H*$_3$)$_3$), 0.13 – 0.02 (m, 12H, Si(C*H*$_3$)$_2$).
**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 178.17 (C=O), 159.58 (C6), 147.77 (C4), 145.68 (C2), 118.02 (C8), 116.40 (C7), 104.67 (C5), 87.69 (C4'), 83.08 (C1'), 73.08 (C11), 72.58 (C3'), 65.16 (C12), 63.29 (C5'), 41.21 (C2'), 36.81 (*C*H($CH_3$)$_2$), 30.06 (C7-*C*$H_2$), 26.08 (SiC(*C*$H_3$)$_3$), 25.93 (SiC(*C*$H_3$)$_3$), 19.15 (CH(*C*$H_3$)$_2$), 18.55 (Si*C*($CH_3$)$_3$), 18.18 (Si*C*($CH_3$)$_3$), -4.52 (Si($CH_3$)$_2$), -4.61 Si($CH_3$)$_2$, -5.20 Si($CH_3$)$_2$, -5.37 Si($CH_3$)$_2$.
**HR-MS** (ESI$^+$): Exact mass calculated for $C_{30}H_{54}N_4NaO_7Si_2$ [M+Na]$^+$: 661.34232, found: 661.34249

**7-deaza-7-(2,3-diacetoxy-propyl)-N(2)-isobutyryl-3',5'-di-(O-tert-butyl-dimethylsilyl)- 2'-deoxyguanosine (78)**



**78**

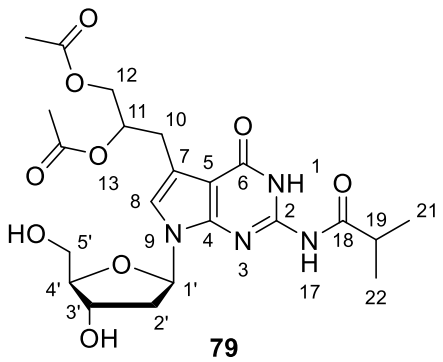
Compound **77** (210 mg, 329 µmol, 1.0 eq) was dissolved in anhydrous pyridine (5 mL) and cooled to 0 °C. Acetic anhydride (311 µL, 3.29 mmol, 10 eq) was added dropwise and the reaction was stirred for 3 h at room temperature. The solution was evaporated under reduced pressure and the crude product was purified by column chromatography (*n*-hexane:EtOAc = 2:1) to yield the product (**78**, 235 mg, 325 µmol, 98%).

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 11.56 (s, 1H, NH), 7.95 (d, $J$ = 2.4 Hz, 1H, C2-NH), 6.78 (d, $J$ = 2.7 Hz, 1H, H-8), 6.37 (ddd, $J$ = 8.4, 5.8, 3.2 Hz, 1H, H-1'), 5.45 (dddd, $J$ = 10.6, 7.3, 6.3, 3.5 Hz, 1H, CH, H11), 4.55 – 4.47 (m, 1H, H-3'), 4.32 (ddd, $J$ = 12.0, 10.6, 3.2 Hz, 1H, $CH_2$, H-12), 4.09 (ddd, $J$ = 12.0, 6.5, 4.4 Hz, 1H, $CH_2$, H-12), 3.92 (dtd, $J$ = 5.2, 3.4, 2.2 Hz, 1H, H-4'), 3.73 (dd, $J$ = 11.0, 3.4 Hz, 1H, H-5'), 3.66 (ddd, $J$ = 11.0, 5.1, 1.7 Hz, 1H, H-5'), 3.17 (dddd, $J$ = 34.0, 14.6, 6.3, 0.9 Hz, 1H, C7-$CH_2$), 3.00 (dddd, $J$ = 40.4, 14.5, 7.3, 0.8 Hz, 1H, C7-$CH_2$), 2.56 (p, $J$ = 6.9 Hz, 1H, C$H$($CH_3$)$_2$), 2.34 (ddd, $J$ = 13.5, 8.1, 5.7 Hz, 1H, H- 2'), 2.25 – 2.15 (m, 1H, H-2'), 2.05 (s, 3H, $CH_3$, acetyl), 2.00 (s, 3H, $CH_3$, acetyl), 1.30 – 1.22 (m, 6H, CH(C$H_3$)$_2$), 0.92 (d, $J$ = 0.6 Hz, 18H, SiC(C$H_3$)$_3$), 0.14 – 0.04 (m, 12H, Si($CH_3$)$_2$).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 177.94 (C=O), 170.96 (C=O, acetyl), 170.43 (C=O, acetyl), 157.84 (C6), 147.38 (C4), 146.02 (C2), 117.04 (C8), 115.47 (C7), 104.46 (C5), 87.51 (C4'), 82.92 (C1'), 72.62 (C3'), 71.67 (C11), 64.83 (C12), 63.41 (C5'), 40.73 (C2'), 36.82 ($C$H($CH_3$)$_2$), 27.55 (C7-$C$H$_2$), 26.09 (SiC($C$H$_3$)$_3$), 25.93 (SiC($C$H$_3$)$_3$), 21.26 ($CH_3$, acetyl), 20.99 ($CH_3$, acetyl), 19.18 (CH($C$H$_3$)$_2$), 19.14 (CH($C$H$_3$)$_2$), 18.54 (Si$C$($CH_3$)$_3$), 18.17 (Si$C$($CH_3$)$_3$), -4.55 (Si($CH_3$)$_2$), -4.61 (Si($CH_3$)$_2$), -5.22 (Si($CH_3$)$_2$), -5.36 (Si($CH_3$)$_2$).

**HR-MS** (ESI$^+$): Exact mass calculated for $C_{34}H_{58}N_4NaO_9Si_2$ [M+Na]$^+$: 745.36345, found: 745.36338

**7-deaza-7-(2,3-diacetoxy-propyl)-*N*(2)-isobutyryl-2'-deoxyguanosine (79)**




**79**

Compound **78** (100 mg, 138 µmol, 1.0 eq) in anhydrous THF (2 mL) was mixed with TBAF (131 mg, 415 µmol, 3.0 eq) and acetic acid (19.3 µL, 337 µmol, 2.44 eq). The reaction was stirred for 15 h at room temperature. The solvent was removed under reduced pressure and the residue was purified by column chromatography (DCM:MeOH = 20:1) to yield the product (**79**, 68.0 mg, 137 µmol, 99%).

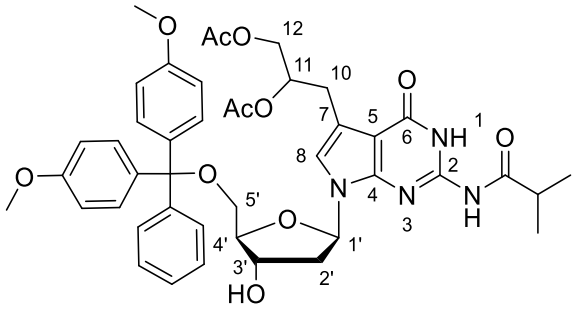
**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 11.73 (s, 1H, NH), 8.60 (d, $J$ = 9.7 Hz, 1H, C2-NH), 6.67 (d, $J$ = 2.5 Hz, 1H, H-8), 6.15 (ddd, $J$ = 9.0, 5.8, 3.6 Hz, 1H, H-1'), 5.47 – 5.33 (m, 1H, CH, H-11), 4.67 (dq, $J$ = 5.0, 2.4 Hz, 1H, H-3'), 4.32 (ddd, $J$ = 16.0, 12.0, 3.1 Hz, 1H, $CH_2$, H-12), 4.10 – 4.09 (m, 1H, CH, H-4'), 4.14 – 4.03 (m, 1H, H-12), 3.90 (ddd, $J$ = 12.2, 2.7, 1.4 Hz, 1H, H-5'), 3.76 (dd, $J$ = 12.2, 2.6 Hz, 1H, H-5'), 3.21 – 3.11 (m, 1H, C7-$CH_2$), 3.10 – 2.94 (m, 1H, C7-$CH_2$), 2.78 – 2.68 (m,

1H, H-2'), 2.64 (p, *J* = 6.9 Hz, 1H, C*H*($CH_3$)$_2$), 2.28 (dtd, *J* = 13.4, 5.7, 2.2 Hz, 1H, H-2'), 2.06 (s, 3H, $CH_3$, acetyl), 2.02 (d, *J* = 4.7 Hz, 3H, $CH_3$, acetyl), 1.27 – 1.22 (m, 6H, CH(C*H*$_3$)$_2$).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): *δ* (ppm) = 178.68 (NH-C=O), 171.20 (C=O, acetyl), 170.76 (C=O, acetyl), 157.72 (C=O, C6), 146.26 (C4), 146.09 (C2), 119.55 (C8), 114.90 (C7), 105.78 (C5), 87.79 (C4'), 87.18 (C1'), 73.03 (C3'), 71.69 (C11), 64.95 (C12), 63.40 (C5'), 40.86 (C2'), 36.59 (*C*H($CH_3$)$_2$), 27.23 (C7-*C*$H_2$), 21.31 ($CH_3$, acetyl), 21.03 ($CH_3$, acetyl), 19.14 (CH(*C*$H_3$)$_2$), 19.06 (CH(*C*$H_3$)$_2$).

**HR-MS** (ESI$^+$): Exact mass calculated for $C_{22}H_{30}N_4NaO_9$ [M+Na]$^+$: 517.19050 found: 517.19168

**7-deaza-7-(2,3-diacetoxy-propyl)-5'-*O*-(4,4'-dimethoxytrityl)-N(2)-isobutyryl-2'-deoxyguanosine (80)**

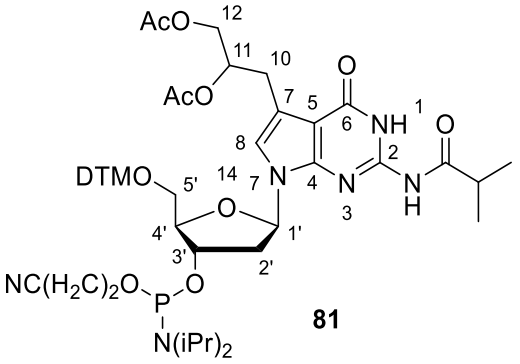



**80**

Compound **79** (264 mg, 534 µmol, 1.0 eq) was dissolved in anhydrous DCM (6 mL). DMT-Cl (217 mg, 641 µmol, 1.2 eq) was added in two portions within one hour and stirred for 1.5 h at room temperature. The reaction was stopped by the addition of MeOH and the solvent was evaporated under reduced pressure. The residue was re-dissolved in DCM and washed with 5% HCl, $H_2O$ and $NaHCO_3$, respectively. The reaction was purified by column chromatography (DCM:MeOH, 0 to 10% MeOH + 1% $NEt_3$) to yield the product (**80**, 321 mg, 403 µmol, 75%) as a colorless foam.

**$^1$H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 11.57 (s, 1H, NH), 7.80 (s, 1H, C2-NH), 7.46 – 7.39 (m, 2H, DMT), 7.35 – 7.27 (m, 6H, DMT), 7.24 – 7.19 (m, 1H, DMT), 6.81 (ddd, *J* = 9.0, 2.5, 1.5 Hz, 4H, DMT), 6.66 (d, *J* = 5.1 Hz, 1H, H-8), 6.33 (td, *J* = 7.4, 6.8, 3.7 Hz, 1H, H-1'), 5.38 (qd, *J* = 6.9, 3.0 Hz, 1H, CH, H-11), 4.61 – 4.56 (m, 1H, H-3'), 4.26 (td, *J* = 11.6, 3.1 Hz, 1H, $CH_2$, H-12), 4.04 (ddd, *J* = 12.0, 6.5, 1.8 Hz, 1H, H-12), 3.99 (d, *J* = 3.8 Hz, 1H, H-4'), 3.80 – 3.76 (m, 6H, $OCH_3$), 3.32 (d, *J* = 4.6 Hz, 2H, H-5'), 3.18 – 2.83 (m, 2H, C7-$CH_2$), 2.68 – 2.60 (m, 1H, H-2'), 2.38 – 2.24 (m, 2H, H-2', C*H*($CH_3$)$_2$), 2.01 (d, *J* = 4.2 Hz, 3H, $CH_3$, acetyl), 1.91 (d, *J* = 7.1 Hz, 3H, $CH_3$, acetyl), 1.19 (dd, *J* = 6.9, 2.2 Hz, 3H, CH(C*H*$_3$)$_2$), 1.14 (dd, *J* = 6.9, 4.2 Hz, 3H, CH(C*H*$_3$)$_2$).

**7-deaza-7-(2,3-diacetoxy-propyl)-5'-*O*-(4,4'-dimethoxytrityl-N(2)-isobutyryl)-2'-deoxyguanosine-3'-[(2-cyanoethyl)-*N*,*N*-diisopropylphosphoramidite (81)**




**81**

Compound **80** (59 mg, 74.0 µmol, 1.0 eq) was dissolved in anhydrous DCM (2 mL) and treated with $^{i}Pr_2NEt$ (50.4 µL, 296 µmol, 4.0 eq) and CEP-Cl (26.3 mg, 111 µmol, 1.5 eq). The reaction was stirred for 2 h at room temperature. The solvent was removed under reduced pressure and the residue purified by column chromatography (n-hexane:acetone = 3:1 + 1% $NEt_3$) to yield the product **81** (68.0 mg, 68.2 µmol, 92%) as a colorless foam.

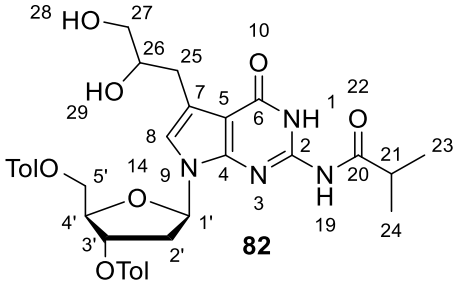
**$^{1}$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 11.62 (s, 1H, NH), 8.47 – 8.91(2s, 1H, C2-NH), 7.47 – 7.40 (m, 2H, DMT), 7.35 – 7.29 (m, 4H, DMT), 7.30 – 7.16 (m, 3H, DMT), 6.80 (dtd, *J* = 8.1, 2.4, 1.2 Hz, 4H, DMT), 6.72 – 6.66 (m, 1H, H-8), 6.37 – 6.25 (m, 1H, H-1'), 5.47 – 5.34 (m, 1H, CH, H-11), 4.73 – 4.58 (m, 1H, H-3'), 4.29 – 4.23 (m, 1H, $CH_2$, H-12), 4.23 – 4.19 (m, 1H, H-4'), 4.19 – 4.10 (m, 2H, $POCH_2$), 4.09 – 4.01 (m, 1H, $CH_2$, H-12), 3.93 – 3.66 (m, 1H, $POCH_2$), 3.78 – 3.76 (m, 6H, $OCH_3$), 3.67 – 3.56 (m, 1H, N(C*H*$(CH_3)_2)_2$), 3.51 (ddt, *J* = 18.4, 13.7, 6.8 Hz, 3H, N(C*H*$(CH_3)_2)_2$), 3.33 (ddd, *J* = 30.7, 10.2, 4.2 Hz, 1H, H-5'), 3.24 (dt, *J* = 10.2, 3.9 Hz, 1H, H-5'), 3.14 – 3.03 (m, 1H, C7-$CH_2$), 3.03 – 2.87 (m, 1H, C7-$CH_2$), 2.79 – 2.73 (m, 2H, $CH_2CN$), 2.73 – 2.63 (m, 1H, $CH_2CN$, H-2'), 2.52 (dd, *J* = 6.9, 3.9 Hz, 1H, H-2'), 2.48 (t, *J* = 6.4 Hz, 1H, $CH_2CN$), 2.42 – 2.34 (m, 1H, H-2'), 2.33 – 2.13 (m, 1H, C*H*$(CH_3)_2$), 2.04 – 1.98 (3s, 3H, $CH_3$, acetyl), 1.92 – 1.88 (3s, 3H, $CH_3$, acetyl), 1.27 (dd, *J* = 6.8, 5.6 Hz, 15H, N(CH(C*H*$_3)_2)_2$), 1.21 – 1.06 (m, 18H, CH(C*H*$_3)_2$).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 178.58 (C2-NH-*C*=O), 178.29 (C2-NH-*C*=O), 171.00 (C=O, acetyl), 170.46 (C=O, acetyl), 158.70 ($C_q$-$OCH_3$, DMT), 147.49 (C4), 144.86 ($C_q$, DMT), 135.86 ($C_q$, DMT), 130.25 (DMT), 130.18 (DMT), 128.38 (DMT), 128.27 (DMT), 128.03 (DMT), 128.01 (DMT), 127.08 (DMT), 115.25 (C7), 113.29 (DMT), 113.25 (DMT), 105.06 (C5), 104.68 (C5), 86.52 ($C_q$-O, DMT), 86.40 ($C_q$-O), 85.19 (C4'), 84.47 (C1'), 84.32 (C1'), 74.54 (C3'), 73.88 (C11), 71.78 (C11), 64.83 (C12), 63.89 (C5'), 58.36 ($POCH_2$), 58.31 ($POCH_2$), 55.37 ($OCH_3$, DMT), 55.36 ($OCH_3$, DMT), 45.48 (N(*C*H$(CH_3)_2)_2$), 45.42 (N(*C*H$(CH_3)_2)_2$), 39.37 (C2'), 39.00 (C2'), 36.40 (*C*H$(CH_3)_2$), 36.28 (*C*H$(CH_3)_2$), 27.48 (C10), 23.13 (N(CH(*C*H$_3)_2)_2$), 23.05 (N(CH(*C*H$_3)_2)_2$), 21.20 ($CH_3$, acetyl), 20.93 ($CH_3$, acetyl), 20.22 ($CH_2CN$), 19.08 (CH(*C*H$_3)_2$), 18.96 (CH(*C*H$_3)_2$).

**$^{31}$P-NMR** (162 MHz, $CDCl_3$): $\delta$ (ppm) = 147.85, 147.49

**HR-MS** ($ESI^+$): Exact mass calculated for $C_{52}H_{65}N_6NaO_{12}P$ $[M+Na]^+$: 1019.4290 found: 1019.4250

**7-deaza-7-(2,3-dihydroxy-propyl)-*N*(2)-isobutyryl-3',5'-di-*O*-p-toluoyl-2'-deoxyguanosine (82)**



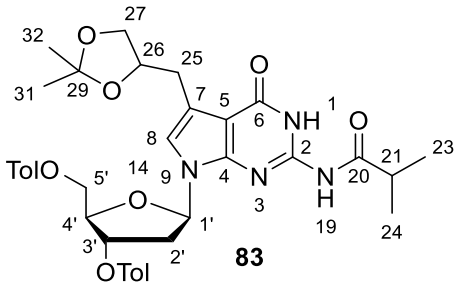

Compound **74** (300 mg, 490 µmol, 1.0 eq) was dissolved in a THF/water solution (7.5 mL/750 µL) and cooled to 0 °C. A 2.5 % $OsO_4$-solution in $^t$BuOH (647 µL, 63.7 µmol, 0.13 eq) and NMMO (98.1 mg, 837 µmol, 1.71 eq) were added and the reaction mixture was stirred for 4 h at 0°C. The mixture was diluted with ethyl acetate and washed with 5% HCl, saturated $NaHCO_3$-, saturated $Na_2SO_3$- and saturated NaCl-solution, respectively. The organic phase was dried over $Na_2SO_4$, and the solvent was removed under reduced pressure. The crude residue was purified by column chromatography (DCM:MeOH = 99:1 to 98:2) to obtain the product (**82**, 188 mg, 291 µmol, 59%) as a colorless foam.

**$^1$H-NMR** ($CDCl_3$, 400 MHz): *δ* (ppm) = 11.89 (s, 1H, NH), 9.39 (s, 1H, NH), 7.95 (d, *J* = 8.0 Hz, 2H, tol), 7.88 (dd, *J* = 8.2, 2.0 Hz, 2H, tol), 7.30 – 7.23 (m, 2H, tol), 7.25 – 7.21 (m, 2H, tol), 6.68 (d, *J* = 4.7 Hz, 1H, H-8), 6.32 (td, *J* = 7.9, 6.1 Hz, 1H, H-1'), 5.78 (dd, *J* = 5.8, 3.1 Hz, 1H, H-3'), 5.01 (ddd, *J* = 11.2, 7.4, 4.0 Hz, 1H, H-5'), 4.64 – 4.55 (m, 2H, H-5', H-4'), 3.77 – 3.73 (m, 1H, C*H*-OH), 3.56 – 3.46 (m, 2H, $CH_2$-OH), 2.98 (tdd, *J* = 20.2, 9.8, 6.2 Hz, 1H, H-2'), 2.90 – 2.80 (m, 2H, C7-$CH_2$), 2.74 (pd, *J* = 7.0, 5.3 Hz, 1H, C(=O)C*H*($CH_3$)$_2$), 2.62 (ddt, *J* = 14.1, 5.7, 2.5 Hz, 1H, H-2'), 2.43 (s, 3H, $CH_3$, tol), 2.41 (s, 3H, $CH_3$, tol), 1.32 – 1.27 (m, 6H, 2x $CH_3$).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): *δ* (ppm) = 179.23 (C=O, C20), 166.71 (C=O, tol), 166.16 (C=O, tol), 158.65 (C6), 146.09 (C4), 144.71 ($C_q$, tol, $C_q$-$CH_3$), 129.97 (tol), 129.76 (tol), 129.49 (tol), 129.42 (tol), 126.62 ($C_q$, tol), 126.54 ($C_q$, tol), 119.91 (C8), 116.53 (C7), 105.21 (C5), 86.15 (C1'), 82.55 (C4'), 75.02 (C3'), 72.75 (CH-OH, C26), 65.42 ($CH_2$-OH, C27), 63.72 (C5'), 37.46 (C2'), 36.74 (*C*H($CH_3$)$_2$), 29.66 ($CH_2$, C25), 21.89 ($CH_3$, tol), 21.86 ($CH_3$, tol), 19.05 ($CH_3$, C23), 19.01 ($CH_3$, C24).

**HR-MS** (ESI$^+$): Exact mass calculated for $C_{34}H_{38}N_4NaO_9$ $[M+Na]^+$: 669.2531 found: 669.2543.

**7-deaza-7-(2,3-(2,2-dimethyl-1,3-dioxolan)propyl)-*N*(2)-isobutyryl-3',5'-di-*O*-p-toluoyl-2'-deoxyguanosine (83)**



**83**

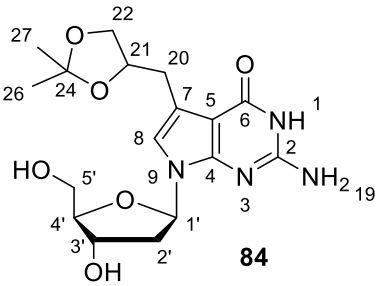
Compound **82** (290 mg, 450 µmol, 1.0 eq) was suspended in anhydrous DCM (2 mL) and mixed with 2,2-dimethoxypropane (550 µL, 4.50 mmol, 10.0 eq) and p-toluenesulfonic acid (94.0 mg, 490 µmol, 1.1 eq). The reaction was stirred for 2 h at room temperature, before the solvent was removed in vacuum. The residue was taken up in DCM and washed with saturated $NaHCO_3$-solution. The organic phase was dried over $Na_2SO_4$, and the solvent was removed under reduced pressure. The crude residue was purified by column chromatography (*n*-hexane:EtOAc = 1:1 to 3:2) to obtain the product (**83**, 260 mg, 400 µmol, 89%) as a colorless foam.

**$^1$H-NMR** ($(CD_3)_2CO$, 400 MHz): $\delta$ (ppm) = 11.78 (s, 1H, NH), 10.29 (s, 1H, NH), 8.04 – 7.95 (m, 4H, tol), 7.43 – 7.32 (m, 4H, tol), 6.93 (s, 1H, H-8), 6.50 (dd, *J* = 8.9, 5.7 Hz, 1H, H-1'), 5.77 (dd, *J* = 5.3, 3.3 Hz, 1H, H-3'), 4.76 – 4.69 (m, 1H, H-5'), 4.64 – 4.58 (m, 1H, H-5'), 4.57 (dd, *J* = 4.4, 2.2 Hz, 1H, H-4'), 4.43 – 4.34 (m, 1H, H-26), 3.84 (ddd, *J* = 8.1, 6.0, 2.1 Hz, 1H, $CH_2$, H-27), 3.57 (ddd, *J* = 8.1, 6.7, 4.8 Hz, 1H, $CH_2$, H-27), 3.03 – 2.97 (m, 1H, $CH_2$, H-25), 2.97 – 2.92 (m, 1H, H-2'), 2.92 – 2.89 (m, 1H, C*H*$(CH_3)_2$), 2.80 – 2.73 (m, 1H, $CH_2$, H-25), 2.73 – 2.66 (m, 1H, H-2'), 2.43 (d, *J* = 9.1 Hz, 6H, $CH_3$, tol), 1.38 – 1.14 (m, 12H, 4x $CH_3$, H-31, H-32, H-23, H-24).

**$^{13}$C-NMR** ($(CD_3)_2CO$, 100 MHz): $\delta$ (ppm) = 180.61 (C=O, C20), 166.59 (C=O, tol), 166.43 (C=O, tol), 158.10 (C6), 148.71 (C4), 144.99 ($C_q$-$CH_3$, tol), 144.96 ($C_q$-$CH_3$, tol), 130.51 (tol), 130.43 (tol), 130.21 (tol), 130.15 (tol), 128.12 ($C_q$, tol), 128.01 ($C_q$, tol), 117.75 (C8), 117.17 (C7), 109.20 ($C_q$-$(CH_3)_2$, C29), 104.69 (C5), 84.33 (C1'), 82.84 (C4'), 76.61 (H-3'), 76.54 (C26), 69.41 (C27), 65.09 (C5'), 38.03 (C2'), 36.49 (*C*H$(CH_3)_2$), 31.23 ($CH_2$, C25), 27.38 ($CH_3$, C31), 26.00 ($CH_3$, C32), 21.61 ($CH_3$, tol), 19.35 (CH(*C*$H_3)_2$), 19.32 (CH(*C*$H_3)_2$).

**HR-MS** ($ESI^+$): Exact mass calculated for $C_{37}H_{42}N_4NaO_9$ $[M+Na]^+$: 709.2844 found: 709.2879.

**7-deaza-7-(2,3-(2,2-dimethyl-1,3-dioxolan)propyl)-*N*(2)-isobutyryl-2'-deoxyguanosine (84)**



**84**

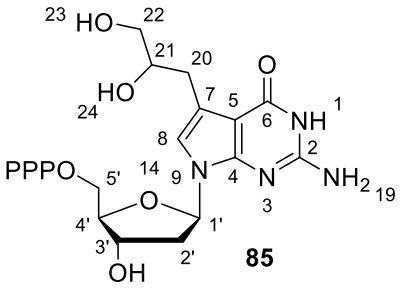
Compound **83** (100 mg, 150 µmol, 1.0 eq) was suspended in $MeNH_2$ (40% in $H_2O$, 20 mL) and stirred under reflux conditions for 2 h. The solvent was removed under reduced pressure and the crude residue was purified by column chromatography (DCM:MeOH = 15:1 to 10:1) to obtain the product (**84**, 49.0 mg, 129 µmol, 86%) as a colorless foam.

**$^1$H-NMR** (DMSO-d6, 400 MHz): $\delta$ (ppm) = 10.33 (s, 1H, NH, N1), 6.72 (s, 1H, H-8), 6.25 (dd, *J* = 8.7, 5.9 Hz, 1H, H-1'), 6.23 (s, 2H, $NH_2$), 5.19 (s, 1H, HO-C3'), 4.87 (s, 1H, HO-C5'), 4.47 – 4.35 (m, 1H, CH, H-21), 4.28 – 4.22 (m, 1H, H-3'), 3.85 (ddd, *J* = 8.2, 6.1, 2.4 Hz, 1H, $CH_2$, H-22), 3.72 (dddd, *J* = 5.0, 3.5, 2.5, 1.0 Hz, 1H, H-4'), 3.58 (ddd, *J* = 8.1, 6.6, 3.0 Hz, 1H, $CH_2$, H-22), 3.52 – 3.41 (m, 2H, H-5'), 2.90 (dddd, *J* = 13.8, 5.6, 3.5, 1.0 Hz, 1H, $CH_2$, H-20), 2.72 – 2.64 (m, 1H, $CH_2$, H-20), 2.29 (dddd, *J* = 16.4, 8.3, 4.2, 2.1 Hz, 1H, H-2'), 2.03 (ddt, *J* = 13.0, 8.2, 2.4 Hz, 1H, H-2'), 1.34 – 1.22 (2s, 6H, $CH_3$).

**$^{13}$C-NMR** (DMSO-d6, 100 MHz): $\delta$ (ppm) = 159.05 (C6), 152.51 (C2), 150.68 (C4), 114.53 (C8), 114.49 (C7), 108.00 (C24), 98.95 (C5), 86.84 (C4'), 81.93 (C1'), 75.47 (C21), 71.01 (C3'), 68.19 (C22), 62.08 (C5'), 39.29 (C2'), 30.63 (C20), 27.01 ($CH_3$), 25.65 ($CH_3$).

**HR-MS** (ESI$^+$): Exact mass calculated for $C_{17}H_{24}N_4NaO_6$ $[M+Na]^+$: 403.1588 found: 403.1588.

**7-deaza-7-(2,3-dihydroxy-propyl)-2'-deoxyguanosine-5'-triphosphate (85)**



**85**

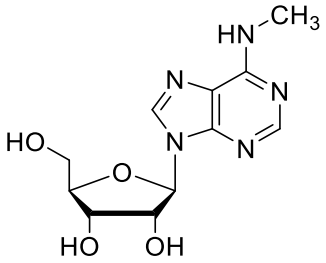
Compound **84** (16 mg, 42.1 µmol, 1.0 eq) was dissolved in anhydrous trimethyl phosphate (1 mL) and proton-sponge® (12.6 mg, 59.0 µmol, 1.4 eq) was added. The reaction mixture was cooled to -20 °C, mixed with freshly distilled $POCl_3$ (7.00 µL, 46.0 µmol, 1.1 eq) and stirred for 3 h at -20 °C. Then, tributylamine (70.0 µL, 294 µmol, 7.0 eq) and bis(tributylammonium) pyrophosphate (189 mg, 286 µmol, 6.8 eq) in anhydrous DMF (470 µL) were added to the reaction mixture and the solution was stirred for 1.5 h at 0 °C. A freshly prepared TEAB buffer (pH 7.5, 3 mL) was added

to the reaction and the aqueous solution was extracted with ethyl acetate. The aqueous phase was evaporated under reduced pressure (water bath < 30°C). The crude residue was purified by HPLC (DEAE and RP-HPLC, respectively).

**$^{1}$H-NMR** ($D_2O$, 400 MHz): *δ* (ppm) = 6.96 (s, 1H, H-8), 6.41 (dd, *J* = 8.3, 6.1 Hz, 1H, H-1'), 4.73 – 4.65 (m, 1H, H-3'), 4.16 (d, *J* = 4.0 Hz, 2H, H-4', H-5'), 4.14 – 4.04 (m, 1H, H-5'), 3.98 (p, *J* = 6.9 Hz, 1H, H-21), 3.60 (dd, *J* = 11.8, 3.9 Hz, 1H, H-22), 3.48 (ddd, *J* = 11.7, 6.6, 4.6 Hz, 1H, H-22), 2.91 (dd, *J* = 14.6, 5.3 Hz, 1H, H-20), 2.82 – 2.74 (m, 1H) (m, 1H, H-20), 2.71 – 2.59 (m, 1H, H-2'), 2.33 (ddd, *J* = 14.0, 6.2, 2.8 Hz, 1H, H-2').
**$^{31}$P-NMR** (162 MHz, $CDCl_3$): *δ* (ppm) = -10.86 (d, *J* = 19.8 Hz), -11.40 (d, *J* = 20.0 Hz), -23.31 (t, *J* = 20.0 Hz).
**HR-MS** ($ESI^+$): Exact mass calculated for $C_{14}H_{22}N_4O_{15}P_3$ $[M-H]^-$: 579.0300 found: 579.0302.

5.1.2.13 Preparation of $m^6v^8A$ phosphoramidite (**91**)[182]

***N*$^6$-methyladeneosine (86)**

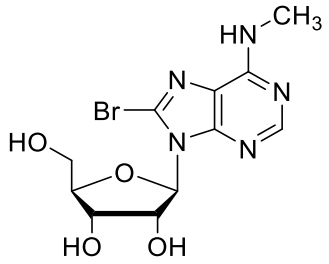


**86**

6-Chloropurine riboside (2.00 g, 7.00 mmol, 1.00 eq.) was dissolved in methylamine (33 % in EtOH, 48 mL). $NEt_3$ (1.20 mL, 8.60 mmol, 1.20 eq.) was added and the reaction was stirred for 5 h at 80 °C. The solvent was removed under reduced pressure and the crude product was purified my column chromatography (DCM:MeOH = 9:1) to obtain the product (**86**, 1.69 g, 6.01 mmol, 86%) as a slightly beige solid.

**$^{1}$H-NMR** (DMSO-d6, 400 MHz): *δ* (ppm) = 8.35 (s, 1H, H-C8), 8.23 (s, 1H, H-2), 7.84 (s, 1H, N6-H), 5.89 (d, J = 6.2 Hz, 1H, H-1'), 5.47–5.45 (m, 1H, HO-C2'), 5.44 (d, J = 4.5 Hz, 1H, HO-C5'), 5.20 (d, J = 4.6 Hz, 1H HO-C3'), 4.61 (td, *J* = 6.2, 4.9 Hz, 1H, H-2'), 4.15 (td, *J* = 4.8, 3.0 Hz, 1H, H-3'), 3.97 (q, J = 3.4 Hz, 1H, H-4'), 3.68 (ddd, J = 12.1, 4.5, 3.6 Hz, 1 H, H-5'), 3.55 (ddd, J = 12,1, 7.3, 3.6 Hz, 1H, H-5'), 2.96 (s, 3H, N6-$CH_3$).
**$^{13}$C-NMR** ($DMSO-d_6$, 100 MHz): *δ* (ppm) = 155.1 (C6), 152.4 (C2), 148.1 (C4), 139.7 (C8), 119.9 (C5), 87.9 (C1'), 85.9 (C4'), 73.5 (C2'), 70.7 (C3'), 61.7 (C5'), 27.0 (N6-$CH_3$).
**HR-MS** ($ESI^+$): Exact mass calculated for $C_{11}H_{15}N_5NaO_4$ $[M+Na]^+$: 304.10162found: 304.10170.

**8-Bromo-*N*$^6$-methyladeneosine (87)**

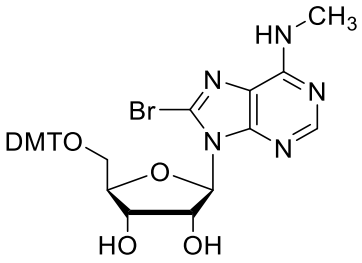


**87**

*N*$^6$-methyladenosine (**86**, 610 mg, 2.17 mmol, 1.0 eq) was suspended in sodium acetate buffer (15.0 mL, pH 4). An aqueous bromine solution (1.04 g, 340 µL, 6.51 mmol, 3.0 eq in 32.6 mL $H_2O$) was added dropwise to the suspension and the reaction was stirred for 72 h at room temperature. The reaction was stopped by the addition of a sodium hydrogen sulfite solution, until the reaction turned colorless. Then the reaction mixture was neutralized with a sodium hydroxide solution (2 M), resulting in precipitation of the product. The solid was filtered off and dried in high vacuum. The product (**87**, 637 mg, 1.77 mmol, 82 %) was obtained as a colorless solid.[505]

**$^1$H-NMR** (DMSO-d6, 400 MHz): *δ* (ppm) = 8.22 (s, 1H, H-2), 8.10 (d, J = 4.9 Hz, 1H, N6-H), 5.83 (d, J = 6.8 Hz, 1H, H-1'), 5.52 (dd, *J* = 8.7, 3.9 Hz, 1H, HO-C5'), 5.48 (d, J = 6.3 Hz, 1H, HO-C2'), 5.26 (d, J = 4.5 Hz, 1H, HO-C3'), 5.08 (td, J = 6.6, 5.2 Hz, 1H, H-2'), 4.19 (td, J = 4.8, 2.3 Hz, 1H, H-3'), 3.98 (td, *J* = 4.0, 2.3 Hz, 1H, H-4'), 3.68 (dt, *J* = 12.1, 3.9 Hz, 1H, H1-5'), 3.52 (ddd, J = 12.1, 8.7, 4.1 Hz, 1H, H1- C5'), 2.94 (d, J = 4.5 Hz, 3H, N6-$CH_3$).
**$^{13}$C-NMR** (DMSO-d6, 100 MHz): *δ* (ppm) = 154.0 (C6), 152.5 (C2), 148.9 (C4), 126.9 (C8), 120.3 (C5), 90.4 (C1'), 86.8 (C4'), 71.2 (C2'), 70.9 (C3'), 62.1 (C5'), 27.0 (N6-$CH_3$).
**HR-MS** (ESI$^+$): Exact mass calculated for $C_{11}H_{15}N_5BrO_4$ $[M+H]^+$: 360.03019 found: 360.03086.

**8-Bromo-5'-*O*-(4,4'-dimethoxytrityl-*N*$^6$-methyladeneosine (88)**



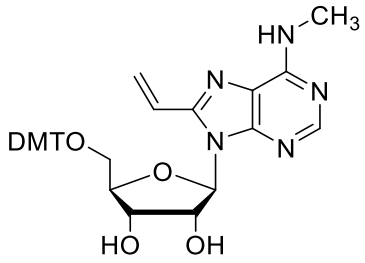
**88**

8-Bromo-*N*$^6$-methyladenosin (**87**, 538 mg, 1.49 mmol, 1.0 eq) was co-evaporated with anhydrous pyridine (2x, 4 mL). The dried residue was dissolved in anhydrous pyridine (5 mL) and DMAP (37.0 mg, 304 µmol, 0.2 eq) was added. DMT-Cl (821 mg, 2.43 mmol, 1.63 eq) was added in four portions within 3 h at room temperature. After an additional hour, MeOH (4 mL) was added to stop the reaction. The solvent was removed under reduced pressure. The residue was taken up in

EtOAc and washed with a saturated NaCl-solution. The organic phase was dried over $Na_2SO_4$, and the solvent was removed under reduced pressure. The reaction mixture was purified by column chromatography (DCM:MeOH = 98:2 + 1 % $NEt_3$) to obtain the product (**88**, 775 mg, 1.16 mmol, 78 %) as a colorless foam.

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 8.14 (s, 1H, H-2), 7.34 – 7.39 (m, 2H, DMT), 7.22 – 7.31 (m, 4H, DMT), 7.18 – 7.23 (m, 2H, DMT), 7.13 – 7.18 (m, 1H, DMT), 6.66-6.78 (m, 4H, DMT), 5.97 (d, *J* = 4.8 Hz, 1H, H-1'), 5.69 (s, 1H, H-N6), 5.44-5.50 (m, 1H, H-2'), 4.75 (dd, *J* = 5.5, 4.5 Hz, 1H, H-3'), 4.20 (q, J = 4.6 Hz, 1H, H-C4'), 3.76 (d, *J* = 0.7 Hz, 6H, $OCH_3$), 3.44 (dd, J = 10.2, 4.8 Hz, 1H, H-5'), 3.32 (dd, J = 10.2, 4.7 Hz, 1H, H-5'), 3.14 (s, 3H, N6-$CH_3$).
**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 158.5 ($C_q$, DMT), 154.2 (C6), 153.0 (C2), 144.8 ($C_q$, DMT), 136.1 ($C_q$, DMT), 136.0 ($C_q$, DMT), 130.2 (DMT), 130.1 (DMT), 128.3 (DMT), 127.9 (DMT), 127.3 (C8), 126.9 (DMT), 113.2 (DMT), 90.8 (C1'), 86.4 ($C_q$, DMT), 83.7 (C4'), 72.1 (C2'), 71.6 (C3'), 63.5 (C5'), 55.3 ($OCH_3$).
**HR-MS** ($ESI^+$): Exact mass calculated for $C_{32}H_{32}N_5BrNaO_6$ $[M+H]^+$: 684.14282 found: 684.14118.

**5'-*O*-(4,4'-dimethoxytrityl-*N*$^6$-8-vinyl-methyladeneosine (89)**




**89**

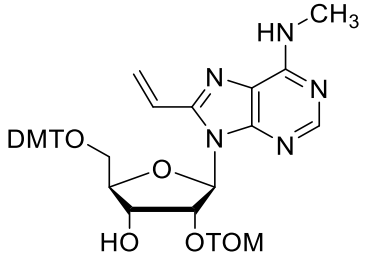
Compound **88** (570 mg, 860 µmol, 1.0 eq) was dissolved in anhydrous THF (8 mL). To this mixture $Pd(OAc)_2$ (39.0 mg, 170 µmol, 0.2 eq), $PPh_3$ (135 mg, 520 µmol, 0.6 eq), CuI (50.0 mg, 260 µmol, 0.3 eq) and tributyl(vinyl)tin (377 µL, 1.29 mmol, 1.5 eq) were added. The reaction was stirred at 60 °C for 24 h. The solvent was removed in vacuum and the crude residue was purified by column chromatography (DCM + 1% $NEt_3$ to DCM:MeOH = 99:1) to obtain the product (**89**, 396 mg, 650 µmol, 75%) as a colorless foam.[491]

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 8.30 (s, 1H, H-2), 7.23-7.34 (m, 2H, DMT), 7.24-7.10 (m, 9H, DMT, v*inyl*-CH), 6.75-6.65 (m, 4H, DMT), 6.45 (dd, *J* = 17.2, 1.7 Hz, 1H, *vinyl*-$CH_2$), 5.95 (d, *J* = 6.7 Hz, 1H, H-1'), 5.89 (s, 1H, H-N6), 5.57 (dd, *J* = 11.1, 1.7 Hz, 1H, *vinyl* $CH_2$), 5.31 (dd, *J* = 6.6, 5.2 Hz, 1H, H-2'), 4.46 (dd, *J* = 5.3, 2.1 Hz, 1H, H-3'), 4.32 (q, *J* = 3.4 Hz, 1H, H-4'), 3.75 (d, *J* = 0.7 Hz, 6H, $OCH_3$), 3.43 (dd, *J* = 10.4, 3.7 Hz, 1H, H-5'), 3.19 (br, s, 3H, N6-$CH_3$), 3.14 (dd, *J* = 10.4, 3.4 Hz, 1H, H-5').

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 158.5 ($C_q$, DMT), 155.1 (C6), 152.4 (C2), 149.3 (C8), 144.6 (DMT), 135.9 (DMT), 135.6 (DMT), 130.1 (DMT), 130.0 (DMT), 128.2 (DMT), 127.9 (DMT), 125.4 (*vinyl*-CH), 122.6 (*vinyl*-CH2), 113.2 (DMT), 89.7 (C1'), 86.3 ($C_q$, DMT), 85.2 (C4'), 74.2 (C2'), 72.5 (C3'), 63.7 (C5'), 55.3 ($OCH_3$).

**HR-MS** ($ESI^+$): Exact mass calculated for $C_{34}H_{35}N_5NaO_6$ $[M+Na]^+$: 632.24795 found: 632.24758.

**5'-*O*-(4,4'-dimethoxytrityl-*N*$^6$-8-vinyl-methyladeneosine (90)**




**90**

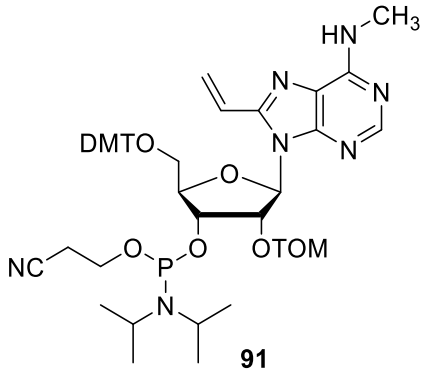
Compound **89** (140 mg, 230 µmol, 1.0 eq) in 1,2-DCE (1 mL) was mixed with $^i$Pr$_2$NEt (140 µL, 805 µmol, 3.5 eq) and $^t$Bu$_2$SnCl$_2$ (77.1 mg, 253 µmol, 1.1 eq) and stirred for 1 h at ambient temperature. After heating the reaction to 80 °C, TOM-Cl (**7**, 56.4 mg, 253 µmol, 1.10 eq) was added and the reaction mixture was stirred for 20 min at 80 °C. The mixture was diluted with DCM and washed with a saturated $NaHCO_3$-solution. The organic phase was dried over $Na_2SO_4$ and the solvent was removed under reduced pressure. The crude residue was purified by column chromatography (*n*-hexane:EtOAc = 6:4 + 1% $NEt_3$, to EtOAc + 1% $NEt_3$) to obtain the product (**90**, 48.1 mg, 60.4 µmol, 26 %) as a colorless foam.

**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 8.25 (s, 1H, H-2), 7.32 – 7.27 (m, 4H, DMT), 7.25 – 7.19 (2H, DMT), 7.19 – 7.14 (m, 1H, DMT), 7.01 (dd, *J*= 17.2, 11.1 Hz, 1H, *vinyl*-CH), 6.73 – 6.78 (m, 4H, DMT), 6.40 (dd, *J* = 17.1, 1.6 Hz, 1H, *vinyl*-$CH_2$), 6.14 (d, *J* = 6.2 Hz, 1H, H-1'), 5.68 - 5.77 (m, 1H, N6-H), 5.43 – 5.49 (m, 2H, H-2', *vinyl*-$CH_2$), 5.06 (d, *J* = 4.7 Hz, 1H, $OCH_2O$), 4.92 (d, *J* = 4.7 Hz, 1H, $OCH_2O$), 4.69 (p, *J* = 3.0 Hz, 1H, H-3'), 4.21 (q, *J* = 4.0 Hz, 1H, H-4'), 3.77 (d, *J* = 0.5 Hz, 6H, $OCH_3$), 3.47 (dd, *J* = 10.4, 4.0 Hz, 1H, H-5'), 3.33 (dd, *J* = 10.4, 4.4 Hz, 1H, H-5'), 3.18 (d, *J* = 5.0 Hz, 3H, N6-$CH_3$), 3.01 (d, *J* = 3.4 Hz, 1H, HO-C3'), 0.96-1.12 (m, 21H, $^i$Pr$_3$Si).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 158.5 ($C_q$, DMT), 155.2 (C6), 152.9 (C2), 148.7 (C8), 144.9 ($C_q$, DMT), 136.0 ($C_q$, DMT), 130.2 (DMT), 128.4 (DMT), 127.9 (DMT), 126.8 (DMT), 124.0 (*vinyl*-CH), 123.6 (*vinyl*-CH2), 119.9 ($C_q$, DMT), 113.2 (DMT), 91.0 ($OCH_2O$), 86.3 (C1'), 84.1 (C4'), 79.9 (C2'), 70.8 (C3'), 63.3 (C5'), 55.3 ($OCH_3$), 17.9 (*i*Pr$_3$Si), 12.0 (*i*Pr$_3$Si).

**HR-MS** ($ESI^+$): Exact mass calculated for $C_{44}H_{58}N_5O_7Si$ $[M+H]^+$: 796.4100 found: 796.41030.

**5'-*O*-(4,4'-dimethoxytrityl-$N^6$-2'-*O*-(triisopropylsilyloxy)methyl-8-vinyl-methyladeneosine (91)**




**91**

Compound **90** (42.0 mg, 52.8 µmol, 1.0 eq) was dissolved in anhydrous DCM (1 mL) and mixed with $Me_2NEt$ (57.1 µL, 528 µmol, 10.0 eq) and CEP-Cl (18.7 mg, 17.7 µL, 79.1 µmol, 1.5 eq). The reaction was stirred for 2 h at ambient temperature, followed by the evaporation of the solvent under reduced pressure. The crude residue was purified by column chromatography (*n*-hexane:EtOAc = 3:1 + 1% $NEt_3$) to obtain the product (**91**, 40.0 mg, 40.1 µmol, 76%) as a colorless foam.

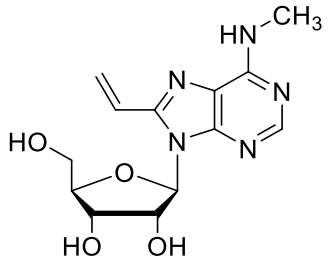
**$^1$H-NMR** ($CDCl_3$, 400 MHz): $\delta$ (ppm) = 8.26 (s, 1H, H-2), 8.23 (s, 1H, H-2), 7.33 – 7.28 (m, 4H, DMT), 7.24 – 7.17 (m, 12H, DMT), 7.17 – 7.13 (m, 2H, DMT), 6.98 (dd, *J* = 17.3, 11.0 Hz, 2H, *vinyl*-CH), 6.77 – 6.68 (m, 8H, DMT), 6.43 (dt, *J* = 17.2, 1.7 Hz, 2H, *vinyl*-CH2), 6.19 (d, *J* = 4.3, 4.6 Hz, 1H, H-1'), 6.16 (d, *J* = 4.3 Hz, 1H, H-1'), 5.69 (t, *J* = 4.8 Hz, 2H, NH), 5.65 – 5.62, (m, 2H, H-2'), 5.59 (dt, *J* = 11.1, 1.8 Hz, 2H, *vinyl*-$CH_2$), 5.09 – 5.04 (m, 2H, H-3'), 5.03 – 5.00 (2m, 2H, $OCH_2O$), 4.96, 4.93, (2d, J = 4.8, 4.6 Hz, 2H, $OCH_2O$), 4.28 (p, *J* = 4.3 Hz, 2H, H-4'), 4.00 – 3.86 (m, 2H, $POCH_2$), 3.76 (2s, 12H, $OCH_3$), 3.68 – 3.58 (m, 3H, $((CH_3)CH)_2N$, H-5', $POCH_2$), 3.58 – 3.48 (m, 2H, H-5'), 3.22 (dt, *J* = 10.5, 4.3 Hz, 2H, H-5'), 3.19 – 3.13 (m, 6H, N6-$CH_3$), jj2.67 (td, J = 6.5, 4.2 Hz, 2H, $CH_2CN$), 2.35 (t, *J* = 6.6 Hz, 2H, $CH_2CN$), 1.24-1.06 (m, 24 H, $((CH_3)_2CH)_2N$), 0.90-0.87 (m, 42H, $^i$Pr$_3$Si).

**$^{13}$C-NMR** ($CDCl_3$, 100 MHz): $\delta$ (ppm) = 158.4 ($C_q$, DMT), 148.74 (C8), 136.2 ($C_q$, DMT), 136.0 (DMT), 130.2 (DMT), 128.4 (DMT), 127.8 (DMT), 126.8 (DMT), 124.0 (*vinyl*-CH), 117.2 (CN), 113.1 (DMT), 87.8 (C1'), 86.1 ($C_q$, DMT), 83.3 (C4'), 70.7 ($OCH_2O$), 62.3 (C5'), 58.6 ($POCH_2$), 55.3 ($OCH_3$), 43.4 ($(((CH_3)_3CH)_2N$), 27.7 (N6-$CH_3$), 24.8 ($((CH_3)_2CH)_2N$), 20.2 ($CH_2CN$), 17.8 ($^i$Pr$_3$Si), 11.9 ($^i$Pr$_3$Si).

**$^{31}$P-NMR** (162 MHz, $CDCl_3$): $\delta$ (ppm) = 150.54, 149.75.

**HR-MS** ($ESI^+$): Exact mass calculated for $C_{53}H_{74}N_7NaO_8PSi$ $[M+Na]^+$: 1018.49980 found: 1018.49980.

**5'-*O*-(4,4'-dimethoxytrityl-$N^6$-8-vinyl-methyladeneosine (99)**

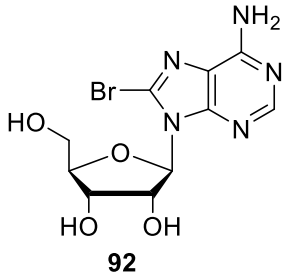


**99**

Compound **89** (88.5 mg, 145 µmol) was mixed with a formic acid-solution (2% in DCM, 2 mL). Methanol was added dropwise, until the orange color disappeared. The solvent was removed under reduced pressure. MeOH was repeatedly added and the solvent was removed, until the residue did not turn orange anymore. The crude residue was purified by column chromatography (DCM:MeOH = 9:1) to obtain the product (29.7 mg, 96 µmol, 66%) as a colorless solid.

**$^1$H-NMR** (DMSO-d6, 400 MHz): $\delta$ (ppm) = 8.19 (s, 1H, H-2), 7.91 (s, 1H, N4-H), 7.08 (dd, *J* = 17.0, 11.1 Hz, 1H, vinyl-CH), 6.33 (dd, *J* = 17.0, 2.0 Hz, 1H, vinyl-$CH_2$), 5.95 (d, *J* = 7.2 Hz, 1H, H-1'), 5.72 (dd, *J* = 8.3, 3.8 Hz, 1H, HO-C5'), 5.68 (dd, *J* = 11.1, 2.0 Hz, 1H, vinyl-$CH_2$), 5.37 (d, *J* = 6.8 Hz, 1H, HO-C2'), 5.21 (d, *J* = 4.6 Hz, 1H, HO-C3'), 4.79 (td, *J* = 6.9, 5.3 Hz, 1H, H-2'), 4.16 (td, *J* = 4.9, 2.3 Hz, 1H, H-3'), 3.98 (q, *J* = 3.2 Hz, 1H, H-4'), 3.68 (dt, *J* = 12.2, 3.6 Hz, 1H, H-5'), 3.56 (ddd, *J* = 12.0, 8.3, 3.5 Hz, 1H, H-5'), 2.96 (s, 3H, $CH_3$).
**$^{13}$C-NMR** (DMSO-d6, 100 MHz): $\delta$ (ppm) = 154.79 (C6), 152.10 (C2), 148.74 (C4), 147.48 (C8), 124.14 (vinyl-CH), 122.89 (vinyl-$CH_2$), 87.79 (C1'), 86.57 (C4'), 72.35 (C2'), 70.69 (C3'), 62.00 (C5'), 27.08 ($CH_3$).
**HR-MS** (ESI$^+$): Exact mass calculated for $C_{13}H_{18}N_5O_4$ $[M+H]^+$: 308.13533 found: 308.13557.

**5'-*O*-(4,4'-dimethoxytrityl-8-vinyl-adeneosine (92)**



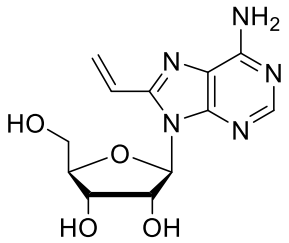
**92**

Adenosine (3.00 g, 11.2 mmol, 1.0 eq) was suspended in sodium acetate buffer (75 mL, pH 4). An aqueous bromine solution (3.58 g, 1.15 mL, 22.4 mmol, 2.0 eq, in 100 mL $H_2O$) was added dropwise to the suspension and the reaction was stirred for 72 h at room temperature. The reaction was stopped by the addition of a sodium hydrogen sulfite solution, until the reaction turned colorless. Then the reaction mixture was neutralized with a sodium hydroxide solution (2 M), resulting

in precipitation of the product. The solid was filtered off and dried in high vacuum. The product (**92**, 3.15 g, 9.10 mmol, 81 %) was obtained as a colorless solid.[505]

**$^1$H-NMR** (DMSO-d6, 400 MHz): *δ* (ppm) = 8.12 (s, 1H, H-2), 7.66 – 7.50 (m, 2H, $NH_2$), 5.83 (d, *J* = 6.8 Hz, 1H, H-1'), 5.52 (s, 1H, HO-C5'), 5.48 (d, *J* = 6.3 Hz, 1H, HO-C2'), 5.25 (s, 1H, HO-C3'), 5.09 (q, *J* = 5.7 Hz, 1H, H-2'), 4.19 (d, *J* = 5.7 Hz, 1H, H-3'), 3.98 (td, *J* = 4.0, 2.3 Hz, 1H, H-4'), 3.68 (dd, *J* = 12.1, 4.0 Hz, 1Hm H-5'), 3.52 (d, *J* = 11.8 Hz, 1H, H-5').
**$^{13}$C-NMR** (DMSO-d6, 100 MHz): *δ* (ppm) = 155.17 (C6), 152.44 (C2), 149.85 (C4), 127.23 (C8), 119.69 (C5), 90.39 (C1'), 86.75 (C4'), 71.12 (C2'), 70.91 (C3'), 62.14 (C5').
**HR-MS** (ESI$^+$): Exact mass calculated for $C_{10}H_{12}N_5BrNaO_4$ [M+Na]$^+$: 367.99649 found: 367.99636.

**5'-*O*-(4,4'-dimethoxytrityl)-8-vinyl-adeneosine (93)**



**93**

Compound **92** (100 mg, 289 µmol, 1.0 eq) was dissolved in anhydrous DMF (1 mL). To this mixture $Pd(PPh_3)_4$ (50.0 mg, 43.3 µmol, 0.15 eq) and tributyl(vinyl)tin (169 µL, 578 µmol, 2.0 eq) were added. The reaction was stirred at 100 °C for 6 h. The solvent was removed in vacuum and the crude residue was purified by column chromatography (DCM:MeOH = 10:1 + 1% $NEt_3$) to obtain the product (**93**, 30.0 mg, 102 µmol, 35%) as a colorless foam.[506]

**$^1$H-NMR** (DMSO-d6, 400 MHz): *δ* (ppm) = 8.09 (s, 1H, H-2), 7.42 (s, 2H, $NH_2$), 7.08 (dd, *J* = 17.0, 11.1 Hz, 1H, vinyl-CH), 6.34 (dd, *J* = 17.0, 2.0 Hz, 1H, vinyl-$CH_2$), 5.94 (d, *J* = 7.2 Hz, 1H, H-1'), 5.72 (dd, *J* = 7.5, 3.0 Hz, 1H, HO-C5'), 5.70 – 5.67 (m, 1H, vinyl-$CH_2$), 5.38 (d, *J* = 6.8 Hz, 1H, HO-C2'), 5.22 (d, *J* = 4.6 Hz, 1H, HO-C3'), 4.80 (td, *J* = 6.9, 5.3 Hz, 1H, H-2'), 4.15 (td, *J* = 4.9, 2.3 Hz, 1H, H-3'), 3.98 (q, *J* = 3.2 Hz, 1H, H-4'), 3.68 (dt, *J* = 12.3, 3.6 Hz, 1H, H-5'), 3.56 (ddd, *J* = 12.0, 8.2, 3.5 Hz, 1H, H-5').
**$^{13}$C-NMR** (DMSO-d6, 100 MHz): *δ* (ppm) = 155.94 (C6), 152.05 (C2), 149.74 (C4), 147.72 (C8), 124.16 (vinyl-CH), 123.07 (vinyl-$CH_2$), 118.91 (C5), 87.77 (C1'), 86.54 (C4'), 72.26 (C2'), 70.69 (C3'), 62.00 (C5').
**HR-MS** (ESI$^+$): Exact mass calculated for $C_{12}H_{15}N_5NaO_4$ [M+Na]$^+$: 316.10162 found: 316.10164.

## 5.2 Oligonucleotide Syntheses and characterization

### 5.2.1 General Remarks

Unmodified DNA and 2'-*O*-TOM-protected RNA phosphoramidites and the controlled pore glass (CPG) solid support (pore size 1000 Å, 20 – 30 µmol/g for RNA and 25 – 35 µmol/g for DNA) were purchased from Sigma Proligo, ChemGenes, or Glen Research. 3'-phosphate CPG and 2-[2-(4,4'-Dimethoxytrityloxy)ethylsulfonyl]ethyl-[(2-cyanoethyl)-(*N*,*N*-diisopropyl)]phosphoramidite for 5'-phosphorylation was prepared according to chapter 5.1.2.8 and chapter 5.1.2.9. Native amino CPG support (1000 Å, 74 µmol/g) was purchased from ChemGenes. The prepared phosphoramidite solutions in anhydrous acetonitrile were additionally dried over molecular sieve (3 Å) and stored in the fridge overnight.

### 5.2.2 Automated solid phase synthesis

RNA and DNA oligonucleotides were prepared on CPG solid support using Pharmacia LKB Gene Assembler Plus or Applied Biosystems ABI 392 DNA/RNA synthesizer on 0.6 µmol scale. The solid support (25-30 mg) was filled into the column and both ends were covered with filters.

The following solutions were used for automated solid-phase synthesis (Pharmacia): 100 mM solutions in anhydrous acetonitrile of the phosphoramidites, 0.25 M ethylthiotetrazole (ETT, activator) in anhydrous acetonitrile, 3% dichloro acetic acid in 1,2-dichloroethane (detritylation), 0.5 M DMAP in acetonitrile (Cap A), acetic acid anhydride/sym-collidine/acetonitrile = 20/30/50 (Cap B), 10 mM $I_2$ in acetonitrile/sym-collidine/$H_2O$ = 10/1/5 (oxidation). A coupling time of 4 min (07CH-A4, C4, G4, T4) for unmodified and 12 min (07CH-X12) for modified phosphoramidites was chosen.

The following solutions were used for automated solid-phase synthesis (ABI): 70 mM solutions in anhydrous acetonitrile of the phosphoramidites, 250 mM ETT (acetonitrile), 3% dichloro acetic acid in DCM (detritylation RNA), 3% trichloro acetic acid in DCM (detritylation DNA), 20 mM iodine in THF/water/pyridine (66/12/22, oxidation), pyridine/acetic anhydride/THF (10/10/80, Cap A) and NMI (16% in THF, Cap B). The standard protocol “0.6uM DNA lv1cap Test Jan” was used for the synthesis of DNA oligonucleotides with 2 min coupling time, while “0.6uM RNA Test Jan” was used for RNA oligonucleotides with 4 min coupling time. For both, DNA and RNA, the option “DMT-off” was used for the last step. The coupling yield for each step was followed by measuring the absorbance of the resulting trityl cation during the detritylation step at 498 nm.

### 5.2.3 Deprotection and work-up

The resulting oligonucleotides were cleaved off from the solid support and deprotected under basic conditions. The conditions are summarized in Table 20.

Table 20. Deprotection conditions for modified- and unmodified RNA and DNA oligonucleotides.

| DNA/RNA oligonucleotides | Condition |
|---|---|
| **unmodified RNA/DNA** | 1 mL $MeNH_2$ in EtOH/$H_2O$ (1:1), 37°C, 5 – 6 h<br>1 mL $MeNH_2$ in $H_2O$ and $NH_4OH$ (1:1), 37°C, 5 – 6 h<br>1 mL $NH_4OH$ in EtOH (3:1), 55 °C, 5 – 6 h |
| **$m^3C$** | 1 mL $NH_4OH$ in EtOH (3:1), 55 °C, 5 – 6 h |
| **$(ms^2)i^6A$** | 1 mL $MeNH_2$ in EtOH/$H_2O$ (1:1), 37°C, 5 – 6 h |
| **$m^3C/(ms^2)i^6A$** | 1 mL $NH_4OH$ in EtOH (3:1), 55 °C, 5 – 6 h |
| **$t^6A$** | 1. 10% DBU in THF, r. t., 2 h, then discard supernatant, and dry CPG<br>2. $NH_4OH$ and $MeNH_2$ in $H_2O$ (1:1), r. t., 1.5 h |
| **c7dG crosslinker** | $NH_4OH$, 50 °C, 14 h |
| **$m^6v^8A$** | 1 mL $H_2O$:MeOH:$NEt_3$ (5:4:1), 25 °C, 48 h |
| **$m^{4,4}C$** | 1. 700 µL $Me_2NH$ (40% in $H_2O$), 37 °C, 5 h (convertible nucleoside)<br>2. 0.5 mL $NH_4OH$, 55 °C, 2 h |
| **NHC** | 1 mL $MeNH_2$ in EtOH/$H_2O$ (1:1), 37°C, 5 – 6 h |

After the basic deprotection, the RNA oligonucleotides were stirred in a freshly prepared 1 M TBAF solution in THF (500 µL) for 12 h at room temperature (This part can be left out for DNA oligonucleotides). Then, the reaction was mixed with 500 µL of 1 M Tris·HCl (pH 8.0) and the volume was reduced to 500 µL by evaporation of THF in vacuum. The oligonucleotide sample was desalted by size exclusion chromatography using an Äkta start purification system (Sephadex G 10, 3 × 5 mL HiTrap desalting columns, flow rate: 2 mL/min). The water was removed under reduced pressure and the crude oligonucleotide was re-dissolved in 400 – 600 µL nanopure water. The concentration of the crude sample was measured by UV absorption and the synthesis quality was checked by anion-exchange chromatography.

### 5.2.4 PAGE purification

The crude oligonucleotides were purified by denaturing PAGE. Depending on the oligonucleotide length, acrylamide gels (7 M urea, 1x TBE) with either 20% (≤ 20 nt), 15% (20 – 50 nt) or 10% (≥ 50 nt) concentration of acrylamide were prepared. Therefore, 45 mL of the corresponding stock solution was mixed with 125 µL of 25% APS solution (radical initiator) and 30 µL *N*, *N*, *N*′, *N*′-tetramethylethylenediamine (TEMED, polymerization catalyst) to initiate the polymerization. The solution was allowed to polymerize for 30 min. For oligonucleotide purification, polyacrylamide gels with a size of 20 x 30 cm x 0.7 mm and four wells were prepared. For one gel 35 W and for two gels 60 W must be applied. After a pre-run of 30 min with 1x TBE buffer (89 mM TRIS, 89 mM boronic acid, 2 mM EDTA (pH 8)), the wells were washed. Up to 20 nmol crude sample, which was mixed with high dye (80% formamide, 1x TBE, 50 mM EDTA, 0.25 mg/mL bromphenol blue, 0.25 mg/mL xylene cyanol) in a ratio of 1:1, was loaded into one pocket. The oligonucleotide bands were visualized after 2.5 h by UV shadowing. The product bands were cut out from the gel, transferred into fresh Eppendorf tubes, covered with 400 µL TEN buffer (10 mM TRIS·HCl, pH 8, 1 mM EDTA, pH 8, 300 mM NaCl) and stored over night at 4 °C. The solution with the extracted RNA/DNA was transferred into new Eppendorf tubes and 300 µL TEN buffer was added to the gel pieces.

After the incubation for 2 h at 37 °C, the supernatant was again transferred into Eppendorf Tubes. The solutions were mixed with cold absolute ethanol in a ratio of 1:3, to get a final concentration of 75% ethanol for precipitation of the RNA/DNA. The Eppendorf tubes with the ethanolic solution were frozen in liquid nitrogen for 5 min and then centrifuged for 30 min at 4 °C and 15000 rpm. The supernatant was removed, and the pellets were covered with 75 µL of a cold ethanol solution (70%) and centrifuged for 10 more minutes. The supernatant was removed, and the pellets were dried in vacuum. The dried oligonucleotides were re-dissolved in 200 – 400 µL of nanopure water. For concentration determination, 1 µL of the oligonucleotide solutions were measured by UV absorbance (NanoPhotometer® P-Class, P360, Implen).

For analytical PAGE of the size 20 x 30 cm x 0.4 mm, 25 W for 1.5 h was used.

Small analytical gels with 10 pockets for Mini-PROTEAN® Tetra Vertical Electrophoresis were prepared by mixing 10 µL of acrylamide solution with 25 µL 25% APS and 10 µL of TEMED. The gels were directly loaded after polymerization with 6 µL of the respective sample in High dye (1:1). The run time in 1x TBE buffer was 30 – 60 min at 200 V constant, depending on the acrylamide concentration and oligonucleotide length. For 15% native PAGE (without urea) 60 min run time at 100 V was used, while cooling the gel-stand with ice. Non-fluorescent double stranded oligonucleotides were stained with SYBR green, whereas single strands were stained with SYBR gold. Fluorescently labeled oligonucleotides were visualized with the ChemiDoc® MP Imaging System Universal Hood III from BioRad using the fluorescein channel (530/28 filter, blue epi illumination).

### 5.2.5 HPLC

To validate the quality of the crude and pure oligonucleotide samples, anion exchange chromatography analysis was performed on ÄKTApurifier system from GE Healthcare with an DNAPac PA200 column (2×250 mm and 4x250) from Thermo Scientific using Dionex buffer (buffer A: 25 mM Tris·HCl, pH 8.0, 6 M urea, buffer B: 25 mM Tris·HCl, pH 8.0, 6 M urea, 0.5 M $NaClO_4$). A sample concentration of 50 pmol oligonucleotide in 15 µL nanopure water was injected to the column. The HPLC runs were performed at 60 °C with the gradient 0 – 48% B over twelve column columns (cv). A flow rate of 0.5 mL/min was used for the 2x250 column, whereas 2 mL/min for the 4x250 column.

After the chemical synthesis of the triphosphate, the first purification step was performed by anion-exchange chromatography on an ÄKTAprime plus HPLC system using a Sephadex column (DEAE) and triethylammonium bicarbonate (TEAB) buffer (buffer A: 0.1 M TEAB, buffer B: 1 M TEAB, pH 7.5-8).

For further purification of the triphosphate and one RNA oligonucleotide reverse-phase (RP) HPLC was used on an ÄKTA micro system from GE Healthcare with a Nucleosil 100-7 C18 column (8x) from Macherey-Nagel at 30 °C and triethylammonium acetate (TEAA) buffer (buffer A: 100 mM

TEAA in $H_2O$, pH 7.0, buffer B: 100 mM TEAA in 10 % $H_2O$ and 90% acetonitrile, pH 7.0). Gradient for the RNA sample: 5 – 20 % B in 4.5 cv. Gradient for the triphosphate: 0.1 – 10 % B in 5 cv with a flow rate of 2 mL/min.

After the HPLC runs the collected fractions were lyophilized and re-dissolved in nanopure water.

## 5.3 Spectroscopy and Crystallography

### 5.3.1 Mass spectrometry

The pure oligonucleotides were validated by HR-ESI/MS spectrometry. 8 µL of the oligonucleotide solution (25 µM in $H_2O$) was mixed with 30 µL HFIP and 30 µL acetonitrile (final concentration ~ 3 µM) and subjected to the Bruker Daltonics micrOTOF focus system (micrOTOF-Q III) for Electrospray Ionization (ESI) in the negative mode. A standard method was used for the measurement in a range of 50 – 4000 m/z with the nitrogen nebulizer set to 0.4 bar, the dry heater set to 200 °C and the dry gas to 4.0 l/min. As calibrant, sodium formate was injected prior to each measurement. Data evaluation was performed with Compass Data Analysis by Bruker Daltonics.

ESI-MS of the small molecules were also measured on the same Bruker Daltonics micrOTOF focus system (micrOTOF-Q III), mainly in the positive-ion mode, expect for the triphosphate building block. The monoisotopic mass is obtained after charge deconvolution of the measured raw spectra.

For MALDI analysis of oligonucleotides 0.5 µL of a 3-hydroxypicolinic acid (3-HPA, 20 mg/mL) and diammonium hydrogen citrate (DAC 10 mg/mL) mixture was applied to the grid as the matrix. After drying, the oligonucleotide sample (1 µL, 12 pmol in MeCN/$H_2O$ (1:1)) was applied on the matrix and was subjected to the Bruker Daltonics UltrafleXtreme Symstem for MALDI measurement.

### 5.3.2 NMR spectroscopy

Oligonucleotides for NMR spectroscopy were further purified after PAGE purification and ethanol precipitation: 50 nmol of the respective oligonucleotide was lyophilized using the Christ Alpha 2-4 LSCbasic lyophilizer, re-dissolved in 20 µL Milli-Q water and precipitated with 100 µL acetone/$LiClO_4$ (2% w/v). Therefore, the sample was stored at -20 °C for 4-6 h and then centrifuged at 4 °C for 30 min with 15000 rpm. The supernatant was removed, and the residue was dried in vacuum. The resulting pellet was dissolved in 100 µL Milli-Q water and lyophilized. After dissolving in 400 µL water the concentration was determined and 36 nmol were transferred for the preparation of a 200 µM hairpin NMR sample. For duplex structures, higher concentrations were prepared with 28-36 nmol of each strand in an equimolar ratio to obtain 160 – 200 µM duplex concentration. To this solution 9 µL sodium phosphate buffer (200 mM, pH 7) and 18 µL NaCl (1 mM) were added to generate a final concentration of 10 µM and 100 µM, respectively. The solution was again lyophi-

lized and re-dissolved in 162 water and 18 µL $D_2O$ (9:1), adding 0.4 µL sodium trimethylsilylpropanesulfonate (DSS, 7 mM) as internal standard. The final volume of 180 µL was heated to 95°C for 2 min and cooled down for 15 min on ice to anneal the hairpin structures, whereas for 30 min at ambient temperature for duplex structures. The correct folding of the oligonucleotide samples was examined by a 15% native PAGE stained with SYBR gold (see chapter 5.2.4). The oligonucleotide sample was transferred to a 3 mm NMR tube. NMR spectra were recorded on a Bruker Avance III 600 NMR spectrometer equipped with a DCH $^{13}C$ / $^{1}H$ cryoprobe or on a on a Bruker Avance III 600 NMR spectrometer equipped with a BBFO room temperature probe at 283, 298, 308, 318 and 329K, performing water suppression by excitation sculpting[507] or jump-return-echo (excitation maximum on the middle of the imino region)[508]. The obtained spectra were processed with Topspin 3.2 (Bruker BioSpin, Germany) and analyzed with the NMR Assignment and Integration software Sparky 3.114.[509] The $^{15}N$-labeled Dickerson Drew sequences were analyzed by 1D $^{1}H$, $^{15}N$-filtered and 2D $^{1}H$-$^{15}N$ HSQC spectra (referenced to DSS).[510-513] Imino-water exchange rates ($k_{EX}$) derived from 1D version of CLEANEX-PM experiments[514, 515] following a protocol published earlier.[516, 517] Exchange experiments were performed with different mixing times $\tau_m$ (5, 25, 50, 50, 100, 100, 150, 200, 300, 400, 400, 500 ms) using the Bruker pulse program zgcxesgp (standard), employing excitation sculpting with gradients as water suppression scheme.[507] For each experiment 1024 scans were recorded with an interscan delay of 2.0 s and 16 dummy scans. A line broadening factor of 10 Hz was used to process the spectra and the peak intensities were used for the data analysis. A published procedure for a saturation recovery experiment was used to determine the water $^{1}H$ longitudinal relaxation rate $R_{1w}$ ($s^{-1}$).[518] The intensity of the respective imino signals from the CLEANEX-PM experiment (I) and the reference experiment ($I_0$) was calculated and plotted as a function of the mixing time ($\tau_m$). To obtain the imino proton water exchange rate $k_{EX}$ ($s^{-1}$) and the imino proton relaxation rate $R_{1A}$ the resulting curve was fitted to the following equation (Mathematica 8.0):

$$\frac{I}{I_0} = \frac{k_{Ex}}{(R_{1A} + k_{EX} - R_{1w})} \cdot \left(e^{-R_{1w} \cdot \tau_m} - e^{(-R_{1A}+k_{EX}) \cdot \tau_m}\right) \qquad (4)$$

With I (intensity of the imino peak at the mixing time $\tau_m$), $I_0$ (intensity of the imino peak in the reference experiment), $k_{EX}$ (imino proton to water exchange rate in $s^{-1}$), $\tau_m$ (mixing time in $s^{-1}$), $R_{1w}$ (water $^{1}H$ longitudinal relaxation rate in $s^{-1}$) and $R_{1A}$ (imino $^{1}H$ apparent relaxation rate as a combination of longitudinal and transverse relaxation rate in $s^{-1}$).

### 5.3.3 UV melting curves

UV thermal melting curves of the oligonucleotide samples were recorded on a Varian Cary 100 spectrometer with a Peltier temperature controller for the determination of hairpin or duplex stability after incorporation of a modified nucleoside. Five concentrations of the RNA oligonucleotides were

prepared in 10 mM sodium phosphate buffer (pH 7.0) containing 100 mM NaCl and supplied in quartz cuvettes. 300 µL of samples with 20 µM and 10 µM duplex concentration were filled in macro cuvettes (Hellma®) with 0.1 cm path length, whereas the samples with of 5 µM, 2 µM and 1 µM duplex concentration were filled in semi-micro cuvettes with 1 cm path length. To prevent evaporation of the solvent, a layer of silicon oil was applied to cover the aqueous solution. Furthermore, the chamber which contained the cuvettes was purged with a constant nitrogen flow, to circumvent the condensation of the air moisture on the glass surface arising from the temperature changes. Two heating and two cooling ramps with a heating rate of 0.5 °C/min were measured in a temperature interval of 10 to 90 °C and monitored at 250, 260 and 280 nm. For analysis, the melting curves were shown as hyperchromicity versus temperature or normalized absorbance versus temperature profiles. The melting temperatures were obtained from the inflection point of the sigmoidal melting curves and are given as the mean values of the two heating and two cooling ramps. They were determined by fitting a two-state transition model with lower and upper baseline, assuming an accuracy of ± 0.5 °C.[58]

The hyperchromicity (Hyp) is the increase in absorption (Abs) upon heating the sample, thus the absorption changes due to denaturation and changes in the base stacking. This absorption change of the hybridized duplex or hairpin structures ($Abs_{min}$) and the denaturing single strands (Abs) is given in percentage according to the following equation:

$$Hyp = \frac{Abs - \mathrm{Abs_{min}}}{\mathrm{Abs_{min}}} \cdot 100 \qquad (5)$$

The 50% value between the maximum and minimum absorption reflects the thermal melting point ($T_m$) of the oligonucleotide (Figure 41). The $T_m$ values were obtained at the inflection point of the sigmoidal curve after plotting the hyperchromicity vs. temperature by applying a sigmoidal fit towards the lower and higher base line. For duplex structures, enthalpy change ($\Delta H^0$), entropy change ($\Delta S^0$), and free energy change ($\Delta G^0$) were calculated from the van't Hoff equation:[519]

$$\frac{1}{T_m} = \frac{(n-1)\cdot R}{\Delta H^0} \cdot ln(c_{tot}) + \frac{[\Delta S^0 - (n-1)\cdot R \cdot ln(2n)]}{\Delta H^0} \qquad (6)$$

where n is the molecularity, R is the molar gas constant (8.31448 J/mol·K) and $c_{tot}$ is the total oligonucleotide concentration.[519] As for duplex structures the molecularity is two, equation 6 can be simplified accordingly:

$$\frac{1}{T_m} = \frac{R}{\Delta H^0} \cdot ln(c_{tot}) + \frac{[\Delta S^0 - R \cdot ln(4)]}{\Delta H^0} \qquad (7)$$

By plotting the $1/T_m$ values against $\ln(c_{tot})$, the enthalpy changes ($\Delta H^0$) of the duplex denaturation can be determined from the slope (m) of the linear fit:

$$\Delta H^0 = \frac{R}{m} \qquad (8)$$

With the obtained enthalpy values and the intercept (b) of the linear fit the change of entropy ($\Delta S^0$) can be calculated according to:

$$\Delta S^0 = b \cdot \Delta H^0 + R \cdot ln(4) \qquad (9)$$

Finally, the free energy can be determined by inserting the calculated $\Delta H^0$ and $\Delta S^0$ values in the Gibbs-Helmholtz equation with T = 298 K:

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 \qquad (10)$$

For hairpin structures these values cannot be determined with those equations, as the molecularity n=1 would lead to m = 0, as the melting of a hairpin is concentration independent.[519]

### 5.3.4 CD spectroscopy

CD spectra were recorded on a JASCO J-715 Spectropolarimeter with a JASCO Power Supply PS-150J and a Xenon Lamp. The oligonucleotide solution (5 µM) in sodium phosphate buffer (pH 7) containing 1 mM $MgCl_2$ was annealed by heating to 95°C for 2 min and subsequent cooling on ice for 30 min for a hairpin structure or room temperature for a duplex structure. The sample was transferred to a semi-micro quartz cuvette (Hellma Analytics) with 1 cm path length and was recorded at room temperature with a band width of 10.0 nm, a Data pitch of 0.1 nm and a Scanning speed of 500 nm/min. The spectra were recorded as duplicates in a range of 350 nm to 190 nm and the mean values were plotted after the base line correction with the CD spectrum of the control sample, which only contained the buffer solution.

### 5.3.5 Fluorescence

Fluorescence spectra were recorded on a JASCO FP-8300 spectrofluorometer equipped with a water thermostatted automatic 8-position turret cell changer (FCT-817S) and a Julabo F12 temperature control system. 10 µM samples were transferred into ultra-micro quartz cuvettes (Hellma®) with 1 cm path length. A photomultiplier tube (PMT)-Voltage of 300 V, a scan speed of 500 nm/min, an Excitation Bandwidth of 5 nm, an Emission Bandwidth of 5 nm and a Data interval of 0.2 nm was chosen to obtain a high intensity of the absorption and emission spectra, with a low signal to noise ratio and a short exposure time of the sample.

### 5.3.6 Crystallography

25 nmol of the respective oligonucleotides were concentrated to around 20 µL using Vivaspin® 500 µL Satorius columns. HEPES and NaCl solutions were added to get a final concentration of 10 mM and 20 mM, respectively, with a total volume of 25 µL (1 mM RNA concentration). The initial

screening was performed with the HELIX™ HT-96 screen. Conditions which resulted in the growth of crystals were chosen for refinement, as well as one condition from a report by Li *et al.* with a similar sequence context (Table 21).[520] 0.4 µL of the respective buffer solution was mixed with 0.4 µL of the oligonucleotide in the hanging drop method. It was stored at room temperature.

Table 21. Refinement of the crystallography conditions obtained from the inital screening (HELIX).

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| **A** | 200 µL 1 M Bis-Tris pH = 6 [100 mM] 914.3 µL 3.5 M $(NH_4)_2SO_4$ [1.6 mM] 885.7 µL $H_2O$ | 200 µL 1 M Bis-Tris pH = 6 [100 mM] 971.4 µL 3.5 M $(NH_4)_2SO_4$ [1.7 mM] 828.6 µL $H_2O$ | 200 µL 1 M Bis-Tris pH = 6 [100 mM] 1028.6 µL 3.5 M $(NH_4)_2SO_4$ [1.8 mM] 771.4 µL $H_2O$ | 200 µL 1 M Bis-Tris pH = 6 [100 mM] 1085.7 µL 3.5 M $(NH_4)_2SO_4$ [1.9 mM] 714.3 µL $H_2O$ | 200 µL 1 M Bis-Tris pH = 6 [100 mM] 1142.9 µL 3.5 M $(NH_4)_2SO_4$ [2.0 mM] 657.1 µL $H_2O$ | 200 µL 1 M Bis-Tris pH = 6 [100 mM] 1200 µL 3.5 M $(NH_4)_2SO_4$ [2.1 mM] 600 µL $H_2O$ |
| **B** | 100 µL 1 M Bis-Tris pH = 7 [50 mM] 100 µL 100% PEG 400 [5%] 914.3 µL 3.5 M $(NH_4)_2SO_4$ [1.6 mm] 885.7 µL $H_2O$ | 100 µL 1 M Bis-Tris pH = 7 [50 mM] 100 µL 100% PEG 400 [5%] 971.4 µL 3.5 M $(NH_4)_2SO_4$ [1.7 mM] 828.6 µL $H_2O$ | 100 µL 1 M Bis-Tris pH = 7 [50 mM] 100 µL 100% PEG 400 [5%] 1028.6 µL 3.5 M $(NH_4)_2SO_4$ [1.8 mM] 771.4 µL $H_2O$ | 100 µL 1 M Bis-Tris pH = 7 [50 mM] 100 µL 100% PEG 400 [5%] 1085.7 µL 3.5 M $(NH_4)_2SO_4$ [1.9 mM] 714.3 µL $H_2O$ | 100 µL 1 M Bis-Tris pH = 7 [50 mM] 100 µL 100% PEG 400 [5%] 1142.9 µL 3.5 M $(NH_4)_2SO_4$ [2.0 mM] 657.1 µL $H_2O$ | 100 µL 1 M Bis-Tris pH = 7 [50 mM] 100 µL 100% PEG 400 [5%] 1200 µL 3.5 M $(NH_4)_2SO_4$ [2.1 mM] 600 µL $H_2O$ |
| **C** | 40 µL 5 M NaCl [100 mM] 50 µL KCl [100 mM] 100 µL 1 M Bis-Tris pH = 6.5 [50 mM] 914.3 µL 3.5 M $(NH_4)_2SO_4$ [1.6 mM] 895.7 µL $H_2O$ | 40 µL 5 M NaCl [100 mM] 50 µL KCl [100 mM] 100 µL 1 M Bis-Tris pH = 6.5 [50 mM] 971.4 µL 3.5 M $(NH_4)_2SO_4$ [1.7 mM] 838.6 µL $H_2O$ | 40 µL 5 M NaCl [100 mM] 50 µL KCl [100 mM] 100 µL 1 M Bis-Tris pH = 6.5 [50 mM] 1028.6 µL 3.5 M $(NH_4)_2SO_4$ [1.8 mM] 781.4 µL $H_2O$ | 40 µL 5 M NaCl [100 mM] 50 µL KCl [100 mM] 100 µL 1 M Bis-Tris pH = 6.5 [50 mM] 1085.7 µL 3.5 M $(NH_4)_2SO_4$ [1.9 mM] 724.3 µL $H_2O$ | 40 µL 5 M NaCl [100 mM] 50 µL KCl [100 mM] 100 µL 1 M Bis-Tris pH = 6.5 [50 mM] 1142.9 µL 3.5 M (NH4)2SO4 [2.0 mM] 667.1 µL $H_2O$ | 40 µL 5 M NaCl [100 mM] 50 µL KCl [100 mM] 100 µL 1 M Bis-Tris pH = 6.5 [50 mM] 1200 µL 3.5 M (NH4)2SO4 [2.1 mM] 610 µL $H_2O$ |
| **D** | 100 µL 5 M NaCl [250 mM] 200 µL 1 M Bis-Tris pH = 7 [100 mM] 100 µL 100 mM Spermine-HCl [5 mM] 461.1 µL 43% PEG 4000 [10%] 1234.9 µL $H_2O$ | 100 µL 5 M NaCl [250 mM] 200 µL 1 M Bis-Tris pH = 7 [100 mM] 100 µL 100 mM Spermine-HCl [5 mM] 558.1 µL 43% PEG 4000 [12%] 1141.9 µL $H_2O$ | 100 µL 5 M NaCl [250 mM] 200 µL 1 M Bis-Tris pH = 7 [100 mM] 100 µL 100 mM Spermine-HCl [5 mM] 651.2 µL 43% PEG 4000 [14%] 1048.8 µL $H_2O$ | 100 µL 5 M NaCl [250 mM] 200 µL 1 M Bis-Tris pH = 7 [100 mM] 100 µL 100 mM Spermine-HCl [5 mM] 744.2 µL 43% PEG 4000 [16%] 955.8 µL $H_2O$ | 100 µL 5 M NaCl [250 mM] 200 µL 1 M Bis-Tris pH = 7 [100 mM] 100 µL 100 mM Spermine-HCl [5 mM] 837.2 µL 43% PEG 4000 [18%] 862.8 µL $H_2O$ | 100 µL 5 M NaCl [250 mM] 200 µL 1 M Bis-Tris pH = 7 [100 mM] 100 µL 100 mM Spermine-HCl [5 mM] 930.2 µL 43% PEG 4000 [20%] 769.8 µL $H_2O$ |

## 5.4 Enzymatic reactions

### 5.4.1 General Remarks

T4 Polynucleotide Kinase (PNK) and DreamTaq polymerase were purchased from ThermoFischer Scientific, Calf Intestinal Alkaline Phosphatase (CIAP) and Pfu DNA polymerase from Fermentas, Fluorescein-thiosemicarbazide (FAM-TSC) from Sigma, 6-FAM-azide and deoxyribonucleotides (dNTPs) from Jena Bioscience, Therminator™ DNA Polymerase and RNase Inhibitor (Human Placenta) from New England BioLabs Inc., Glycogen for mol. Biol. From Roche, Taq-polymerase was prepared by Ann-Kathrin Lenz based on an established Taq-purification procedure.

### 5.4.2 Copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) of 5' alkynes

To a solution of the respective oligonucleotide (10.0 nmol in 10 µL $H_2O$), containing 5' alkyne, 6 µL DMSO/$^t$BuOH (3:1 v/v) and 1 µL of a 6-FAM-azid-solution (50 mM in DMSO) were added. 1 µL of a freshly prepared CuBr-solution (100 mM) in DMSO/$^t$BuOH (3:1 v/v) was mixed with 2 µL of TBTA (100 mM) in DMSO/$^t$BuOH (3:1 v/v). This CuBr/TBTA solution was added to the oligonucleotide reaction mixture and incubated for 3 h at 37 °C in the dark. To stop the reaction, the mixture was mixed with 20 µL of $H_2O$ and 40 µL of stop solution and the labeled product was purified by PAGE.

### 5.4.3 Labeling of non-functionalized 3'-ends with $NaIO_4$

Native RNA (5 nmol in 30 µL $H_2O$) was mixed with 8 µL sodium phosphate buffer (250 mM, pH 7.4) and 2 µL of a freshly prepared $NaIO_4$-solution (400 mM). The reaction was incubated for 10 min at

37 °C and then stopped by the addition of 4 µL $Na_2SO_3$ (1 M). After incubation for 5 min at 37 °C, 4 µL Fluorescein-5-thiosemicarbazide (100 mM in DMF) were added to the reaction and incubated for 2 h at 37 °C in the dark. The reaction was mixed with 40 µL of STOP-solution and purified by PAGE and analyzed after ethanol purification by MALDI-MS.

### 5.4.4 Phosphorylation of 5'-OH ends with T4 PNK

To a solution of the respective oligonucleotide (1-5 nmol in $H_2O$) 5.0 µL of 10x PNK buffer A (500 mM Tris-HCl (pH 7.4), 100 mM $MgCl_2$, 50 mM DTT, 1 mM spermidine), 5 µL ATP (10 mM) and 3 µL T4 PNK (10 u/µL) were added. The reaction was adjusted to a total volume of 50 µL with $H_2O$ and incubated for 30 min at 37 °C. The enzyme was heat inactivated by incubating for 20 min at 65 °C.

### 5.4.5 Dephosphorylation of 3'-phosphate ends with CIAP

A solution of 3' phosphorylated RNA (10 nmol in 40 µL $H_2O$) was mixed with 5 µL CIAP 10x buffer (500 mM Tris-HCl (pH-Wert 8,5), 1 mM EDTA) and 5 µL CIAP (1 u/µL) and incubated for 1 h at 37 °C. RNA was extracted with a premixed phenol/chloroform/isoamylalcohol solution and subsequent ethanol precipitation.

### 5.4.6 Primer Extension

A solution of the DNA template (40 pmol) and fluorescently labeled primer (20 pmol) was annealed by heating to 95 °C for 2 min and slowly cooling to room temperature. This solution was mixed with 0.1 µL dNTP-mix (10 mM), 2 µL 10x ThermoPol buffer (200 mM Tris-HCl, 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 20 mM $MgSO_4$, 1% Triton®X-100, pH 8.8) and 0.5 µL Terminator polymerase (2 u/µL) and incubated for 5 min at 72 °C. The total volume of this reaction was 20 µL. 3 µL aliquots were mixed with 3 µL of stop-solution for denaturing PAGE analysis (analytical PAGE).

Primer Extension with modified dNTP (c7dGTP):

A solution of the DNA template (40 pmol) and fluorescently labeled primer (20 pmol) was annealed by heating to 95 °C for 2 min and slowly cooling to room temperature. This solution was mixed with 0.1 µL dATP/dCTP/TTP-Mix (10 mM), 1.85 µL c7dGTP (108 µM), 2 µL 10x ThermoPol buffer (200 mM Tris-HCl, 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 20 mM $MgSO_4$, 1% Triton®X-100, pH 8.8) and 0.5 µL Terminator polymerase (2 u/µL) and incubated for 5 min at 72 °C. The total volume of this reaction was 20 µL. The product band was visualized in with the ChemiDoc imager in the Fluorescein channel.

### 5.4.7 PCR

0.26 µL of DNA template (7.57 µM, 2 pmol), 0.14 µL forward primer (73.1 µM, 10 pmol), 0.14 µL reverse primer (73.9 µM, 10 pmol), 0.4 µL dNTP-mix (5 mM), 2 µL 10x DreamTaq buffer and 0.1 µL DreamTaq polymerase (5 u/µL) in a total volume of 20 µL was prepared in a PCR tube. The PCR reaction was performed over seven steps with the following conditions: 95 °C for 4 minutes, 29× (95 °C for 30 sec, 50 °C for 30 sec, 72 °C for 25 sec) and 72 °C for 5 minutes, followed by cooling to 4 °C. An aliquot (3 µL) was mixed with stop solution (3 µL) and analyzed on denaturing PAGE (15%, mini gel). The reaction was ethanol precipitated.

PCR with modified dNTP (c7dGTP):

0.26 µL of DNA template (7.57 µM, 2 pmol), 0.14 µL forward primer (73.1 µM, 10 pmol), 0.14 µL reverse primer (73.9 µM, 10 pmol), 0.4 µL dATP/dCTP/TTP-mix (5 mM), 1.85 µL c7dGTP (108 µM), 2 µL 10x DreamTaq buffer and 0.1 µL DreamTaq polymerase (5 u/µL) in a total volume of 20 µL was prepared in a PCR tube. The PCR reaction was performed over seven steps with the following conditions: 95 °C for 4 minutes, 29× (95 °C for 30 sec, 50 °C for 30 sec, 72 °C for 25 sec) and 72 °C for 5 minutes, followed by cooling to 4 °C. An aliquot (3 µL) was mixed with stop solution (3 µL) and analyzed on denaturing PAGE. The reaction was ethanol precipitated.

### 5.4.8 Demethylation assay with ALKBH1/ALKBH3

A mixture of RNA oligonucleotide (100 pmol in $H_2O$), 0.5 µL10x oxidation buffer (500 mM HEPES pH 7, 50 mM $MgCl_2$, 40 mM ascorbic acid), 0.5 µL α-ketoglutarate (1 mM), 0.5 µL ammonium iron(II) sulfate (1 mM), 0.25 µL Ribolock and 0.77 µL of ALKBH1 or 0.95 µL of ALKBH3 in a total volume of 5 µL were incubated for 1 h at 22 °C. To stop the reaction 1 µL of Proteinase K was added and the reaction was incubated for 30 min at 22 °C. 130 µL $H_2O$ was added and the RNA was extracted once with a premixed phenol/chloroform/isoamyl alcohol solution and once with a chloroform/isoamyl alcohol solution. The supernatant was transferred to a fresh tube and 10 µL of NaOAc (3 M, pH 5.3), 1.0 µL glycogen and 300 µL ethanol were added for ethanol precipitation. The pellet was dissolved in 15.1 µL $H_2O$ and subjected to Anion-exchange chromatography or HR-ESI/MS. The ratio between methylated and demethylated sample was determined by peak-integration.

### 5.4.9 ZipTip$_{C18}$ desalting

The ethanol precipitated sample from primer extension was desalted with Merck ZipTip$^{TM}$ 0.6µL-C18-pipette tips according to the protocol from Merckmillipore “Sample Preparation of Oligonucleotides Prior to MALDI-TOF MS Using ZipTip$_{C18}$ and ZipTip$_{µ\text{-}C18}$ Pipette Tips.”. The tips were equilibrated by prewetting with acetonitrile/Milli-Q water (1:1) followed by washing the tip with 0.1 M triethylammonium acetate (TEAA, pH 7) for three times. Then the oligonucleotide it bound to ZipTip$_{C18}$ by aspirating and dispensing the oligonucleotide solution (10 µL, in 0.1 M TEAA) five to

ten cycles. The tip is washed with 10 µL of 0.1 M TEAA for three times and the solution was dispensed to waste. Subsequently, the tip is washed with 10 µL of Milli-Q water for three times and the solution was dispensed to waste to remove the salt. The pure oligonucleotide was eluted from the $ZipTip_{C18}$ with 10 µL acetonitrile/Milli-Q water (1:1) in three portions.

### 5.4.10 Splint-Ligation

5 nmol of the ASLs (R599, R816 and R825) were 5'-phosphorylated (see 5.4.4) with PNK. 1 nmol of the three RNA fragments were mixed with 1 nmol of the DNA splint and 2 µL of 10x annealing buffer (40 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 8.0). After annealing for 3 min at 95 °C and cooling at ambient temperature for 15 min and at 4 °C for 10 min, 2.5 µL of ligation buffer (400 mM Tris- HCl, 100 mM $MgCl_2$, 100 mM DTT, 5 mM ATP, pH = 7.8) and 2.5 µL of T4-DNA ligase (5 u/µL) were added. The reaction was incubated for 18 h at 37 °C. The Ligation product was separated and isolated by 10% denaturing PAGE.

### 5.4.11 Crosslinker functionalization

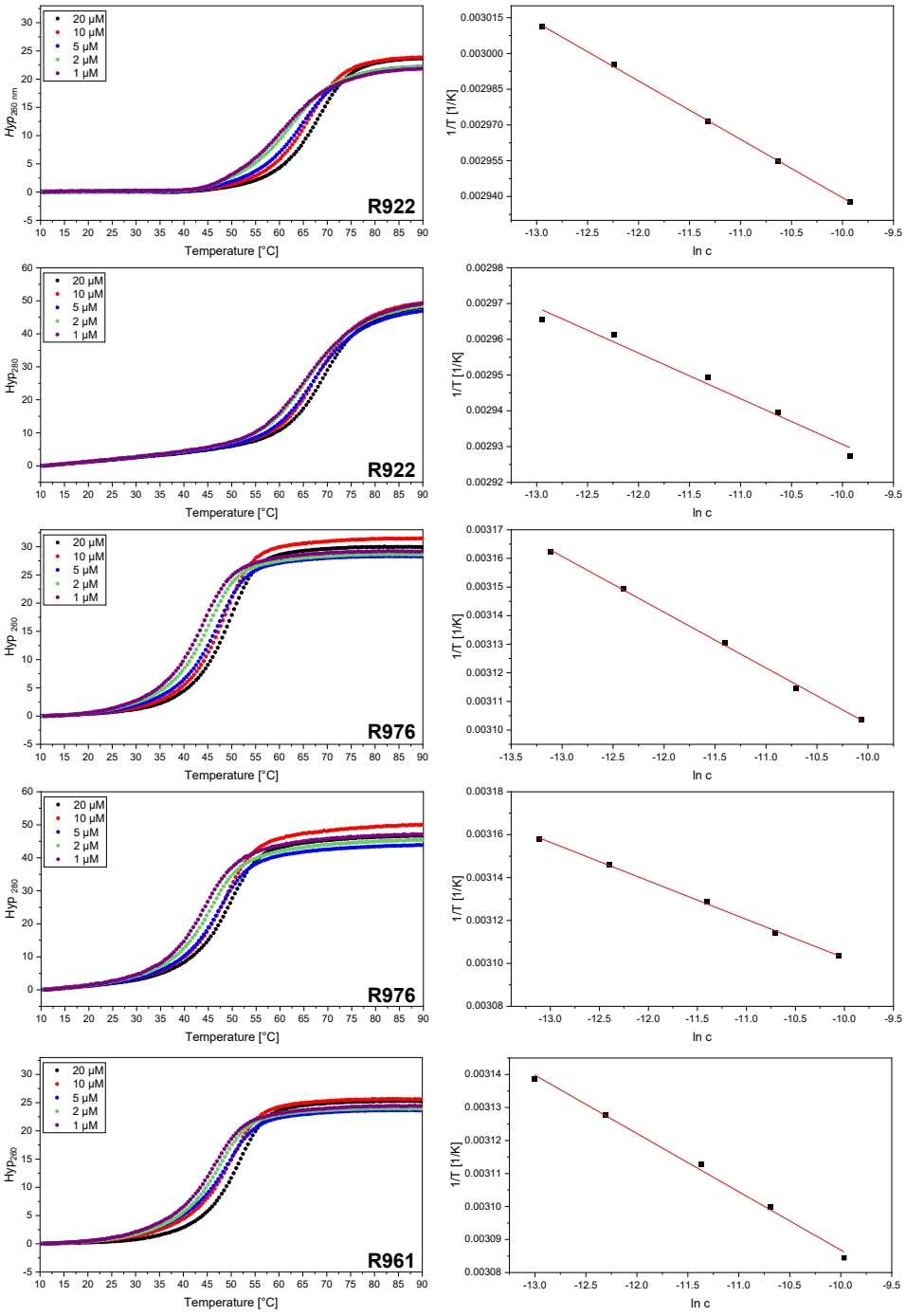
The c7dG-containing (crosslinker) DNA oligonucleotide which was obtained by PCR (chapter 5.4.7) was ethanol precipitated. The pellet was dissolved in 4.2 µL $H_2O$ and mixed with 0.5 µL sodium phosphate buffer (1 M) and 0.3 µL $NaIO_4$-solution (50 mM). The reaction was incubated for 1 h at 37 °C in the dark. After ethanol precipitation the pellet was dissolved in 4 µL $H_2O$ and mixed with 0.5 µL sodium cacodylate (250 mM) and 0.5 µL Flu-TSC (3 mM) and incubated for 3 h at 37 °C in the dark. The reaction was ethanol precipitated and pellet was dissolved in 6 µL stop solution and analyzed by denaturing PAGE. The product band was visualized in with the ChemiDoc imager in the Fluorescein channel and subsequently stained by SYBR gold to visualize the other, non-labeled, bands.

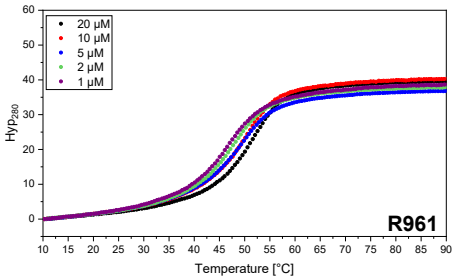
# 6. Appendix

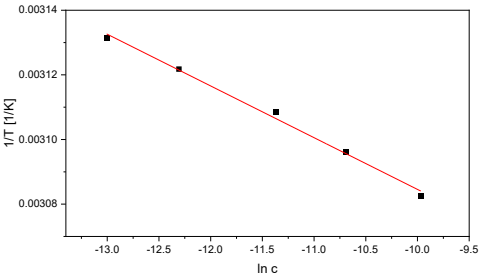
## 6.1 Supplementary melting curves

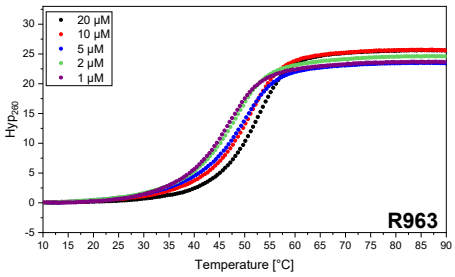
Melting curves were measured with 20, 10, 5, 2, and 1 µM samples in 10 mM sodium phosphate buffer (pH 7) containing 100 mM NaCl. The hyperchromicity of the third ramp was plotted against the temperature. For the duplex forming sequences the linear fits of the van't Hoff plot, which was used for the determination of the thermodynamic parameters, are shown.

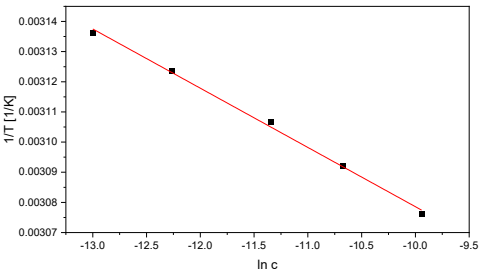
### 6.1.1 Dickerson Drew - duplex

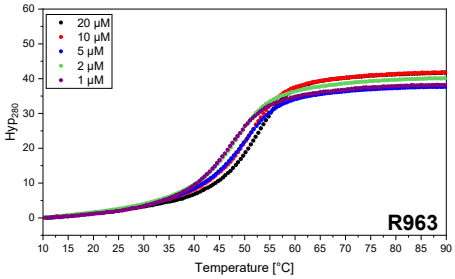


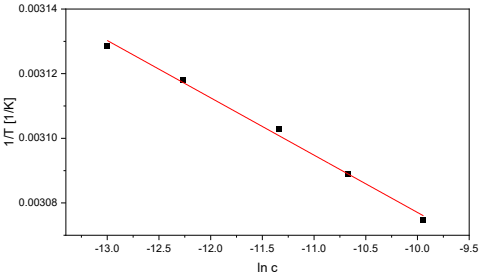

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp260
Temperature [°C]
R961

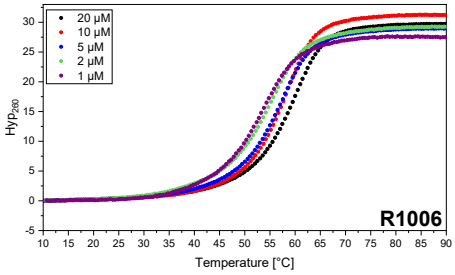

1/T [1/K]
ln c

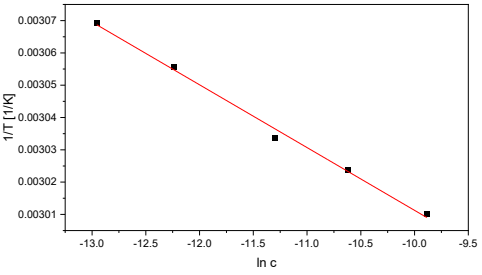

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp260
Temperature [°C]
R963

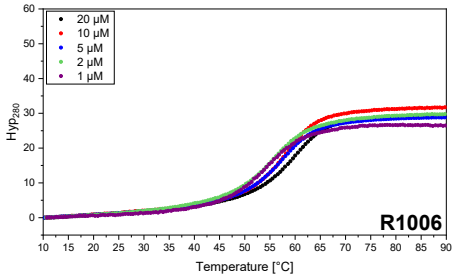

1/T [1/K]
ln c

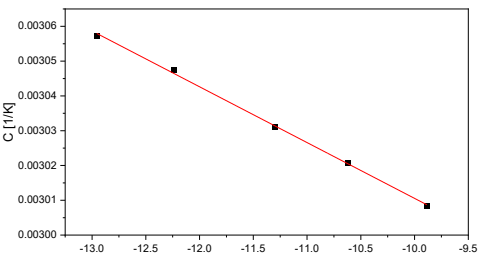

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp260
Temperature [°C]
R963

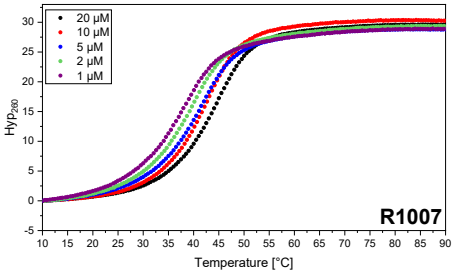

1/T [1/K]
ln c

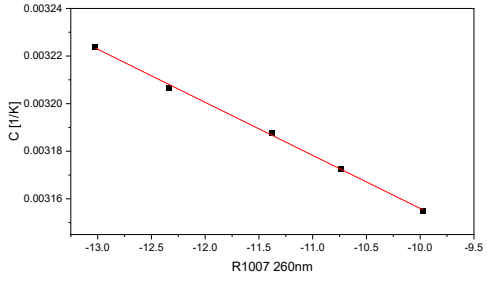

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp260
Temperature [°C]
R1006

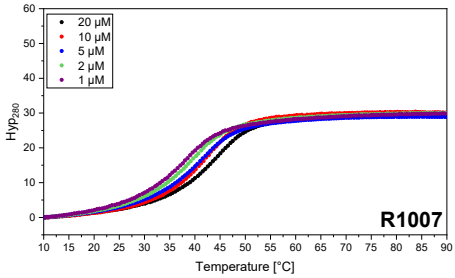

1/T [1/K]
ln c

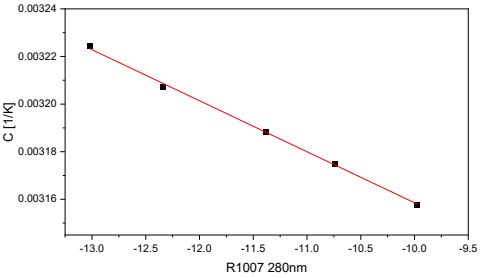

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp260
Temperature [°C]
R1006


C [1/K]

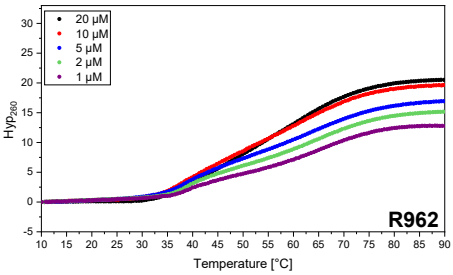

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp260
Temperature [°C]
R1007

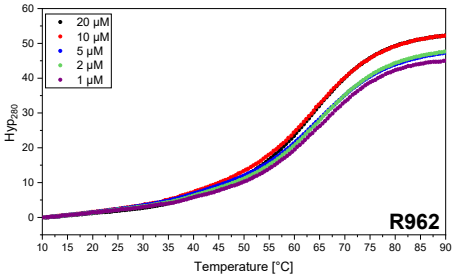

C [1/K]
R1007 260nm

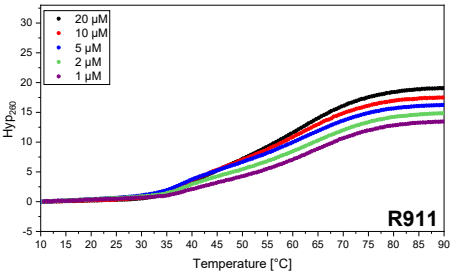


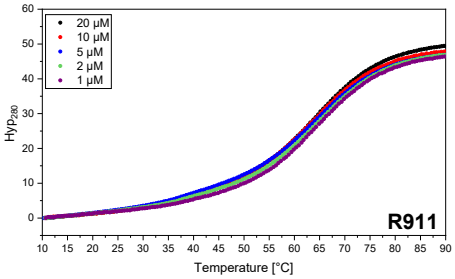



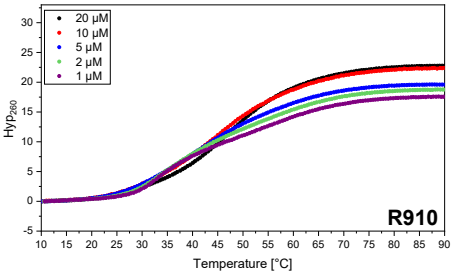

## 6.1.2 Dickerson Drew - hairpin

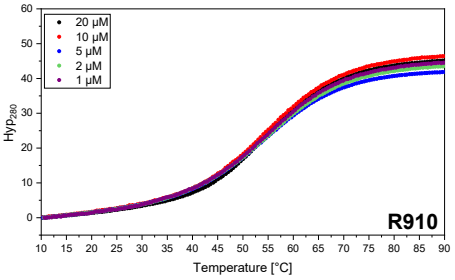


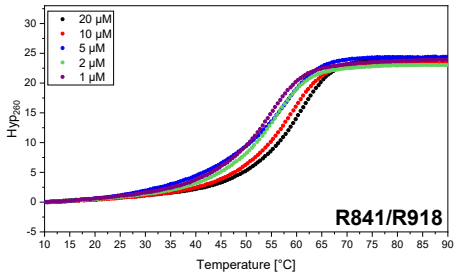



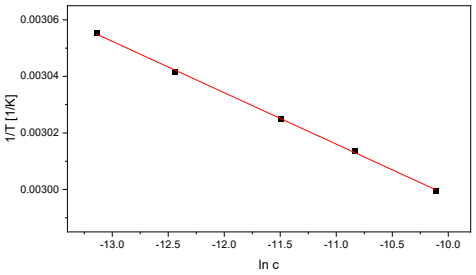



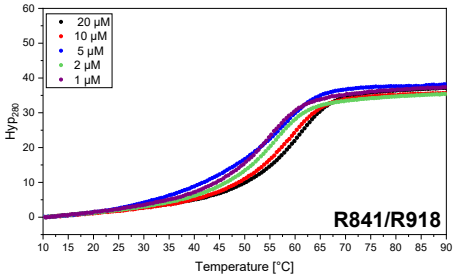



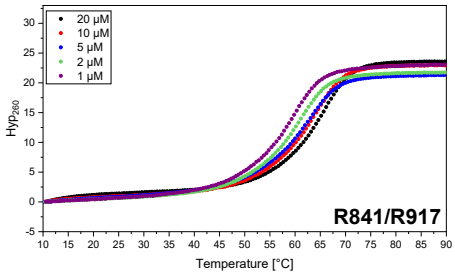

## 6.1.3 NHC in another sequence context.

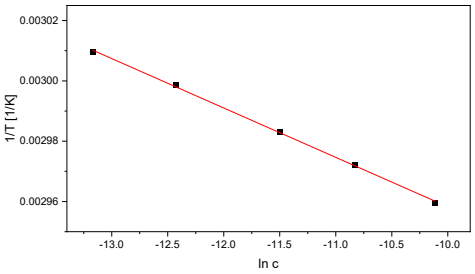
For this sequence context, the van't Hoff plot was only determined at 260 nm.

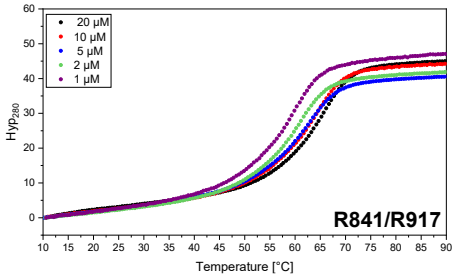


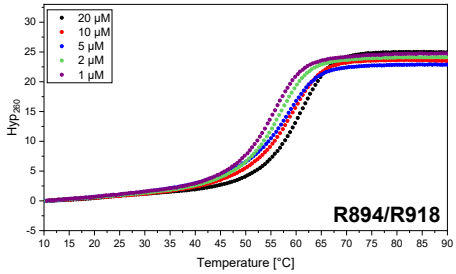



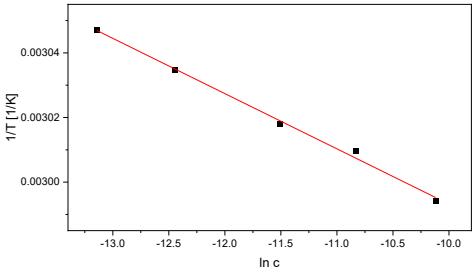

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R841/R918

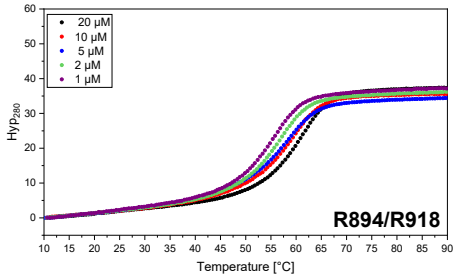

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R841/R917

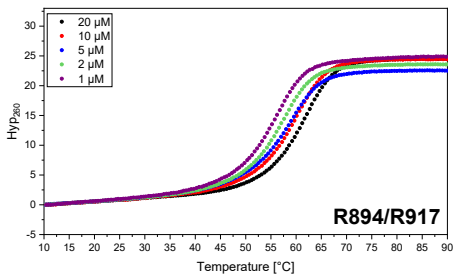

1/T [1/K]
ln c

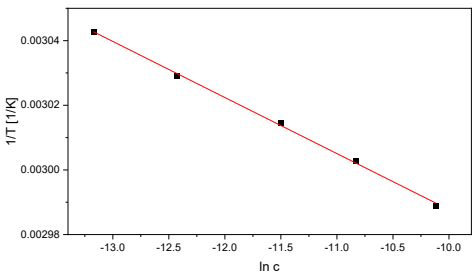

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R841/R917

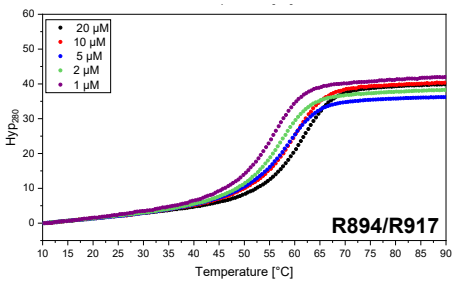

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R894/R918

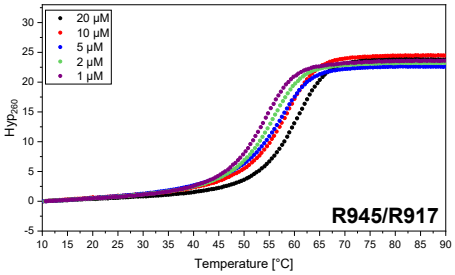

1/T [1/K]
ln c

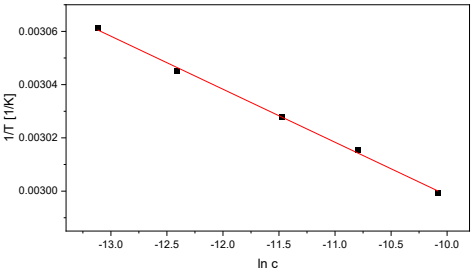

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R894/R918

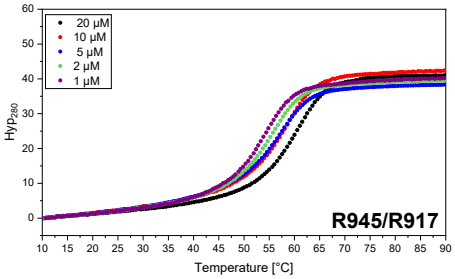

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R894/R917

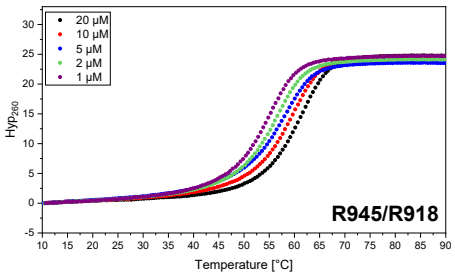

1/T [1/K]
ln c

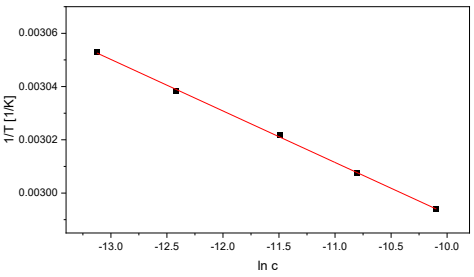

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R894/R917

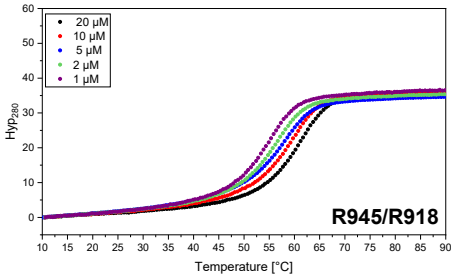

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R945/R917

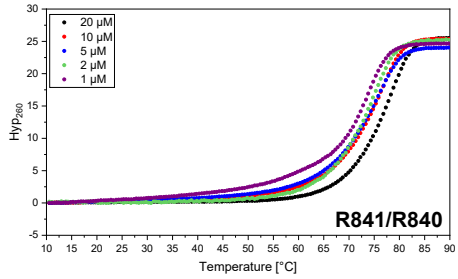

1/T [1/K]
ln c

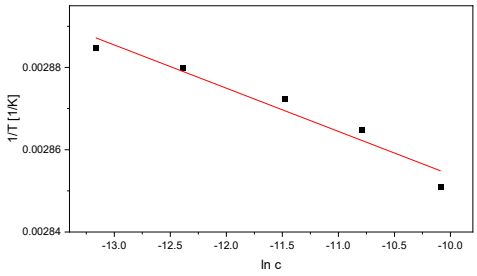

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R945/R917

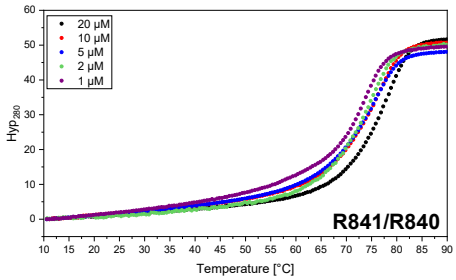

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R945/R918

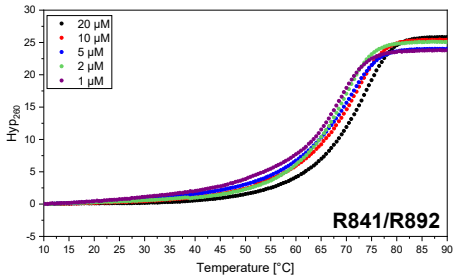

1/T [1/K]
ln c

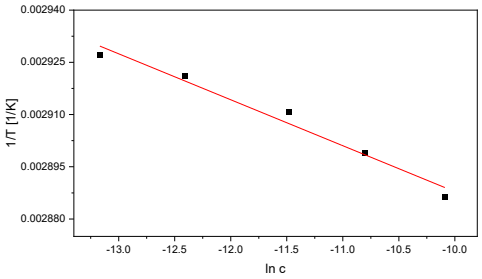

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R945/R918

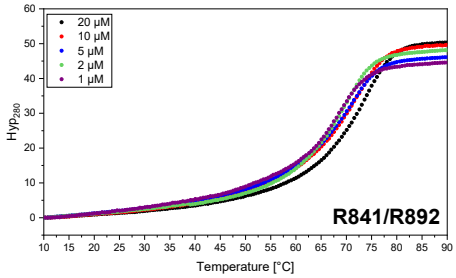

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R841/R840

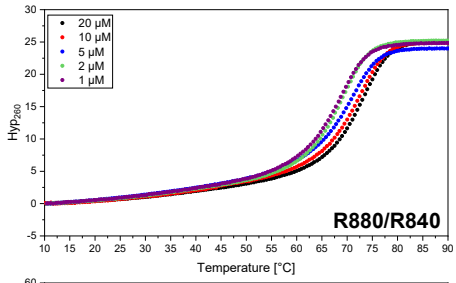

1/T [1/K]
ln c

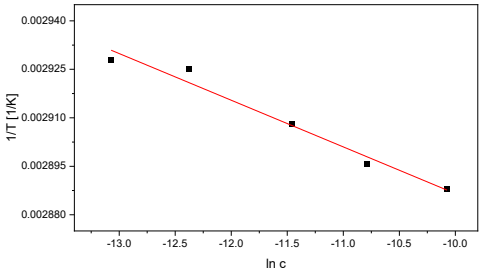

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R841/R840

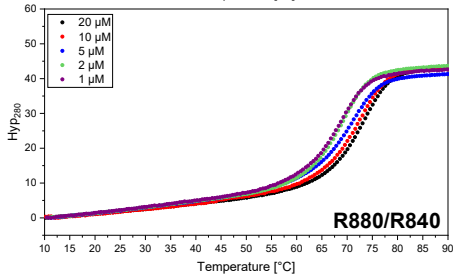

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R841/R892

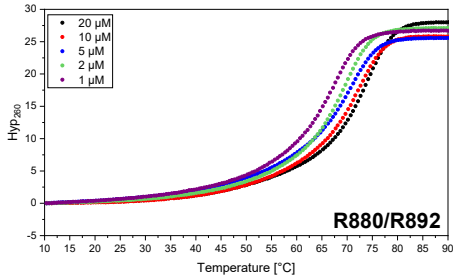

1/T [1/K]
ln c

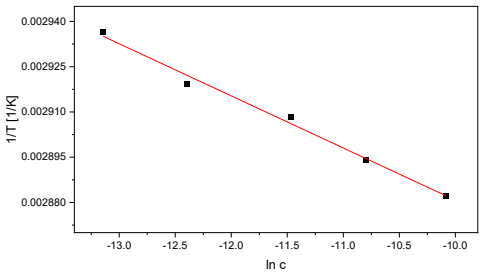

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R841/R892

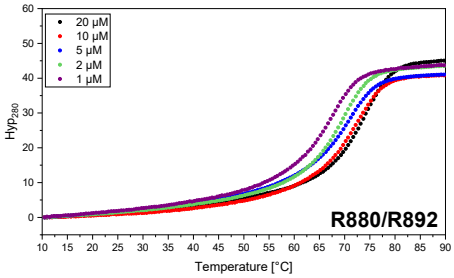

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R880/R840

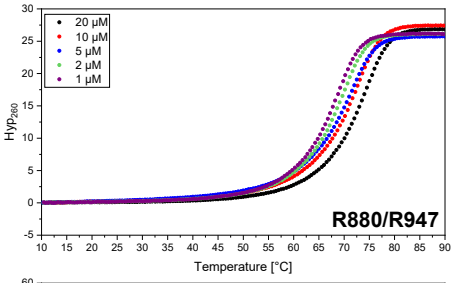

1/T [1/K]
ln c

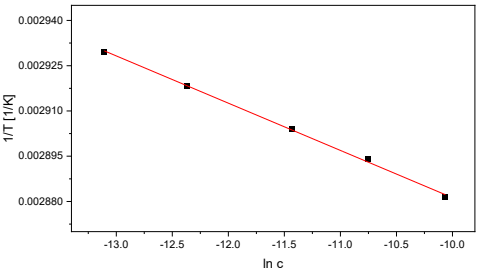

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R880/R840

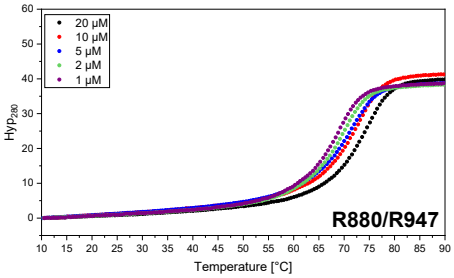

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R880/R892

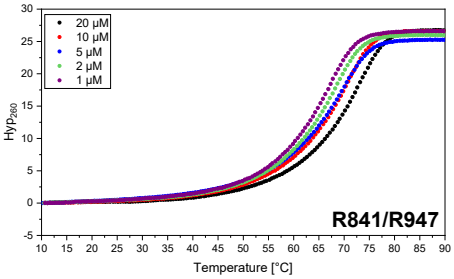

1/T [1/K]
ln c

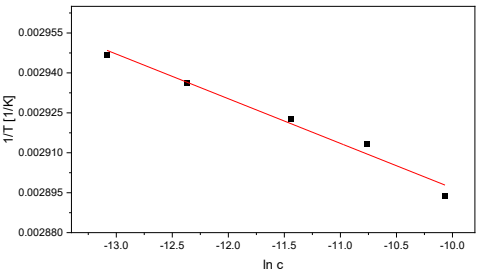

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R880/R892

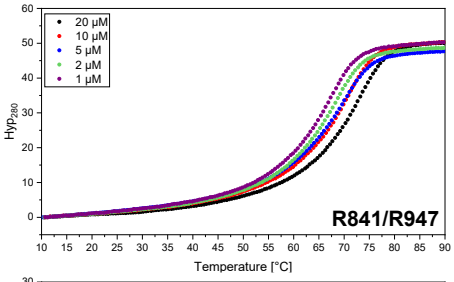

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R880/R947

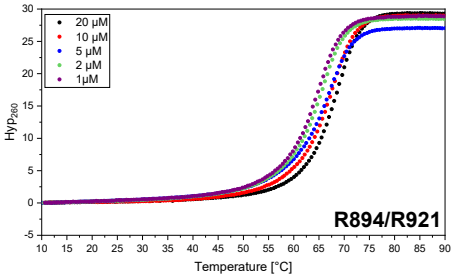

1/T [1/K]
ln c

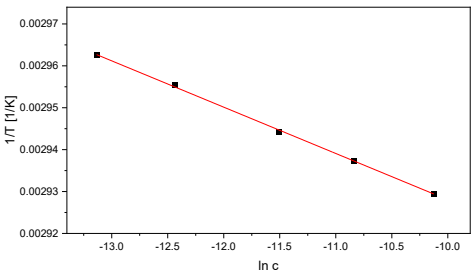

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R880/R947

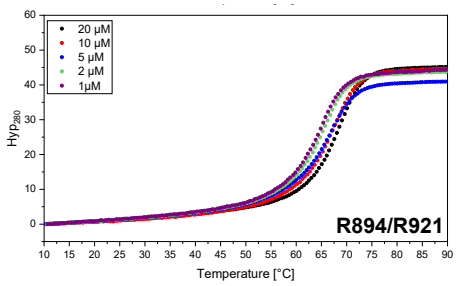

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R841/R947

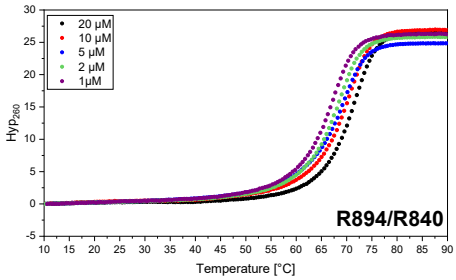

1/T [1/K]
ln c

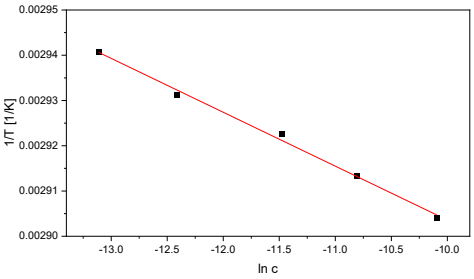

20 µM
10 µM
5 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R841/R947

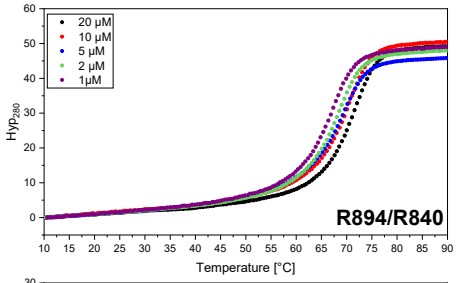

20 µM
10 µM
5 µM
2 µM
1µM
Hyp280
Temperature [°C]
R894/R921

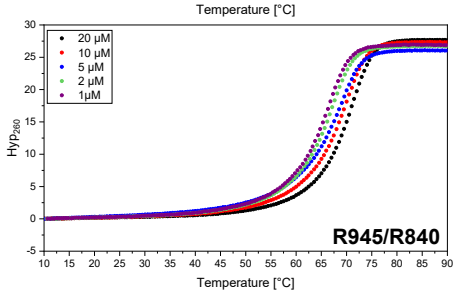

1/T [1/K]
ln c

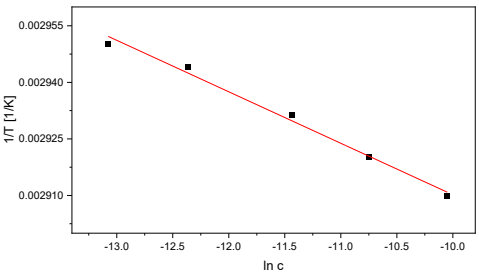

20 µM
10 µM
5 µM
2 µM
1µM
Hyp280
Temperature [°C]
R894/R921

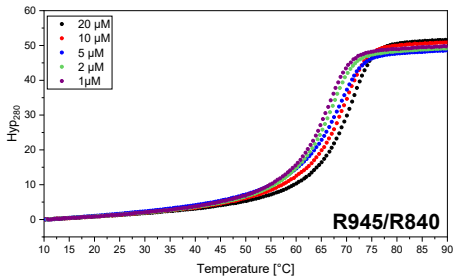

20 µM
10 µM
5 µM
2 µM
1µM
Hyp280
Temperature [°C]
R894/R840

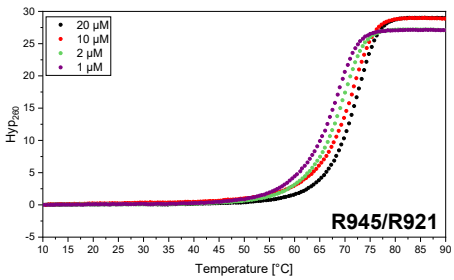

1/T [1/K]
ln c

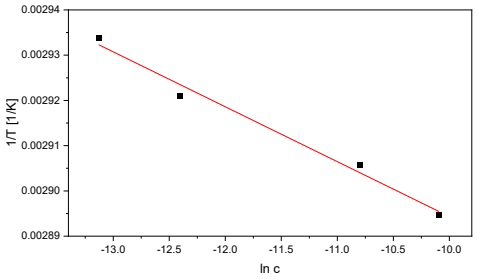

20 µM
10 µM
5 µM
2 µM
1µM
Hyp280
Temperature [°C]
R894/R840

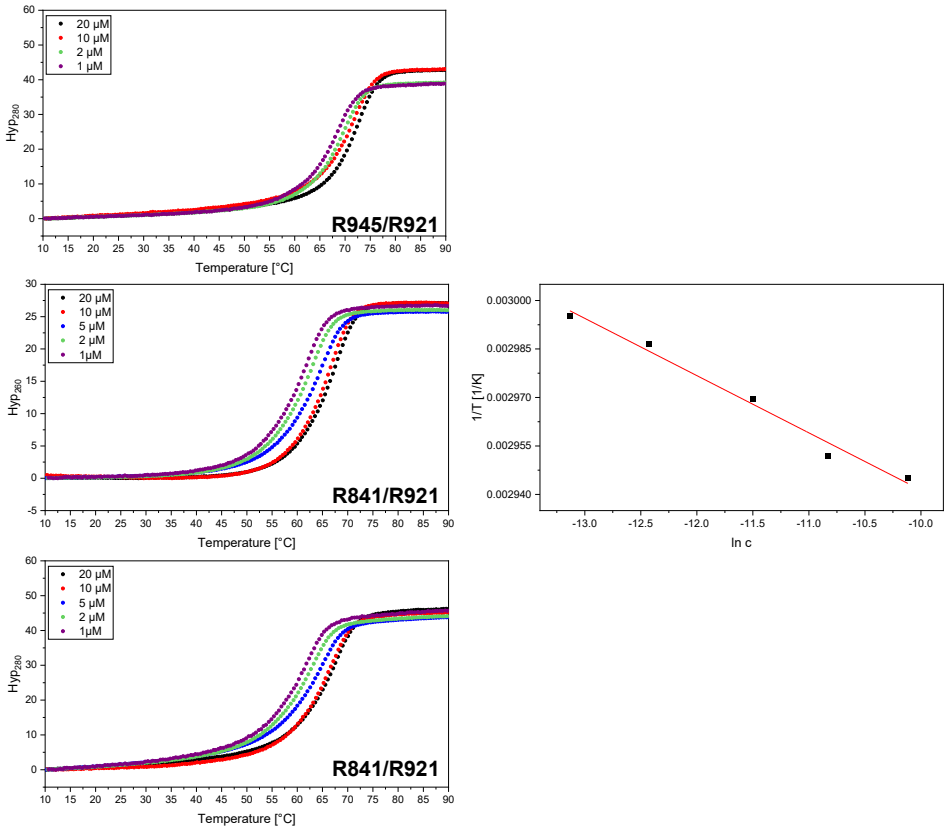

20 µM
10 µM
5 µM
2 µM
1µM
Hyp280
Temperature [°C]
R945/R840


1/T [1/K]
ln c


20 µM
10 µM
5 µM
2 µM
1µM
Hyp280
Temperature [°C]
R945/R840


20 µM
10 µM
2 µM
1 µM
Hyp280
Temperature [°C]
R945/R921

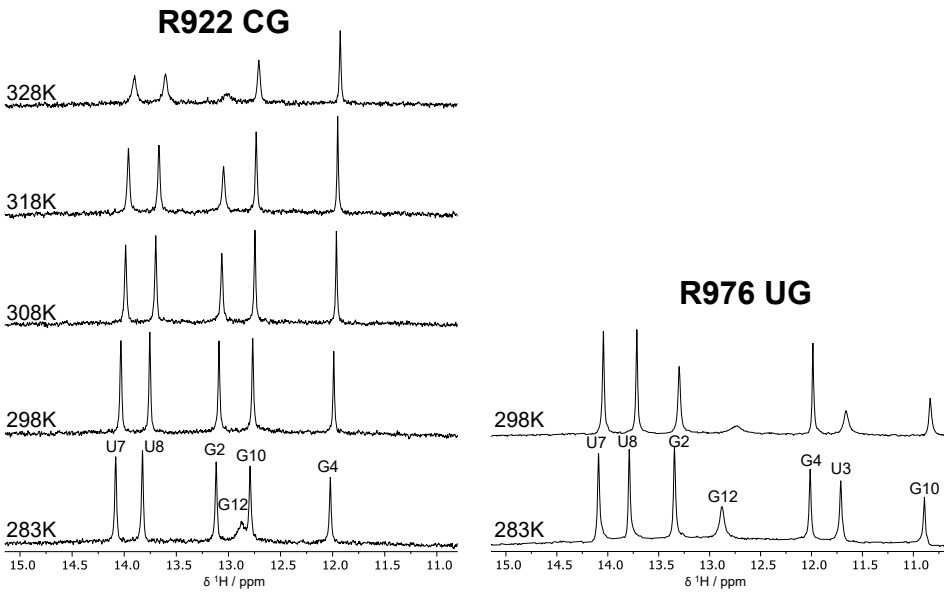

1/T [1/K]
ln c

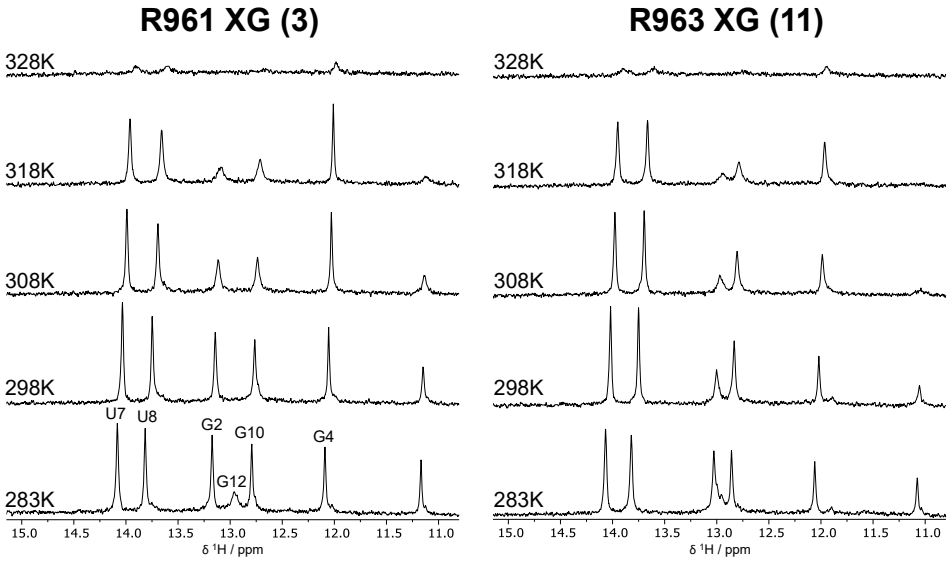



## 6.2 Supplementary NMR data

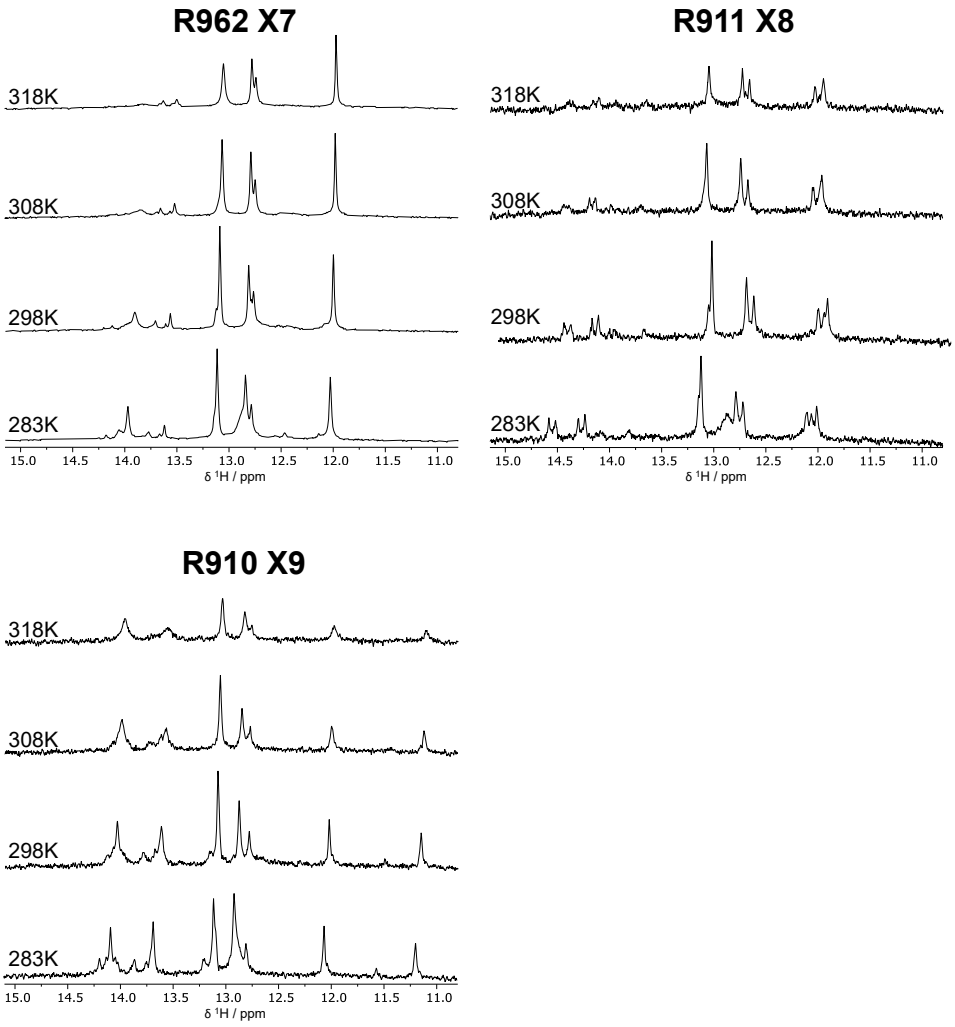
### 6.2.1 Dickerson Drew – duplex

$^{1}$H-imino region of the Dickerson Drew sequences at up to five different temperatures (283K, 298K, 308K, 318K and 328K). The assignment for some sequences is indicated by the numbers.

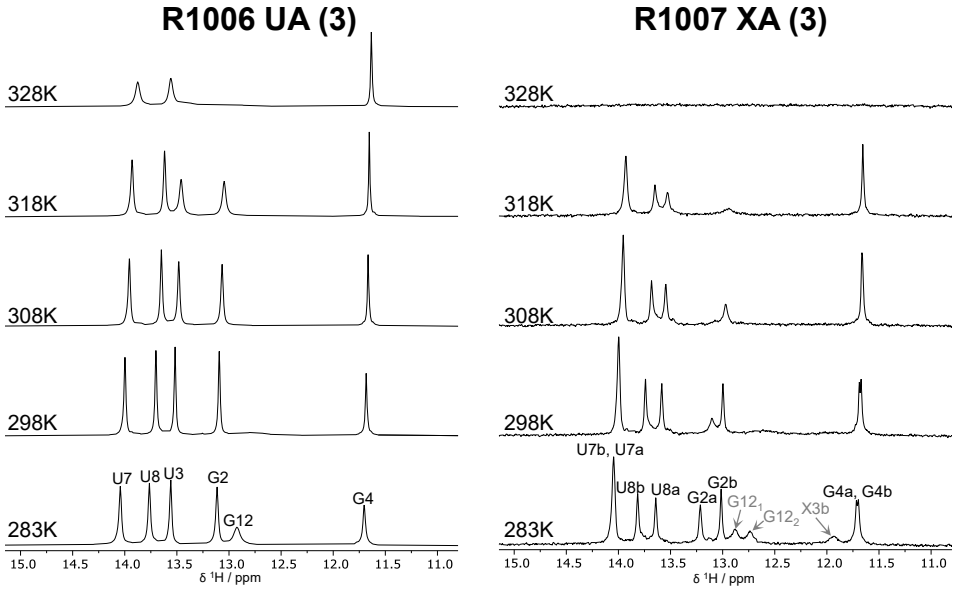




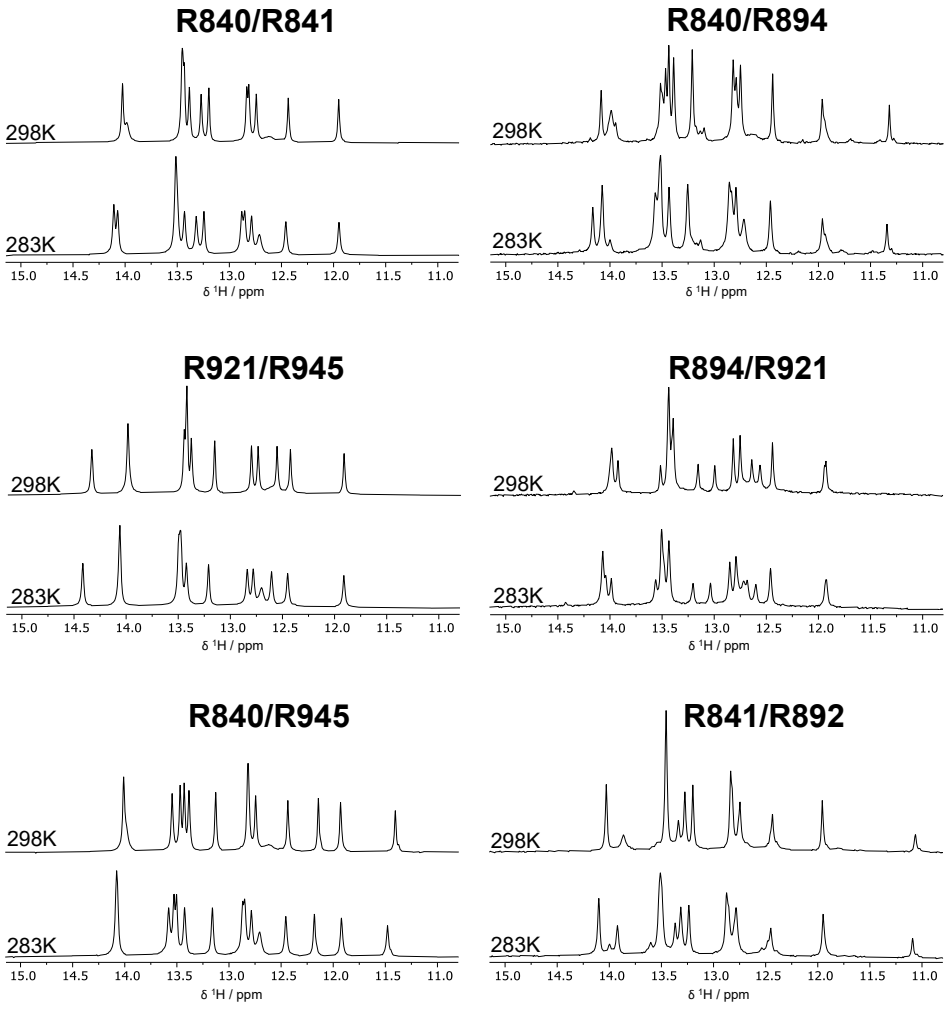

## 6.2.2 Dickerson Drew – hairpin

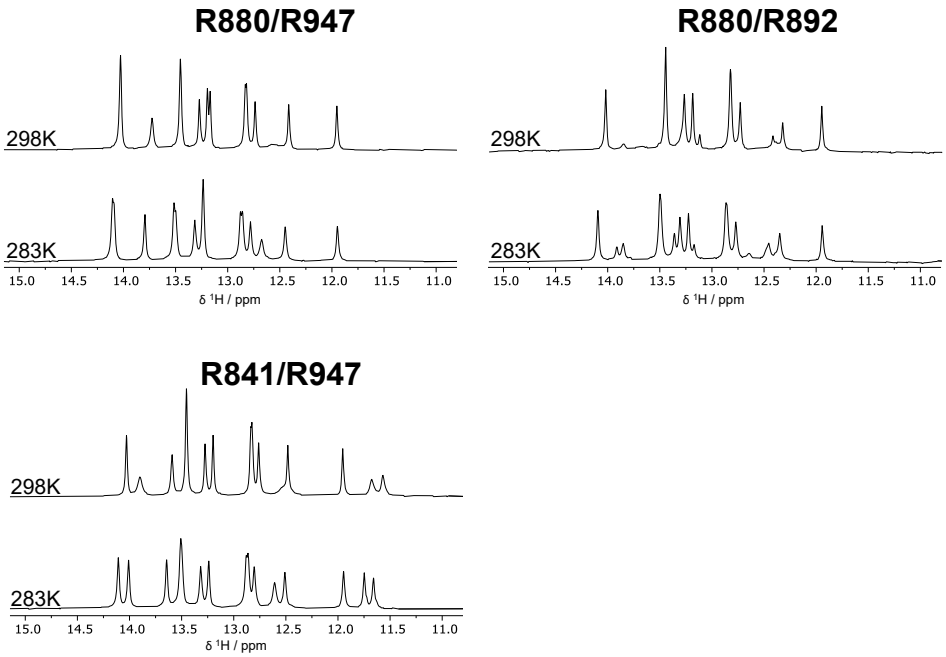


### 6.2.3 Dickerson Drew – duplex ($^{15}$N)




### 6.2.4 NHC in another sequence context

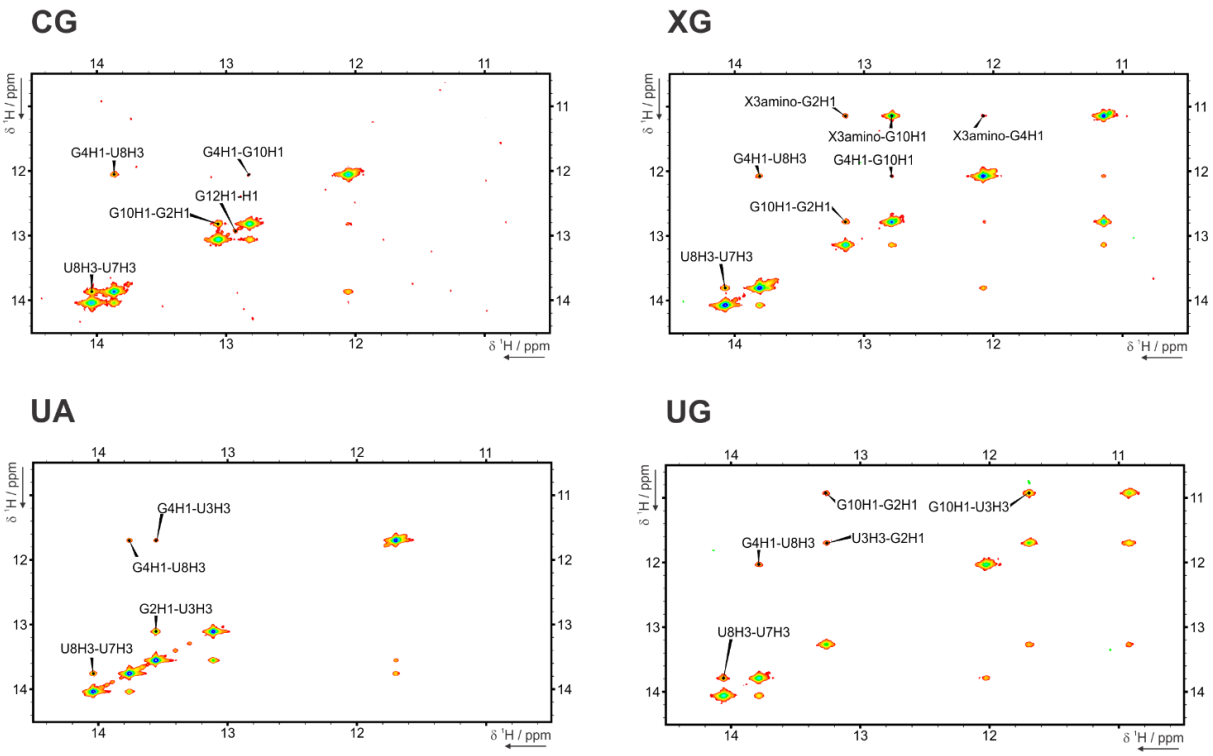




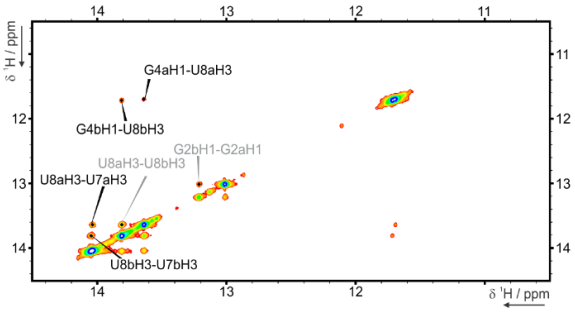

### 6.2.5 2D-spectra of the Dickerson Drew Sequences

The 2D-spectra were measured and assigned by Dr. Irene Bessi.

NOESY of R922 (CG), R961 (XG), R1006 (UA), R976 (UG) and R1007 (XA) (imino region, 283K, 600 MHz):

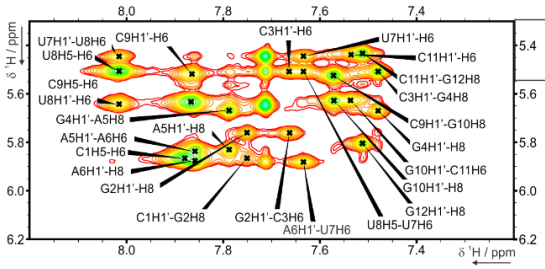


XA

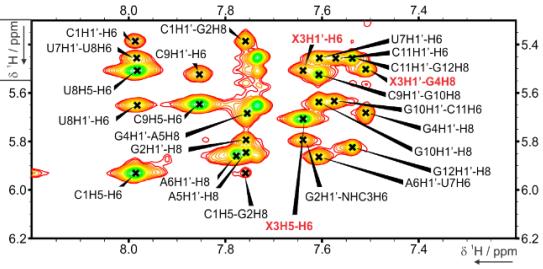



NOESY of R922 (CG), R961 (XG), R1006 (UA), R976 (UG) and R1007 (XA) (Aromatic-anomeric region, 283K, 600 MHz):

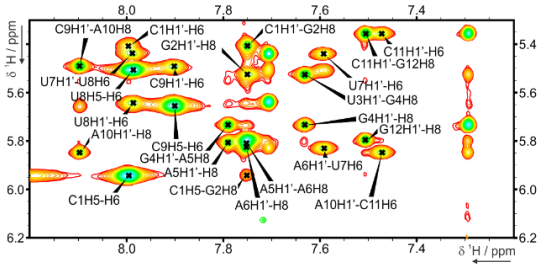
CG



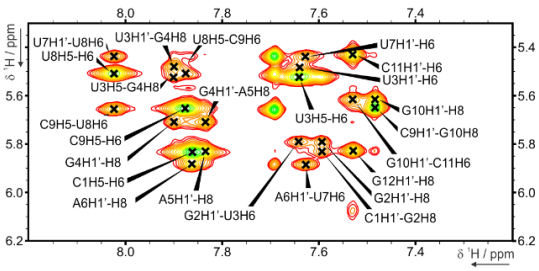

XG



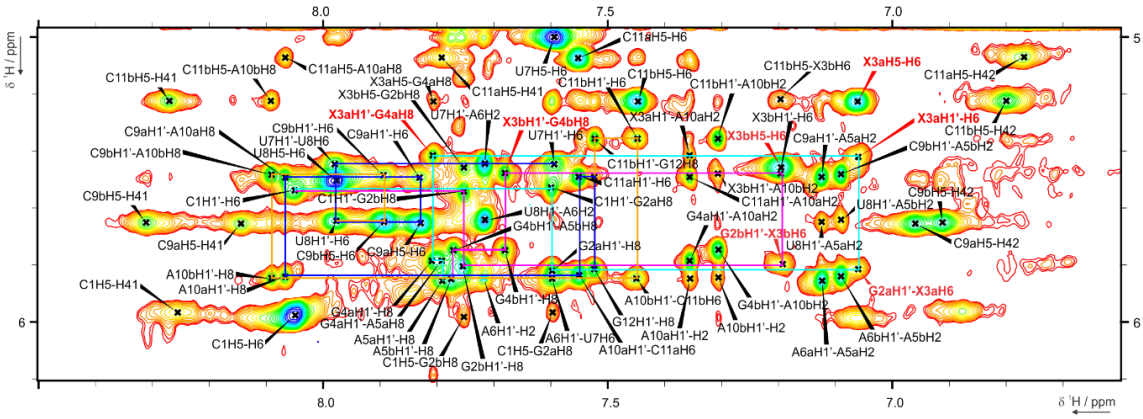

UA



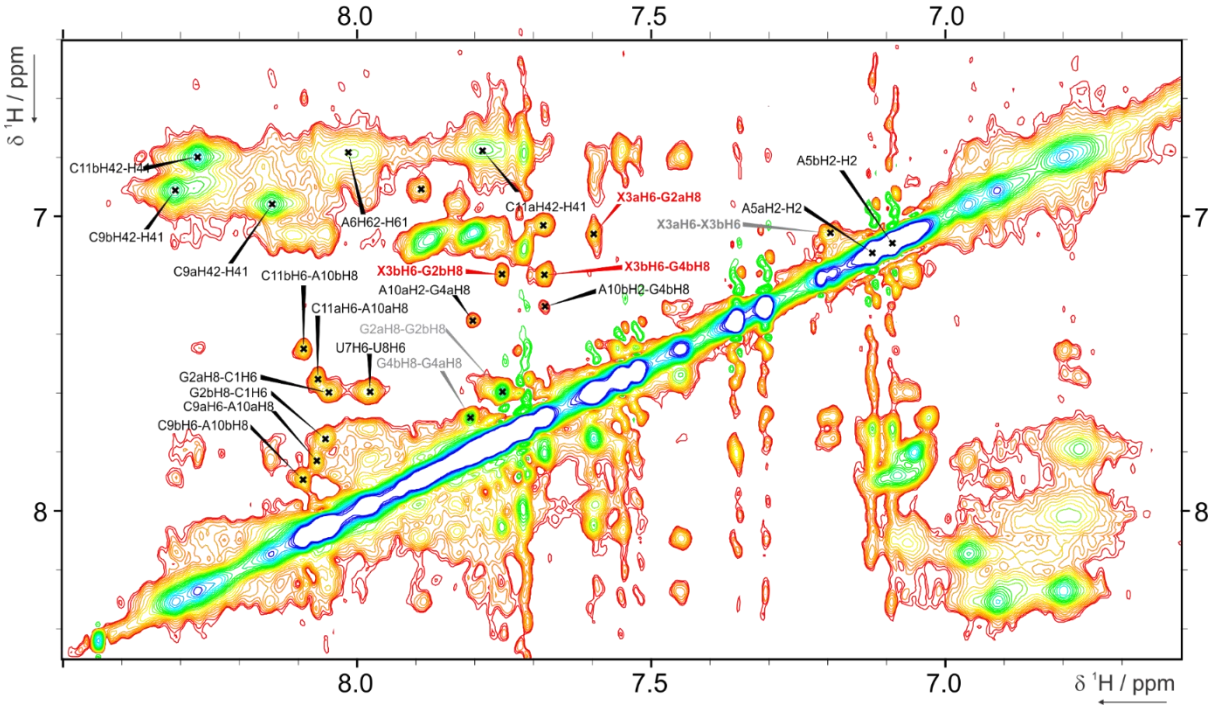

UG



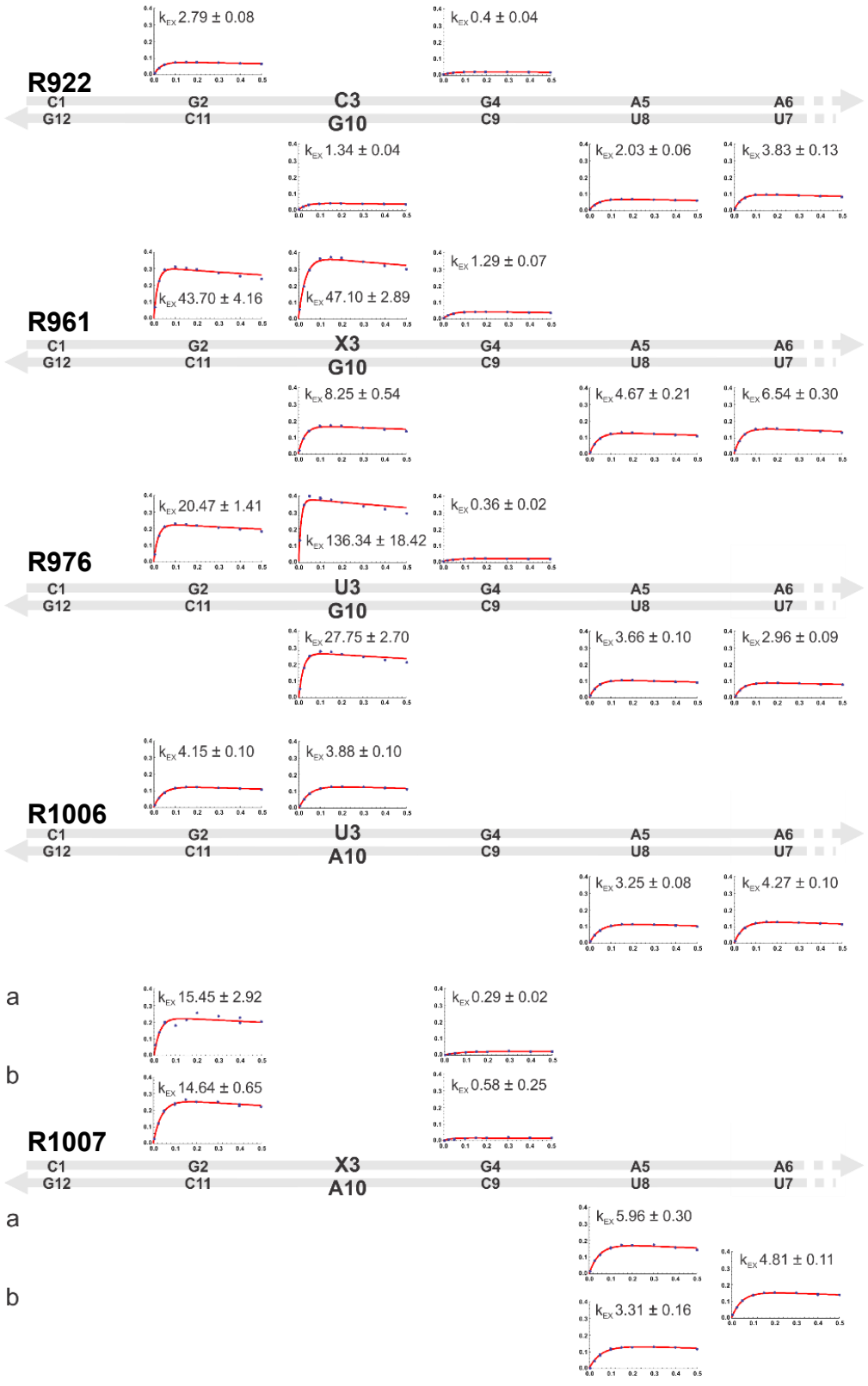

XA



NOESY (Aromatic region of **R1007** (XA), 283K, 600 MHz, exchange peaks in grey):

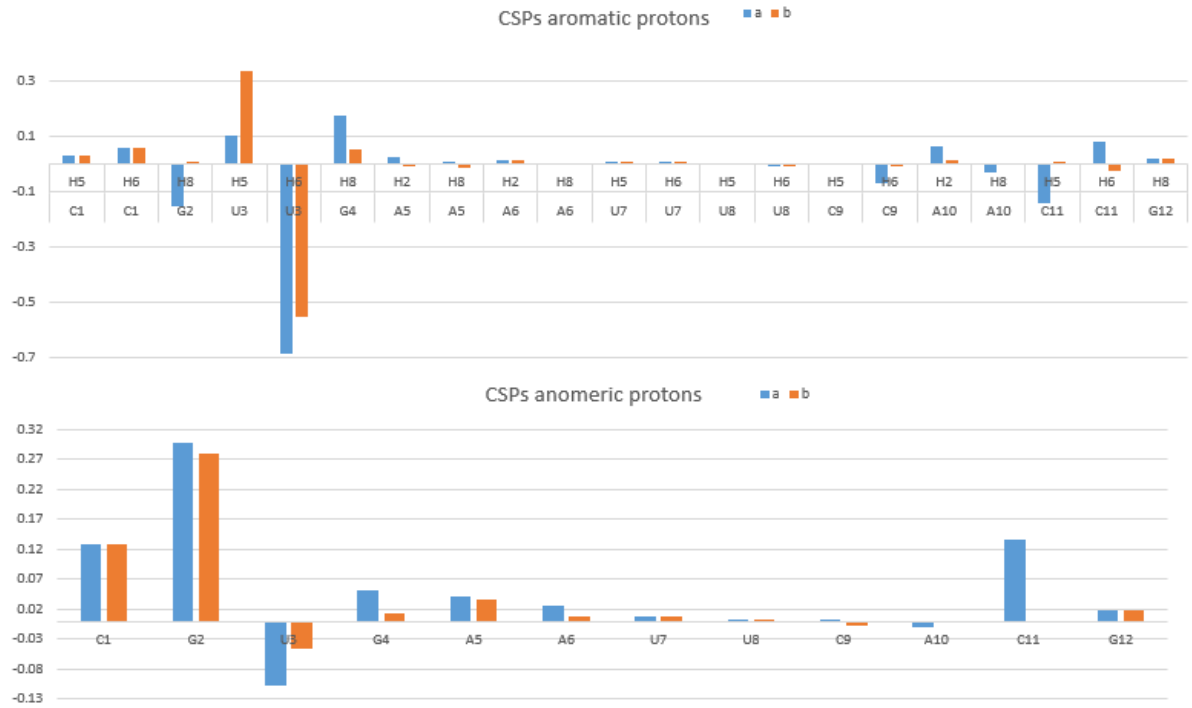


## 6.2.6 Imino-water exchange rates



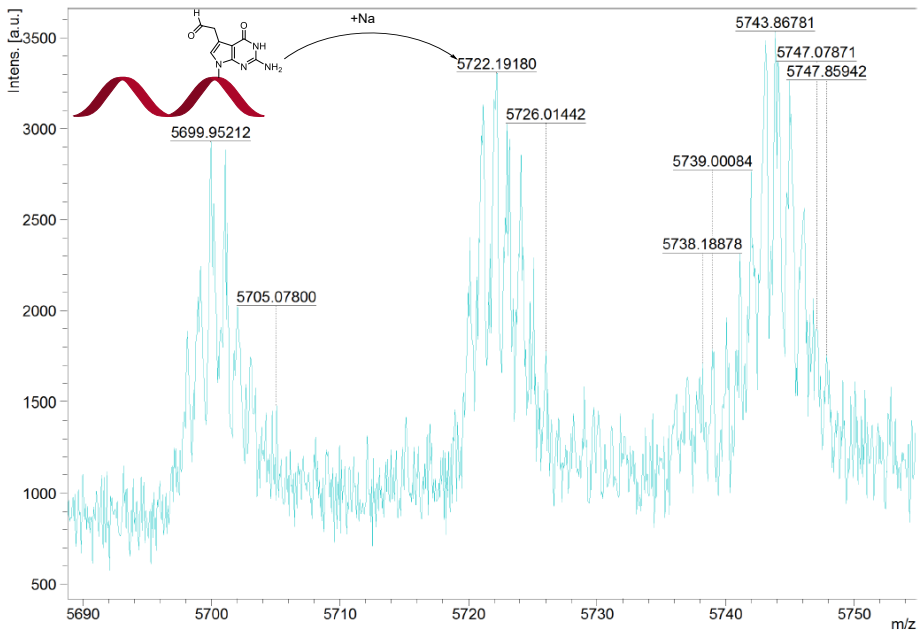
### 6.2.7 Chemical shift perturbation (CSP)

CSP = δ (R**1007**) – δ (**R1006**). The two different conformations of R1007 are indicated by ‘a’ and ‘b’.

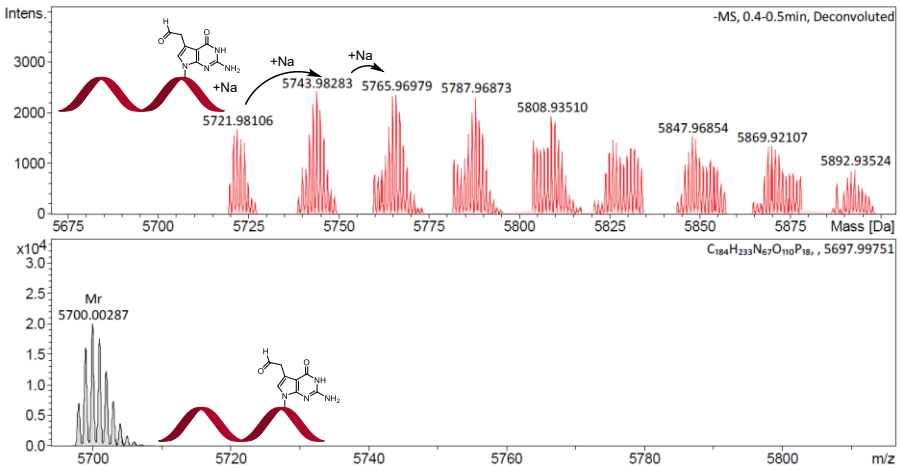



## 6.3 Crosslinker

The diol function of the 7-deazacrosslinker was oxidized with $NaIO_4$ and analyzed by MALDI (chapter 3.3.2; Alk-GATGTGCAGCTXCACATC, X=c7dG-CL). The aldehyde-modified oligonucleotide exhibits a mass of 5697.9 g/mol whereas the starting material would have a mass of 5730.02 g/mol.



The same sample was also analyzed by ESI-MS:




## 6.4 Abbreviations

| | |
|---|---|
| α-KG | α-ketoglutarate |
| 2-AP | 2-aminopurine |
| aaRS | aminoacyl tRNA-synthetases |
| aa-tRNA | aminoacyl-tRNA |
| ABN | C-linked 8-(diethylamino)benzo[b][1,8] naphthyridin-2(1*H*)-one nucleoside |
| Abs | absorption |
| Ac | Acetyl |
| ACE2 | angiotensin converting enzyme 2 |
| ACL | anticodon loop |
| AP | apurinic/apyrimidinic |
| APS | ammonium peroxodisulfate |
| ASL | anticodon stem loop |
| BOP | benzotriazolyloxytris(dimethylamino)phosphoniumhexafluorophosphate |
| Bz | benzoyl |
| $ca^5C$ | 5-carboxylcytosine |
| CD | circular dichroism |
| CENU's | chloroethylnitrosoureas |
| CEP-Cl | 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite |
| $cm^5U$ | 5-carboxymethyluridine |
| COSY | correlation spectroscopy |
| CPG | controlled pore glass |

| | |
|---|---|
| CPM | count per minute |
| cryo-EM | Cryogenic electron microscope |
| CSP | chemical shift perturbation |
| $ct^6A$ | cyclic $t^6A$ |
| CuAAC | copper-catalyzed azide-alkyne cycloaddition |
| cv | column columns |
| cyt-tRNA | cytoplasmic tRNAs |
| 1,2-DCE | 1,2-dichloroethane |
| DAAs | direct-acting antivirals |
| d | doublet |
| dA | deoxyadenosine |
| DAC | diammonium hydrogen citrate |
| DALRD3 | DALR anticodon binding domain containing 3 |
| DAPI | 4',6-diaminidino-2-phenylindole |
| dC | deoxycytosine |
| DCC | *N,N'*-dicyclohexylcarbodiimide |
| DCE | 1,2-dichloroethane |
| DCM | dichloromethane |
| DCI | 4,5-dicyanoimidazole |
| DD | Dickerson Drew |
| ddhCTP | 3'-deoxy-3′,4'-didehydro-CTP |
| dG | deoxyguanosine |
| DIPEA | *N,N*-di-*iso*-propylethylamine |
| DMAP | 4-(dimethylamino)pyridine |
| DMF-DMA | dimethylformamide dimethyacetal |
| DMSO | dimethylsulfoxide |
| DMT-Cl | 4,4'-dimethoxytritylchloride |
| DNA | deoxyribonucleic acids |
| dsDNA | double-stranded DNA |
| dPTP | Pyrene-triphosphate |
| dsDNA | double-stranded DNA |
| DSS | trimethylsilylpropanesulfonate |
| DTBS-ditriflat | Di-*tert*-butylsilyl-bis(trifluormethansulfonat) |
| E | envelope protein |

| | |
|---|---|
| *E.coli* | *Escherichia coli* |
| EDTA | ethylenediaminetetraacetic acid |
| ESI | Electrospray Ionization |
| ESI-MS | electrospray ionization mass spectrometry |
| EtBr | ethidium bromide |
| ETT | 5-ethylthio-1*H*-tetrazole |
| $f^5C$ | 5-formylcytosine |
| FAM-TSC | fluorescein-5-thiosemicarbazide |
| FTO | fat mass and obesity-associated |
| HAC-seq | Hydrazine-Aniline Cleavage Sequencing |
| HEPES | 2-(4-(2-hydroxyethyl)piperazin-1-yl)ethanesulfonic acid |
| HFIP | hexafluoro-*iso*-propanol |
| HMBC | heteronuclear multiple bond correlation |
| $hm^5C_m$ | 5-hydroxymethyl-2'-*O*-methylcytidine |
| $hm^5C$ | 5-hydroxymethylcytidine |
| HPLC | high-performance liquid chromatography |
| HR-ESI-MS | High-resolution electrospray ionization mass spectrometry |
| HSQC | heteronuclear single quantum coherence spectroscopy |
| Hyp | hyperchromicity |
| I | inosine |
| $i^6A$ | $N^6$-isopentenyl adenosine |
| lnRNA | long noncoding RNAs |
| IFN-I1 | Type I interferon |
| ISCs | interstrand crosslinks |
| 6-MI | 6-methyl isoxanthopterin |
| M | membrane glycoprotein |
| $m^6A$ | $N^6$-methyladenosine |
| $m^{6,6}A$ | $N^6$,$N^6$-dimethyladenosine |
| $m^4C$ | $N^4$-methylcytidine |
| $m^5C$ | 5-methylcytosine |
| $m^6v^8A$ | $N^6$-methyl-8-vinyladenosine |
| MBP | Maltose-binding protein |
| $mchm^5U$ | 5-methoxycarbonylhydroxymethyluridine |
| $mcm^5U$ | 5-methoxycarbonylmethyluridine |

| | |
|---|---|
| *m*CPBA | *meta*-chloroperoxybenzoic acid |
| MERS-CoV | Middle East respiratory syndrome coronavirus |
| METTL | methyltransferase-like |
| $m^7G$ | 7-methylguanosine |
| $m_2^2G$ | $N^2$,$N^2$-dimethylguanosine |
| mitoRNA | mitochondrial RNA |
| mRNA | messenger RNA |
| $ms^2i^6A$ | 2-methylthio-$N^6$-isopentenyl adenosine |
| MTS | mitochondrial targeting sequence |
| $NaIO_4$ | sodium periodate |
| NGS | next-generation sequencing |
| NiRAN | nidovirus RdRp-associated nucleotidyltransferase |
| NIS | *N*-iodosuccinimide |
| NMMO | *N*-methylmorpholin-*N*-oxid |
| NMR | nuclear magnetic resonance |
| NOESY | nuclear Overhauser enhancement spectroscopy |
| npe | p-nitrophenyl ester |
| npe-OH | 2-(4-nitrophenyl)ethanol |
| nsp | nonstructural protein |
| nt | nucleotides |
| NTP | nucleoside triphosphate |
| ONT | Oxford Nanopore Technologies |
| ORD | optical rotatory dispersion |
| ORFs | open reading frames |
| $OsO_4$ | osmium tetroxide |
| PAGE | polyacrylamide gel electrophoresis |
| PCI | phenol:chloroform:isoamyl |
| $Pd(PPh_3)_4$ | tetrakis(triphenylphosphine)palladium |
| PET | photo-induced charge transfer |
| PFPE | perfluoropolyether |
| PMT | photomultiplier tube |
| RdRp | RNA-dependent RNA polymerase |
| RNA | ribonucleic acid |
| rNTP | ribonucleotide triphosphate |

| | |
|---|---|
| RP | reversed phase and reverse primer |
| RRM | RNA recognition motif |
| RT | reverse transcription |
| s | singlet |
| S | S protein |
| SAM | *S*-adenosylmethionine |
| SARS | seryl-tRNA synthetase |
| SARS-CoV-2 | severe acute respiratory syndrome coronavirus 2 |
| *S. cerevisiae* | *Saccharomyces cerevisiae* |
| SMC | structural maintenance of chromosomes |
| snorRNA | small nucleolar RNA |
| snRNA | small nuclear |
| S. pombe | *Schizosaccharomyces pombe* |
| SPS | solvent purification system |
| ssDNA | single-stranded DNA |
| sub | substitution |
| t | triplet |
| $t^6A$ | $N^6$-threonylcarbamyl adenosine |
| TBAF | tetra-*n*-butylammonium fluoride |
| TBDMS | t*ert*-butyldimethylsilyl |
| TBE | TRIS/borate/EDTA |
| *T. brucei* | *Trypanosoma brucei* |
| TDA-1 | Tris[2-(2-methoxyethoxy)ethyl]amine |
| TEAA | triethylammonium acetate |
| TEAB | triethylammonium bicarbonate |
| TEMED | *N*, *N*, *N*′, *N*′-tetramethylethylenediamine |
| TEN | Tris/EDTA/NaCl |
| TGS | long-read single molecule sequencing |
| $^{th}A$ | thieno-A |
| THF | tetrahydrofuran |
| $^{th}I$ | thieno-Inosine |
| thieno-U | thiophene expanded uridine derivative |
| TLC | thin layer chromatography |
| $T_m$ | melting temperature |

| | |
|---|---|
| TMSE-OH | 2-trimethylsilylethanol |
| TOM-Cl | (Triisopropylsiloxy)-methylchlorid |
| Trisyl-Cl | 2,4,6-triisopropyl-benzolsulfonylchlorid |
| TRIS | tris(hydroxymethyl)aminomethane |
| tRNA | transfer RNA |
| TsOH | *p*-toluenesulfonic acid |
| U | uridine |
| UV | ultraviolet |
| $v^8A$ | 8-vinyladenosine |
| 8vdA | 8-vinyl-deoxyadenosine |
| WC | Watson-Crick |
| WT | wild-type |
| -Δ | knockout |
| -1 PRF | programmed -1 ribosomal frameshifting |
| +ssRNAs | positive-sense single stranded RNA |
| Ψ | pseudo uridine |

# References


1. Franklin, R. E.; Gosling, R. G., The structure of sodium thymonucleate fibres. I. The influence of water content. *Acta Crystallographica* **1953,** *6* (8-9), 673-677.
2. Watson, J. D.; Crick, F. H. In *The structure of DNA*, Cold Spring Harbor symposia on quantitative biology, Cold Spring Harbor Laboratory Press: 1953; pp 123-131.
3. Pray, L., Discovery of DNA structure and function: Watson and Crick. *Nature Education* **2008,** *1* (1), 100.
4. Crick, F., Central dogma of molecular biology. *Nature* **1970,** *227* (5258), 561-563.
5. Crick, F. H. In *On protein synthesis*, Symp Soc Exp Biol, 1958; p 8.
6. Carell, T.; Brandmayr, C.; Hienzsch, A.; Muller, M.; Pearson, D.; Reiter, V.; Thoma, I.; Thumbs, P.; Wagner, M., Structure and function of noncanonical nucleobases. *Angew. Chem. Int. Ed. Engl.* **2012,** *51* (29), 7110-31.
7. Berg, J. M.; Tymoczko, J. L., *Stryer biochemie*. Springer: 2018; Vol. 8.
8. Levene, P., The structure of yeast nucleic acid. *Studies from the Rockefeller Institute for Medical Research* **1919,** *30*, 221.
9. Potaman, V. N.; Sinden, R. R., DNA: alternative conformations and biology. In *Madame Curie Bioscience Database [Internet]*, Landes Bioscience: 2013.
10. Altona, C. t.; Sundaralingam, M., Conformational analysis of the sugar ring in nucleosides and nucleotides. New description using the concept of pseudorotation. *J. Am. Chem. Soc.* **1972,** *94* (23), 8205-8212.
11. Wing, R.; Drew, H.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E., Crystal structure analysis of a complete turn of B-DNA. *Nature* **1980,** *287* (5784), 755-8.
12. Weeks, K. M.; Crothers, D. M., Major groove accessibility of RNA. *Science* **1993,** *261* (5128), 1574-1577.
13. Tinoco Jr, I.; Bustamante, C., How RNA folds. *J. Mol. Biol.* **1999,** *293* (2), 271-281.
14. Hunter, C. A., Sequence-dependent DNA structure. The role of base stacking interactions. *J. Mol. Biol.* **1993,** *230* (3), 1025-54.
15. Mattick, J. S., RNA driving the epigenetic bus. *EMBO J.* **2012,** *31* (3), 515-6.
16. Altman, S., Nobel lecture. Enzymatic cleavage of RNA by RNA. *Biosci. Rep.* **1990,** *10* (4), 317-37.
17. Mercer, T. R.; Gerhardt, D. J.; Dinger, M. E.; Crawford, J.; Trapnell, C.; Jeddeloh, J. A.; Mattick, J. S.; Rinn, J. L., Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.* **2012,** *30* (1), 99-104.
18. Cech, T. R., The efficiency and versatility of catalytic RNA: implications for an RNA world. *Gene* **1993,** *135* (1), 33-36.
19. Guttman, M.; Rinn, J. L., Modular regulatory principles of large non-coding RNAs. *Nature* **2012,** *482* (7385), 339-346.
20. Ng, S. Y.; Johnson, R.; Stanton, L. W., Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *The EMBO journal* **2012,** *31* (3), 522-533.
21. Clark, B.; Doctor, B.; Holmes, K. C.; Klug, A.; Marcker, K.; Morris, S.; Paradies, H., Crystallization of transfer RNA. *Nature* **1968,** *219* (5160), 1222-1224.
22. Kim, S.; Suddath, F.; Quigley, G.; McPherson, A.; Sussman, J.; Wang, A.; Seeman, N.; Rich, A., Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science* **1974,** *185* (4149), 435-440.
23. Robertus, J.; Ladner, J. E.; Finch, J.; Rhodes, D.; Brown, R.; Clark, B.; Klug, A., Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature* **1974,** *250* (5467), 546-551.
24. Shi, H.; Moore, P. B., The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited. *Rna* **2000,** *6* (8), 1091-1105.
25. Ibba, M.; Söll; Dieter, Quality control mechanisms during translation. *Science* **1999,** *286* (5446), 1893-1897.
26. Arnez, J. G.; Moras, D., Structural and functional considerations of the aminoacylation reaction. *Trends Biochem. Sci* **1997,** *22* (6), 211-6.
27. Ringer, D.; Chladek, S., Inhibition of the peptidyl transferase A-site function by 2'-O-aminoacyloligonucleotides. *Biochem. Biophys. Res. Commun.* **1974,** *56* (3), 760-766.
28. Eriani, G.; Delarue, M.; Poch, O.; Gangloff, J.; Moras, D., Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **1990,** *347* (6289), 203-206.
29. Giegé, R.; Sissler, M.; Florentz, C., Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.* **1998,** *26* (22), 5017-5035.

30. Nissen, P.; Kjeldgaard, M.; Thirup, S.; Polekhina, G.; Reshetnikova, L.; Clark, B. F.; Nyborg, J., Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog. *Science* **1995,** *270* (5241), 1464-1472.
31. Becker, H. D.; Kern, D., Thermus thermophilus: a link in evolution of the tRNA-dependent amino acid amidation pathways. *Proc Natl Acad Sci U S A* **1998,** *95* (22), 12832-7.
32. STANZEL, M.; SCHÖN, A.; SPRINZL, M., Discrimination against misacylated tRNA by chloroplast elongation factor Tu. *Eur. J. Biochem.* **1994,** *219* (1-2), 435-439.
33. Rodnina, M.; Fricke, R.; Kuhn, L.; Wintermeyer, W., Codon-dependent conformational change of elongation factor Tu preceding GTP hydrolysis on the ribosome. *The EMBO journal* **1995,** *14* (11), 2613-2619.
34. Pape, T.; Wintermeyer, W.; Rodnina, M., Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. *The EMBO journal* **1999,** *18* (13), 3800-3807.
35. Yoshizawa, S.; Fourmy, D.; Puglisi, J. D., Recognition of the codon-anticodon helix by ribosomal RNA. *Science* **1999,** *285* (5434), 1722-1725.
36. McCown, P. J.; Ruszkowska, A.; Kunkler, C. N.; Breger, K.; Hulewicz, J. P.; Wang, M. C.; Springer, N. A.; Brown, J. A., Naturally occurring modified ribonucleosides. *Wiley Interdiscip Rev RNA* **2020,** *11* (5), e1595.
37. Cohn, W. E.; Volkin, E., Nucleoside-5′-phosphates from ribonucleic acid. *Nature* **1951,** *167* (4247), 483-484.
38. Cohn, W. E., 5-Ribosyl uracil, a carbon-carbon ribofuranosyl nucleoside in ribonucleic acids. *Biochim. Biophys. Acta* **1959,** *32*, 569-71.
39. Hamma, T.; Ferre-D'Amare, A. R., Pseudouridine synthases. *Chem Biol* **2006,** *13* (11), 1125-35.
40. Davis, F. F.; Allen, F. W., Ribonucleic acids from yeast which contain a fifth nucleotide. *J. Biol. Chem.* **1957,** *227* (2), 907-15.
41. Ge, J.; Yu, Y. T., RNA pseudouridylation: new insights into an old modification. *Trends Biochem. Sci* **2013,** *38* (4), 210-8.
42. Schwartz, S.; Bernstein, D. A.; Mumbach, M. R.; Jovanovic, M.; Herbst, R. H.; Leon-Ricardo, B. X.; Engreitz, J. M.; Guttman, M.; Satija, R.; Lander, E. S.; Fink, G.; Regev, A., Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **2014,** *159* (1), 148-162.
43. Carlile, T. M.; Martinez, N. M.; Schaening, C.; Su, A.; Bell, T. A.; Zinshteyn, B.; Gilbert, W. V., mRNA structure determines modification by pseudouridine synthase 1. *Nat Chem Biol* **2019,** *15* (10), 966-974.
44. Baudin-Baillieu, A.; Fabret, C.; Liang, X. H.; Piekna-Przybylska, D.; Fournier, M. J.; Rousset, J. P., Nucleotide modifications in three functionally important regions of the Saccharomyces cerevisiae ribosome affect translation accuracy. *Nucleic Acids Res.* **2009,** *37* (22), 7665-77.
45. Arnez, J. G.; Steitz, T. A., Crystal structure of unmodified tRNAGln complexed with glutaminyl-tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. *Biochem. Cell Biol.* **1994,** *33* (24), 7560-7567.
46. Davis, D. R.; Veltri, C. A.; Nielsen, L., An RNA model system for investigation of pseudouridine stabilization of the codon-anticodon interaction in tRNALys, tRNAHis and tRNATyr. *J. Biomol. Struct. Dyn.* **1998,** *15* (6), 1121-32.
47. Harrington, K. M.; Nazarenko, I. A.; Dix, D. B.; Thompson, R. C.; Uhlenbeck, O. C., In vitro analysis of translational rate and accuracy with an unmodified tRNA. *Biochemistry* **1993,** *32* (30), 7617-7622.
48. Kierzek, E.; Malgowska, M.; Lisowiec, J.; Turner, D. H.; Gdaniec, Z.; Kierzek, R., The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Res.* **2014,** *42* (5), 3492-501.
49. Boo, S. H.; Kim, Y. K., The emerging role of RNA modifications in the regulation of mRNA stability. *Exp Mol Med* **2020,** *52* (3), 400-408.
50. Phizicky, E. M.; Alfonzo, J. D., Do all modifications benefit all tRNAs? *FEBS Lett.* **2010,** *584* (2), 265-271.
51. Lorenz, C.; Lünse, C. E.; Mörl, M., tRNA modifications: impact on structure and thermal adaptation. *Biomolecules* **2017,** *7* (2), 35.
52. Arimbasseri, A. G.; Iben, J.; Wei, F. Y.; Rijal, K.; Tomizawa, K.; Hafner, M.; Maraia, R. J., Evolving specificity of tRNA 3-methyl-cytidine-32 (m3C32) modification: a subset of tRNAsSer requires N6-isopentenylation of A37. *RNA* **2016,** *22* (9), 1400-10.
53. Phizicky, E. M.; Hopper, A. K., tRNA biology charges to the front. *Genes & development* **2010,** *24* (17), 1832-1860.
54. Jühling, F.; Mörl, M.; Hartmann, R. K.; Sprinzl, M.; Stadler, P. F.; Pütz, J., tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res.* **2009,** *37* (suppl_1), D159-D162.

55. Noma, A.; Yi, S.; Katoh, T.; Takai, Y.; Suzuki, T.; Suzuki, T., Actin-binding protein ABP140 is a methyltransferase for 3-methylcytidine at position 32 of tRNAs in Saccharomyces cerevisiae. *Rna* **2011,** *17* (6), 1111-1119.
56. Suzuki, T.; Yashiro, Y.; Kikuchi, I.; Ishigami, Y.; Saito, H.; Matsuzawa, I.; Okada, S.; Mito, M.; Iwasaki, S.; Ma, D.; Zhao, X.; Asano, K.; Lin, H.; Kirino, Y.; Sakaguchi, Y.; Suzuki, T., Complete chemical structures of human mitochondrial tRNAs. *Nat Commun* **2020,** *11* (1), 4269.
57. Marchand, V.; Ayadi, L.; Ernst, F. G.; Hertler, J.; Bourguignon-Igel, V.; Galvanin, A.; Kotter, A.; Helm, M.; Lafontaine, D. L.; Motorin, Y., AlkAniline-Seq: Profiling of m7G and m3C RNA Modifications at Single Nucleotide Resolution. *Angew. Chem. Int. Ed.* **2018,** *57* (51), 16785-16790.
58. Kleiber, N.; Lemus-Diaz, N.; Stiller, C.; Heinrichs, M.; Mai, M. M.; Hackert, P.; Richter-Dennerlein, R.; Hobartner, C.; Bohnsack, K. E.; Bohnsack, M. T., The RNA methyltransferase METTL8 installs m(3)C32 in mitochondrial tRNAs(Thr/Ser(UCN)) to optimise tRNA structure and mitochondrial translation. *Nat Commun* **2022,** *13* (1), 209.
59. Keith, G., The primary structures of two arginine tRNAs (anticodous CCU and mcm5a2U-C-ψ) and of glutamine tRNA (anticodon CUG) from bovine liver. *Nucleic Acids Res.* **1984,** *12* (5), 2543-2547.
60. D'Silva, S.; Haider, S. J.; Phizicky, E. M., A domain of the actin binding protein Abp140 is the yeast methyltransferase responsible for 3-methylcytidine modification in the tRNA anti-codon loop. *Rna* **2011,** *17* (6), 1100-1110.
61. Clark, W. C.; Evans, M. E.; Dominissini, D.; Zheng, G.; Pan, T., tRNA base methylation identification and quantification via high-throughput sequencing. *RNA* **2016,** *22* (11), 1771-1784.
62. de Crécy-Lagard, V.; Boccaletto, P.; Mangleburg, C. G.; Sharma, P.; Lowe, T. M.; Leidel, S. A.; Bujnicki, J. M., Matching tRNA modifications in humans to their known and predicted enzymes. *Nucleic Acids Res.* **2019,** *47* (5), 2143-2159.
63. Cui, J.; Liu, Q.; Sendinc, E.; Shi, Y.; Gregory, R. I., Nucleotide resolution profiling of m3C RNA modification by HAC-seq. *Nucleic Acids Res.* **2021,** *49* (5), e27.
64. Xu, L.; Liu, X.; Sheng, N.; Oo, K. S.; Liang, J.; Chionh, Y. H.; Xu, J.; Ye, F.; Gao, Y. G.; Dedon, P. C.; Fu, X. Y., Three distinct 3-methylcytidine (m(3)C) methyltransferases modify tRNA and mRNA in mice and humans. *J. Biol. Chem.* **2017,** *292* (35), 14695-14703.
65. Ma, C.-J.; Ding, J.-H.; Ye, T.-T.; Yuan, B.-F.; Feng, Y.-Q., AlkB homologue 1 demethylates N 3-methylcytidine in mRNA of mammals. *ACS Chemical Biology* **2019,** *14* (7), 1418-1425.
66. Persson, B. C.; Esberg, B.; Olafsson, O.; Bjork, G. R., Synthesis and function of isopentenyl adenosine derivatives in tRNA. *Biochimie* **1994,** *76* (12), 1152-60.
67. Thiaville, P. C.; Iwata-Reuyl, D.; de Crécy-Lagard, V., Diversity of the biosynthesis pathway for threonylcarbamoyladenosine (t6A), a universal modification of tRNA. *RNA biology* **2014,** *11* (12), 1529-1539.
68. Yamaizumi, Z.; Nishimura, S.; Limburg, K.; Raba, M.; Gross, H. J.; Crain, P.; McCloskey, J. A., Structure elucidation by high resolution mass spectrometry of a highly modified nucleoside from mammalian transfer RNA. N-[(9-. beta.-D-Ribofuranosyl-2-methylthiopurin-6-yl) carbamoyl] threonine. *J. Am. Chem. Soc.* **1979,** *101* (8), 2224-2225.
69. Agris, P. F.; Armstrong, D. J.; Schafer, K. P.; Soll, D., Maturation of a hypermodified nucleoside in transfer RNA. *Nucleic Acids Res.* **1975,** *2* (5), 691-8.
70. Thiaville, P. C.; El Yacoubi, B.; Köhrer, C.; Thiaville, J. J.; Deutsch, C.; Iwata-Reuyl, D.; Bacusmo, J. M.; Armengaud, J.; Bessho, Y.; Wetzel, C., Essentiality of threonylcarbamoyladenosine (t6 A), a universal t RNA modification, in bacteria. *Mol. Microbiol.* **2015,** *98* (6), 1199-1221.
71. Boccaletto, P.; Stefaniak, F.; Ray, A.; Cappannini, A.; Mukherjee, S.; Purta, E.; Kurkowska, M.; Shirvanizadeh, N.; Destefanis, E.; Groza, P.; Avsar, G.; Romitelli, A.; Pir, P.; Dassi, E.; Conticello, S. G.; Aguilo, F.; Bujnicki, J. M., MODOMICS: a database of RNA modification pathways. 2021 update. *Nucleic Acids Res.* **2022,** *50* (D1), D231-D235.
72. Anderson, J.; Phan, L.; Cuesta, R.; Carlson, B. A.; Pak, M.; Asano, K.; Björk, G. R.; Tamame, M.; Hinnebusch, A. G., The essential Gcd10p–Gcd14p nuclear complex is required for 1-methyladenosine modification and maturation of initiator methionyl-tRNA. *Genes & development* **1998,** *12* (23), 3650-3662.
73. Madore, E.; Florentz, C.; Giege, R.; Sekine, S.; Yokoyama, S.; Lapointe, J., Effect of modified nucleotides on Escherichia coli tRNAGlu structure and on its aminoacylation by glutamyl-tRNA synthetase. Predominant and distinct roles of the mnm5 and s2 modifications of U34. *Eur. J. Biochem.* **1999,** *266* (3), 1128-35.
74. Motorin, Y.; Helm, M., tRNA stabilization by modified nucleotides. *Biochemistry* **2010,** *49* (24), 4934-44.
75. Alexandrov, A.; Chernyakov, I.; Gu, W.; Hiley, S. L.; Hughes, T. R.; Grayhack, E. J.; Phizicky, E. M., Rapid tRNA decay can result from lack of nonessential modifications. *Molecular cell* **2006,** *21* (1), 87-96.

76. Shin, B.-S.; Kim, J.-R.; Walker, S. E.; Dong, J.; Lorsch, J. R.; Dever, T. E., Initiation factor eIF2γ promotes eIF2–GTP–Met-tRNA i Met ternary complex binding to the 40S ribosome. *Nature structural molecular biology* **2011,** *18* (11), 1227-1234.
77. Lu, J.; Esberg, A.; Huang, B.; Byström, A. S., Kluyveromyces lactis γ-toxin, a ribonuclease that recognizes the anticodon stem loop of tRNA. *Nucleic Acids Res.* **2008,** *36* (4), 1072-1080.
78. Schaefer, M.; Pollex, T.; Hanna, K.; Tuorto, F.; Meusburger, M.; Helm, M.; Lyko, F., RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev* **2010,** *24* (15), 1590-5.
79. Blanco, S.; Dietmann, S.; Flores, J. V.; Hussain, S.; Kutter, C.; Humphreys, P.; Lukk, M.; Lombard, P.; Treps, L.; Popis, M., Aberrant methylation of t RNA s links cellular stress to neuro-developmental disorders. *The EMBO journal* **2014,** *33* (18), 2020-2039.
80. Klassen, R.; Paluszynski, J. P.; Wemhoff, S.; Pfeiffer, A.; Fricke, J.; Meinhardt, F., The primary target of the killer toxin from Pichia acaciae is tRNAGln. *Mol. Microbiol.* **2008,** *69* (3), 681-697.
81. Saikia, M.; Krokowski, D.; Guan, B. J.; Ivanov, P.; Parisien, M.; Hu, G. F.; Anderson, P.; Pan, T.; Hatzoglou, M., Genome-wide identification and quantitative analysis of cleaved tRNA fragments induced by cellular stress. *J. Biol. Chem.* **2012,** *287* (51), 42708-25.
82. Manickam, N.; Joshi, K.; Bhatt, M. J.; Farabaugh, P. J., Effects of tRNA modification on translational accuracy depend on intrinsic codon-anticodon strength. *Nucleic Acids Res.* **2016,** *44* (4), 1871-81.
83. Weixlbaumer, A.; Murphy, F. V. t.; Dziergowska, A.; Malkiewicz, A.; Vendeix, F. A.; Agris, P. F.; Ramakrishnan, V., Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. *Nat Struct Mol Biol* **2007,** *14* (6), 498-502.
84. Bjork, G. R.; Wikstrom, P. M.; Bystrom, A. S., Prevention of translational frameshifting by the modified nucleoside 1-methylguanosine. *Science* **1989,** *244* (4907), 986-9.
85. Hori, H., Methylated nucleosides in tRNA and tRNA methyltransferases. *Front Genet* **2014,** *5*, 144.
86. Yarian, C.; Townsend, H.; Czestkowski, W.; Sochacka, E.; Malkiewicz, A. J.; Guenther, R.; Miskiewicz, A.; Agris, P. F., Accurate translation of the genetic code depends on tRNA modified nucleosides. *J. Biol. Chem.* **2002,** *277* (19), 16391-16395.
87. Crick, F. H., Codon--anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* **1966,** *19* (2), 548-55.
88. Gerber, A. P.; Keller, W., An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science* **1999,** *286* (5442), 1146-1149.
89. Daugeron, M.-C.; Lenstra, T. L.; Frizzarin, M.; El Yacoubi, B.; Liu, X.; Baudin-Baillieu, A.; Lijnzaad, P.; Decourty, L.; Saveanu, C.; Jacquier, A., Gcn4 misregulation reveals a direct role for the evolutionary conserved EKC/KEOPS in the t6A modification of tRNAs. *Nucleic Acids Res.* **2011,** *39* (14), 6148-6160.
90. Lamichhane, T. N.; Mattijssen, S.; Maraia, R. J., Human cells have a limited set of tRNA anticodon loop substrates of the tRNA isopentenyltransferase TRIT1 tumor suppressor. *Molecular and cellular biology* **2013,** *33* (24), 4900-4908.
91. Murphy, F. V.; Ramakrishnan, V.; Malkiewicz, A.; Agris, P. F., The role of modifications in codon discrimination by tRNALysUUU. *Nature structural molecular biology* **2004,** *11* (12), 1186-1191.
92. Lehmann, J.; Libchaber, A., Degeneracy of the genetic code and stability of the base pair at the second position of the anticodon. *Rna* **2008,** *14* (7), 1264-1269.
93. Duechler, M.; Leszczyńska, G.; Sochacka, E.; Nawrot, B., Nucleoside modifications in the regulation of gene expression: focus on tRNA. *Cellular and Molecular Life Sciences* **2016,** *73* (16), 3075-3095.
94. Swinehart, W. E.; Jackman, J. E., Diversity in mechanism and function of tRNA methyltransferases. *RNA Biol* **2015,** *12* (4), 398-411.
95. Ignatova, V. V.; Jansen, P.; Baltissen, M. P.; Vermeulen, M.; Schneider, R., The interactome of a family of potential methyltransferases in HeLa cells. *Sci Rep* **2019,** *9* (1), 6584.
96. Ignatova, V. V.; Kaiser, S.; Ho, J. S. Y.; Bing, X.; Stolz, P.; Tan, Y. X.; Lee, C. L.; Gay, F. P. H.; Lastres, P. R.; Gerlini, R., METTL6 is a tRNA m3C methyltransferase that regulates pluripotency and tumor cell growth. *Science advances* **2020,** *6* (35), eaaz4551.
97. Lentini, J. M.; Alsaif, H. S.; Faqeih, E.; Alkuraya, F. S.; Fu, D., DALRD3 encodes a protein mutated in epileptic encephalopathy that targets arginine tRNAs for 3-methylcytosine modification. *Nature communications* **2020,** *11* (1), 1-14.
98. Cozen, A. E.; Quartley, E.; Holmes, A. D.; Hrabeta-Robinson, E.; Phizicky, E. M.; Lowe, T. M., ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nature methods* **2015,** *12* (9), 879-884.
99. Han, L.; Marcus, E.; D'Silva, S.; Phizicky, E. M., S. cerevisiae Trm140 has two recognition modes for 3-methylcytidine modification of the anticodon loop of tRNA substrates. *Rna* **2017,** *23* (3), 406-419.

100. Jackman, J. E.; Phizicky, E. M., tRNAHis guanylyltransferase adds G− 1 to the 5′ end of tRNAHis by recognition of the anticodon, one of several features unexpectedly shared with tRNA synthetases. *Rna* **2006,** *12* (6), 1007-1014.
101. Hur, S.; Stroud, R. M., How U38, 39, and 40 of many tRNAs become the targets for pseudouridylation by TruA. *Molecular cell* **2007,** *26* (2), 189-203.
102. Tomikawa, J.; Homma, T.; Tajima, S.; Shibata, T.; Inamoto, Y.; Takase, K.; Inoue, N.; Ohkura, S.; Uenoyama, Y.; Maeda, K.; Tsukamura, H., Molecular characterization and estrogen regulation of hypothalamic KISS1 gene in the pig. *Biol Reprod* **2010,** *82* (2), 313-9.
103. Rubio, M. A. T.; Ragone, F. L.; Gaston, K. W.; Ibba, M.; Alfonzo, J. D., C to U editing stimulates A to I editing in the anticodon loop of a cytoplasmic threonyl tRNA in Trypanosoma brucei. *J. Biol. Chem.* **2006,** *281* (1), 115-120.
104. Rubio, M. A. T.; Gaston, K. W.; McKenney, K. M.; Fleming, I.; Paris, Z.; Limbach, P. A.; Alfonzo, J. D., Editing and methylation at a single site by functionally interdependent activities. *Nature* **2017,** *542* (7642), 494-497.
105. Frye, M.; Watt, F. M., The RNA methyltransferase Misu (NSun2) mediates Myc-induced proliferation and is upregulated in tumors. *Current Biology* **2006,** *16* (10), 971-981.
106. Torres, A. G.; Batlle, E.; de Pouplana, L. R., Role of tRNA modifications in human diseases. *Trends in molecular medicine* **2014,** *20* (6), 306-314.
107. Bohnsack, M. T.; Sloan, K. E., The mitochondrial epitranscriptome: the roles of RNA modifications in mitochondrial translation and human disease. *Cell Mol Life Sci* **2018,** *75* (2), 241-260.
108. Abbasi-Moheb, L.; Mertel, S.; Gonsior, M.; Nouri-Vahid, L.; Kahrizi, K.; Cirak, S.; Wieczorek, D.; Motazacker, M. M.; Esmaeeli-Nieh, S.; Cremer, K.; Weissmann, R.; Tzschach, A.; Garshasbi, M.; Abedini, S. S.; Najmabadi, H.; Ropers, H. H.; Sigrist, S. J.; Kuss, A. W., Mutations in NSUN2 cause autosomal-recessive intellectual disability. *Am J Hum Genet* **2012,** *90* (5), 847-55.
109. Igoillo-Esteve, M.; Genin, A.; Lambert, N.; Désir, J.; Pirson, I.; Abdulkarim, B.; Simonis, N.; Drielsma, A.; Marselli, L.; Marchetti, P., tRNA methyltransferase homolog gene TRMT10A mutation in young onset diabetes and primary microcephaly in humans. *PLoS genetics* **2013,** *9* (10), e1003888.
110. Trewick, S. C.; Henshaw, T. F.; Hausinger, R. P.; Lindahl, T.; Sedgwick, B., Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature* **2002,** *419* (6903), 174-178.
111. Aravind, L.; Koonin, E. V., The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol* **2001,** *2* (3), RESEARCH0007.
112. Fedeles, B. I.; Singh, V.; Delaney, J. C.; Li, D.; Essigmann, J. M., The AlkB family of Fe (II)/α-ketoglutarate-dependent dioxygenases: repairing nucleic acid alkylation damage and beyond. *J. Biol. Chem.* **2015,** *290* (34), 20734-20742.
113. Duncan, T.; Trewick, S. C.; Koivisto, P.; Bates, P. A.; Lindahl, T.; Sedgwick, B., Reversal of DNA alkylation damage by two human dioxygenases. *Proceedings of the National Academy of Sciences* **2002,** *99* (26), 16660-16665.
114. Gerken, T.; Girard, C. A.; Tung, Y.-C. L.; Webby, C. J.; Saudek, V.; Hewitson, K. S.; Yeo, G. S.; McDonough, M. A.; Cunliffe, S.; McNeill, L. A., The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **2007,** *318* (5855), 1469-1472.
115. Jia, G.; Fu, Y.; Zhao, X.; Dai, Q.; Zheng, G.; Yang, Y.; Yi, C.; Lindahl, T.; Pan, T.; Yang, Y.-G., N 6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature chemical biology* **2011,** *7* (12), 885-887.
116. Kurowski, M. A.; Bhagwat, A. S.; Papaj, G.; Bujnicki, J. M., Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB. *BMC genomics* **2003,** *4* (1), 1-6.
117. Fu, Y.; Jia, G.; Pang, X.; Wang, R. N.; Wang, X.; Li, C. J.; Smemo, S.; Dai, Q.; Bailey, K. A.; Nobrega, M. A., FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. *Nature communications* **2013,** *4* (1), 1-8.
118. Mauer, J.; Luo, X.; Blanjoie, A.; Jiao, X.; Grozhik, A. V.; Patil, D. P.; Linder, B.; Pickering, B. F.; Vasseur, J.-J.; Chen, Q., Reversible methylation of m6Am in the 5′ cap controls mRNA stability. *Nature* **2017,** *541* (7637), 371-375.
119. Zheng, G.; Dahl, J. A.; Niu, Y.; Fedorcsak, P.; Huang, C.-M.; Li, C. J.; Vågbø, C. B.; Shi, Y.; Wang, W.-L.; Song, S.-H., ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Molecular cell* **2013,** *49* (1), 18-29.
120. Zheng, G.; He, C., RNA demethylation by FTO and ALKBH5. In *2-Oxoglutarate-Dependent Oxygenases*, Royal Society of Chemistry: 2015; pp 263-274.
121. Kaur, S.; Tam, N. Y.; McDonough, M. A.; Schofield, C. J.; Aik, W. S., Mechanisms of substrate recognition and N 6-methyladenosine demethylation revealed by crystal structures of ALKBH5–RNA complexes. *Nucleic Acids Res.* **2022,** *50* (7), 4148-4160.

122. Chen, Z.; Qi, M.; Shen, B.; Luo, G.; Wu, Y.; Li, J.; Lu, Z.; Zheng, Z.; Dai, Q.; Wang, H., Transfer RNA demethylase ALKBH3 promotes cancer progression via induction of tRNA-derived small RNAs. *Nucleic Acids Res.* **2019,** *47* (5), 2533-2545.
123. Aas, P. A.; Otterlei, M.; Falnes, P. Ø.; Vågbø, C. B.; Skorpen, F.; Akbari, M.; Sundheim, O.; Bjørås, M.; Slupphaug, G.; Seeberg, E., Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* **2003,** *421* (6925), 859-863.
124. Ueda, Y.; Ooshio, I.; Fusamae, Y.; Kitae, K.; Kawaguchi, M.; Jingushi, K.; Hase, H.; Harada, K.; Hirata, K.; Tsujikawa, K., AlkB homolog 3-mediated tRNA demethylation promotes protein synthesis in cancer cells. *Scientific reports* **2017,** *7* (1), 1-10.
125. Haag, S.; Sloan, K. E.; Ranjan, N.; Warda, A. S.; Kretschmer, J.; Blessing, C.; Hübner, B.; Seikowski, J.; Dennerlein, S.; Rehling, P., NSUN 3 and ABH 1 modify the wobble position of mt-t RNA Met to expand codon recognition in mitochondrial translation. *The EMBO journal* **2016,** *35* (19), 2104-2119.
126. Westbye, M. P.; Feyzi, E.; Aas, P. A.; Vågbø, C. B.; Talstad, V. A.; Kavli, B.; Hagen, L.; Sundheim, O.; Akbari, M.; Liabakk, N.-B., Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA. *J. Biol. Chem.* **2008,** *283* (36), 25046-25056.
127. Ougland, R.; Lando, D.; Jonson, I.; Dahl, J. A.; Moen, M. N.; Nordstrand, L. M.; Rognes, T.; Lee, J. T.; Klungland, A.; Kouzarides, T., ALKBH1 is a histone H2A dioxygenase involved in neural differentiation. *Stem cells* **2012,** *30* (12), 2672-2682.
128. Müller, T. A.; Andrzejak, M. M.; Hausinger, R. P., A covalent protein–DNA 5′-product adduct is generated following AP lyase activity of human ALKBH1 (AlkB homologue 1). *Biochem. J* **2013,** *452* (3), 509-518.
129. Kawarada, L.; Suzuki, T.; Ohira, T.; Hirata, S.; Miyauchi, K.; Suzuki, T., ALKBH1 is an RNA dioxygenase responsible for cytoplasmic and mitochondrial tRNA modifications. *Nucleic Acids Res.* **2017,** *45* (12), 7401-7415.
130. Xiao, C.-L.; Zhu, S.; He, M.; Chen, D.; Zhang, Q.; Chen, Y.; Yu, G.; Liu, J.; Xie, S.-Q.; Luo, F., N6-methyladenine DNA modification in the human genome. *Molecular cell* **2018,** *71* (2), 306-318. e7.
131. Zhang, Y.; Wang, C., Demethyltransferase AlkBH1 substrate diversity and relationship to human diseases. *Molecular Biology Reports* **2021,** *48* (5), 4747-4756.
132. Fu, Y.; Dai, Q.; Zhang, W.; Ren, J.; Pan, T.; He, C., The AlkB domain of mammalian ABH8 catalyzes hydroxylation of 5-methoxycarbonylmethyluridine at the wobble position of tRNA. *Angew. Chem. Int. Ed.* **2010,** *49* (47), 8885-8888.
133. Van Den Born, E.; Vågbø, C. B.; Songe-Møller, L.; Leihne, V.; Lien, G. F.; Leszczynska, G.; Malkiewicz, A.; Krokan, H. E.; Kirpekar, F.; Klungland, A., ALKBH8-mediated formation of a novel diastereomeric pair of wobble nucleosides in mammalian tRNA. *Nature communications* **2011,** *2* (1), 1-7.
134. Li, M.-M.; Nilsen, A.; Shi, Y.; Fusser, M.; Ding, Y.-H.; Fu, Y.; Liu, B.; Niu, Y.; Wu, Y.-S.; Huang, C.-M., ALKBH4-dependent demethylation of actin regulates actomyosin dynamics. *Nature communications* **2013,** *4* (1), 1-13.
135. Zhao, S.; Devega, R.; Francois, A.; Kidane, D., Human ALKBH6 is required for maintenance of genomic stability and promoting cell survival during exposure of alkylating agents in pancreatic cancer. *Frontiers in Genetics* **2021,** *12*, 468.
136. Huong, T. T.; Ngoc, L. N. T.; Kang, H., Functional characterization of a putative RNA demethylase ALKBH6 in Arabidopsis growth and abiotic stress responses. *International journal of molecular sciences* **2020,** *21* (18), 6707.
137. Zhang, L.-S.; Xiong, Q.-P.; Peña Perez, S.; Liu, C.; Wei, J.; Le, C.; Zhang, L.; Harada, B. T.; Dai, Q.; Feng, X., ALKBH7-mediated demethylation regulates mitochondrial polycistronic RNA processing. *Nature Cell Biology* **2021,** *23* (7), 684-691.
138. Fu, D.; Jordan, J. J.; Samson, L. D., Human ALKBH7 is required for alkylation and oxidation-induced programmed necrosis. *Genes & development* **2013,** *27* (10), 1089-1100.
139. Kellner, S.; Burhenne, J.; Helm, M., Detection of RNA modifications. *RNA biology* **2010,** *7* (2), 237-247.
140. Wiener, D.; Schwartz, S., The epitranscriptome beyond m6A. *Nature Reviews Genetics* **2021,** *22* (2), 119-131.
141. Zheng, G.; Qin, Y.; Clark, W. C.; Dai, Q.; Yi, C.; He, C.; Lambowitz, A. M.; Pan, T., Efficient and quantitative high-throughput tRNA sequencing. *Nature methods* **2015,** *12* (9), 835-837.
142. Cui, J.; Liu, Q.; Sendinc, E.; Shi, Y.; Gregory, R. I., Nucleotide resolution profiling of m3C RNA modification by HAC-seq. *Nucleic Acids Res.* **2021,** *49* (5), e27-e27.
143. Jain, M.; Olsen, H. E.; Paten, B.; Akeson, M., The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome biology* **2016,** *17* (1), 1-11.
144. Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Di Ventra, M.; Garaj, S.; Hibbs, A.; Huang, X.; Jovanovich, S. B.; Krstic, P. S.; Lindsay, S.; Ling, X. S.; Mastrangelo,

C. H.; Meller, A.; Oliver, J. S.; Pershin, Y. V.; Ramsey, J. M.; Riehn, R.; Soni, G. V.; Tabard-Cossa, V.; Wanunu, M.; Wiggin, M.; Schloss, J. A., The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* **2008,** *26* (10), 1146-53.
145. Leggett, R. M.; Clark, M. D., A world of opportunities with nanopore sequencing. *J. Exp. Bot.* **2017,** *68* (20), 5419-5429.
146. Deamer, D.; Akeson, M.; Branton, D., Three decades of nanopore sequencing. *Nat. Biotechnol.* **2016,** *34* (5), 518-524.
147. Wescoe, Z. L.; Schreiber, J.; Akeson, M., Nanopores discriminate among five C5-cytosine variants in DNA. *J. Am. Chem. Soc.* **2014,** *136* (47), 16582-16587.
148. Rand, A. C.; Jain, M.; Eizenga, J. M.; Musselman-Brown, A.; Olsen, H. E.; Akeson, M.; Paten, B., Mapping DNA methylation with high-throughput nanopore sequencing. *Nature methods* **2017,** *14* (4), 411-413.
149. Thomas, N. K.; Poodari, V. C.; Jain, M.; Olsen, H. E.; Akeson, M.; Abu-Shumays, R. L., Direct nanopore sequencing of individual full length tRNA strands. *ACS nano* **2021,** *15* (10), 16642-16653.
150. Schlosser, K.; Li, Y., Biologically inspired synthetic enzymes made from DNA. *Chemistry & biology* **2009,** *16* (3), 311-322.
151. Breaker, R. R.; Joyce, G. F., A DNA enzyme that cleaves RNA. *Chemistry & biology* **1994,** *1* (4), 223-229.
152. Baum, D. A.; Silverman, S. K., Deoxyribozymes: useful DNA catalysts in vitro and in vivo. *Cell Mol Life Sci* **2008,** *65* (14), 2156-74.
153. Li, J.; Zheng, W.; Kwon, A. H.; Lu, Y., In vitro selection and characterization of a highly efficient Zn (II)-dependent RNA-cleaving deoxyribozyme. *Nucleic Acids Res.* **2000,** *28* (2), 481-488.
154. Schlosser, K.; Gu, J.; Sule, L.; Li, Y., Sequence-function relationships provide new insight into the cleavage site selectivity of the 8–17 RNA-cleaving deoxyribozyme. *Nucleic Acids Res.* **2008,** *36* (5), 1472-1481.
155. Santoro, S. W.; Joyce, G. F., A general purpose RNA-cleaving DNA enzyme. *Proceedings of the national academy of sciences* **1997,** *94* (9), 4262-4266.
156. Micura, R.; Höbartner, C., Fundamental studies of functional nucleic acids: aptamers, riboswitches, ribozymes and DNAzymes. *Chem. Soc. Rev.* **2020,** *49* (20), 7331-7353.
157. Cruz, R. P.; Withers, J. B.; Li, Y., Dinucleotide junction cleavage versatility of 8-17 deoxyribozyme. *Chemistry & biology* **2004,** *11* (1), 57-67.
158. Buchhaupt, M.; Peifer, C.; Entian, K.-D., Analysis of 2′-O-methylated nucleosides and pseudouridines in ribosomal RNAs using DNAzymes. *Anal. Biochem.* **2007,** *361* (1), 102-108.
159. Hengesbach, M.; Meusburger, M.; Lyko, F.; Helm, M., Use of DNAzymes for site-specific analysis of ribonucleotide modifications. *Rna* **2008,** *14* (1), 180-187.
160. Meusburger, M.; Hengesbach, M.; Helm, M., A post-labeling approach for the characterization and quantification of RNA modifications based on site-directed cleavage by DNAzymes. In *RNA and DNA Editing*, Springer: 2011; pp 259-270.
161. Sednev, M. V.; Mykhailiuk, V.; Choudhury, P.; Halang, J.; Sloan, K. E.; Bohnsack, M. T.; Höbartner, C., N6-methyladenosine-sensitive RNA-cleaving deoxyribozymes. *Angew. Chem. Int. Ed.* **2018,** *57* (46), 15117-15121.
162. Liaqat, A.; Stiller, C.; Michel, M.; Sednev, M. V.; Höbartner, C., N6-Isopentenyladenosine in RNA Determines the Cleavage Site of Endonuclease Deoxyribozymes. *Angew. Chem.* **2020,** *132* (42), 18786-18790.
163. Sednev, M. V.; Liaqat, A.; Höbartner, C., High-Throughput Activity Profiling of RNA-Cleaving DNA Catalysts by Deoxyribozyme Sequencing (DZ-seq). *J. Am. Chem. Soc.* **2022**.
164. Asseline, U., Development and applications of fluorescent oligonucleotides. *Curr. Org. Chem.* **2006,** *10* (4), 491-518.
165. Tanpure, A. A.; Pawar, M. G.; Srivatsan, S. G., Fluorescent nucleoside analogs: probes for investigating nucleic acid structure and function. *Isr. J. Chem.* **2013,** *53* (6-7), 366-378.
166. Widengren, J.; Kudryavtsev, V.; Antonik, M.; Berger, S.; Gerken, M.; Seidel, C. A., Single-molecule detection and identification of multiple species by multiparameter fluorescence detection. *Anal. Chem.* **2006,** *78* (6), 2039-2050.
167. Dziuba, D.; Didier, P.; Ciaco, S.; Barth, A.; Seidel, C. A.; Mély, Y., Fundamental photophysics of isomorphic and expanded fluorescent nucleoside analogues. *Chem. Soc. Rev.* **2021,** *50* (12), 7062-7107.
168. Wilhelmsson, L. M., Fluorescent nucleic acid base analogues. *Quarterly reviews of biophysics* **2010,** *43* (2), 159-183.
169. Ranasinghe, R. T.; Brown, T., Fluorescence based strategies for genetic analysis. *Chem. Commun.* **2005,** (44), 5487-5502.

170. Xu, W.; Chan, K. M.; Kool, E. T., Fluorescent nucleobases as tools for studying DNA and RNA. *Nature chemistry* **2017,** *9* (11), 1043-1055.
171. Peon, J.; Zewail, A. H., DNA/RNA nucleotides and nucleosides: direct measurement of excited-state lifetimes by femtosecond fluorescence up-conversion. *Chem. Phys. Lett.* **2001,** *348* (3-4), 255-262.
172. Daniels, M.; Hauswirth, W., Fluorescence of the purine and pyrimidine bases of the nucleic acids in neutral aqueous solution at 300 K. *Science* **1971,** *171* (3972), 675-677.
173. Ward, D.; Reich, E.; Stryer, L., Fluorescence studies of nucleotides and polynucleotides: I. Formycin, 2-aminopurine riboside, 2, 6-diaminopurine riboside, and their derivatives. *J. Biol. Chem.* **1969,** *244* (5), 1228-1237.
174. Watanabe, S. M.; Goodman, M. F., On the molecular basis of transition mutations: frequencies of forming 2-aminopurine. cytosine and adenine. cytosine base mispairs in vitro. *Proceedings of the National Academy of Sciences* **1981,** *78* (5), 2864-2868.
175. Sowers, L. C.; Fazakerley, G. V.; Eritja, R.; Kaplan, B. E.; Goodman, M. F., Base pairing and mutagenesis: observation of a protonated base pair between 2-aminopurine and cytosine in an oligonucleotide by proton NMR. *Proceedings of the National Academy of Sciences* **1986,** *83* (15), 5434-5438.
176. Jean, J. M.; Hall, K. B., 2-Aminopurine fluorescence quenching and lifetimes: role of base stacking. *Proceedings of the National Academy of Sciences* **2001,** *98* (1), 37-41.
177. Allan, B. W.; Reich, N. O., Targeted base stacking disruption by the Eco RI DNA methyltransferase. *Biochemistry* **1996,** *35* (47), 14757-14762.
178. Guest, C. R.; Hochstrasser, R. A.; Sowers, L. C.; Millar, D. P., Dynamics of mismatched base pairs in DNA. *Biochemistry* **1991,** *30* (13), 3271-3279.
179. Stivers, J. T., 2-Aminopurine fluorescence studies of base stacking interactions at abasic sites in DNA: metal-ion and base sequence effects. *Nucleic Acids Res.* **1998,** *26* (16), 3837-3844.
180. Menger, M.; Tuschl, T.; Eckstein, F.; Porschke, D., Mg2+-dependent conformational changes in the hammerhead ribozyme. *Biochemistry* **1996,** *35* (47), 14710-14716.
181. Lobsiger, S.; Blaser, S.; Sinha, R. K.; Frey, H.-M.; Leutwyler, S., Switching on the fluorescence of 2-aminopurine by site-selective microhydration. *Nature chemistry* **2014,** *6* (11), 989-993.
182. Gaied, N. B.; Glasser, N.; Ramalanjaona, N.; Beltz, H.; Wolff, P.; Marquet, R.; Burger, A.; Mely, Y., 8-vinyl-deoxyadenosine, an alternative fluorescent nucleoside analog to 2′-deoxyribosyl-2-aminopurine with improved properties. *Nucleic Acids Res.* **2005,** *33* (3), 1031-1039.
183. Hawkins, M. E., Fluorescent pteridine nucleoside analogs. *Cell biochemistry and biophysics* **2001,** *34* (2), 257-281.
184. Srivatsan, S. G.; Greco, N. J.; Tor, Y., A Highly Emissive Fluorescent Nucleoside that Signals the Activity of Toxic Ribosome-Inactivating Proteins. *Angew. Chem.* **2008,** *120* (35), 6763-6767.
185. Berry, D. A.; Jung, K.-Y.; Wise, D. S.; Sercel, A. D.; Pearson, W. H.; Mackie, H.; Randolph, J. B.; Somers, R. L., Pyrrolo-dC and pyrrolo-C: fluorescent analogs of cytidine and 2′-deoxycytidine for the study of oligonucleotides. *Tetrahedron Lett.* **2004,** *45* (11), 2457-2461.
186. Förster, U.; Lommel, K.; Sauter, D.; Grünewald, C.; Engels, J. W.; Wachtveitl, J., 2-(1-Ethynylpyrene)-Adenosine as a Folding Probe for RNA—Pyrene in or out. *ChemBioChem* **2010,** *11* (5), 664-672.
187. Samaan, G. N.; Wyllie, M. K.; Cizmic, J. M.; Needham, L.-M.; Nobis, D.; Ngo, K.; Andersen, S.; Magennis, S. W.; Lee, S. F.; Purse, B. W., Single-molecule fluorescence detection of a tricyclic nucleoside analogue. *Chemical science* **2021,** *12* (7), 2623-2628.
188. Michel, B. Y.; Dziuba, D.; Benhida, R.; Demchenko, A. P.; Burger, A., Probing of nucleic acid structures, dynamics, and interactions with environment-sensitive fluorescent labels. *Frontiers in Chemistry* **2020,** *8*, 112.
189. Saito, Y.; Hudson, R. H., Base-modified fluorescent purine nucleosides and nucleotides for use in oligonucleotide probes. *Journal of Photochemistry and Photobiology C: Photochemistry Reviews* **2018,** *36*, 48-73.
190. Mizrahi, R. A.; Shin, D.; Sinkeldam, R. W.; Phelps, K. J.; Fin, A.; Tantillo, D. J.; Tor, Y.; Beal, P. A., A Fluorescent Adenosine Analogue as a Substrate for an A-to-I RNA Editing Enzyme. *Angew. Chem.* **2015,** *127* (30), 8837-8840.
191. Guckian, K. M.; Schweitzer, B. A.; Ren, R. X.-F.; Sheils, C. J.; Tahmassebi, D. C.; Kool, E. T., Factors contributing to aromatic stacking in water: evaluation in the context of DNA. *J. Am. Chem. Soc.* **2000,** *122* (10), 2213-2222.
192. Kool, E. T., Active site tightness and substrate fit in DNA replication. *Annu. Rev. Biochem* **2002,** *71* (1), 191-219.
193. Matray, T. J.; Kool, E. T., A specific partner for abasic damage in DNA. *Nature* **1999,** *399* (6737), 704-708.

194. Dzantiev, L.; Alekseyev, Y. O.; Morales, J. C.; Kool, E. T.; Romano, L. J., Significance of nucleobase shape complementarity and hydrogen bonding in the formation and stability of the closed polymerase− DNA complex. *Biochemistry* **2001,** *40* (10), 3215-3221.
195. Ren, R. X.-F.; Chaudhuri, N. C.; Paris, P. L.; Rumney IV, S.; Kool, E. T., Naphthalene, phenanthrene, and pyrene as DNA base analogues: synthesis, structure, and fluorescence in DNA. *J. Am. Chem. Soc.* **1996,** *118* (33), 7671-7678.
196. Paris, P. L.; Langenhan, J. M.; Kool, E. T., Probing DNA sequences in solution with a monomer-excimer fluorescence color change. *Nucleic Acids Res.* **1998,** *26* (16), 3789-3793.
197. Coleman, R. S.; Madaras, M. L., Synthesis of a novel coumarin C-riboside as a photophysical probe of oligonucleotide dynamics. *The Journal of Organic Chemistry* **1998,** *63* (16), 5700-5703.
198. Brauns, E. B.; Madaras, M. L.; Coleman, R. S.; Murphy, C. J.; Berg, M. A., Measurement of local DNA reorganization on the picosecond and nanosecond time scales. *J. Am. Chem. Soc.* **1999,** *121* (50), 11644-11649.
199. Morales-Rojas, H.; Kool, E. T., A porphyrin C-nucleoside incorporated into DNA. *Org. Lett.* **2002,** *4* (25), 4377-4380.
200. Beharry, A. A.; Lacoste, S.; O'Connor, T. R.; Kool, E. T., Fluorescence Monitoring of the Oxidative Repair of DNA Alkylation Damage by ALKBH3, a Prostate Cancer Marker. *J. Am. Chem. Soc.* **2016,** *138* (11), 3647-50.
201. Rajski, S. R.; Williams, R. M., DNA cross-linking agents as antitumor drugs. *Chem. Rev.* **1998,** *98* (8), 2723-2796.
202. Hopkins, P. B.; Millard, J. T.; Woo, J.; Weidner, M. F.; Kirchner, J. J.; Sigurdsson, S. T.; Raucher, S., Sequence preferences of DNA interstrand cross-linking agents: importance of minimal DNA structural reorganization in the cross-linking reactions of mechlorethamine, cisplatin and mitomycin C. *Tetrahedron* **1991,** *47* (14-15), 2475-2489.
203. Kohn, K. W., DNA damage in mammalian cells. *Biosci., Biotechnol., Biochem.* **1981,** *31* (8), 593-597.
204. Lawley, P., Alkylation of DNA and its aftermath. *BioEssays* **1995,** *17* (6), 561-568.
205. Goodman, L. S.; Wintrobe, M. M.; Dameshek, W.; Goodman, M. J.; Gilman, A.; McLENNAN, M. T., Nitrogen mustard therapy: Use of methyl-bis (beta-chloroethyl) amine hydrochloride and tris (beta-chloroethyl) amine hydrochloride for hodgkin's disease, lymphosarcoma, leukemia and certain allied and miscellaneous disorders. *Journal of the American Medical Association* **1946,** *132* (3), 126-132.
206. Gilman, A.; Philips, F. S., The biological actions and therapeutic applications of the B-chloroethyl amines and sulfides. *Sci. Technol. Weld. Joining* **1946,** *103* (2675), 409-436.
207. Rink, S. M.; Solomon, M. S.; Taylor, M. J.; Rajur, S. B.; McLaughlin, L. W.; Hopkins, P. B., Covalent structure of a nitrogen mustard-induced DNA interstrand cross-link: an N7-to-N7 linkage of deoxyguanosine residues at the duplex sequence 5'-d (GNC). *J. Am. Chem. Soc.* **1993,** *115* (7), 2551-2557.
208. Millard, J. T.; Raucher, S.; Hopkins, P. B., Mechlorethamine cross-links deoxyguanosine residues at 5'-GNC sequences in duplex DNA fragments. *J. Am. Chem. Soc.* **1990,** *112* (6), 2459-2460.
209. Ojwang, J. O.; Grueneberg, D. A.; Loechler, E. L., Synthesis of a duplex oligonucleotide containing a nitrogen mustard interstrand DNA-DNA cross-link. *Cancer Res* **1989,** *49* (23), 6529-37.
210. Costanza, M. E.; Nathanson, L.; Lenhard, R.; Wolter, J.; Colsky, J.; Oberfield, R. A.; Schilling, A., Therapy of malignant melanoma with an imidazole carboxamide and bis-chloroethyl nitrosourea. *Cancer* **1972,** *30* (6), 1457-61.
211. Prestayko, A. W.; Baker, L. H.; Crooke, S. T., *Nitrosoureas: current status and new developments*. Elsevier: 2013.
212. Crider, A. M.; Kolczynski, T. M.; Yates, K. M., Synthesis and anticancer activity of nitrosourea derivatives of phensuximide. *J. Med. Chem.* **1980,** *23* (3), 324-6.
213. Prakash, A.; Gibson, N., Sequence-selective depurination, DNA interstrand cross-linking and DNA strand break formation associated with alkylated DNA. *Carcinogenesis* **1992,** *13* (3), 425-431.
214. Tong, W. P.; Kirk, M. C.; Ludlum, D. B., Formation of the cross-link 1-[N3-deoxycytidyl),2-[N1-deoxyguanosinyl]ethane in DNA treated with N,N'-bis(2-chloroethyl)-N-nitrosourea. *Cancer Res* **1982,** *42* (8), 3102-5.
215. Noll, D. M.; Mason, T. M.; Miller, P. S., Formation and repair of interstrand cross-links in DNA. *Chem. Rev.* **2006,** *106* (2), 277-301.
216. Kartalou, M.; Essigmann, J. M., Recognition of cisplatin adducts by cellular proteins. *Mutat Res* **2001,** *478* (1-2), 1-21.
217. Malinge, J.-M.; Pérez, C.; Leng, M., Base sequence-independent distortions induced by interstrand cross-links in cis-diamminedichloroplatinum (II)-modified DNA. *Nucleic Acids Res.* **1994,** *22* (19), 3834-3839.

218. Wickramaratne, S.; Mukherjee, S.; Villalta, P. W.; Scharer, O. D.; Tretyakova, N. Y., Synthesis of sequence-specific DNA-protein conjugates via a reductive amination strategy. *Bioconjug Chem* **2013,** *24* (9), 1496-506.
219. Barker, S.; Weinfeld, M.; Murray, D., DNA–protein crosslinks: their induction, repair, and biological consequences. *Mutation Research/Reviews in Mutation Research* **2005,** *589* (2), 111-135.
220. Rohs, R.; Jin, X.; West, S. M.; Joshi, R.; Honig, B.; Mann, R. S., Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem* **2010,** *79*, 233-269.
221. Rohs, R.; Jin, X.; West, S. M.; Joshi, R.; Honig, B.; Mann, R. S., Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem* **2010,** *79*, 233-69.
222. Winnacker, M.; Breeger, S.; Strasser, R.; Carell, T., Novel diazirine-containing DNA photoaffinity probes for the investigation of DNA-protein-interactions. *Chembiochem* **2009,** *10* (1), 109-18.
223. Norris, C. L.; Meisenheimer, P. L.; Koch, T. H., Mechanistic studies of the 5-iodouracil chromophore relevant to its use in nucleoprotein photo-cross-linking. *J. Am. Chem. Soc.* **1996,** *118* (24), 5796-5803.
224. Hicke, B. J.; Willis, M. C.; Koch, T. H.; Cech, T. R., Telomeric Protein-DNA Point Contracts Identified by Photo-Crosslinking Using 5-Bromodeoxyuridine. *Biochemistry* **1994,** *33* (11), 3364-3373.
225. Liebmann, M.; Di Pasquale, F.; Marx, A., A new photoactive building block for investigation of DNA backbone interactions: photoaffinity labeling of human DNA polymerase beta. *Chembiochem* **2006,** *7* (12), 1965-9.
226. Jäger, S.; Rasched, G.; Kornreich-Leshem, H.; Engeser, M.; Thum, O.; Famulok, M., A versatile toolbox for variable DNA functionalization at high density. *J. Am. Chem. Soc.* **2005,** *127* (43), 15071-15082.
227. He, C.; Verdine, G. L., Trapping distinct structural states of a protein/DNA interaction through disulfide crosslinking. *Chemistry & biology* **2002,** *9* (12), 1297-1303.
228. Mishina, Y.; He, C., Probing the structure and function of the Escherichia coli DNA alkylation repair AlkB protein through chemical cross-linking. *J. Am. Chem. Soc.* **2003,** *125* (29), 8730-8731.
229. Ivancova, I.; Pohl, R.; Hubalek, M.; Hocek, M., Squaramate-Modified Nucleotides and DNA for Specific Cross-Linking with Lysine-Containing Peptides and Proteins. *Angew. Chem. Int. Ed. Engl.* **2019,** *58* (38), 13345-13348.
230. Li, F.; Zhang, Y.; Bai, J.; Greenberg, M. M.; Xi, Z.; Zhou, C., 5-Formylcytosine Yields DNA-Protein Cross-Links in Nucleosome Core Particles. *J. Am. Chem. Soc.* **2017,** *139* (31), 10617-10620.
231. Storer, R. I.; Aciro, C.; Jones, L. H., Squaramides: physical properties, synthesis and applications. *Chem. Soc. Rev.* **2011,** *40* (5), 2330-46.
232. Marchetti, L. A.; Kumawat, L. K.; Mao, N.; Stephens, J. C.; Elmes, R. B., The versatility of squaramides: From supramolecular chemistry to chemical biology. *Chem* **2019,** *5* (6), 1398-1485.
233. Fonvielle, M.; Sakkas, N.; Iannazzo, L.; Le Fournis, C.; Patin, D.; Mengin-Lecreulx, D.; El-Sagheer, A.; Braud, E.; Cardon, S.; Brown, T., Electrophilic RNA for Peptidyl-RNA Synthesis and Site-Specific Cross-Linking with tRNA-Binding Enzymes. *Angew. Chem. Int. Ed.* **2016,** *128* (43), 13751-13755.
234. Brazdova, M.; Palecek, J.; Cherny, D. I.; Billova, S.; Fojta, M.; Pecinka, P.; Vojtesek, B.; Jovin, T. M.; Palecek, E., Role of tumor suppressor p53 domains in selective binding to supercoiled DNA. *Nucleic Acids Res.* **2002,** *30* (22), 4966-74.
235. Krömer, M.; Brunderová, M.; Ivancová, I.; Poštová Slavětínská, L.; Hocek, M., 2-Formyl-dATP as Substrate for Polymerase Synthesis of Reactive DNA Bearing an Aldehyde Group in the Minor Groove. *ChemPlusChem* **2020,** *85* (6), 1164-1170.
236. Raiber, E. A.; Portella, G.; Martinez Cuesta, S.; Hardisty, R.; Murat, P.; Li, Z.; Iurlaro, M.; Dean, W.; Spindel, J.; Beraldi, D.; Liu, Z.; Dawson, M. A.; Reik, W.; Balasubramanian, S., 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells. *Nat Chem* **2018,** *10* (12), 1258-1266.
237. Raindlová, V.; Pohl, R.; Hocek, M., Synthesis of aldehyde-linked nucleotides and DNA and their bioconjugations with lysine and peptides through reductive amination. *Chemistry–A European Journal* **2012,** *18* (13), 4080-4087.
238. Pande, P.; Ji, S.; Mukherjee, S.; Scharer, O. D.; Tretyakova, N. Y.; Basu, A. K., Mutagenicity of a Model DNA-Peptide Cross-Link in Human Cells: Roles of Translesion Synthesis DNA Polymerases. *Chem. Res. Toxicol.* **2017,** *30* (2), 669-677.
239. Dadova, J.; Orsag, P.; Pohl, R.; Brazdova, M.; Fojta, M.; Hocek, M., Vinylsulfonamide and acrylamide modification of DNA for cross-linking with proteins. *Angew. Chem. Int. Ed. Engl.* **2013,** *52* (40), 10515-8.
240. Olszewska, A.; Pohl, R.; Brazdova, M.; Fojta, M.; Hocek, M., Chloroacetamide-Linked Nucleotides and DNA for Cross-Linking with Peptides and Proteins. *Bioconjug Chem* **2016,** *27* (9), 2089-94.
241. Leone, D. L.; Pohl, R.; Hubálek, M.; Kadeřábková, M.; Krömer, M.; Sýkorová, V.; Hocek, M., Glyoxal-Linked Nucleotides and DNA for Bioconjugations and Crosslinking with Arginine-Containing Peptides and Proteins. *Chemistry–A European Journal* **2022,** *28* (14), e202104208.

242. Wei, X.; Peng, Y.; Bryan, C.; Yang, K., Mechanisms of DNA-protein cross-link formation and repair. *Biochim Biophys Acta Proteins Proteom* **2021,** *1869* (8), 140669.
243. Leone, D. L.; Hubalek, M.; Pohl, R.; Sykorova, V.; Hocek, M., 1,3-Diketone-Modified Nucleotides and DNA for Cross-Linking with Arginine-Containing Peptides and Proteins. *Angew. Chem. Int. Ed. Engl.* **2021,** *60* (32), 17383-17387.
244. Pujari, S. S.; Wu, M.; Thomforde, J.; Wang, Z. A.; Chao, C.; Olson, N. M.; Erber, L.; Pomerantz, W. C. K.; Cole, P.; Tretyakova, N. Y., Site-Specific 5-Formyl Cytosine Mediated DNA-Histone Cross-Links: Synthesis and Polymerase Bypass by Human DNA Polymerase eta. *Angew. Chem. Int. Ed. Engl.* **2021,** *60* (51), 26489-26494.
245. Monakhova, M. V.; Kubareva, E. A.; Kolesnikov, K. K.; Anashkin, V. A.; Kosaretskiy, E. M.; Zvereva, M. I.; Romanova, E. A.; Friedhoff, P.; Oretskaya, T. S.; Zatsepin, T. S., Reactive Acrylamide-Modified DNA Traps for Accurate Cross-Linking with Cysteine Residues in DNA–Protein Complexes Using Mismatch Repair Protein MutS as a Model. *Molecules* **2022,** *27* (8), 2438.
246. Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L., A pneumonia outbreak associated with a new coronavirus of probable bat origin. *nature* **2020,** *579* (7798), 270-273.
247. Holmes, K. V., CORONAVIRUSES (CORONAVIRIDAE). *Encyclopedia of Virology* **1999**, 291-298.
248. Su, S.; Wong, G.; Shi, W.; Liu, J.; Lai, A. C.; Zhou, J.; Liu, W.; Bi, Y.; Gao, G. F., Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends in microbiology* **2016,** *24* (6), 490-502.
249. de Haan, C. A.; Rottier, P. J., Molecular interactions in the assembly of coronaviruses. *Adv Virus Res* **2005,** *64*, 165-230.
250. Boopathi, S.; Poma, A. B.; Kolandaivel, P., Novel 2019 coronavirus structure, mechanism of action, antiviral drug promises and rule out against its treatment. *J. Biomol. Struct. Dyn.* **2021,** *39* (9), 3409-3418.
251. Ke, Z.; Oton, J.; Qu, K.; Cortese, M.; Zila, V.; McKeane, L.; Nakane, T.; Zivanov, J.; Neufeldt, C. J.; Cerikan, B., Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature* **2020,** *588* (7838), 498-502.
252. Lai, M. M.; Cavanagh, D., The molecular biology of coronaviruses. *Advances in virus research* **1997,** *48*, 1-100.
253. Peiris, J.; Lai, S.; Poon, L.; Guan, Y.; Yam, L.; Lim, W.; Nicholls, J.; Yee, W.; Yan, W.; Cheung, M., Coronavirus as a possible cause of severe acute respiratory syndrome. *The Lancet* **2003,** *361* (9366), 1319-1325.
254. Raj, V. S.; Osterhaus, A. D.; Fouchier, R. A.; Haagmans, B. L., MERS: emergence of a novel human coronavirus. *Current opinion in virology* **2014,** *5*, 58-62.
255. Van Der Hoek, L.; Pyrc, K.; Jebbink, M. F.; Vermeulen-Oost, W.; Berkhout, R. J.; Wolthers, K. C.; Wertheim-van Dillen, P. M.; Kaandorp, J.; Spaargaren, J.; Berkhout, B., Identification of a new human coronavirus. *Nature medicine* **2004,** *10* (4), 368-373.
256. Robson, F.; Khan, K. S.; Le, T. K.; Paris, C.; Demirbag, S.; Barfuss, P.; Rocchi, P.; Ng, W.-L., Coronavirus RNA proofreading: molecular basis and therapeutic targeting. *Molecular cell* **2020,** *79* (5), 710-727.
257. Thiel, V.; Ivanov, K. A.; Putics, A.; Hertzig, T.; Schelle, B.; Bayer, S.; Weißbrich, B.; Snijder, E. J.; Rabenau, H.; Doerr, H. W., Mechanisms and enzymes involved in SARS coronavirus genome expression. *Journal of General Virology* **2003,** *84* (9), 2305-2315.
258. Kim, D.; Lee, J.-Y.; Yang, J.-S.; Kim, J. W.; Kim, V. N.; Chang, H., The architecture of SARS-CoV-2 transcriptome. *Cell* **2020,** *181* (4), 914-921. e10.
259. Plant, E. P.; Pérez-Alvarado, G. C.; Jacobs, J. L.; Mukhopadhyay, B.; Hennig, M.; Dinman, J. D., A three-stemmed mRNA pseudoknot in the SARS coronavirus frameshift signal. *PLoS biology* **2005,** *3* (6), e172.
260. Keum, Y.-S.; Jeong, Y.-J., Development of chemical inhibitors of the SARS coronavirus: viral helicase as a potential target. *Biochem. Pharmacol.* **2012,** *84* (10), 1351-1358.
261. Eckerle, L. D.; Becker, M. M.; Halpin, R. A.; Li, K.; Venter, E.; Lu, X.; Scherbakova, S.; Graham, R. L.; Baric, R. S.; Stockwell, T. B., Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS pathogens* **2010,** *6* (5), e1000896.
262. Hillen, H. S.; Kokic, G.; Farnung, L.; Dienemann, C.; Tegunov, D.; Cramer, P., Structure of replicating SARS-CoV-2 polymerase. *Nature* **2020,** *584* (7819), 154-156.
263. Gao, Y.; Yan, L.; Huang, Y.; Liu, F.; Zhao, Y.; Cao, L.; Wang, T.; Sun, Q.; Ming, Z.; Zhang, L., Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* **2020,** *368* (6492), 779-782.
264. Kirchdoerfer, R. N.; Ward, A. B., Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun* **2019,** *10* (1), 2342.

265. Kokic, G.; Hillen, H. S.; Tegunov, D.; Dienemann, C.; Seitz, F.; Schmitzova, J.; Farnung, L.; Siewert, A.; Höbartner, C.; Cramer, P., Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nature communications* **2021,** *12* (1), 1-7.
266. Simonis, A.; Theobald, S. J.; Fätkenheuer, G.; Rybniker, J.; Malin, J. J., A comparative analysis of remdesivir and other repurposed antivirals against SARS-CoV-2. *EMBO molecular medicine* **2021,** *13* (1), e13105.
267. Jahanshahlu, L.; Rezaei, N., Monoclonal antibody as a potential anti-COVID-19. *Biomedicine & Pharmacotherapy* **2020,** *129*, 110337.
268. Prabakaran, P.; Zhu, Z.; Xiao, X.; Biragyn, A.; Dimitrov, A. S.; Broder, C. C.; Dimitrov, D. S., Potent human monoclonal antibodies against SARS CoV, Nipah and Hendra viruses. *Expert opinion on biological therapy* **2009,** *9* (3), 355-368.
269. Wang, C.; Li, W.; Drabek, D.; Okba, N.; van Haperen, R.; Osterhaus, A. D.; van Kuppeveld, F. J.; Haagmans, B. L.; Grosveld, F.; Bosch, B.-J., A human monoclonal antibody blocking SARS-CoV-2 infection. *Nature communications* **2020,** *11* (1), 1-6.
270. Walls, A. C.; Xiong, X.; Park, Y.-J.; Tortorici, M. A.; Snijder, J.; Quispe, J.; Cameroni, E.; Gopal, R.; Dai, M.; Lanzavecchia, A., Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. *Cell* **2019,** *176* (5), 1026-1039. e15.
271. Pruijssers, A. J.; Denison, M. R., Nucleoside analogues for the treatment of coronavirus infections. *Current opinion in virology* **2019,** *35*, 57-62.
272. Sa Ribero, M.; Jouvenet, N.; Dreux, M.; Nisole, S., Interplay between SARS-CoV-2 and the type I interferon response. *PLoS pathogens* **2020,** *16* (7), e1008737.
273. Mantlo, E.; Bukreyeva, N.; Maruyama, J.; Paessler, S.; Huang, C., Antiviral activities of type I interferons to SARS-CoV-2 infection. *Antiviral research* **2020,** *179*, 104811.
274. Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet* **2020,** *395* (10223), 497-506.
275. Li, H.; Chen, C.; Hu, F.; Wang, J.; Zhao, Q.; Gale, R. P.; Liang, Y., Impact of corticosteroid therapy on outcomes of persons with SARS-CoV-2, SARS-CoV, or MERS-CoV infection: a systematic review and meta-analysis. *Leukemia* **2020,** *34* (6), 1503-1511.
276. Zhu, W.; Shyr, Z.; Lo, D. C.; Zheng, W., Viral Proteases as Targets for Coronavirus Disease 2019 Drug Development. *J. Pharmacol. Exp. Ther.* **2021,** *378* (2), 166-172.
277. Anderson, J.; Schiffer, C.; Lee, S. K.; Swanstrom, R., Viral protease inhibitors. *Handb. Exp. Pharmacol.* **2009,** (189), 85-110.
278. Yuan, S.; Yin, X.; Meng, X.; Chan, J. F.-W.; Ye, Z.-W.; Riva, L.; Pache, L.; Chan, C. C.-Y.; Lai, P.-M.; Chan, C. C.-S., Clofazimine broadly inhibits coronaviruses including SARS-CoV-2. *Nature* **2021,** *593* (7859), 418-423.
279. White, M. A.; Lin, W.; Cheng, X., Discovery of COVID-19 inhibitors targeting the SARS-CoV-2 Nsp13 helicase. *The journal of physical chemistry letters* **2020,** *11* (21), 9144-9151.
280. Wang, Y.; Anirudhan, V.; Du, R.; Cui, Q.; Rong, L., RNA-dependent RNA polymerase of SARS-CoV-2 as a therapeutic target. *Journal of medical virology* **2021,** *93* (1), 300-310.
281. Peng, Q.; Peng, R.; Yuan, B.; Wang, M.; Zhao, J.; Fu, L.; Qi, J.; Shi, Y., Structural basis of SARS-CoV-2 polymerase inhibition by Favipiravir. *The Innovation* **2021,** *2* (1), 100080.
282. Warren, T. K.; Jordan, R.; Lo, M. K.; Ray, A. S.; Mackman, R. L.; Soloveva, V.; Siegel, D.; Perron, M.; Bannister, R.; Hui, H. C., Therapeutic efficacy of the small molecule GS-5734 against Ebola virus in rhesus monkeys. *Nature* **2016,** *531* (7594), 381-385.
283. Drożdżal, S.; Rosik, J.; Lechowicz, K.; Machaj, F.; Kotfis, K.; Ghavami, S.; Łos, M. J., FDA approved drugs with pharmacotherapeutic potential for SARS-CoV-2 (COVID-19) therapy. *Drug resistance updates* **2020,** *53*, 100719.
284. Food, U.; Administration, D. Coronavirus (COVID-19) update: FDA authorizes first oral antiviral for treatment of COVID-19. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-first-oral-antiviral-treatment-covid-19.
285. Rubin, D.; Chan-Tack, K.; Farley, J.; Sherwat, A., FDA approval of remdesivir—a step in the right direction. *New England Journal of Medicine* **2020,** *383* (27), 2598-2600.
286. Grein, J.; Ohmagari, N.; Shin, D.; Diaz, G.; Asperges, E.; Castagna, A.; Feldt, T.; Green, G.; Green, M. L.; Lescure, F.-X., Compassionate use of remdesivir for patients with severe Covid-19. *New England Journal of Medicine* **2020,** *382* (24), 2327-2336.
287. Sheahan, T. P.; Sims, A. C.; Graham, R. L.; Menachery, V. D.; Gralinski, L. E.; Case, J. B.; Leist, S. R.; Pyrc, K.; Feng, J. Y.; Trantcheva, I., Broad-spectrum antiviral GS-5734 inhibits both epidemic and zoonotic coronaviruses. *Science translational medicine* **2017,** *9* (396), eaal3653.

288. de Wit, E.; Feldmann, F.; Cronin, J.; Jordan, R.; Okumura, A.; Thomas, T.; Scott, D.; Cihlar, T.; Feldmann, H., Prophylactic and therapeutic remdesivir (GS-5734) treatment in the rhesus macaque model of MERS-CoV infection. *Proceedings of the National Academy of Sciences* **2020,** *117* (12), 6771-6776.
289. Mulangu, S.; Dodd, L. E.; Davey Jr, R. T.; Tshiani Mbaya, O.; Proschan, M.; Mukadi, D.; Lusakibanza Manzo, M.; Nzolo, D.; Tshomba Oloma, A.; Ibanda, A., A randomized, controlled trial of Ebola virus disease therapeutics. *New England journal of medicine* **2019,** *381* (24), 2293-2303.
290. Lo, M. K.; Jordan, R.; Arvey, A.; Sudhamsu, J.; Shrivastava-Ranjan, P.; Hotard, A. L.; Flint, M.; McMullan, L. K.; Siegel, D.; Clarke, M. O., GS-5734 and its parent nucleoside analog inhibit Filo-, Pneumo-, and Paramyxoviruses. *Scientific reports* **2017,** *7* (1), 1-7.
291. Chen, J.; Malone, B.; Llewellyn, E.; Grasso, M.; Shelton, P. M.; Olinares, P. D. B.; Maruthi, K.; Eng, E. T.; Vatandaslar, H.; Chait, B. T., Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication-transcription complex. *Cell* **2020,** *182* (6), 1560-1573. e13.
292. Agostini, M. L.; Andres, E. L.; Sims, A. C.; Graham, R. L.; Sheahan, T. P.; Lu, X.; Smith, E. C.; Case, J. B.; Feng, J. Y.; Jordan, R., Coronavirus susceptibility to the antiviral remdesivir (GS-5734) is mediated by the viral polymerase and the proofreading exoribonuclease. *MBio* **2018,** *9* (2), e00221-18.
293. Julander, J. G.; Siddharthan, V.; Evans, J.; Taylor, R.; Tolbert, K.; Apuli, C.; Stewart, J.; Collins, P.; Gebre, M.; Neilson, S., Efficacy of the broad-spectrum antiviral compound BCX4430 against Zika virus in cell culture and in a mouse model. *Antiviral research* **2017,** *137*, 14-22.
294. Warren, T. K.; Wells, J.; Panchal, R. G.; Stuthman, K. S.; Garza, N. L.; Van Tongeren, S. A.; Dong, L.; Retterer, C. J.; Eaton, B. P.; Pegoraro, G.; Honnold, S.; Bantia, S.; Kotian, P.; Chen, X.; Taubenheim, B. R.; Welch, L. S.; Minning, D. M.; Babu, Y. S.; Sheridan, W. P.; Bavari, S., Protection against filovirus diseases by a novel broad-spectrum nucleoside analogue BCX4430. *Nature* **2014,** *508* (7496), 402-5.
295. De Clercq, E., Ebola virus (EBOV) infection: therapeutic strategies. *Biochem. Pharmacol.* **2015,** *93* (1), 1-10.
296. Wong, G.; Qiu, X.; Olinger, G. G.; Kobinger, G. P., Post-exposure therapy of filovirus infections. *Trends in microbiology* **2014,** *22* (8), 456-463.
297. Julander, J. G.; Demarest, J. F.; Taylor, R.; Gowen, B. B.; Walling, D. M.; Mathis, A.; Babu, Y., An update on the progress of galidesivir (BCX4430), a broad-spectrum antiviral. *Antiviral research* **2021,** *195*, 105180.
298. Elfiky, A. A., Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): A molecular docking study. *Life Sci.* **2020,** *253*, 117592.
299. Perry, C. M.; Faulds, D., Valaciclovir. A review of its antiviral activity, pharmacokinetic properties and therapeutic efficacy in herpesvirus infections. *Drugs* **1996,** *52* (5), 754-72.
300. Leflore, S.; Anderson, P. L.; Fletcher, C. V., A risk-benefit evaluation of aciclovir for the treatment and prophylaxis of herpes simplex virus infections. *Drug Safety* **2000,** *23* (2), 131-142.
301. Hamza, R. Z.; Al-Talhi, T.; Gobouri, A. A.; Alsanie, W. F.; El-Megharbel, S. M., Are favipiravir and acyclovir with IgG injections supplemented with vitamin D "suggested therapeutic option" can fight against COVID-19. *Adv. Anim. Vet. Sci* **2021,** *9* (4), 549-554.
302. Peters, H. L.; Jochmans, D.; De Wilde, A. H.; Posthuma, C. C.; Snijder, E. J.; Neyts, J.; Seley-Radtke, K. L., Design, synthesis and evaluation of a series of acyclic fleximer nucleoside analogues with anti-coronavirus activity. *Bioorg. Med. Chem. Lett.* **2015,** *25* (15), 2923-2926.
303. Malinoski, F.; Stollar, V., Inhibitors of IMP dehydrogenase prevent Sindbis virus replication and reduce GTP levels in Aedes albopictus cells. *Virology* **1981,** *110* (2), 281-291.
304. Te, H. S.; Randall, G.; Jensen, D. M., Mechanism of action of ribavirin in the treatment of chronic hepatitis C. *Gastroenterology & hepatology* **2007,** *3* (3), 218.
305. Lau, J. Y.; Tam, R. C.; Liang, T. J.; Hong, Z., Mechanism of action of ribavirin in the combination treatment of chronic HCV infection. *Hepatology* **2002,** *35* (5), 1002-1009.
306. Wilson, S. Z.; Knight, V.; Wyde, P. R.; Drake, S.; Couch, R. B., Amantadine and ribavirin aerosol treatment of influenza A and B infection in mice. *Antimicrob. Agents Chemother.* **1980,** *17* (4), 642-648.
307. Morgenstern, B.; Michaelis, M.; Baer, P. C.; Doerr, H. W.; Cinatl Jr, J., Ribavirin and interferon-β synergistically inhibit SARS-associated coronavirus replication in animal and human cell lines. *Biochem. Biophys. Res. Commun.* **2005,** *326* (4), 905-908.
308. Crotty, S.; Maag, D.; Arnold, J. J.; Zhong, W.; Lau, J. Y.; Hong, Z.; Andino, R.; Cameron, C. E., The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. *Nature medicine* **2000,** *6* (12), 1375-1379.
309. Tong, S.; Su, Y.; Yu, Y.; Wu, C.; Chen, J.; Wang, S.; Jiang, J., Ribavirin therapy for severe COVID-19: a retrospective cohort study. *International journal of antimicrobial agents* **2020,** *56* (3), 106114.

310. Gan, L.; Seyedsayamdost, M. R.; Shuto, S.; Matsuda, A.; Petsko, G. A.; Hedstrom, L., The immunosuppressive agent mizoribine monophosphate forms a transition state analogue complex with inosine monophosphate dehydrogenase. *Biochemistry* **2003,** *42* (4), 857-863.
311. Saijo, M.; Morikawa, S.; Fukushi, S.; Mizutani, T.; Hasegawa, H.; Nagata, N.; Iwata, N.; Kurane, I., Inhibitory effect of mizoribine and ribavirin on the replication of severe acute respiratory syndrome (SARS)-associated coronavirus. *Antiviral Res* **2005,** *66* (2-3), 159-63.
312. Jakovac, H., COVID-19 and hypertension: is the HSP60 culprit for the severe course and worse outcome? *American Journal of Physiology-Heart Circulatory Physiology* **2020,** *319* (4), H793-H796.
313. Cinatl, J.; Morgenstern, B.; Bauer, G.; Chandra, P.; Rabenau, H.; Doerr, H. W., Glycyrrhizin, an active component of liquorice roots, and replication of SARS-associated coronavirus. *Lancet* **2003,** *361* (9374), 2045-6.
314. Muhseen, Z. T.; Hameed, A. R.; Al-Hasani, H. M. H.; Ahmad, S.; Li, G., Computational Determination of Potential Multiprotein Targeting Natural Compounds for Rational Drug Design Against SARS-COV-2. *Molecules* **2021,** *26* (3), 674.
315. Barton-Burke, M., Gemcitabine: a pharmacologic and clinical overview. *Cancer Nurs* **1999,** *22* (2), 176-83.
316. Stephenson, A. A.; Cao, S.; Taggart, D. J.; Vyavahare, V. P.; Suo, Z., Design, synthesis, and evaluation of liver-specific gemcitabine prodrugs for potential treatment of hepatitis C virus infection and hepatocellular carcinoma. *Eur. J. Med. Chem.* **2021,** *213*, 113135.
317. Kuivanen, S.; Bespalov, M. M.; Nandania, J.; Ianevski, A.; Velagapudi, V.; De Brabander, J. K.; Kainov, D. E.; Vapalahti, O., Obatoclax, saliphenylhalamide and gemcitabine inhibit Zika virus infection in vitro and differentially affect cellular signaling, transcription and metabolism. *Antiviral Res* **2017,** *139*, 117-128.
318. Denisova, O. V.; Kakkola, L.; Feng, L.; Stenman, J.; Nagaraj, A.; Lampe, J.; Yadav, B.; Aittokallio, T.; Kaukinen, P.; Ahola, T.; Kuivanen, S.; Vapalahti, O.; Kantele, A.; Tynell, J.; Julkunen, I.; Kallio-Kokko, H.; Paavilainen, H.; Hukkanen, V.; Elliott, R. M.; De Brabander, J. K.; Saelens, X.; Kainov, D. E., Obatoclax, saliphenylhalamide, and gemcitabine inhibit influenza a virus infection. *J. Biol. Chem.* **2012,** *287* (42), 35324-35332.
319. Clouser, C. L.; Holtz, C. M.; Mullett, M.; Crankshaw, D. L.; Briggs, J. E.; O'Sullivan, M. G.; Patterson, S. E.; Mansky, L. M., Activity of a novel combined antiretroviral therapy of gemcitabine and decitabine in a mouse model for HIV-1. *Antimicrob. Agents Chemother.* **2012,** *56* (4), 1942-8.
320. Dyall, J.; Coleman, C. M.; Hart, B. J.; Venkataraman, T.; Holbrook, M. R.; Kindrachuk, J.; Johnson, R. F.; Olinger, G. G., Jr.; Jahrling, P. B.; Laidlaw, M.; Johansen, L. M.; Lear-Rooney, C. M.; Glass, P. J.; Hensley, L. E.; Frieman, M. B., Repurposing of clinically developed drugs for treatment of Middle East respiratory syndrome coronavirus infection. *Antimicrob. Agents Chemother.* **2014,** *58* (8), 4885-93.
321. Zhang, Y. N.; Zhang, Q. Y.; Li, X. D.; Xiong, J.; Xiao, S. Q.; Wang, Z.; Zhang, Z. R.; Deng, C. L.; Yang, X. L.; Wei, H. P.; Yuan, Z. M.; Ye, H. Q.; Zhang, B., Gemcitabine, lycorine and oxysophoridine inhibit novel coronavirus (SARS-CoV-2) in cell culture. *Emerg Microbes Infect* **2020,** *9* (1), 1170-1173.
322. Lee, K.; Kim, D.-E.; Jang, K.-S.; Kim, S.-J.; Cho, S.; Kim, C., Gemcitabine, a broad-spectrum antiviral drug, suppresses enterovirus infections through innate immunity induced by the inhibition of pyrimidine biosynthesis and nucleotide depletion. *Oncotarget* **2017,** *8* (70), 115315.
323. Tejero, H.; Montero, F.; Nuño, J. C., Theories of lethal mutagenesis: from error catastrophe to lethal defection. *Quasispecies: From Theory to Experimental Systems* **2015**, 161-179.
324. Nagata, T.; Lefor, A. K.; Hasegawa, M.; Ishii, M., Favipiravir: a new medication for the Ebola virus disease pandemic. *Disaster medicine and public health preparedness* **2015,** *9* (1), 79-81.
325. Furuta, Y.; Gowen, B. B.; Takahashi, K.; Shiraki, K.; Smee, D. F.; Barnard, D. L., Favipiravir (T-705), a novel viral RNA polymerase inhibitor. *Antiviral research* **2013,** *100* (2), 446-454.
326. Rocha-Pereira, J.; Jochmans, D.; Dallmeier, K.; Leyssen, P.; Nascimento, M.; Neyts, J., Favipiravir (T-705) inhibits in vitro norovirus replication. *Biochem. Biophys. Res. Commun.* **2012,** *424* (4), 777-780.
327. Kim, J.; Seong, R.-K.; Kumar, M.; Shin, O. S., Favipiravir and ribavirin inhibit replication of Asian and African strains of Zika virus in different cell models. *Viruses* **2018,** *10* (2), 72.
328. Shannon, A.; Selisko, B.; Le, N.; Huchting, J.; Touret, F.; Piorkowski, G.; Fattorini, V.; Ferron, F.; Decroly, E.; Meier, C., Favipiravir strikes the SARS-CoV-2 at its Achilles heel, the RNA polymerase. *BioRxiv* **2020**.
329. Agudo, R.; Arias, A.; Domingo, E., 5-fluorouracil in lethal mutagenesis of foot-and-mouth disease virus. *Future Med Chem* **2009,** *1* (3), 529-39.
330. Smith, E. C.; Case, J. B.; Blanc, H.; Isakov, O.; Shomron, N.; Vignuzzi, M.; Denison, M. R., Mutations in coronavirus nonstructural protein 10 decrease virus replication fidelity. *J Virol* **2015,** *89* (12), 6418-26.

331. Dapp, M. J.; Clouser, C. L.; Patterson, S.; Mansky, L. M., 5-Azacytidine can induce lethal mutagenesis in human immunodeficiency virus type 1. *Journal of virology* **2009,** *83* (22), 11950-11958.
332. Toots, M.; Yoon, J.-J.; Hart, M.; Natchus, M. G.; Painter, G. R.; Plemper, R. K., Quantitative efficacy paradigms of the influenza clinical drug candidate EIDD-2801 in the ferret model. *Translational Research* **2020,** *218*, 16-28.
333. Stuyver, L. J.; Whitaker, T.; McBrayer, T. R.; Hernandez-Santiago, B. I.; Lostia, S.; Tharnish, P. M.; Ramesh, M.; Chu, C. K.; Jordan, R.; Shi, J., Ribonucleoside analogue that blocks replication of bovine viral diarrhea and hepatitis C viruses in culture. *Antimicrob. Agents Chemother.* **2003,** *47* (1), 244-254.
334. Reynard, O.; Nguyen, X.-N.; Alazard-Dany, N.; Barateau, V.; Cimarelli, A.; Volchkov, V. E., Identification of a new ribonucleoside inhibitor of Ebola virus replication. *Viruses* **2015,** *7* (12), 6233-6240.
335. Painter, G. R.; Bowen, R. A.; Bluemling, G. R.; DeBergh, J.; Edpuganti, V.; Gruddanti, P. R.; Guthrie, D. B.; Hager, M.; Kuiper, D. L.; Lockwood, M. A., The prophylactic and therapeutic activity of a broadly active ribonucleoside analog in a murine model of intranasal venezuelan equine encephalitis virus infection. *Antiviral research* **2019,** *171*, 104597.
336. Barnard, D. L.; Hubbard, V. D.; Burton, J.; Smee, D. F.; Morrey, J. D.; Otto, M. J.; Sidwell, R. W., Inhibition of severe acute respiratory syndrome-associated coronavirus (SARSCoV) by calpain inhibitors and beta-D-N4-hydroxycytidine. *Antivir Chem Chemother* **2004,** *15* (1), 15-22.
337. Cox, R. M.; Wolf, J. D.; Plemper, R. K., Therapeutically administered ribonucleoside analogue MK-4482/EIDD-2801 blocks SARS-CoV-2 transmission in ferrets. *Nature microbiology* **2021,** *6* (1), 11-18.
338. Painter, W. P.; Holman, W.; Bush, J. A.; Almazedi, F.; Malik, H.; Eraut, N. C.; Morin, M. J.; Szewczyk, L. J.; Painter, G. R., Human safety, tolerability, and pharmacokinetics of molnupiravir, a novel broad-spectrum oral antiviral agent with activity against SARS-CoV-2. *Antimicrob. Agents Chemother.* **2021,** *65* (5), e02428-20.
339. Sheahan, T. P.; Sims, A. C.; Zhou, S.; Graham, R. L.; Pruijssers, A. J.; Agostini, M. L.; Leist, S. R.; Schafer, A.; Dinnon, K. H., 3rd; Stevens, L. J.; Chappell, J. D.; Lu, X.; Hughes, T. M.; George, A. S.; Hill, C. S.; Montgomery, S. A.; Brown, A. J.; Bluemling, G. R.; Natchus, M. G.; Saindane, M.; Kolykhalov, A. A.; Painter, G.; Harcourt, J.; Tamin, A.; Thornburg, N. J.; Swanstrom, R.; Denison, M. R.; Baric, R. S., An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. *Sci Transl Med* **2020,** *12* (541), eabb5883.
340. Flavell, R.; Sabo, D.; Bandle, E.; Weissmann, C., Site-directed mutagenesis: Generation of an extracistronic mutation in bacteriophage Qβ RNA. *J. Mol. Biol.* **1974,** *89* (2), 255-272.
341. Hernandez-Santiago, B. I.; Beltran, T.; Stuyver, L.; Chu, C. K.; Schinazi, R. F., Metabolism of the anti-hepatitis C virus nucleoside β-d-N 4-hydroxycytidine in different liver cells. *Antimicrob. Agents Chemother.* **2004,** *48* (12), 4636-4642.
342. Yoon, J.-J.; Toots, M.; Lee, S.; Lee, M.-E.; Ludeke, B.; Luczo, J. M.; Ganti, K.; Cox, R. M.; Sticher, Z. M.; Edpuganti, V., Orally efficacious broad-spectrum ribonucleoside analog inhibitor of influenza and respiratory syncytial viruses. *Antimicrob. Agents Chemother.* **2018,** *62* (8), e00766-18.
343. Sticher, Z. M.; Lu, G.; Mitchell, D. G.; Marlow, J.; Moellering, L.; Bluemling, G. R.; Guthrie, D. B.; Natchus, M. G.; Painter, G. R.; Kolykhalov, A. A., Analysis of the potential for N 4-hydroxycytidine to inhibit mitochondrial replication and function. *Antimicrob. Agents Chemother.* **2020,** *64* (2), e01719-19.
344. Shugar, D.; Kierdaszuk, B., New light on tautomerism of purines and pyrimidines and its biological and genetic implications. *Journal of Biosciences* **1985,** *8* (3), 657-668.
345. Brown, D. M.; Hewlins, M. J.; Schell, P., The tautomeric state of N(4)-hydroxy- and of N(4)-amino-cytosine derivatives. *J Chem Soc Perkin 1* **1968,** *15*, 1925-9.
346. Janion, C., Hydroxylamine mutagenesis: ultraviolet absorption spectra and tautomeric forms of N4-hydroxy, N4-methoxy and N4-methyl, N4-hydroxy derivatives of cytosine, and 5-and 6-methylcytosines. *Acta Biochim. Pol.* **1972,** *19* (3).
347. Kierdaszuk, B.; Shugar, D., Structure of the planar complex of N4-methoxycytosine with adenine, and its relevance to the mechanism of hydroxylamine mutagenesis. *Biophys. Chem.* **1983,** *17* (4), 285-95.
348. KIERDASZUK, B.; STOLARSKI, R.; SHUGAR, D., Hydroxylamine Mutagenesis: Observation of Inverted Watson-Crick Base-Pairing Between N4-Methoxycytosine and Adenine with the Aid of Natural-Abundance High-Resolution 15N NMR Spectroscopy. *Eur. J. Biochem.* **1983,** *130* (3), 559-564.
349. Kulinska, K.; Psoda, A.; Shugar, D., Mechanism of hydroxylamine mutagenesis: an infrared study of the association in non-polar solutions of 5-methyl-N4-hydroxycytosines. *Acta Biochim. Pol.* **1980,** *27* (1), 57-65.
350. Shugar, D.; Huber, C. P.; Birnbaum, G. I., Mechanism of hydroxylamine mutagenesis. Crystal structure and conformation of 1,5-dimethyl-N4-hydroxycytosine. *Biochim. Biophys. Acta* **1976,** *447* (3), 274-84.

351. Leś, A.; Adamowicz, L.; Rode, W., Structure and conformation of N4-hydroxycytosine and N4-hydroxy-5-fluorocytosine. A theoretical ab initio study. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression* **1993,** *1173* (1), 39-48.
352. Sharov, A. V.; Burkhanova, T. M.; Taskın Tok, T.; Babashkina, M. G.; Safin, D. A., Computational analysis of molnupiravir. *International Journal of Molecular Sciences* **2022,** *23* (3), 1508.
353. Fazakerley, G. V.; Gdaniec, Z.; Sowers, L. C., Base-pair induced shifts in the tautomeric equilibrium of a modified DNA base. *J. Mol. Biol.* **1993,** *230* (1), 6-10.
354. Nedderman, A. N.; Stone, M. J.; Williams, D. H.; Lin, P. K.; Brown, D. M., Molecular basis for methoxyamine-initiated mutagenesis: 1H nuclear magnetic resonance studies of oligonucleotide duplexes containing base-modified cytosine residues. *J. Mol. Biol.* **1993,** *230* (3), 1068-76.
355. Stone, M. J.; Nedderman, A. N.; Williams, D. H.; Lin, P. K. T.; Brown, D. M., Molecular basis for methoxyamine initiated mutagenesis: 1H nuclear magnetic resonance studies of base-modified oligodeoxynucleotides. *J. Mol. Biol.* **1991,** *222* (3), 711-723.
356. Caraco, Y.; Crofoot, G. E.; Moncada, P. A.; Galustyan, A. N.; Musungaie, D. B.; Payne, B.; Kovalchuk, E.; Gonzalez, A.; Brown, M. L.; Williams-Diaz, A., Phase 2/3 trial of molnupiravir for treatment of Covid-19 in nonhospitalized adults. *NEJM Evidence* **2022,** *1* (2), EVIDoa2100043.
357. Jayk Bernal, A.; Gomes da Silva, M. M.; Musungaie, D. B.; Kovalchuk, E.; Gonzalez, A.; Delos Reyes, V.; Martín-Quirós, A.; Caraco, Y.; Williams-Diaz, A.; Brown, M. L., Molnupiravir for oral treatment of Covid-19 in nonhospitalized patients. *New England Journal of Medicine* **2022,** *386* (6), 509-520.
358. Merck, Merck and Ridgeback's Investigational Oral Antiviral Molnupiravir Reduced the Risk of Hospitalization or Death by Approximately 50 Percent Compared to Placebo for Patients with Mild or Moderate COVID-19 in Positive Interim Analysis of Phase 3 Study. 2021.
359. Merck, Merck and Ridgeback Biotherapeutics Provide Update on Results from MOVe-OUT Study of Molnupiravir, an Investigational Oral Antiviral Medicine, in At Risk Adults With Mild-to-Moderate COVID-19 2021.
360. FDA, Coronavirus (COVID-19) Update: FDA Authorizes Additional Oral Antiviral for Treatment of COVID-19 in Certain Adults. 2021.
361. Painter, G. R.; BLUEMLING, G. R.; Natchus, M. G.; Guthrie, D., N4-hydroxycytidine and derivatives and anti-viral uses related thereto. Google Patents: 2019.
362. Steiner, A.; Znidar, D.; Otvos, S. B.; Snead, D. R.; Dallinger, D.; Kappe, C. O., A High-Yielding Synthesis of EIDD-2801 from Uridine. *European J Org Chem* **2020,** *2020* (43), 6736-6739.
363. Vasudevan, N.; Ahlqvist, G. P.; McGeough, C. P.; Paymode, D. J.; Cardoso, F. S. P.; Lucas, T.; Dietz, J. P.; Opatz, T.; Jamison, T. F.; Gupton, F. B.; Snead, D. R., A concise route to MK-4482 (EIDD-2801) from cytidine. *Chem Commun (Camb)* **2020,** *56* (87), 13363-13364.
364. Gopalsamuthiram, V.; Williams, C.; Noble, J.; Jamison, T. F.; Gupton, B. F.; Snead, D. R., A Concise Route to MK-4482 (EIDD-2801) from Cytidine: Part 2. *Synlett* **2021,** *32* (03), 326-328.
365. Hou, B.; Zhang, Y.-M.; Liao, H.-Y.; Fu, L.-F.; Li, D.-D.; Zhao, X.; Qi, J.-X.; Yang, W.; Xiao, G.-F.; Yang, L., Target-Based Virtual Screening and LC/MS-Guided Isolation Procedure for Identifying Phloroglucinol-Terpenoid Inhibitors of SARS-CoV-2. *J. Nat. Prod.* **2022**.
366. Brookes, P.; Lawley, P., 254. The methylation of cystosine and cytidine. *Journal of the Chemical Society (Resumed)* **1962**, 1348-1351.
367. Zhu, X.-F.; Williams Jr, H. J.; Ian Scott, A., An improved transient method for the synthesis of N-benzoylated nucleosides. *Synth. Commun.* **2003,** *33* (7), 1233-1243.
368. Smith, M.; Rammler, D.; Goldberg, I.; Khorana, H., Studies on Polynucleotides. XIV. 1 Specific Synthesis of the C3 ″-C5 ″Interribonucleotide Linkage. Syntheses of Uridylyl-(3 ″→ 5 ″)-Uridine and Uridylyl-(3 ″→ 5 ″)-Adenosine 2. *J. Am. Chem. Soc.* **1962,** *84* (3), 430-440.
369. Pitsch, S.; Weiss, P. A.; Jenny, L.; Stutz, A.; Wu, X., Reliable chemical synthesis of oligoribonucleotides (RNA) with 2′-O-[(triisopropylsilyl) oxy] methyl (2′-O-tom)-protected phosphoramidites. *Helv. Chim. Acta* **2001,** *84* (12), 3773-3795.
370. Mao, S.; Haruehanroengra, P.; Ranganathan, S. V.; Shen, F.; Begley, T. J.; Sheng, J., Base Pairing and Functional Insights into N 3-Methylcytidine (m3C) in RNA. *ACS Chemical Biology* **2020,** *16* (1), 76-85.
371. Moreno, S.; Flemmich, L.; Micura, R., Synthesis of N4-acetylated 3-methylcytidine phosphoramidites for RNA solid-phase synthesis. *Monatshefte für Chemie-Chemical Monthly* **2022,** *153* (3), 285-291.
372. Höbartner, C.; Kreutz, C.; Flecker, E.; Ottenschläger, E.; Pils, W.; Grubmayr, K.; Micura, R., The synthesis of 2′-O-[(triisopropylsilyl) oxy] methyl (TOM) phosphoramidites of methylated ribonucleosides (m 1 G, m 2 G, m 2 2 G, m 1 I, m 3 U, m 4 C, m 6 A, m 6 2 A) for use in automated RNA solid-phase synthesis. *Monatshefte für Chemie/Chemical Monthly* **2003,** *134* (6), 851-873.

373. Zhang, X. H.; Xu, Y. Z., NMR studies on 4-thio-5-furan-modified and 4-thio-5-thiophene-modified nucleosides. *Magn. Reson. Chem.* **2016,** *54* (11), 887-892.
374. Büttner, L.; Seikowski, J.; Wawrzyniak, K.; Ochmann, A.; Höbartner, C., Synthesis of spin-labeled riboswitch RNAs using convertible nucleosides and DNA-catalyzed RNA ligation. *Biorg. Med. Chem.* **2013,** *21* (20), 6171-6180.
375. Yu, J.; Zhang, W.; Li, K., Synthesis of N^ 4-alkyl-5-methyl-2'-deoxycytidines and their incorporation into oligonucleotides. *CHINESE JOURNAL OF MEDICINAL CHEMISTRY* **2007,** *17* (4), 199.
376. Allerson, C. R.; Chen, S. L.; Verdine, G. L., A chemical method for site-specific modification of RNA: the convertible nucleoside approach. *J. Am. Chem. Soc.* **1997,** *119* (32), 7423-7433.
377. Tararov, V. I.; Kolyachkina, S. V.; Alexeev, C. S.; Mikhailov, S. N., N6-Acetyl-2′, 3′, 5′-tri-O-acetyladenosine; a convenient,'Missed Out'substrate for regioselective N6-alkylations. *Synthesis* **2011,** *2011* (15), 2483-2489.
378. Fleysher, M.; Hakala, M.; Bloch, A.; Hall, R., Synthesis and biological activity of some N6-alkyladenosines. *J. Med. Chem.* **1968,** *11* (4), 717-720.
379. Fleysher, M., N6-Substituted adenosines. Synthesis, biological activity, and some structure-activity relations. *J. Med. Chem.* **1972,** *15* (2), 187-191.
380. Timofeev, E. N.; Mikhailov, S. N.; Zuev, A. N.; Efimtseva, E. V.; Herdewijn, P.; Somers, R. L.; Lemaitre, M. M., Oligodeoxynucleotides containing 2′-deoxy-1-methyladenosine and Dimroth rearrangement. *Helv. Chim. Acta* **2007,** *90* (5), 928-937.
381. Vorbrüggen, H., Einfache neue Synthese von N6-substituierten Adenosinen und 2-Amino-adenosinen. *Angew. Chem.* **1972,** *84* (8), 347-348.
382. Wan, Z.-K.; Binnun, E.; Wilson, D. P.; Lee, J., A highly facile and efficient one-step synthesis of N 6-adenosine and N 6-2 '-deoxyadenosine derivatives. *Org. Lett.* **2005,** *7* (26), 5877-5880.
383. Lescrinier, E.; Pannecouque, C.; Rozenski, J.; Van Aerschot, A.; Kerremans, L.; Herdewijn, P., Synthesis of N 6-alkylated adenosine derivatives. *Nucleosides, Nucleotides & Nucleic Acids* **1996,** *15* (11-12), 1863-1869.
384. Zhang, J.; Chang, H.-M.; Kane, R. R., Reductive monoalkylation of aromatic amines via amidine intermediates. *Synlett* **2001,** *2001* (05), 0643-0645.
385. Aritomo, K.; Wada, T.; Sekine, M., Alkylation of 6-N-acylated adenosine derivatives by the use of phase transfer catalysis. *J. Chem. Soc., Perkin Trans. 1* **1995,** (14), 1837-1844.
386. Porcher, S.; Pitsch, S., Synthesis of 2′-O-[(triisopropylsilyl) oxy] methyl (= tom)-protected ribonucleoside phosphoramidites containing various nucleobase analogues. *Helv. Chim. Acta* **2005,** *88* (10), 2683-2704.
387. Kierzek, E.; Kierzek, R., The synthesis of oligoribonucleotides containing N 6-alkyladenosines and 2-methylthio-N 6-alkyladenosines via post-synthetic modification of precursor oligomers. *Nucleic Acids Res.* **2003,** *31* (15), 4461-4471.
388. Reese, C. B., Protection of 2′-hydroxy functions of ribonucleosides. *Current Protocols in Nucleic Acid Chemistry* **2000,** (1), 2.2. 1-2.2. 24.
389. Amarnath, V.; Broom, A., Chemical synthesis of oligonucleotides. *Chem. Rev.* **1977,** *77* (2), 183-217.
390. Ghosh, S.; Khan, T. S.; Ghosh, A.; Chowdhury, A. H.; Haider, M. A.; Khan, A.; Islam, S. M., Utility of silver nanoparticles embedded covalent organic frameworks as recyclable catalysts for the sustainable synthesis of cyclic carbamates and 2-oxazolidinones via atmospheric cyclizative CO2 capture. *ACS Sustainable Chemistry & Engineering* **2020,** *8* (14), 5495-5513.
391. Kimura-Harada, F.; Von Minden, D.; McCloskey, J. A.; Nishimura, S., N-[(9-β-D-Ribofuranosylpurin-6-yl)-N-methylcarbamoyl] threonine, a modified nucleoside isolated from Escherichia coli threonin transfer ribonucleic acid. *Biochem. Cell Biol.* **1972,** *11* (21), 3910-3915.
392. Debiec, K.; Sochacka, E., Efficient access to 3'-O-phosphoramidite derivatives of tRNA related N (6)-threonylcarbamoyladenosine (t(6)A) and 2-methylthio-N (6)-threonylcarbamoyladenosine (ms(2)t(6)A). *RSC Adv* **2021,** *11* (4), 1992-1999.
393. Miyauchi, K.; Kimura, S.; Suzuki, T., A cyclic form of N6-threonylcarbamoyladenosine as a widely distributed tRNA hypermodification. *Nature chemical biology* **2013,** *9* (2), 105-111.
394. Matuszewski, M.; Debiec, K.; Sochacka, E., Efficient conversion of N 6-threonylcarbamoyladenosine (t 6 A) into a tRNA native hydantoin cyclic form (ct 6 A) performed at nucleoside and oligoribonucleotide levels. *Chem. Commun.* **2017,** *53* (56), 7945-7948.
395. Matuszewski, M.; Wojciechowski, J.; Miyauchi, K.; Gdaniec, Z.; Wolf, W. M.; Suzuki, T.; Sochacka, E., A hydantoin isoform of cyclic N 6-threonylcarbamoyladenosine (ct6A) is present in tRNAs. *Nucleic Acids Res.* **2017,** *45* (4), 2137-2149.
396. Kang, B.-i.; Miyauchi, K.; Matuszewski, M.; D'Almeida, G. S.; Rubio, M. A. T.; Alfonzo, J. D.; Inoue, K.; Sakaguchi, Y.; Suzuki, T.; Sochacka, E., Identification of 2-methylthio cyclic N 6-

threonylcarbamoyladenosine (ms2ct6A) as a novel RNA modification at position 37 of tRNAs. *Nucleic Acids Res.* **2017,** *45* (4), 2124-2136.
397. Nagao, A.; Ohara, M.; Miyauchi, K.; Yokobori, S. I.; Yamagishi, A.; Watanabe, K.; Suzuki, T., Hydroxylation of a conserved tRNA modification establishes non-universal genetic code in echinoderm mitochondria. *Nat Struct Mol Biol* **2017,** *24* (9), 778-782.
398. Chheda, G. B.; Hong, C. I., Synthesis of naturally occurring 6-ureidopurines and their nucleosides. *J. Med. Chem.* **1971,** *14* (8), 748-753.
399. Boudou, V.; Langridge, J.; Van Aerschot, A.; Hendrix, C.; Millar, A.; Weiss, P.; Herdewijn, P., Synthesis of the Anticodon Hairpin tRNAfMet Containing N-{[9-(β-D-Ribofuranosyl)-9H-purin-6-yl] carbamoyl}-L-threonine (= N6-{{[(1S, 2R)-1-Carboxy-2-hydroxypropyl] amino} carbonyl} adenosine, t6A). *Helv. Chim. Acta* **2000,** *83* (1), 152-161.
400. Sundaram, M.; Crain, P. F.; Davis, D. R., Synthesis and Characterization of the Native Anticodon Domain of E. c oli tRNALys: Simultaneous Incorporation of Modified Nucleosides mnm5s2U, t6A, and Pseudouridine Using Phosphoramidite Chemistry. *The Journal of Organic Chemistry* **2000,** *65* (18), 5609-5614.
401. Nainytė, M.; Müller, F.; Ganazzoli, G.; Chan, C. Y.; Crisp, A.; Globisch, D.; Carell, T., Amino Acid Modified RNA Bases as Building Blocks of an Early Earth RNA-Peptide World. *Chemistry–A European Journal* **2020,** *26* (65), 14856-14860.
402. Himmelsbach, F.; Schulz, B. S.; Trichtinger, T.; Charubala, R.; Pfleiderer, W., The p-nitrophenylethyl (NPE) group: a versatile new blocking group for phosphate and aglycone protection in nucleosides and nucleotides. *Tetrahedron* **1984,** *40* (1), 59-72.
403. Matuszewski, M.; Sochacka, E., Stability studies on the newly discovered cyclic form of tRNA N6-threonylcarbamoyladenosine (ct6A). *Bioorg. Med. Chem. Lett.* **2014,** *24* (12), 2703-2706.
404. Rublack, N. Synthese von Nukleosidanaloga zur Funktionalisierung und Strukturaufklärung von RNA. 2014.
405. Hanessian, S.; Grzyb, J. A.; Cengelli, F.; Juillerat-Jeanneret, L., Synthesis of chemically functionalized superparamagnetic nanoparticles as delivery vectors for chemotherapeutic drugs. *Biorg. Med. Chem.* **2008,** *16* (6), 2921-2931.
406. Adamiak, R. W., A HIGHLY EFFECTIVE ROUTE TO N, N'-DISUBSTITUTED UREAS UNDER MILD CONDITIONS. AN APPLICATION TO THE SYNTHESIS OF TRNA ANTICODON LOOP FRAGMENTS CONTAINING UREIDONUCLEOSIDES. **1977**.
407. Debiec, K.; Matuszewski, M.; Podskoczyj, K.; Leszczynska, G.; Sochacka, E., Chemical Synthesis of Oligoribonucleotide (ASL of tRNALys T. brucei) Containing a Recently Discovered Cyclic Form of 2-Methylthio-N6-threonylcarbamoyladenosine (ms2ct6A). *Chemistry–A European Journal* **2019,** *25* (58), 13309-13317.
408. Adamiak, R.; Biała, E.; Grzeskowiak, K.; Kierzek, R.; Kraszewski, A.; Markiewicz, W.; Okupniak, J.; Stawiński, J.; Wiewiorowski, M., The chemical synthesis of the anticodon loop of an eukaryotic initiator tRNA containing the hypermodified nucleoside N 6-/N-threonylcarbonyl/-adenosine/t 6 A/1. *Nucleic Acids Res.* **1978,** *5* (6), 1889-1905.
409. Horn, T.; Urdea, M. S., A chemical 5′-phosphorylation of oligodeoxyribonucleotides that can be monitored by trityl cation release. *Tetrahedron Lett.* **1986,** *27* (39), 4705-4708.
410. Tang, Q.; Cai, A.; Bian, K.; Chen, F.; Delaney, J. C.; Adusumalli, S.; Bach, A. C.; Akhlaghi, F.; Cho, B. P.; Li, D., Characterization of byproducts from chemical syntheses of oligonucleotides containing 1-methyladenine and 3-methylcytosine. *ACS omega* **2017,** *2* (11), 8205-8212.
411. Leutzinger, E. E.; Miller, P. S.; Kan, L.-S., Studies on the hydrolysis of 3-methyl-2′-deoxycytidine in aqueous solution A synthesis of 3-methyl-2′-deoxyuridine. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression* **1982,** *697* (2), 243-251.
412. Reddy, M.; Hanna, N.; Farooqui, F., Fast cleavage and deprotection of oligonucleotides. *Tetrahedron Lett.* **1994,** *35* (25), 4311-4314.
413. Johnsson, R. A.; Bogojeski, J. J.; Damha, M. J., An evaluation of selective deprotection conditions for the synthesis of RNA on a light labile solid support. *Bioorg. Med. Chem. Lett.* **2014,** *24* (9), 2146-9.
414. Lin, H.; Miyauchi, K.; Harada, T.; Okita, R.; Takeshita, E.; Komaki, H.; Fujioka, K.; Yagasaki, H.; Goto, Y.-i.; Yanaka, K., CO2-sensitive tRNA modification associated with human mitochondrial disease. *Nature communications* **2018,** *9* (1), 1-17.
415. Mao, X.-L.; Li, Z.-H.; Huang, M.-H.; Wang, J.-T.; Zhou, J.-B.; Li, Q.-R.; Xu, H.; Wang, X.-J.; Zhou, X.-L., Mutually exclusive substrate selection strategy by human m3C RNA transferases METTL2A and METTL6. *Nucleic Acids Res.* **2021,** *49* (14), 8309-8323.
416. Mergny, J. L.; Lacroix, L., Analysis of thermal melting curves. *Oligonucleotides* **2003,** *13* (6), 515-37.

417. Schildkraut, C.; Lifson, S., Dependence of the melting temperature of DNA on salt concentration. *Biopolymers: Original Research on Biomolecules* **1965,** *3* (2), 195-208.
418. Kierzek, E.; Kierzek, R., The thermodynamic stability of RNA duplexes and hairpins containing N 6-alkyladenosines and 2-methylthio-N 6-alkyladenosines. *Nucleic Acids Res.* **2003,** *31* (15), 4472-4480.
419. Agris, P. F., The importance of being modified: roles of modified nucleosides and Mg2+ in RNA structure and function. *Progress in nucleic acid research molecular biology* **1996,** *53*, 79-129.
420. Prescott, B.; Steinmetz, W.; Thomas Jr, G., Characterization of DNA structures by laser Raman spectroscopy. *Biopolymers: Original Research on Biomolecules* **1984,** *23* (2), 235-256.
421. Berova, N.; Nakanishi, K.; Woody, R. W., *Circular dichroism: principles and applications*. John Wiley & Sons: 2000.
422. Sang, Y.; Han, J.; Zhao, T.; Duan, P.; Liu, M., Circularly polarized luminescence in nanoassemblies: generation, amplification, and application. *Adv. Mater.* **2020,** *32* (41), 1900110.
423. Beychok, S., Circular Dichroism of Biological Macromolecules: Circular dichroism spectra of proteins and nucleic acids provide insights into solution conformations. *Science* **1966,** *154* (3754), 1288-1299.
424. Eliel, E. L.; Wilen, S. H., *Stereochemistry of organic compounds*. John Wiley & Sons: 1994.
425. Bishop, G. R.; Chaires, J. B., Characterization of DNA structures by circular dichroism. *Current protocols in nucleic acid chemistry* **2002,** *11* (1), 7.11. 1-7.11. 8.
426. Kypr, J.; Kejnovská, I.; Renčiuk, D.; Vorlíčková, M., Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* **2009,** *37* (6), 1713-1725.
427. Warshaw, M. M.; Cantor, C. R., Oligonucleotide interactions. IV. Conformational differences between deoxy-and ribodinucleoside phosphates. *Biopolymers: Original Research on Biomolecules* **1970,** *9* (9), 1079-1103.
428. Shockett, P. E.; Schatz, D. G., DNA hairpin opening mediated by the RAG1 and RAG2 proteins. *Molecular and cellular biology* **1999,** *19* (6), 4159-4166.
429. Schöller, E.; Marks, J.; Marchand, V.; Bruckmann, A.; Powell, C. A.; Reichold, M.; Mutti, C. D.; Dettmer, K.; Feederle, R.; Hüttelmaier, S., Balancing of mitochondrial translation through METTL8-mediated m3C modification of mitochondrial tRNAs. *Molecular Cell* **2021,** *81* (23), 4810-4825. e12.
430. Zhang, L.-H.; Zhang, X.-Y.; Hu, T.; Chen, X.-Y.; Li, J.-J.; Raida, M.; Sun, N.; Luo, Y.; Gao, X., The SUMOylated METTL8 induces R-loop and tumorigenesis via m3C. *Iscience* **2020,** *23* (3), 100968.
431. Liu, F.; Clark, W.; Luo, G.; Wang, X.; Fu, Y.; Wei, J.; Wang, X.; Hao, Z.; Dai, Q.; Zheng, G., ALKBH1-mediated tRNA demethylation regulates translation. *Cell Biochem. Funct.* **2016,** *167* (3), 816-828. e16.
432. The Human Protein Atlas. https://www.proteinatlas.org/ENSG00000166199-ALKBH3/subcellular (accessed 27.09.2022).
433. Müller, T. A.; Meek, K.; Hausinger, R. P., Human AlkB homologue 1 (ABH1) exhibits DNA lyase activity at abasic sites. *DNA repair* **2010,** *9* (1), 58-65.
434. Cheng, H. P.; Yang, X. H.; Lan, L.; Xie, L. J.; Chen, C.; Liu, C.; Chu, J.; Li, Z. Y.; Liu, L.; Zhang, T. Q., Chemical Deprenylation of N6-Isopentenyladenosine (i6A) RNA. *Angew. Chem.* **2020,** *132* (26), 10732-10737.
435. Tian, L.-F.; Liu, Y.-P.; Chen, L.; Tang, Q.; Wu, W.; Sun, W.; Chen, Z.; Yan, X.-X., Structural basis of nucleic acid recognition and 6mA demethylation by human ALKBH1. *Cell research* **2020,** *30* (3), 272-275.
436. Sundheim, O.; Vågbø, C. B.; Bjørås, M.; Sousa, M. M.; Talstad, V.; Aas, P. A.; Drabløs, F.; Krokan, H. E.; Tainer, J. A.; Slupphaug, G., Human ABH3 structure and key residues for oxidative demethylation to reverse DNA/RNA damage. *The EMBO journal* **2006,** *25* (14), 3389-3397.
437. Liaqat, A.; Sednev, M. V.; Stiller, C.; Höbartner, C., RNA-Cleaving Deoxyribozymes Differentiate Methylated Cytidine Isomers in RNA. *Angew. Chem.* **2021,** *133* (35), 19206-19210.
438. Ghaem Maghami, M.; Scheitl, C. P.; Höbartner, C., Direct in vitro selection of trans-acting ribozymes for posttranscriptional, site-specific, and covalent fluorescent labeling of RNA. *J. Am. Chem. Soc.* **2019,** *141* (50), 19546-19549.
439. Kabinger, F.; Stiller, C.; Schmitzova, J.; Dienemann, C.; Kokic, G.; Hillen, H. S.; Hobartner, C.; Cramer, P., Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis. *Nat Struct Mol Biol* **2021,** *28* (9), 740-746.
440. Lu, J.; Li, N.-S.; Koo, S. C.; Piccirilli, J. A., Efficient synthesis of N4-methyl-and N4-hydroxycytidine phosphoramidites. *Synthesis* **2010,** *2010* (16), 2708-2712.
441. Jena, N., Role of different tautomers in the base-pairing abilities of some of the vital antiviral drugs used against COVID-19. *PCCP* **2020,** *22* (48), 28115-28122.
442. Gordon, C. J.; Tchesnokov, E. P.; Schinazi, R. F.; Götte, M., Molnupiravir promotes SARS-CoV-2 mutagenesis via the RNA template. *J. Biol. Chem.* **2021,** *297* (1).
443. Zhou, S.; Hill, C. S.; Sarkar, S.; Tse, L. V.; Woodburn, B. M.; Schinazi, R. F.; Sheahan, T. P.; Baric, R. S.; Heise, M. T.; Swanstrom, R., β-d-N 4-hydroxycytidine inhibits SARS-CoV-2 through lethal

mutagenesis but is also mutagenic to mammalian cells. *The Journal of infectious diseases* **2021,** *224* (3), 415-419.
444. Seela, F.; Kehne, A., Palindromic octa- and dodecanucleotides containing 2'-deoxytubercidin: synthesis, hairpin formation, and recognition by the endodeoxyribonuclease EcoRI. *Biochemistry* **1987,** *26* (8), 2232-8.
445. Soderberg, T.; Poulter, C. D., Escherichia coli Dimethylallyl Diphosphate: tRNA Dimethylallyltransferase: Essential Elements for Recognition of tRNA Substrates Within the Anticodon Stem− Loop. *Biochemistry* **2000,** *39* (21), 6546-6553.
446. NMR Prediction. http://rainbow.life.tsinghua.edu.cn/csmotif/.
447. Isaksson, J.; Plashkevych, O.; Pradeepkumar, P. I.; Pathmasiri, W.; Petit, C.; Chattopadhyaya, J., 3'-endo/4'-exo Locked Thymidine in the Dickerson-Drew Dodecamer Causes Local Base Pairing Distortions–An NMR Structure and Hydration Study. *Nucleic Acids Res.* **2005**.
448. Micura, R.; Pils, W.; Höbartner, C.; Grubmayr, K.; Ebert, M.-O.; Jaun, B., Methylation of the nucleobases in RNA oligonucleotides mediates duplex–hairpin conversion. *Nucleic Acids Res.* **2001,** *29* (19), 3997-4005.
449. Cornish, P. V.; Giedroc, D. P.; Hennig, M., Dissecting non-canonical interactions in frameshift-stimulating mRNA pseudoknots. *J. Biomol. NMR* **2006,** *35* (3), 209-23.
450. Lambert, J. B.; Binsch, G.; Roberts, J. D., Nitrogen-15 Magnetic Resonance Spectroscopy. I. Chemical Shifts. *Proc Natl Acad Sci U S A* **1964,** *51* (5), 735-7.
451. L. D. Field, H. L. L., A. M. Magill, *Organic Structures from 2D NMR Spectra* John Wiley & Sons Inc: 2015.
452. Witanowski, M.; Stefaniak, L.; Januszewski, H., Nitrogen chemical shifts in organic compounds. In *Nitrogen NMR*, Springer: 1973; pp 163-260.
453. Berger, S.; Braun, S.; Kalinowski, H.-O., NMR Spectroscopy of the Non-metallic Elements. *Ed. John Wiley* **1997**.
454. Sharov, A. V.; Burkhanova, T. M.; Taskin Tok, T.; Babashkina, M. G.; Safin, D. A., Computational Analysis of Molnupiravir. *Int J Mol Sci* **2022,** *23* (3), 1508.
455. Hu, T.; Xie, Y.; Liu, Y.; Xue, H.; Zhu, F.; Aisa, H. A.; Shen, J., A convenient and cost efficient route suitable for “one-pot” synthesis of molnupiravir. **2021**.
456. Markowski, V.; Sullivan, G. R.; Roberts, J. D., Nitrogen-15 nuclear magnetic resonance spectroscopy of some nucleosides and nucleotides. *J. Am. Chem. Soc.* **1977,** *99* (3), 714-718.
457. Rangadurai, A.; Kremser, J.; Shi, H.; Kreutz, C.; Al-Hashimi, H. M., Direct evidence for (G)O6...H2-N4(C)(+) hydrogen bonding in transient G(syn)-C(+) and G(syn)-m(5)C(+) Hoogsteen base pairs in duplex DNA from cytosine amino nitrogen off-resonance R1rho relaxation dispersion measurements. *J Magn Reson* **2019,** *308*, 106589.
458. Nixon, P. L.; Rangan, A.; Kim, Y. G.; Rich, A.; Hoffman, D. W.; Hennig, M.; Giedroc, D. P., Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot. *J. Mol. Biol.* **2002,** *322* (3), 621-33.
459. Cash, D. D.; Cohen-Zontag, O.; Kim, N.-K.; Shefer, K.; Brown, Y.; Ulyanov, N. B.; Tzfati, Y.; Feigon, J., Pyrimidine motif triple helix in the Kluyveromyces lactis telomerase RNA pseudoknot is essential for function in vivo. *Proceedings of the National Academy of Sciences* **2013,** *110* (27), 10970-10975.
460. Rudisser, S.; Pelton, J. G.; Tinoco, I., Jr., Assignment of cytosine N3 resonances in nucleic acids via intrabase three-bond coupling to amino protons. *J. Biomol. NMR* **1999,** *15* (2), 173-6.
461. Sowers, L. C.; Boulard, Y.; Fazakerley, G. V., Multiple structures for the 2-aminopurine-cytosine mispair. *Biochemistry* **2000,** *39* (25), 7613-20.
462. Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z., BioMagResBank. *Nucleic Acids Res.* **2007,** *36* (suppl_1), D402-D408.
463. Andersson, H.; Carlsson, A.-C. C.; Nekoueishahraki, B.; Brath, U.; Erdelyi, M., Solvent effects on nitrogen chemical shifts. *Annual reports on NMR spectroscopy* **2015,** *86*, 73-210.
464. Figueroa, N.; Keith, G.; Leroy, J.-L.; Plateau, P.; Roy, S.; Gueron, M., NMR study of slowly exchanging imino protons in yeast tRNAasp. *Proceedings of the National Academy of Sciences* **1983,** *80* (14), 4330-4333.
465. Snoussi, K.; Leroy, J.-L., Imino proton exchange and base-pair kinetics in RNA duplexes. *Biochemistry* **2001,** *40* (30), 8898-8904.
466. Englander, J.; Kallenbach, N.; Englander, S., Hydrogen exchange study of some polynucleotides and transfer RNA. *J. Mol. Biol.* **1972,** *63* (1), 153-169.
467. Ivanov, V. I.; Minchenkova, L.; Schyolkina, A.; Poletayev, A., Different conformations of double-stranded nucleic acid in solution as revealed by circular dichroism. *Biopolymers: Original Research on Biomolecules* **1973,** *12* (1), 89-110.
468. Hall, K.; Cruz, P.; Tinoco, I.; Jovin, T. M.; Van De Sande, J. H., ‘Z-RNA’—a left-handed RNA double helix. *Nature* **1984,** *311* (5986), 584-586.

469. Miyahara, T.; Nakatsuji, H.; Sugiyama, H., Similarities and differences between RNA and DNA double-helical structures in circular dichroism spectroscopy: a SAC–CI study. *The Journal of Physical Chemistry A* **2016,** *120* (45), 9008-9018.
470. Riazance, J. H.; Baase, W. A.; Johnson Jr, W. C.; Hall, K.; Cruz, P.; Tinoco Jr, I., Evidence for Z-form RNA by vacuum UV circular dichroism. *Nucleic Acids Res.* **1985,** *13* (13), 4983-4989.
471. Vorlíčková, M.; Kejnovská, I.; Bednářová, K.; Renčiuk, D.; Kypr, J., Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes. *Chirality* **2012,** *24* (9), 691-698.
472. Hanging Drop Vapor Diffusion Crystallization. hamptonresearch.com/uploads/cg_pdf/CG101_Hanging_Drop_Vapor_Diffusion.pdf (accessed 25.09.2022).
473. Cobbe, N.; Heck, M. M., SMCs in the world of chromosome biology—from prokaryotes to higher eukaryotes. *Journal of structural biology* **2000,** *129* (2-3), 123-143.
474. Hirano, T., At the heart of the chromosome: SMC proteins in action. *Nature reviews Molecular cell biology* **2006,** *7* (5), 311-322.
475. Cobbe, N.; Heck, M. M., The evolution of ATPase activity in SMC proteins. *Proteins: Structure, Function, and Bioinformatics* **2006,** *63* (3), 685-696.
476. Guacci, V.; Koshland, D.; Strunnikov, A., A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in S. cerevisiae. *Cell* **1997,** *91* (1), 47-57.
477. Michaelis, C.; Ciosk, R.; Nasmyth, K., Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell* **1997,** *91* (1), 35-45.
478. Hirano, T.; Kobayashi, R.; Hirano, M., Condensins, chromosome condensation protein complexes containing XCAP-C, XCAP-E and a Xenopus homolog of the Drosophila Barren protein. *Cell* **1997,** *89* (4), 511-521.
479. Strunnikov, A. V.; Hogan, E.; Koshland, D., SMC2, a Saccharomyces cerevisiae gene essential for chromosome segregation and condensation, defines a subgroup within the SMC family. *Genes & development* **1995,** *9* (5), 587-599.
480. Fousteri, M. I.; Lehmann, A. R., A novel SMC protein complex in Schizosaccharomyces pombe contains the Rad18 DNA repair protein. *The EMBO Journal* **2000,** *19* (7), 1691-1702.
481. Hirano, T., Condensins: organizing and segregating the genome. *Current Biology* **2005,** *15* (7), R265-R275.
482. Hassler, M.; Shaltiel, I. A.; Haering, C. H., Towards a unified model of SMC complex function. *Current Biology* **2018,** *28* (21), R1266-R1281.
483. Ganji, M.; Shaltiel, I. A.; Bisht, S.; Kim, E.; Kalichava, A.; Haering, C. H.; Dekker, C., Real-time imaging of DNA loop extrusion by condensin. *Science* **2018,** *360* (6384), 102-105.
484. Angelov, T.; Guainazzi, A.; Scharer, O. D., Generation of DNA interstrand cross-links by post-synthetic reductive amination. *Org. Lett.* **2009,** *11* (3), 661-4.
485. Seela, F.; Westermann, B.; Bindig, U., Liquid–liquid and solid–liquid phase-transfer glycosylation of pyrrolo [2, 3-d] pyrimidines: stereospecific synthesis of 2-deoxy-β-D-ribofuranosides related to 2′-deoxy-7-carbaguanosine. *J. Chem. Soc., Perkin Trans. 1* **1988,** (3), 697-702.
486. Nawale, G. N.; Gore, K. R.; Höbartner, C.; Pradeepkumar, P., Incorporation of 4′-C-aminomethyl-2′-O-methylthymidine into DNA by thermophilic DNA polymerases. *Chem. Commun.* **2012,** *48* (77), 9619-9621.
487. Kschonsak, M.; Merkel, F.; Bisht, S.; Metz, J.; Rybin, V.; Hassler, M.; Haering, C. H., Structural basis for a safety-belt mechanism that anchors condensin to chromosomes. *Cell* **2017,** *171* (3), 588-600. e24.
488. Gardner, A. F.; Jackson, K. M.; Boyle, M. M.; Buss, J. A.; Potapov, V.; Gehring, A. M.; Zatopek, K. M.; Corrêa Jr, I. R.; Ong, J. L.; Jack, W. E., Therminator DNA polymerase: modified nucleotides and unnatural substrates. *Frontiers in Molecular Biosciences* **2019,** *6*, 28.
489. Tindall, K. R.; Kunkel, T. A., Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase. *Biochemistry* **1988,** *27* (16), 6008-6013.
490. Patil, D. P.; Pickering, B. F.; Jaffrey, S. R., Reading m6A in the transcriptome: m6A-binding proteins. *Trends in cell biology* **2018,** *28* (2), 113-127.
491. Tu, C.; Keane, C.; Eaton, B. E., Palladium catalysis in the synthesis of 8-position modified adenosine, 2′-deoxyadenosine and guanosine. *Nucleosides, Nucleotides & Nucleic Acids* **1995,** *14* (8), 1631-1638.
492. Espinet, P.; Echavarren, A. M., The mechanisms of the Stille reaction. *Angew. Chem. Int. Ed.* **2004,** *43* (36), 4704-4734.
493. Scott, N. W.; Ford, M. J.; Schotes, C.; Parker, R. R.; Whitwood, A. C.; Fairlamb, I. J., The ubiquitous cross-coupling catalyst system ‘Pd (OAc) 2’/2PPh 3 forms a unique dinuclear Pd I complex: an

important entry point into catalytically competent cyclic Pd 3 clusters. *Chemical Science* **2019,** *10* (34), 7898-7906.
494. Cheong, A.; Low, J. J.; Lim, A.; Yen, P. M.; Woon, E. C., A fluorescent methylation-switchable probe for highly sensitive analysis of FTO N 6-methyladenosine demethylase activity in cells. *Chemical science* **2018,** *9* (36), 7174-7185.
495. Mailler, E.; Paillart, J. C.; Marquet, R.; Smyth, R. P.; Vivet-Boudou, V., The evolution of RNA structural probing methods: From gels to next-generation sequencing. *Wiley Interdisciplinary Reviews: RNA* **2019,** *10* (2), e1518.
496. Neuner, S.; Santner, T.; Kreutz, C.; Micura, R., The "Speedy" Synthesis of Atom-Specific 15N Imino/Amido-Labeled RNA. *Chemistry–A European Journal* **2015,** *21* (33), 11634-11643.
497. Gizzi, A. S.; Grove, T. L.; Arnold, J. J.; Jose, J.; Jangra, R. K.; Garforth, S. J.; Du, Q.; Cahill, S. M.; Dulyaninova, N. G.; Love, J. D., A naturally occurring antiviral ribonucleotide encoded by the human genome. *Nature* **2018,** *558* (7711), 610-614.
498. Brookes, P.; Lawley, P. D., The methylation of cytosine and cytidine. *J. Chem. Soc.* **1962**, 1348-1351.
499. Pitsch, S.; Weiss, P. A.; Jenny, L.; Stutz, A.; Wu, X., Reliable Chemical Synthesis of Oligoribonucleotides (RNA) with 2'-O-[(Triisopropylsilyl)oxy]methyl(2'-O-tom)-Protected Phosphoramidites. *Helv. Chim. Acta* **2001,** *84* (12), 3773-3795.
500. Ali, M. A.; Yao, X.; Sun, H.; Lu, H., [RhCp* Cl2] 2-catalyzed directed N-Boc amidation of arenes "on water". *Org. Lett.* **2015,** *17* (6), 1513-1516.
501. García-Echeverría, C., A base labile handle for solid phase organic chemistry. *Tetrahedron Lett.* **1997,** *38* (52), 8933-8934.
502. Steinmetzger, C.; Bäuerlein, C.; Höbartner, C., Supramolecular fluorescence resonance energy transfer in nucleobase-modified fluorogenic RNA aptamers. *Angew. Chem.* **2020,** *132* (17), 6826-6830.
503. Rothenbühler, S.; Iacovache, I.; Langenegger, S. M.; Zuber, B.; Häner, R., Supramolecular assembly of DNA-constructed vesicles. *Nanoscale* **2020,** *12* (41), 21118-21123.
504. Painter, G.; Bluemling, G.; Natchus, M.; Guthrie, D. WO2019113462. 2018.
505. Kohyama, N.; Katashima, T.; Yamamoto, Y., Synthesis of novel 2-aryl AICAR derivatives. *Synthesis* **2004,** *2004* (17), 2799-2804.
506. Strohmeier, J., *Synthese strukturell und funktionell modifizierter Nukleoside und Nukleotide zur Untersuchung von DNA-Konformationen und Protein-DNA-Interaktionen*. Cuvillier Verlag: 2012.
507. Hwang, T.-L.; Shaka, A., Water suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *Journal of Magnetic Resonance, Series A* **1995,** *112* (2), 275-279.
508. Sklenář, V.; Bax, A., Spin-echo water suppression for the generation of pure-phase two-dimensional NMR spectra. *Journal of Magnetic Resonance (1969)* **1987,** *74* (3), 469-479.
509. Goddard, T.; Kneller, D., University of California; San Francisco: 2001. *SPARKY Manual*.
510. Bodenhausen, G.; Ruben, D. J., Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **1980,** *69* (1), 185-189.
511. Piotto, M.; Saudek, V.; Sklenar, V., Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **1992,** *2* (6), 661-5.
512. Sklenar, V.; Piotto, M.; Leppik, R.; Saudek, V., Gradient-tailored water suppression for 1H-15N HSQC experiments optimized to retain full sensitivity. *Journal of Magnetic Resonance, Series A* **1993,** *102* (2), 241-245.
513. Mori, S.; Abeygunawardana, C.; Johnson, M. O. n.; Vanzijl, P. C., Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *Journal of Magnetic Resonance, Series B* **1995,** *108* (1), 94-98.
514. Hwang, T.-L.; van Zijl, P.; Mori, S., Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *J. Biomol. NMR* **1998,** *11* (2), 221-226.
515. Hwang, T.-L.; Mori, S.; Shaka, A.; van Zijl, P. C., Application of phase-modulated CLEAN chemical EXchange spectroscopy (CLEANEX-PM) to detect water− protein proton exchange and intermolecular NOEs. *J. Am. Chem. Soc.* **1997,** *119* (26), 6203-6204.
516. Strebitzer, E.; Rangadurai, A.; Plangger, R.; Kremser, J.; Juen, M. A.; Tollinger, M.; Al-Hashimi, H. M.; Kreutz, C., 5-Oxyacetic Acid Modification Destabilizes Double Helical Stem Structures and Favors Anionic Watson–Crick like cmo5U-G Base Pairs. *Chemistry–A European Journal* **2018,** *24* (71), 18903-18906.
517. Bereiter, R.; Himmelstoss, M.; Renard, E.; Mairhofer, E.; Egger, M.; Breuker, K.; Kreutz, C.; Ennifar, E.; Micura, R., Impact of 3-deazapurine nucleobases on RNA properties. *Nucleic Acids Res.* **2021,** *49* (8), 4281-4293.

518. Szulik, M. W.; Voehler, M.; Stone, M. P., NMR analysis of base-pair opening kinetics in DNA. *Curr Protoc Nucleic Acid Chem* **2014,** *59* (1), 7 20 1-18.
519. Plum, G. E., Optical methods. *Current Protocols in Nucleic Acid Chemistry* **2000,** (1), 7.3. 1-7.3. 17.
520. Li, F.; Pallan, P. S.; Maier, M. A.; Rajeev, K. G.; Mathieu, S. L.; Kreutz, C.; Fan, Y.; Sanghvi, J.; Micura, R.; Rozners, E., Crystal structure, stability and in vitro RNAi activity of oligoribonucleotides containing the ribo-difluorotoluyl nucleotide: insights into substrate requirements by the human RISC Ago2 enzyme. *Nucleic Acids Res.* **2007,** *35* (19), 6424-6438.