Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Investigation of active DNA demethylation utilizing synthetic nucleosides, nucleotides and oligonucleotides
and
Synthesis of oxazole-containing building blocks for therapeutical peptidomimetics

Anni Eveliina Schön (geb. Ponkkonen)
aus
Hattula, Finnland
2025

Erklärung

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Frau Prof. Dr. Anja Hoffmann-Röder betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 21.02.2025

Anni Eveliina Schön

Dissertation eingereicht am:	04.03.2025
1. Gutachterin:	Prof. Dr. Anja Hoffmann-Röder
2. Gutachter:	Prof. Dr. Oliver Trapp
Mündliche Prüfung am:	17.04.2025

“Vanhojen totuuksien täytyy väistyä uusien tieltä”

Minna Canth

Parts of this work have already been published in academic journals or presented on scientific conferences.

List of Publications

- *"Analysis of an active deformylation mechanism of 5-formyl-deoxycytidine (fdC) in stem cells"* Alexander Schön#, Ewelina Kaminska#, Florian Schelter#, Eveliina Ponkkonen, Eva Korytiaková, Sarah Schiffers, Thomas Carell*, *Angew. Chem. Int. Ed.* **2020**, *59*, 5591-5594.

- *"Direct and Base Excision Repair-Mediated Regulation of a GC-Rich cis-Element in Response to 5-Formylcytosine and 5-Carboxycytosine"* Nadine Müller, Eveliina Ponkkonen, Thomas Carell and Andriy Khobta*, *Int. J. Mol. Sci.* **2021**, *22*, 11025.

- *"Unified Description of Ultrafast Excited State Decay Processes in Epigenetic Deoxycytidine Derivatives"* Piotr Kabaciński#, Marco Romanelli#, Eveliina Ponkkonen, Vishal Kumar Jaiswal, Thomas Carell, Marco Garavelli*, Giulio Cerullo*, and Irene Conti*, *Phys. Chem. Lett.* **2021**, *12*, 11070−11077.

#: These Authors contributed equally to this study
*: Corresponding author

Conference participations

- ORChem conference, poster presentation, in Regensburg DE, 2024
- Symposium: New Frontier in Chemical Biology, Munich DE, 2024
- Light Dynamics week, online seminars 2020
- Technology training camp in Vienna AT, 2020
- Light Dynamics week in Bologna IT, 2019
- Light Dynamics introduction week in Naples IT, 2019

Table of Contents

1. Summary

Epigenetic modifications in the 5-position of genomic 2'-deoxycytidine (dC) have a strong functional influence on DNA of living organisms. Epigenetics play an important role in the developmental stages of a stem cell into a multicellular organism. Even though all the highly specialized and differentiated cells contain the same genetic material, their differences in morphology and function are enormous. A second layer of information on DNA controls the strictly organized gene-expression at specific time points and sequences, creating a huge variety of cell functions. The most abundant and best-understood chemical modification, the methylation of the C5-position of genomic 2'-deoxycytidine (dC) to 5-methyl-2'-deoxycytidine (mdC) in CpG-rich sequences, is known to deactivate transcriptional activity. Accordingly, it is known that the corresponding demethylation of genomic mdC can reactivate the transcription of the genes.

One focus of this dissertation was to investigate the underlying mechanisms of active DNA demethylation of mdC to dC. Two independent demethylation mechanisms are of interest, one postulating a direct chemical genomic demethylation via a C-C-bond cleavage reaction of the oxidized derivatives 5-formyl-2'-deoxycytidine (fdC) and 5-carboxy-2'-deoxycytidine (cadC) by a deformylation or decarboxylation reaction, whereas the other pathway, the so-called base excision repair (BER) pathway, is known to involve enzymatic cleavage of the glycosidic bonds in modified nucleosides. To this aim, chemically modified nucleosides, nucleotides, and oligonucleotides were synthesized and used for *in vivo* experiments to study their biological metabolism applying modern analytical detection techniques.

As a result, evidence which could support the presence of an active decarboxylation pathway of genomic cadC via direct C-C bond cleavage was obtained when investigating the 2'-fluorinated analogue of cadC (F-cadC) in different cell lines. To enable these experiments, an optimized DNA hydrolysis protocol and an optimized nucleotide extraction protocol were developed. Also, a suitable UHPLC-MS/MS method to measure highly ion suppressed F-cadC was elucidated. MS-analysis based thereon revealed the F-cadC to F-dC turnover to be very

fast *in vivo* and also subsequent remethylation of F-dC to F-mdC was detected. Additionally, cell feeding experiments using the 2'-fluorinated analogue of fdC (F-fdC) showed a slower turnover rate to F-dC than the experiments with F-cadC. Sequence dependent active demethylation studies furthermore revealed, that in the absence of BER, cadC strongly inhibits promoter activity while fdC showed no significant effect, similarly to mdC and hmdC. On the other hand, BER of cadC caused significant promoter reactivation. These results might indicate cadC to act as a repressory mark to prevent premature activation of promoters undergoing the final stages of DNA demethylation, when the symmetric CpG methylation has already been lost.

To gain more information about the enzymatic entity behind the decarboxylation process, potential decarboxylation inhibitors were designed, synthesized and their effects in cell feeding experiments were investigated. Preliminary feeding experiments of the nitrated compound 5-NO_2-F-dC gave an idea of its potential inhibitory properties.

The active demethylation of fdC to dC has already been shown in mESCs and in several other different cell lines, and in the course of this work a mechanistic study of the deformylation reaction *in vivo* was performed to investigate if this process requires nucleophilic activation of the C6-position in fdC. 6-aza-fdC, a reactivity-inverted probe molecule, was fed to different cell lines and it could be shown that a nucleophilic attack at the C6-position and consequently *in vivo* deformylation did not take place. In addition, kinetic studies showed that the 6-position reactivity of F-fdC and fdC is similar and that the 2'-F-substituent does not remarkably influence this process.

In a further part of this work, the design and synthesis of a derivatization reagent to quantify abasic sites from the BER pathway or genomic fdC was investigated. For the UHPLC-MS/MS analysis, the development of a sensitivity-improved hydroxylamine-based reagent was of interest. However, after the successful synthesis of a new fluorinated reagent, it was found that neither its reactivity towards 2'-deoxyribose nor towards fdC meets with the requirements for highly sensitive, quantitative UHPLC-MS/MS analysis.

The second major part of this thesis focusses on the chemical synthesis of new highly functionalized oxazole-based building blocks which can be utilized for the subsequent synthesis of oxazole-containing peptidomimetics. Peptidomimetics are compounds in which the existing peptide sequence typically of therapeutically active peptides is modified so as to improve their *in vivo* stability and reactivity. Exemplary modifications can comprise a substitution of a natural amino acid in the peptide chain by a chemically altered amino acid or other chemical or enzymatical modifications, such as L to D isomerization, cyclization or linkage to cell penetrating peptides. In the present work, oxazole-based building blocks were developed which allow the introduction of at least one oxazole heterocycle into the peptidomimetic compounds to be synthesized. The orthogonal protecting group strategy of the core building blocks which was established in the course of this work allowed successful orthogonal deprotection of the target compounds under mild conditions and subsequent coupling of an exemplary L-alanine derivative to the deprotected positions of interest thereby providing new and versatile building blocks that enable the development of novel synthetic methodologies for a regioselective assembly of oxazole-containing peptidomimetics.

Part 1:

Investigation of active DNA demethylation utilizing synthetic nucleosides, nucleotides and oligonucleotides

2. Epigenetics and DNA methylation

While the discovery of the storage of genetic information, its structure, and the understanding of the importance of the primary information layer in DNA were huge benchmarks in life sciences, new techniques and technologies have been developed since then leading to the discovery of a new feature of DNA, a second layer of information. Despite cells having the same genetic code, they are extremely specialized in performing different and specific functions. How can this dramatic difference in function and structure between for instance neurons and erythrocytes be explained? The differences in their functions lay in patterns of gene expression that differs greatly among different cell types. Mechanisms of gene expression, involving controlled activation and silencing of specific genes, that are not associated with changes in the nucleotide sequence, are based on epigenetically relevant chemical modifications of DNA.

Epigenetics is a field focusing on defining the mechanisms of heritable changes in gene expression that do not involve changes in the corresponding DNA sequence - a change in phenotype without a change in genotype. Epigenetic change can be of regular and natural occurrence but can also be influenced by several factors including age, environment, lifestyle, and disease state. Previous studies highlighted the importance of epigenetic regulation in maintaining the pluripotent state and the reprogramming process in embryonic stem cells (ESC).[1] This state can be transferred to new cells through hundreds of cell divisions, or the state can be abolished creating different, epigenetically maintained, patterns of expression leading to different paths of differentiation into distinct cell types.

5-Methyl-2'-deoxycytidine (mdC) was pronounced as an additional element in DNA in 1950.[2–4] After its role being intensively studied for 53 years, it was recognized as a regulatory epigenetic element in 2003.[5] S-adenosyl methionine (SAM) is the primary methyl source for hundreds of transmethylases that can methylate DNA, RNA, histones, proteins and small biological molecules. For the methylation of genomic dC, three SAM-dependent DNA methyltransferases (DNMT) are known: DNMT1[157,58], DNMT3A and DNMT3B[59,60]. DNMT3A and 3B initiate DNA methylation *de novo* in cytosine-phosphate-guanine (CpG) dinucleotides (Figure 1).[61,62] DNMT1, however, plays an important role maintaining the methylation state of

hemi-methylated sites in daughter strands.[63–65] Methylation of dC in certain promoter regions, mainly in the CpG context, can lead to transcriptional silencing of the corresponding genes. In addition, it is known that DNA can be wrapped around histones to form nucleosomes, and these structures can be also chemically modified as well. This can regulate the accessibility of DNA which allows cells to control the transcriptional activity of genes.[6]

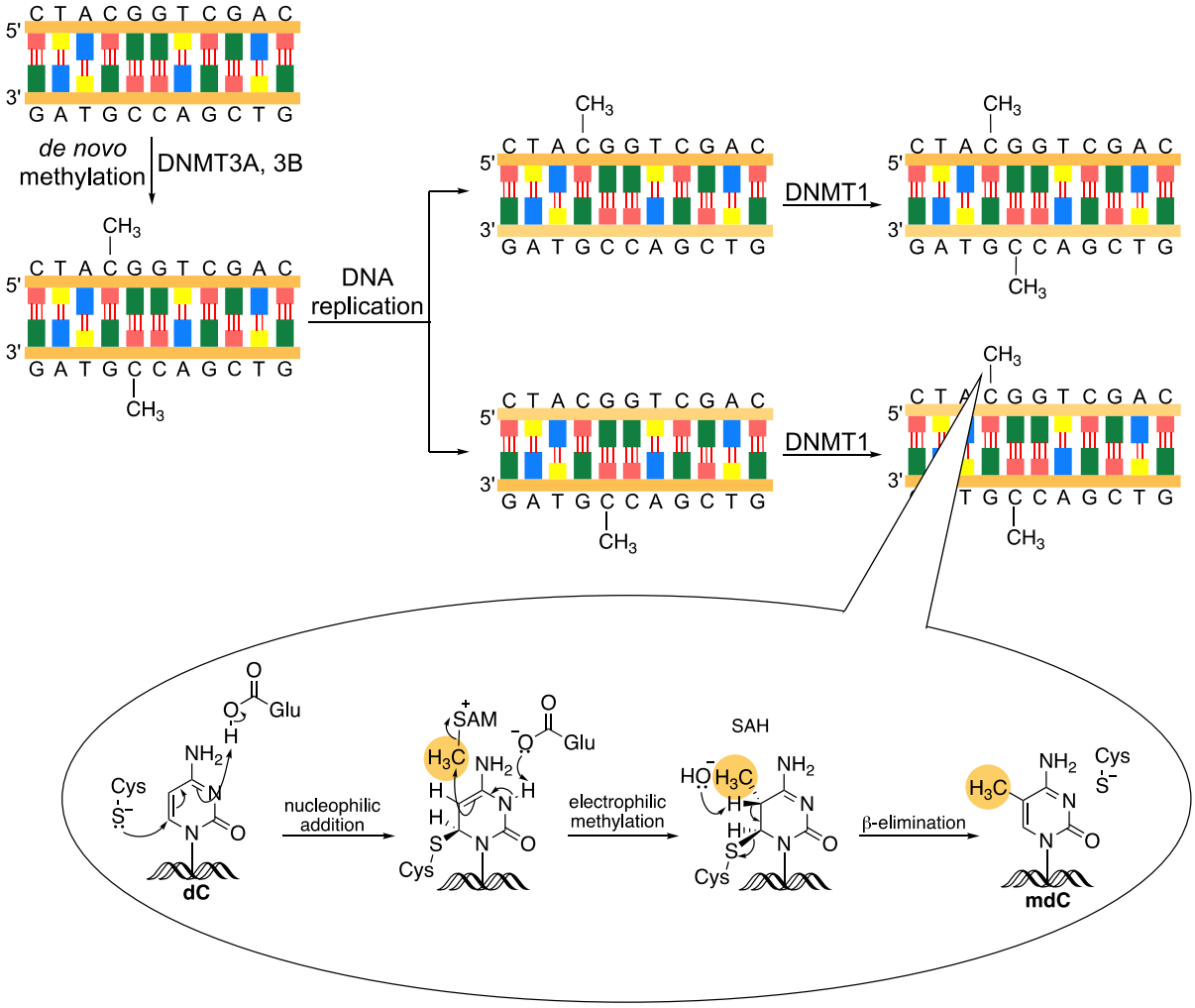


Figure 1. Methylation of genomic dC to mdC with DNMT methyltransferases as well as the proposed methylation mechanism.

2.1. Epigenetically relevant dC modifications in DNA

The potential presence of 5-hydroxymethyl-2'-deoxycytidine (hmdC) as an additional component in the DNA was already reported in 1972[7], however, the lack of proper methodology at the time prohibited the researchers from definitive results.[8] Establishment of

the synthetic route to isotope-labelled mdC and hmdC and the use of quantitative LC-MS techniques in 2009, allowed two independent research groups to detect hmdC in DNA. Kriaucionis and Heintz *et al.*[9] reported significant amounts of hmdC being present in Purkinje neurons (0.6%/dC), but also in granule cells (0.2%/dC) and in stem cells (0.1%/dC). At the same time Rao *et al.*[10] reported that hmdC is formed from mdC by oxidation which is catalysed by the ten-eleven translocation (TET) enzymes.[10] TET enzymes, that are α-ketoglutarate (α-KG) dependent oxidases, utilize an active site Fe^{2+} and molecular oxygen to perform the oxidation via a radical based mechanism.[11]

After the discovery of hmdC, scientists made a great effort to discover potentially existing further oxidized cytidine DNA modifications. A first study on neuronal cells, utilizing HPLC coupled with ESI-MS, provided no evidence on other oxidized cytidines.[12] However, in 2011, Pfaffeneder *et al.* detected in a different MS-based study, utilizing synthetic 5-formyl-2'-deoxycytidine (fdC) and a stable isotopically labelled standard with two incorporated ^{15}N-atoms, significant levels of fdC in the genome of mESCs. Furthermore, it was shown that fdC is indeed derived from mdC.[13] Subsequently after this discovery and also in 2011, it was discovered that the TET enzymes are catalysing the whole oxidative cascade from mdC to hmdC, to fdC and moreover to 5-carboxy-2'-deoxycytidine (cadC) (Figure 2).[14]

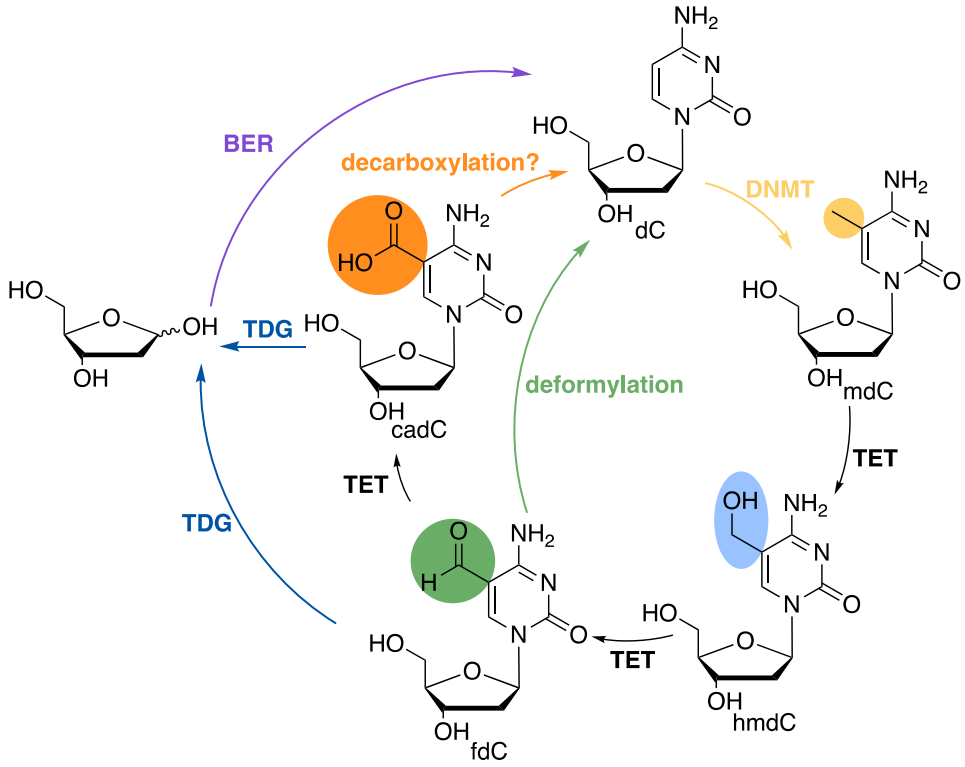


Figure 2. Formation of epigenetic dC modifications and the potential active demethylation pathways.

Another key finding in the same year was achieved using isotope standards, wherein cadC was found in substantially elevated levels when studying stem cells with a thymine DNA glycosylase (TDG) knockout.[15] Right after this, a basis for the so-called base excision repair (BER) pathway was established by Maiti and Drohat showing that TDG recognizes and cleaves genomic cadC and fdC.[15] However, the question of potentially harmful abasic sites formed by the TDG-catalysed base excision has made scientists doubt whether BER is the only pathway to remove fdC and cadC. Indeed, recent advances allowed it to show that fdC is directly removed via direct C-C bond cleavage in vivo.[16] Figure 2 gives an overview over the epigenetically relevant dC nucleosides and demethylation pathways

2.2. Biological function and detection of dC modifications

Several studies of the oxidized dC-derivatives have demonstrated how these modifications do not interfere with Watson-Crick base pairing but introduce chemical functionalities and thereby influence DNA structure and nucleosomal organization as well as modification of specific proteins. The possibility of various combinations of dC-modifications within a DNA strand can create chemically diverse environments and thereby various interaction surfaces.

Proteins interacting with both strands including chemically modified nucleobases can interpret the interaction sites differently. This view is supported by several reports showing altered binding of certain proteins for oxidized dC-derivatives. So far, only few specific hmdC binding proteins have been discovered[17–19], and it could be shown that fdC and cadC are recruiting more proteins than hmdC.[17] More importantly, proteins that bind mdC cannot bind hmdC, which might indicate that the amount of binding partners is irrelevant as can be seen for cadC that has the most binding partners but little to no function.

Understanding the function of the epigenetic bases is of high interest and valuable information has already been gained by determination of their respective abundance in different tissue types, cell types as well as at different developmental stages. At the moment, DNA isolation and enzymatic hydrolysis followed by high-performance liquid chromatography–mass spectrometry (HPLC-MS/MS) is the predominantly utilized method in quantification of DNA modifications.

In studies utilizing liquid chromatography-mass spectrometry (LC-MS) technology, Carell and co-workers introduced in 2009 synthetic isotopologues to study the distribution and absolute levels of epigenetically relevant mdC oxidation products.[20] To perform the HPLC-MS analysis, DNA samples were enzymatically digested into individual nucleoside mixtures. These were then separated chromatographically and eventually detected by the mass spectrometer. This method in combination with synthetic standard compounds was used to study the distribution of hmdC in mouse brain[20] and subsequently in many other organs.[12] In the murine central nervous system, the hmdC levels were found the highest with 630-1470 modifications per 10^6 nucleotides. In other organs like heart, kidney and lungs only 315-420 modifications per 10^6 nucleotides were detected and in liver, spleen and testis the values were even lower. In 2015, the first study on age-dependent global hmdC levels in humans over the entire lifespan[21] showed that the highest abundance of hmdC is found in the central nervous system and, in contrast thereto, another study focusing on cancer showed the hmdC levels to be strongly reduced in brain tumor cells and mdC levels were found to be constant in most of the organs.[22] Additionally, Wagner *et al.*[21] showed that while hmdC levels are increasing within age and reach a steady state level of about 2520 modifications per 10^6 nucleotides in fully developed human brain tissue, fdC levels decline during early developmental stages. Increasing hmdC

values and decreasing fdC values in brain tissue do not correlate with each other which might indicate, that an active demethylation process is slowing down within age.[21] Nevertheless, a recent study, utilizing a mouse model, agrees with the age dependence of hmdC levels, but suggests liver cells to be a more suitable model for studying the age-dependence of hmdC due to the continuous increase of hmdC levels by increasing age of adult mice.[23] Another study compared the levels of hmdC, fdC and cadC in human brain and in mouse brain DNA.[24] It was shown that the levels of hmdC, fdC and cadC were 1550, 1.7 and 0.15 modifications per 10^6 nucleosides respectively, whereas the values in mouse brain were 560, 1.4, 0.12 modifications per 10^6 nucleotides, respectively. Regardless of the low global levels of fdC in mESCs being, 2.3-3.4 modifications per 10^6 nucleotides[25,26], it was found that fdC can exists at specific genomic loci in relatively high levels compared to hmdC and mdC levels.[27]

Nevertheless, the extremely low abundance of hmdC, fdC and cadC in biological samples in combination with biological matrix effects like poor ionization abilities and potential coelution properties with other nucleosides during LC makes their detection challenging. As a result of these disadvantages, efforts to develop more advanced and at the same time simple, fast, and efficient techniques to increase the detection sensitivity have been made. An early LC-MS/MS study of modified cytidines, reported LODs for hmdC, fdC and cadC obtained from 30-80 μg of DNA to be 0.056, 0.098 and 0.14 fmol respectively.[24] By the time, the detection limits of non-derivatized nucleosides are now reduced to sub-femtomole amounts when obtained from 500 ng DNA (<1 ppm of modification per nucleoside).[28] By reaching the limits of sensitivity of the equipment, the poor ionization ability of the oxidized dC modifications was tackled by the development of more sensitive methods based on chemical derivatization. Specific chemical derivatization of hmdC, fdC and cadC was shown to improve the ionization efficiency in ESI[29–32] and also shifted the retention times during the *rp* LC making it possible for the modifications no longer co-elute with the canonical nucleosides. As a result, analytes were found to be more efficiently ionized under higher organic content due to higher spraying and desolvation efficiency.[33] This reduced the LODs of fdC and cadC to 110 amol and 230 amol when using 10 μg of DNA, respectively.[29] However, these first derivatization reagents were not completely compatible with biological samples, and in addition the required amount of DNA was still considerably high.[29,32] In 2019, a report utilizing a hydrazine-based reagent reduced the LODs of fdC and cadC down to 10 amol and 25 amol, respectively, requiring only

600 ng of DNA.[25] Based on the data reported therein and with the help of this hydrazine-based derivatization reagent, quantification of fdC and cadC in different mouse tissues revealed that fdC and cadC did not have tissue specific distribution and that the levels in different tissues were generally remarkably low. As a reference, the authors also analysed the levels of mdC and hmdC, which agreed with the data published by other groups.[12] Figure 3 illustrates the tissue distribution of second layer of information, namely mdC, hmdC, fdC and cadC from recent publications.

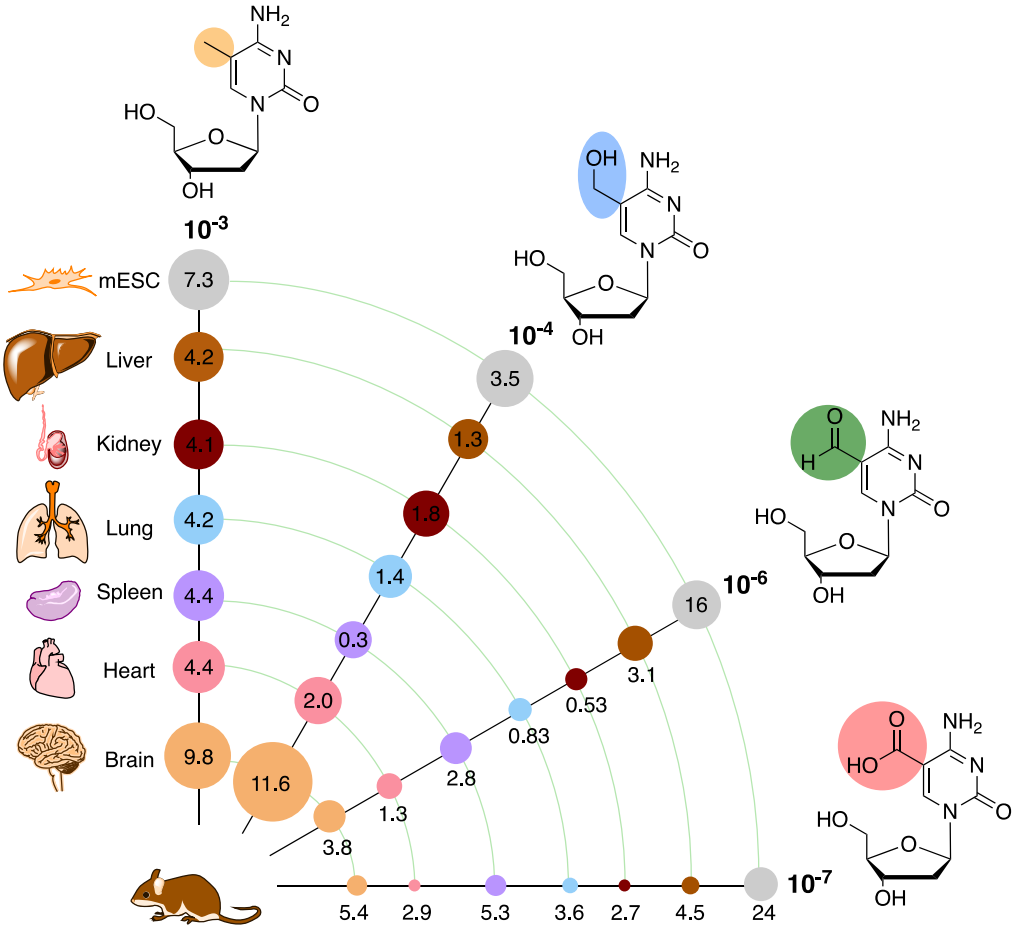


Figure 3. Tissue and embryonic stem cell distribution of mdC, hmdC[12,21] and fdC and cadC[25] as nucleoside per dG. Figure adapted from the review of Carell *et al.*[26]

Utilizing the same derivatization reagent, another study on hmdC levels in human plasma, showed an age dependent increase of hmdC levels that was in agreement with previously published data from Wagner *et al.*[21]. The approach, a MnO_2 oxidation of hmdC followed by subsequent derivatization method named 5hmC-MIQuant- claims to achieve ultrasensitive HPLC-MS/MS quantification of hmdC in low-input DNA samples. They reported an oxidation of hmdC to fdC followed by derivatization to provide a LOD (HPLC-MS/MS) up to 14 amol,

whereas the respective LOD investigating unlabelled samples is known to be 2.5 fmol. However, the study did not elaborate how the levels of natural fdC were taken into account in data analysis, but claimed its low abundance not to influence on exact quantification of hmdC.[34] All in all, these studies nicely demonstrate how combination of organic and analytical chemistry can develop powerful tools to study biological processes.

Instead of chemical derivatization to improve mass spectrometric detection, efforts to enhance the detection of modified cytosines with different buffer-systems, different purification methods on DNA hydrolysis products have been investigated. A very recent publication demonstrated how ion suppression of modified cytosines during UHPLC-MS/MS analysis can be overcome by changing the buffer system. Ammonium bicarbonate (NH_4HCO_3) buffer solutions were shown to act as superior proton sources and to facilitate proton transfer in ESI.[23] The commonly used buffer for positive ionization mode, diluted formic acid (FA) was shown to have a suppression ratio of 6.3 (internal standard in H_2O vs. in digested DNA sample), whereas with NH_4HCO_3 buffer the suppression ratio was detected with only 1.1. The detection of natural hmdC levels in digested mouse DNA samples increased 15.5 times, but only 3.2 times for fdC. cadC having the most ion-like character of the modified cytosines was, however, not detected in this study. The pKa of its carboxylic acid group is estimated extraordinarily low with less than 1.5.[35] This might, in turn, cause a detection problem when protonated cadC is being investigated. Acidification of sensitive DNA samples to pH 1 might cause decomposition of the samples and, further, the accessible pH range for UHPLC columns is usually set between 2.0-8.0.

Moreover, already established DNA digestion methods for DNA containing canonical nucleosides are known to not always be compatible with all modified cytidine bases. The traditional DNA hydrolysis approach, a nuclease P1 protocol[36] is a time-consuming two-step, triple-enzymatic digestion method. By the time, several one-step protocols have been developed and optimized for different purposes.[14,37–39] Total DNA digestion requires endonucleases (e.g. DNase I), that digest the DNA into oligonucleotides, followed by an exonuclease treatment, for example using snake venom phosphodiesterase I (PDE1) that digests oligonucleotides further to nucleosides. In addition, when measuring DNA samples, residual RNA is to be considered. RNA contaminations need to be properly removed to avoid

detection problems during MS. It was also shown that common DNA hydrolysis protocols do not completely digest the already very low abundant and difficultly detectable cadC nucleotides and even differences in the digestion efficiencies between synthetic oligonucleotides containing cadC and genomic DNA were observed.[40] It is clear that, when studying different modifications in the same sample, a digestion method with different efficiencies towards the individual modifications may lead to misinterpretation of the data.

2.3. Demethylation pathways

Compared to other epigenetic modifications hmdC, fdC and cadC, DNA methylation is believed to be relatively stable. Despite its role in long-term gene silencing, the discovery of TET related oxidation of mdC made it apparent that the regulation of DNA methylation can also be a dynamic process, where demethylation might occur very rapidly.[41] The stable carbon-carbon bond between C5 of the cytidine heterocycle and the methyl substituent makes the direct cleavage chemically demanding and efforts to reveal the mechanisms via oxidized forms have been made for several years. Besides to that, demethylation can also occur via a passive mechanism when methylation patterns are not maintained on newly synthesized DNA during replication. On the other hand, the active mechanism seems to occur, when pluripotent stem cells differentiate in early mammalian development, or also to be gene-specifically in somatic cells.[42] Caused by the increased interest in the dynamics of the second information layer of DNA, many structural and proteomic investigations as well as studies of viable mechanisms of active DNA demethylation have been reported.[43–46]

The absence of DNMT1 or conditions that impair DNMT1-mediated maintenance of CpG methylation are processes that induce passive demethylation. Under these conditions, over the course of multiple DNA replications, the levels of mdC in the genome drop. Since DNMT1 is known to act on hemi-methylated CpG sites, the presence of oxidation products hmdC, fdC and cadC is shown to have an inhibitory effect on maintenance methylation. This suggests that genomic areas which require passive demethylation, are marked with hmdC before replication starts.[47] Through multiple rounds of DNA replication and lower DNMT1 activity towards hmdC:dG, fdC:dG and cadC:dG base pairs, the DNA is demethylated (Figure 4).[48,49] To reactivate silenced genes experimentally, demethylation has been achieved with chemically

synthesized 5-aza-dC, which has been shown to inhibit DNMT1 and thereby could reduce the global mdC levels.[50,51]

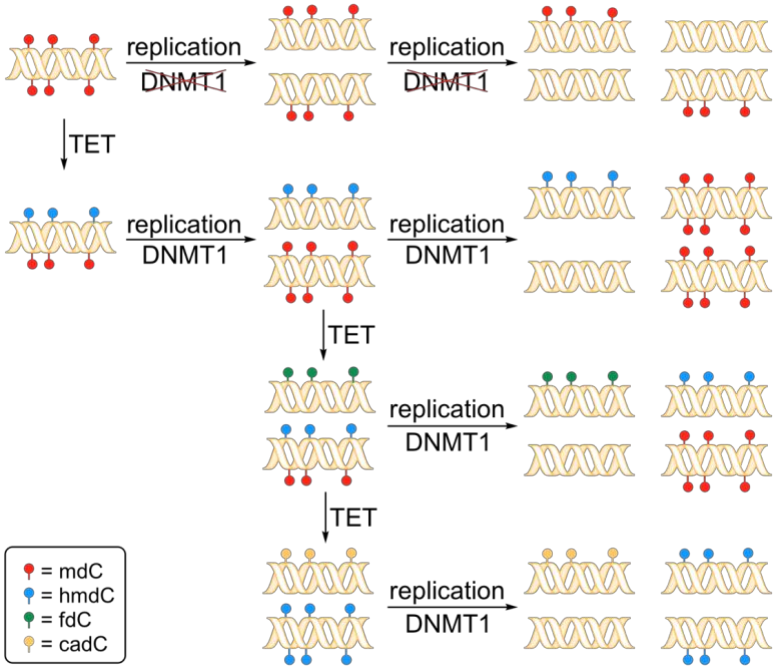


Figure 4. Overview of passive dilution of mdC and its oxidized bases through DNA replication.

On the other hand, high interest in solving the mechanism and dynamics of reactivation of gene reading has resulted in various proposals of active demethylation pathways. The most studied mdC oxidation dependent, TDG mediated removal of fdC and cadC by base excision repair (BER) is considered as the main pathway.[15,52–54] However, the question whether the TDG-based BER is the only active pathway, became more relevant after the observation of TDG independent active demethylation in zygotes.[55] In addition, the BER pathway involves the formation of reactive abasic sites in the DNA that are known to cause potentially harmful single strand breaks or even double strand breaks. Thereby, an abasic site circumventing pathway via direct C-C bond cleavage has also been postulated (Figure 5).[16,43,56]

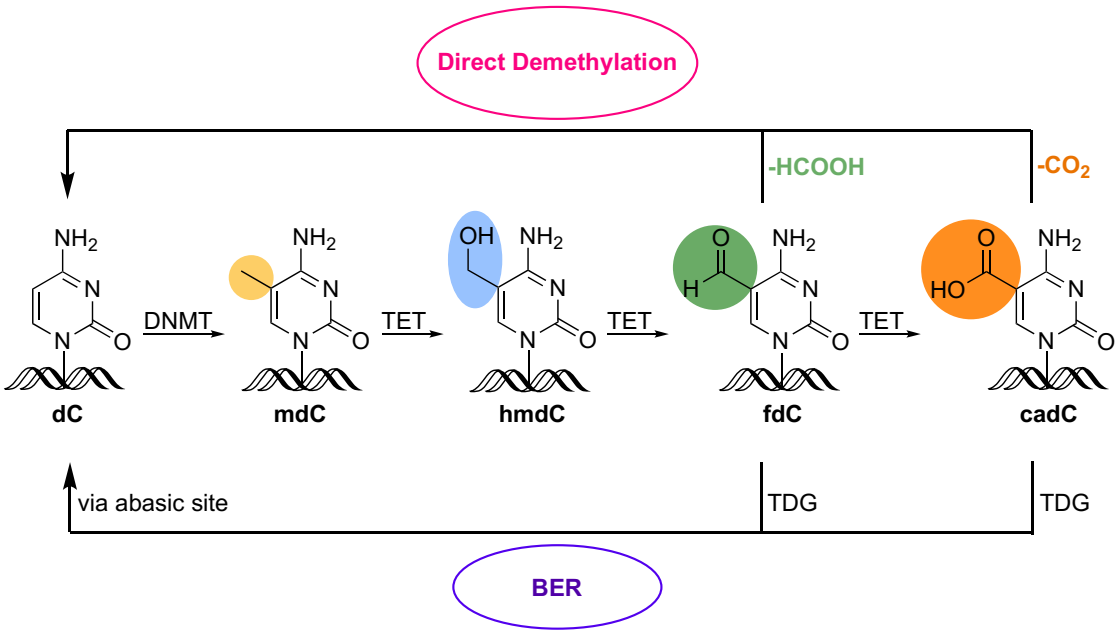


Figure 5. Active demethylation pathways.

2.3.1. Base Excision Repair

Thymine DNA glycosylase (TDG), one of the proteins that has been shown to bind to genomic fdC and cadC, but not to hmdC[52] utilizes the same base flipping mechanism as the other enzymes in the DNA glycosylases family. These special DNA glycosylases recognize their targets and start a two-step BER process to remove them from the genome: the first step is an initial hydrolysis step carried out by the DNA glycosylases.[57–59] The glycosylases can be divided into monofunctional and bifunctional glycosylases based on their excision behavior. Monofunctional BER enzymes are DNA glycosylases that hydrolyse the N-glycosidic bond, creating an abasic site (see Figure 6). This abasic site is then further processed by an AP endonuclease in the second step. The endonuclease cleaves the DNA backbone on the 5'-side and a phosphodiesterase cuts the 3'-phosphate leaving a single nucleotide gap in the DNA strand. The gap is filled with a new nucleotide introduced by a DNA polymerase and a DNA ligase.[58] Bifunctional BER enzymes have both glycosylase and lyase activities – first they remove the modified nucleobase by a glycosyl transfer reaction with an amine nucleophile and then the formed Schiff base (imine) intermediate is removed by beta-elimination which cleaves the DNA backbone on the 3'-side. This is followed by cleavage of sugar-phosphate backbone by AP endonuclease resulting in single strand break. This gap is further processed via short-patch or long-patch repair including DNA polymerases and ligases.[59]

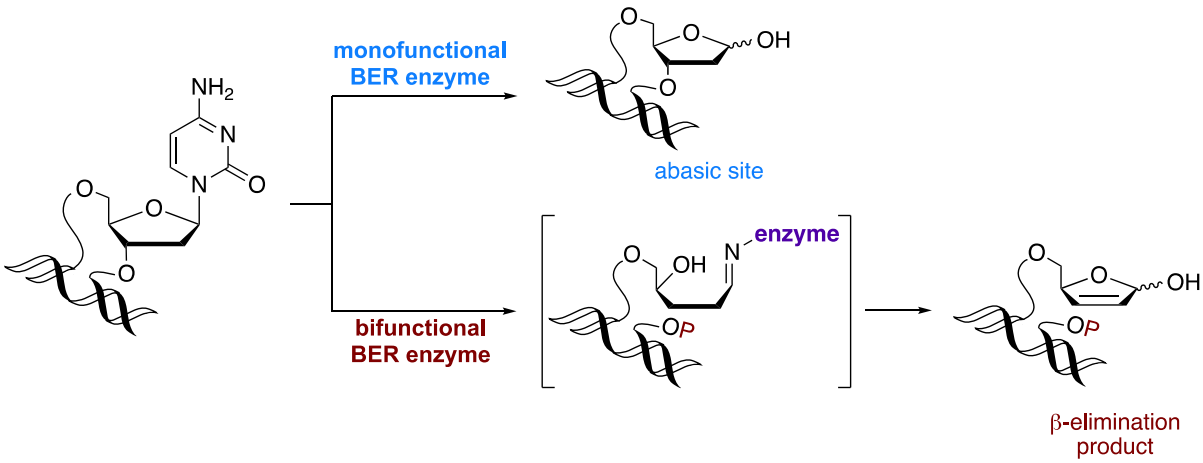


Figure 6. Overview over the known BER pathways.

The pioneering structural investigations on TDG by Maiti and Drohat[52] showed already in 2011 the specific ability of TDG to excise fdC and cadC but being unable to remove hmdC. Later on in 2019, a study on TDG specificity on fdC and cadC suggested that the specificity was caused by their influences on the DNA flexibility.[45] Regardless of the geometry alteration of the DNA grooves in the double stranded DNA (dsDNA), it was shown that neither formyl- nor carboxyl groups as substituents in the 5-position of cytosine sterically interfere the regular G-C base paring. The importance of base pairing with G is related to strong specificity of TDG to excise G paired -and followed bases.[60–62] The presence of fdC, however, is known to cause a displacement alteration at the fdC-G site and leads to an opening of the major groove and a narrowing the minor groove.[45,63] In case of cadC, the repulsive negative charge towards the phosphodiester backbone alters the cadC base location with respect to the backbone and widens the minor groove. Furthermore, the fdC context is considered to require less energy for DNA bending and unstacking during the base-flipping mechanism thereby accelerating the removal of fdC.[44]

Whilst BER is considered to cause harmful single or double strand breaks (DSBs),[64,65] recent studies are more confidently claiming the interaction between TETs and TDG to be so highly coordinated and that harmful DSB formation is minimized by the tight binding of TDG to the AP site.[46] Supporting this finding, the chemical character of fdC and cadC was shown to mediate a significant geometry change in dsDNA. Due to the sp^2 hybridization of the formyl and carboxyl groups, they adopt planar conformation in the same plane as the pyrimidine ring. That is stabilized by the intramolecular hydrogen bond between the carbonyl oxygen and

exocyclic amino group of the pyrimidine.[45] This can decrease the DNA stability and support the formation of DSBs, however, the close interaction between TETs and TDG can hinder the exposure of free fdC and cadC sites thereby preventing DSBs. Since hmdC, unlike fdC and cadC, has a sp^3-hybridized carbon at the hydroxymethyl group, it was postulated to lead to a distinct orientation and an energy state that TDG cannot recognize.[45] Furthermore, Rahimoff *et al.*[66] showed fdC and cadC are indeed removed without the accumulation of potentially harmful BER intermediates in the DNA of stem cells.

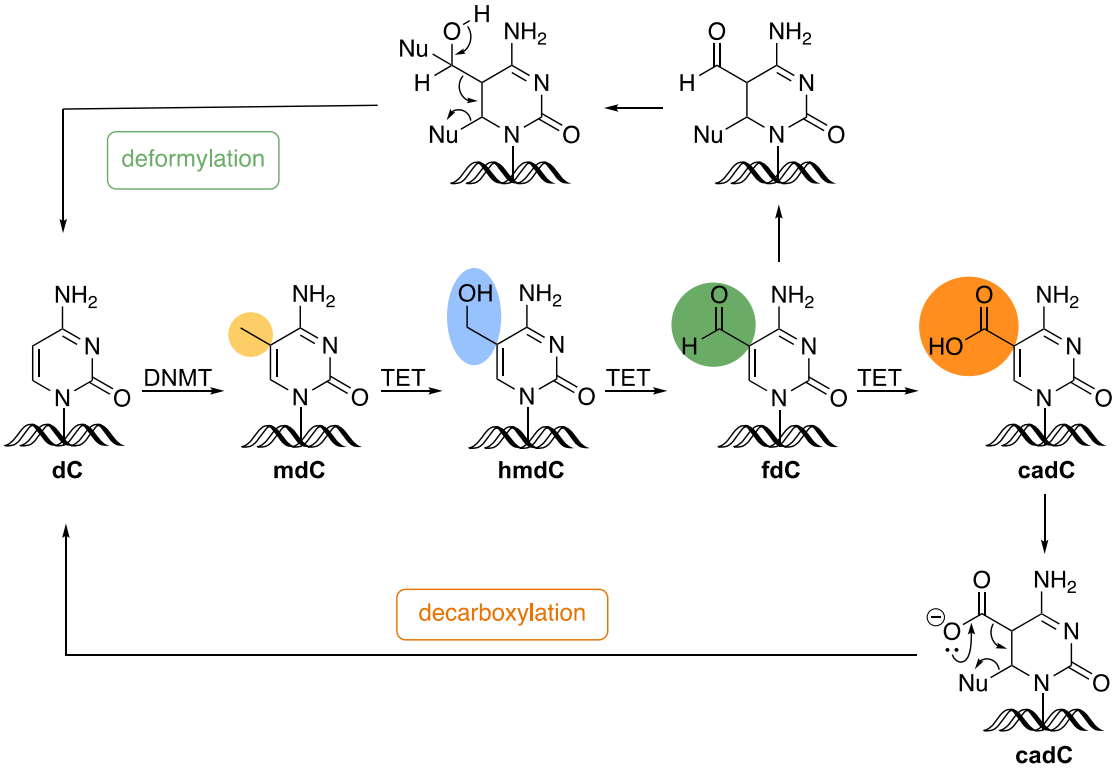
2.3.2. C-C bond cleavage

A proposed alternative to TDG-mediated BER is a direct C-C bond cleavage of the formyl or carboxyl group to respectively reform dC.[67–71] The first indication of deformylation of FdC was obtained in 2013 by Schiesser *et al.*[71] and in 2018, deformylation of genomic fdC was indeed reported to occur in mESCs *in vivo*.[16] Isotopically labelled $^{15}N_2$-fdC and 2'-fluorinated fdC (F-fdC) were fed to mESCs and their respective metabolism was monitored via UHPLC/MS-MS. The deformylation products 2'-fluorinated dC (F-dC) and isotopically labelled $^{15}N_2$-dC were detected giving strong evidence of *in vivo* deformylation.

To specifically study the direct C-C bond cleavage *in vivo*, F-dC derivatives were found to be well suitable probe molecules. Due to their 2'-fluorine substituent, these analogues resist BER and yet bear a similar chemical structure and character as the naturally occurring epigenetic bases. Schröder *et al.*[72] developed a synthetic pathway towards F-mdC and its oxidized derivatives and demonstrated that the 2'-fluorine substituent did not have a measurable influence on methylation and oxidation was possible while base excision was completely abolished. However, deamination properties of 2´-fluorinated compounds were not investigated. Oligonucleotides containing dC or F-dC were incubated with the methyltransferase M.SssI and the resulting methylation rates of dC to mdC and F-dC to F-mdC were respectively found to be almost the same. In addition, their stability towards human TDG (hTDG) was studied with oligonucleotides containing fdC or F-fdC. hTDG induced strand cleavage products were observed in fdC containing oligonucleotides but not in oligonucleotides containing the F-fdC derivative.[72] Furthermore, a follow-up study showed that F-mdC can be oxidized to F-hmdC by the TET enzymes.[73] The detection and quantification suitability of these probe molecules proved to be impeccable, whereby the 2'-fluorine

substituent provided a sharp signal in the UHPLC-MS/MS due to the more labile glycosidic bond in the mass spectrometric fragmentation step. Also, fluorinated compounds exhibit different retention times compared to the natural nucleosides and they are 18 atom units heavier than their naturally existing counterparts.[16]

The chemical principle for direct deformylation of fdC and decarboxylation of cadC was established almost 10 years ago[70,71] and the postulated mechanism behind it is described in Scheme 1. Thiols are found as active site residues in many DNA modifying enzymes making them suitable molecules to potentially initiate deformylation or decarboxylation. Therefore, it was believed that the C-C bond cleavage reaction requires the presence of thiol groups and to be acid-catalyzed. An *in vitro* study showed hmdC to dehydroxymethylate very slowly, but fdC and especially cadC to convert fast to dC.[69,71] These reactions were believed to proceed via a nucleophilic attack of a thiol to the C-6 position of fdC or cadC[71], via a comparable mechanism to the SAM-dependent methylation of dC to mdC catalyzed by the DNMTs.[74,75]



Scheme 1. Proposed active demethylation mechanisms via C-C bond cleavage of fdC and cadC.

Also, several studies on decarboxylation have been published.[69–71] In 2012, when the presence of cadC in DNA was shown and doubt of BER being the only active demethylation pathway

was raised, the synthesis of a $^{15}N_2$-labelled cadC oligonucleotide was reported. It was subsequently used in an isotope tracing experiment in order to investigate the active demethylation via decarboxylation of cadC, opening the direction towards a potential decarboxylation pathway.[69] Thereby, mechanistic insight was provided and a chemical model was suggested wherein the decarboxylation reaction of cadC was considered to be accelerated by the saturation of the C5-C6 double bond. Interestingly, a follow up study showed cadC containing oligonucleotides to decarboxylate under thiol-mediated, acid catalyzed conditions by a factor of 11 more efficiently than fdC was deformylating.[71]

A recent *in vivo* study utilizing synthetic cadC with seven phosphodiester replaced by phosphorthioates in the backbone of an oligonucleotide demonstrated a direct C-C bond cleavage in isotopically enriched human cells.[43] This approach utilized metabolically labelled cells, yielding 85% isotopically labeled $^{15}N_3$-dC. The cells were then transfected with synthetic, unlabeled oligonucleotides and after DNA isolation and digestion, sulfur containing monophosphates of cadC (S-cadCMP) and dC (S-dCMP) were analyzed. After 24h, MS analysis revealed sulfur containing decarboxylation product S-dCMP, whereby S-cadCMP was undetectable.

3. Aim of the Work

The main focus of this work is to study active demethylation of mdC to dC *in vivo* that involves direct C-C bond cleavage, concentrating on decarboxylation of cadC. It has already been shown that fdC can be converted into dC via direct C-C bond cleavage *in vivo*.[16] Subsequently many new questions were raised: does the fdC first oxidize to cadC and then decarboxylate or are these two independent processes? What is the actual chemical mechanism, is it sequence dependent, and furthermost what is the enzymatic entity behind this process?

The first step is to investigate if measurable decarboxylation of genomic cadC to dC occurs *in vivo* (Chapter 5.1). This study involves cell feeding experiments of synthetic nucleoside and nucleotide probe molecules and the analysis of their metabolism via UHPLC-MS/MS. To specifically study the C-C bond cleavage of cadC, a probe molecule, the 2'-fluorinated analogue of cadC (F-cadC), that is known to resist TDG mediated BER, is used. Additionally, various other routes of administration of the F-cadC nucleoside into the genomic DNA in cell feeding experiments are investigated by using a prodrug or a triphosphate of F-cadC in combination with a special triphosphate transporter.

To gain information about the enzyme catalyzing the potential direct C-C bond cleavage decarboxylation reaction of cadC to dC, the synthesis of a new probe molecule, 5-nitro-2'-fluoro-deoxycytidine (NO_2-F-dC), a structural and electronic analogue to F-cadC is studied. The expected strong electron withdrawing character is considered to potentially trap the corresponding decarboxylase enzyme and thereby to inhibit *in vivo* decarboxylation thus potentially lead to altered genomic fdC and cadC levels. To this aim, NO_2-F-dC is to be chemically synthesized, extensively purified and fed to cells of interest to subsequently analyze the levels of fdC and cadC via UHPLC-MS/MS.

To study the BER process *in vivo* utilizing UHPLC-MS/MS, a highly sensitive derivatization reagent is of interest. An already established hydroxylamine-based derivatization reagent has been used to detect and quantify genomic abasic sites[76] and lately, new reagents have been reported and used to detect rare modifications in the gDNA increasing the sensitivity of the measurements remarkably.[25] In order to enhance the analytical sensitivity, to improve the

fragmentation pattern in MS and to decrease the required collision energy (CE), a new advanced tetra-fluorinated derivatization reagent is synthesized. Its substitution pattern is expected to fulfil these requirements and also to be suitable to detect low abundant fdC and cadC from biological samples.

On the other hand, in the specificity of these processes in certain specific gene elements is of interest. Therefore, 8 different oligonucleotides with incorporations of fdC, cadC, F-fdC and F-cadC are to be synthesized. These oligonucleotides are then further processed into DNA vectors and transported into HeLa cells to investigate their effect on protein expression levels in the presence (fdC/cadC-vectors) or absence (F-fdC/F-cadC-vectors) of BER. This project is implemented in collaboration with research group Prof. Dr. Andriy Khobta (Mainz).

It is furthermore of interest to study if the oxidized dC modifications in the genome can also be formed as a DNA photodamage. In collaboration with the Cerullo group (Milano) and the Garavelli group (Bologna) ultrafast TA spectroscopy in combination with sub-30-fs transient absorption spectroscopy and correlated multiconfigurational CASPT2/MM computations should be used to study the photophysical decay mechanisms of mdC, hmdC, fdC and cadC.

4. Published Work

4.1. Mechanistic study of deformylation in stem cells

"Analysis of an active deformylation mechanism of 5-formyl-deoxycytidine (fdC) in stem cells"
Alexander Schön[#], Ewelina Kaminska[#], Florian Schelter[#], Eveliina Ponkkonen, Eva Korytiaková, Sarah Schiffers, Thomas Carell*, *Angew. Chem. Int. Ed.* **2020**, *59*, 5591-5594

#: These authors contributed equally
*: Corresponding author

Prologue

Earlier in vitro studies using synthetic oligonucleotides have shown that deformylation of fdC on the oligonucleotide level takes place in the presence of basic amino acids and thiols and a mechanism has been postulated, wherein an activation of the 6-position of the modified pyrimidine heterocycle is the first step. To investigate this mechanism *in vivo*, 6-aza-5-formyl-2'-deoxycytidine (a-fdC), a reactivity inverted probe molecule, was synthesized and fed to various somatic cell lines and induced mouse embryonic stem cells, together with 2'-fluorinated fdC (F-fdC). While deformylation of F-fdC was clearly observed *in vivo*, it did not occur with a-fdC, thus suggesting that the C-C bond-cleaving deformylation is initiated by nucleophilic activation.

Contribution to this study

For this work, I conducted experiments to study the reaction kinetics of fdC, F-fdC and a-fdC with bisulfite to determine their respective reactivity with a nucleophile expected to react with the 6-position of the modified pyrimidine ring. The analytics for this study were done utilizing HPLC and LC/MS. Based on the measurements, we could conclude that the reactivity of fdC and F-fdC with bisulfite does not show remarkable differences, and this supported that the conclusions of this work were substantiated.

License

Epigenetics

International Edition: DOI: 10.1002/anie.202000414
German Edition: DOI: 10.1002/ange.202000414

Analysis of an Active Deformylation Mechanism of 5-Formyl-deoxycytidine (fdC) in Stem Cells

*Alexander Schön[+], Ewelina Kaminska[+], Florian Schelter[+], Eveliina Ponkkonen, Eva Korytiaková, Sarah Schiffers, and Thomas Carell**

Dedicated to Dr. Klaus Römer on the occasion of his 80th birthday

Abstract: *The removal of 5-methyl-deoxycytidine (mdC) from promoter elements is associated with reactivation of the silenced corresponding genes. It takes place through an active demethylation process involving the oxidation of mdC to 5-hydroxymethyl-deoxycytidine (hmdC) and further on to 5-formyl-deoxycytidine (fdC) and 5-carboxy-deoxycytidine (cadC) with the help of α-ketoglutarate-dependent Tet oxygenases. The next step can occur through the action of a glycosylase (TDG), which cleaves fdC out of the genome for replacement by dC. A second pathway is proposed to involve C–C bond cleavage that converts fdC directly into dC. A 6-aza-5-formyl-deoxycytidine (a-fdC) probe molecule was synthesized and fed to various somatic cell lines and induced mouse embryonic stem cells, together with a 2′-fluorinated fdC analogue (F-fdC). While deformylation of F-fdC was clearly observed in vivo, it did not occur with a-fdC, thus suggesting that the C–C bond-cleaving deformylation is initiated by nucleophilic activation.*

The nucleobase modification 5-formyl-deoxycytidine (fdC, **1**) is found in stem cells during early development and in the brain.[1–5] These tissues are particularly rich in 5-hydroxymethyl-deoxycytidine (hmdC) from which fdC (**1**) is produced.[6,7] The formation of hmdC and fdC requires oxidation reactions that are performed by α-ketoglutarate-dependent Tet enzymes, with 5-methyl-deoxycytidine (mdC) being the initial starting molecule.[8–10] This cascade of oxidation reactions is a part of an active demethylation process, in which mdC as a silencer of transcription is replaced by unmodified dC.[11] The central molecule that is removed seems to be fdC.[12,13] It can be cleaved out of the genome by a dedicated DNA glycosylase, which creates an abasic site that is further processed, leading to the insertion of an unmodified dC.[14] Because abasic sites are harmful DNA-repair intermediates that can cause genome instability, it was suggested early on that fdC might be directly deformylated to dC by C–C bond cleavage.[15,16] Evidence for the existence of such a direct deformylation process was recently reported.[17] Model studies showed that direct deformylation of fdC and potentially also decarboxylation of 5-carboxy-deoxycytidine (cadC) are indeed possible.[15] Nevertheless, it requires activation of the nucleobase by a nucleophilic addition to the C6 position. For fdC, an additional hydrate formation on the formyl group seems to be necessary, as depicted in Figure 1 A. Although activation with a helper nucleophile is well known as the central mechanistic process during the methylation of dC to mdC by DNA methyltransferases (Dnmts),[18,19] it remains to be confirmed whether such activation occurs in vivo as well.

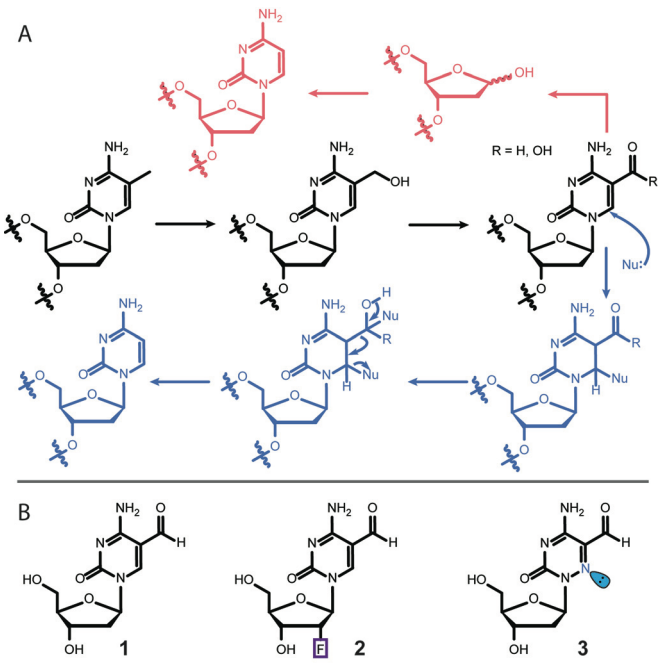


Figure 1. A) The mdC removal pathways that involve oxidation to hmdC, fdC, and cadC followed by either base-excision repair (magenta) or C–C bond cleavage (blue). B) Structures of fdC (**1**) and the two probe molecules **2** and **3** used for this study.

[*] M. Sc. A. Schön,[+] E. Kaminska,[+] M. Sc. F. Schelter,[+] M. Sc. E. Ponkkonen, M. Sc. E. Korytiaková, Dr. S. Schiffers, Prof. Dr. T. Carell
Department of Chemistry, Ludwig-Maximilians Universität München
Butenandtstr. 5–13, 81377 München (Germany)
E-mail: thomas.carell@lmu.de
Homepage: https://www.carellgroup.de

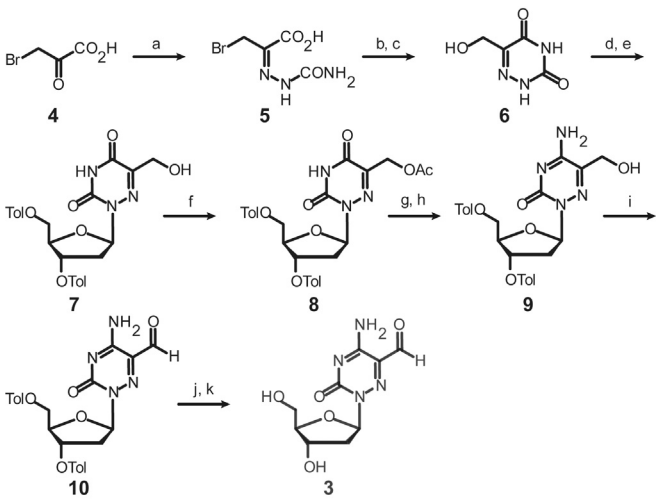
[+] These authors contributed equally to this work.

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.202000414.

In this work, we investigated this hypothesis with two probe molecules, 2′-fluorinated-fdC (F-fdC, **2**) and 6-aza-fdC (a-fdC, **3**). The two compounds were simultaneously fed to different cell types, including primed stem cells. This led to random incorporation of these bases at the "C" sites in the respective genomes. Furthermore, it led to the presence of F-fdC and a-fdC not only at CpG sites. Ultrasensitive UHPLC-QQQ-MS^2 was subsequently used to interrogate the chemical processes that occur at F-fdC and a-fdC in the genomes. The data show that while F-fdC is efficiently deformylated, this does not occur for a-fdC. The only difference between the two nucleobases is the presence of an in-ring nitrogen atom (6-aza atom), which features a lone pair that prohibits nucleophilic addition. These results thus provide strong evidence that nucleophilic activation is the central governing mechanistic event that is required for C–C bond cleavage in vivo.

The fluorinated nucleoside F-fdC (**2**) was recently introduced by us as a deformylation probe.[17] Compound **2** is an antimetabolite that is effectively incorporated into the genomes of growing cells. The 2′-fluoro group is required to block all types of glycosylases, so that base-excision repair is efficiently inhibited. This ensures high levels of F-fdC (**2**) in the genome, as required to observe potential deformylation processes.

The synthesis of the novel nucleoside a-fdC (**3**) is depicted in Scheme 1. The synthesis was started with bromo pyruvic acid (**4**), which we first converted into the semicarbazone **5**, followed by conversion into the acid chloride, subsequent cyclization, and hydrolysis to give hydroxymethylated 6-aza-uracil (**6**).[20] Vorbrüggen nucleosidation with Hoffers' chlorosugar subsequently provided the nucleoside **7** as a mixture of the α- and β-anomers, which could be separated by recrystallization. Next, we acetyl-protected the hydroxymethyl group to give **8**, and then used a standard procedure to convert the U base **8** into the C-derived base **9** by amination of the 4-triazole intermediate with ammonium hydroxide. This led to the concomitant cleavage of the acetyl protecting group. Dess–Martin oxidation of **9** to **10** and final removal of the toluoyl groups furnished the 6-aza-5-formyl-deoxycytidine nucleoside (a-fdC) **3** in a good total yield of 22% with respect to **6** (Supporting Information).



Scheme 1. Synthesis of the probe molecule a-fdC (**3**). a) semicarbazide·HCl, NaOAc, HOAc, H_2O, 0°C to r.t., 2.5 h, 49%. b) pyridine, $SOCl_2$, 80°C, 75 min. c) H_2O, 110°C, 17 h, 74% over 2 steps. d) TMSCl, HMDS, 135°C, 75 min, *then* e) Hoffer's chlorosugar, $CHCl_3$, r.t., 17 h, 56% over 2 steps. f) Ac_2O, pyridine, r.t., 22 h, 96%. g) 1,2,4-triazole, $POCl_3$, NEt_3, MeCN, 0°C to r.t., 18 h, then h) NH_4OH, 1,4-dioxane, 40°C, 5 h, 84%. i) Dess–Martin periodinane, CH_2Cl_2, −15°C to r.t., 1 h, 89%. j) NaOMe, MeOH, benzene, r.t., 1.5 h, then k) reversed-phase HPLC, 54%.

Compound **3** features a nitrogen atom instead of a carbon atom at the 6-position, which possesses a lone pair that blocks any nucleophilic addition to this position. Compound **3** is consequently a perfect model system to investigate whether such a nucleophilic activation is required for the deformylation, as mechanistically postulated (Figure 1A).

The nucleosides **2** and **3** were subsequently added at a concentration of 350 μM to the media of Neuro-2a, RBL-2H3, CHO-K1 cells for 72 hours (see the Supporting Information). During this time, the nucleosides are converted in vivo into the corresponding triphosphates and then incorporated into the genome of the dividing cells. Initial studies in which we fed the nucleosides individually allowed us to determine that neither compound decreases cell viability up to a concentration of 400 μM, thus the experiments were conducted below the toxicity level. In addition, we tested **2** and **3** at 350 μM on E14 TDG +/− and −/− mouse embryonic stem cells (mESCs) under a three-day priming process with C/R media. This system allowed us to exclude the BER pathway, leading to a detectable and quantifiable accumulation of natural fdC (see the Supporting Information). After three days, the cells were harvested and lysed, and the genomic DNA was extracted using an optimised protocol (see the Supporting Information). This was followed by an enzymatic digestion of gDNA to single nucleosides and analysed according to a method that we reported recently in detail.[21] The obtained nucleoside mixture containing mostly the canonical nucleosides dA, dC, dG, and dT, plus the non-canonical nucleosides mdC, hmdC, and fdC, as well as the incorporated molecules F-fdC and a-fdC and their potential downstream products (F-dC, F-mdC, a-dC, a-mdC). Nucleosides were separated by ultra-HPL chromatography and characterized by coupling of the UHPLC system to a triple-quadrupole mass spectrometer. For exact quantification of the nucleosides by isotope dilution, isotopically labelled standards of F-fdC and of the product F-dC were spiked into the analysis mixture as internal standards (see the Supporting Information). To enable exact quantification, calibration curves using these standards were determined (see the Supporting Information). Quantification was performed in the linear region.

During the analysis, we noted that an unusually low amount of a-fdC (**3**) was detected because it showed a broad elution profile with very low intensity (Figure 2B). All attempts to sharpen the elution profile in order to gain sensitivity failed. NMR analysis of compound **3** showed the reason for broad elution profile (see the Supporting Information). Due to the additional electron-withdrawing in-ring nitrogen atom, compound **3** exists partially as its hydrate in aqueous solution (20%, see the Supporting Information).

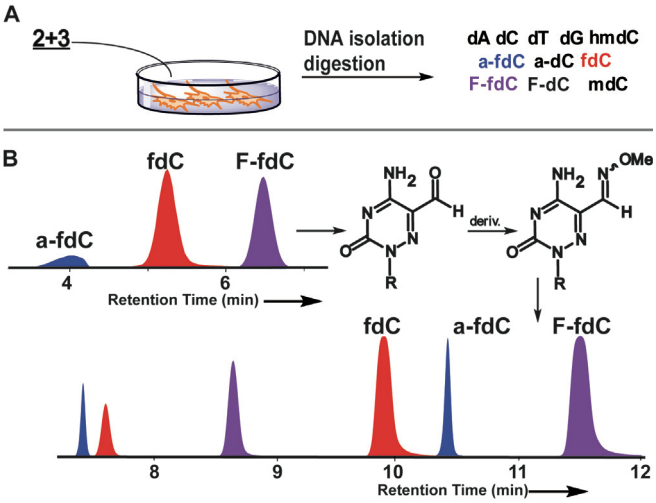


Figure 2. A) Overview of the experimental steps with the feeding and analysis. B) Analysis scheme and the reaction of a-fdC with methoxyamine to block hydrate formation and of a typical UHPL-chromatogram before (C-8 column) and after derivatization (C-18 column) for exact quantification. Peak splitting is due isomerization (blue peaks: a-fdC, red peaks: fdC, and purple peaks: F-fdC).

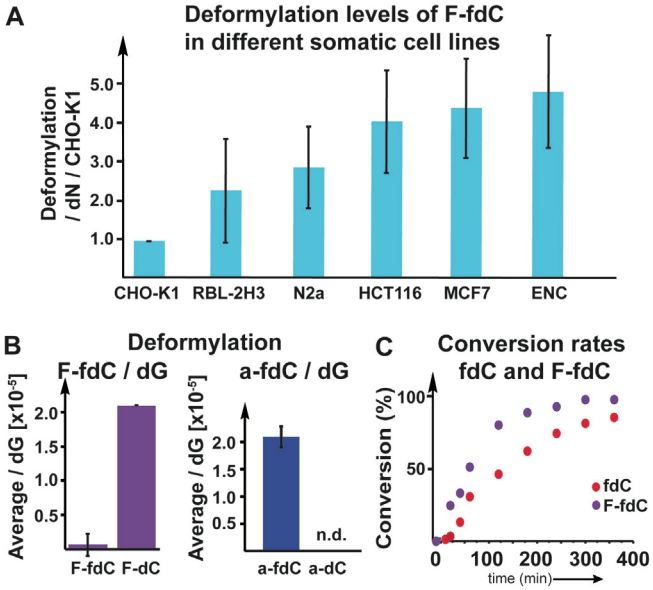


Figure 3. A) Deformylation data for F-fdC in different cell types, showing that F-fdC is deformylated in very different cells. Deformylation rate was calculated by the F-dC + F-mdC/dN per F-fdC/dN, then the values were normalized to the cell line with the lowest deformylation level (CHO-K1 = 1). B) The deformylation of F-fdC/dG and a-fdC/dG, showing the induced differences due to C6-carbon-to-nitrogen exchange. C) The bisulfite data show that the deformylation of fdC and F-fdC is comparable, thus showing that the 2′-F substitution has only a small accelerating effect, whereas the reaction of a-fdC could not be detected.

Although the ease of hydrate formation may foster deformylation, the hydrate/carbonyl equilibrium makes efficient detection of compound **3** basically impossible. In order to circumvent the problem, we started to derivatize a-fdC (**3**) before analysis with methoxyamine. Addition of CH_3ONH_2 (150 mM) to the digestion solution indeed provided the methoxyoxime of a-fdC in quantitative yield after just 15 min at 25 °C and pH 10. The naturally present fdC (**1**) and the probe molecule F-fdC (**2**), however, react as well, but unfortunately not quantitatively. To reduce impurities during MS measurements, we decided against using a catalyst for oxime formation. We therefore decided to analyse the digested DNA in two batches. The first one contained the digested untreated DNA to quantify all bases other than a-fdC. In the second batch, we treated the digested DNA with methoxyamine for a-fdC quantification. For quantification of the derivatized a-fdC, we constructed an external calibration curve (see the Supporting Information).

With this method in hand, we next quantified all nucleosides present in the genome of the cells treated with a mixture of **2** + **3**.

Figure 3 A shows that we indeed detected the fluorinated F-dC (**2**), thus confirming very efficient deformylation activity. We tested different cell types and found different levels of deformylation activity. But in all cases, the conversion of F-fdC into F-dC was clearly detectable. Most interesting is that we observed the highest deformylation activity in cells associated with neuronal properties. This is in line with neurons featuring the highest levels of hmdC and fdC. In contrast, Figure 3 B shows that for a-fdC (**3**), we were unable to detect any formation of the deformylated compound a-dC despite the high propensity of **3** to exist in the hydrated form, which is one prerequisite for efficient C–C bond cleavage. This result suggests that the ability to react with a nucleophile at the 6-position is also required in vivo for efficient deformylation.

In order to substantiate this result, we next performed in vitro studies with bisulfite. Bisulfite is a strong nucleophile that has been reported to cause deformylation of fdC by first attacking the C6 position, followed by conversion of the C5, C6-saturated fdC adduct into the bisulfite adduct, which then undergoes deformylation.[22] The deformylated product dC is then further converted into dU by the well-known bisulfite-induced deamination reaction of dC (see the Supporting Information). Indeed, when we reacted fdC with bisulfite, we observed efficient deformylation and deamination to dU. We then studied to what extent the reaction is influenced by the 2′-F atom present in F-fdC, in order to estimate whether the in vivo deformylation could be just the result of the 2′-F atom. Treatment of F-fdC with bisulfite also led to deformylation and deamination to F-dU, and indeed the reaction is a little faster compared to fdC (see Figure 3 C). Although the difference is measurable, it is in total rather small. With these data in hand, we can conclude that we may overestimate the amount of deformylation that can occur with fdC lacking the 2′-F atom. We can certainly exclude that deformylation in vivo occurs only with F-fdC. It is unfortunate that we are unable to measure the direct deformylation of fdC because of the presence of efficient BER processes. A TDG −/− cell line showed a huge increase in fdC compared to the TDG +/−, whereas a-fdC and F-fdC stay constant, thus showing that these compounds are indeed not repaired by the

TDG protein (see the Supporting Information). The bisulfite studies, however, show that the F-fdC compound is not a perfect but sufficient reporter of this C–C bond cleavage. Treatment of a-fdC (**3**) with bisulfite did not provide the deformylated product a-dC under any circumstances, showing that the inability to react with a nucleophile at the 6-position totally blocks the C–C bond cleavage. We can therefore conclude that the deformylation of fdC during active demethylation requires oxidation of mdC to fdC. fdC can undergo a direct C–C bond cleavage to dC, but this reaction requires a helper nucleophile to attack the C6-position, which is blocked in the case of a-fdC by the lone pair introduced by the C6-carbon-to-nitrogen exchange. While the chemistry that allows the transformation of fdC into dC is now elucidated, we next need to find the nucleophiles that perform the reaction in vivo.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) GRK2338 (Project ID 321812289), SFB1309 (PID 325871075), SFB1361 (PID 393547839) and SPP1784 (PID 255344185). This project has received additional funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° EPiR 741912) and through a H2020 Marie Skłodowska-Curie Action (LightDyNAmics, 765866).

Conflict of interest

The authors declare no conflict of interest.

Keywords: demethylation · DNA modifications · epigenetics · formylcytidine

How to cite: *Angew. Chem. Int. Ed.* **2020**, *59*, 5591 – 5594
Angew. Chem. **2020**, *132*, 5639 – 5643

[1] M. Wagner, J. Steinbacher, T. F. J. Kraus, S. Michalakis, B. Hackner, T. Pfaffeneder, A. Perera, M. Müller, A. Giese, H. A. Kretzschmar, T. Carell, *Angew. Chem. Int. Ed.* **2015**, *54*, 12511 – 12514; *Angew. Chem.* **2015**, *127*, 12691 – 12695.

[2] S. Liu, J. Wang, Y. Su, C. Guerrero, Y. Zeng, D. Mitra, P. J. Brooks, D. E. Fisher, H. Song, Y. Wang, *Nucleic Acids Res.* **2013**, *41*, 6421 – 6429.

[3] C. X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S. Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G. L. Xu, P. Jin, C. He, *Cell* **2013**, *153*, 678 – 691.

[4] T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008 – 7012; *Angew. Chem.* **2011**, *123*, 7146 – 7150.

[5] M. Bachman, S. Uribe-Lewis, X. Yang, H. E. Burgess, M. Iurlaro, W. Reik, A. Murrell, S. Balasubramanian, *Nat. Chem. Biol.* **2015**, *11*, 555 – 557.

[6] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929 – 930.

[7] D. Q. Shi, I. Ali, J. Tang, W. C. Yang, *Front. Genet.* **2017**, *8*, 100.

[8] M. Münzel, D. Globisch, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 6460 – 6468; *Angew. Chem.* **2011**, *123*, 6588 – 6596.

[9] C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. T. C. Jansen, C. Bauer, M. Münzel, M. Wagner, M. Müller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Müller, J. Walter, R. Boelens, H. Van Ingen, H. Leonhardt, T. Carell, M. Vermeulen, *Cell* **2013**, *152*, 1146 – 1159.

[10] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300 – 1303.

[11] F. Neri, D. Incarnato, A. Krepelova, S. Rapelli, F. Anselmi, C. Parlato, C. Medana, F. DalBello, S. Oliviero, *Cell Rep.* **2015**, *10*, 674 – 683.

[12] M. Su, A. Kirchner, S. Stazzoni, M. Müller, M. Wagner, A. Schröder, T. Carell, *Angew. Chem. Int. Ed.* **2016**, *55*, 11797 – 11800; *Angew. Chem.* **2016**, *128*, 11974 – 11978.

[13] T. Fu, L. Liu, Q. L. Yang, Y. Wang, P. Xu, L. Zhang, S. Liu, Q. Dai, Q. Ji, G. L. Xu, C. He, C. Luo, L. Zhang, *Chem. Sci.* **2019**, *10*, 7407 – 7417.

[14] R. Rahimoff, O. Kosmatchev, A. Kirchner, T. Pfaffeneder, F. Spada, V. Brantl, M. Müller, T. Carell, *J. Am. Chem. Soc.* **2017**, *139*, 10359 – 10364.

[15] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593 – 14599.

[16] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, C. Hagemeier, M. Truss, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 6516 – 6520; *Angew. Chem.* **2012**, *124*, 6622 – 6626.

[17] K. Iwan, R. Rahimoff, A. Kirchner, F. Spada, A. S. Schröder, O. Kosmatchev, S. Ferizaj, J. Steinbacher, E. Parsa, M. Müller, T. Carell, *Nat. Chem. Biol.* **2018**, *14*, 72 – 78.

[18] A. Jeltsch, *ChemBioChem* **2002**, *3*, 274 – 293.

[19] Q. Du, Z. Wang, V. L. Schramm, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 2916 – 2921.

[20] I. V. Alekseeva, A. S. Shalamai, V. S. Shalamai, V. P. Chernetski, *Ukr. Khim. Zh. (Ukr. Ed.)* **1976**, *42*, 398 – 401.

[21] F. R. Traube, S. Schiffers, K. Iwan, S. Kellner, F. Spada, M. Müller, T. Carell, *Nat. Protoc.* **2019**, *14*, 283 – 312.

[22] E. Kriukienė, Z. Liutkevičiūtėa, S. Klimašauskas, *Chem. Soc. Rev.* **2012**, *41*, 6916 – 6930.

Manuscript received: January 9, 2020
Accepted manuscript online: January 30, 2020
Version of record online: February 25, 2020

4.2. Studies on sequence dependent deformylation and decarboxylation

"Direct and Base Excision Repair-Mediated Regulation of a GC-Rich cis-Element in Response to 5-Formylcytosine and 5-Carboxycytosine" Nadine Müller, Eveliina Ponkkonen, Thomas Carell and Andriy Khobta*, *Int. J. Mol. Sci.* **2021**, *22*, 11025

#: These authors contributed equally

*: Corresponding author

Prologue

To gain knowledge of the role of fdC and cadC at defined genetic elements and their effect on the expression of reporter constructs, eight 18-mer oligonucleotides respectively containing one incorporation of fdC, cadC, F-fdC or F-cadC were synthesized. These were further incorporated into the CpG-dinucleotide of the GC box *cis*-element that was upstream from the RNA polymerase II core promoter. This modified gene element retains its full activity when the CpG dinucleotide is hemi-methylated and oxidation to hmdC and fdC is well tolerated. However, in the absence of BER or if protected from the *N*-glycosylase activity in case of the 2´-fluorinated analogue, cadC was shown to cause strong inhibition of the promoter activity suggesting a role as a repressory mark. The presence of TDG restored the GC box activity almost completely and BER of cadC caused significant promoter activation. Thus, cadC was suggested to act as a repressory mark to prevent premature activation of promoters that are undergoing the final stages of DNA demethylation, when the symmetric CpG methylation has been lost. However, after few hours of reactivation, the silencing response returned which was shown to be connected to the strand cleavage at AP lesion that requires 5´-endonuclease activity, likely APE1. Thereby it could be concluded that downstream promoter activation or repression responses are regulated by two separate BER steps, where TDG and APE1 can act as potential switches.

Contribution to this study

For this work, I synthesized F-fdC and F-cadC phosphoramidites and incorporated them each into four different oligonucleotides via solid phase synthesis. The strands were synthesized in big amounts, extensively purified via HPLC, characterized and further processed by the group of Prof. Andriy Khobta.

License

International Journal of
Molecular Sciences

MDPI

Article

Direct and Base Excision Repair-Mediated Regulation of a GC-Rich *cis*-Element in Response to 5-Formylcytosine and 5-Carboxycytosine

Nadine Müller [1], Eveliina Ponkkonen [2], Thomas Carell [2] and Andriy Khobta [1,3,*]

[1] Institute of Toxicology, University Medical Center Mainz, 55131 Mainz, Germany; nadgeier@uni-mainz.de
[2] Department of Chemistry and Biochemistry, Ludwig-Maximilians University Munich, 81377 Munich, Germany; eveliina.ponkkonen@cup.uni-muenchen.de (E.P.); thomas.carell@cup.uni-muenchen.de (T.C.)
[3] Institute of Nutritional Sciences, Friedrich Schiller University Jena, 07743 Jena, Germany
* Correspondence: a.khobta@uni-jena.de

Abstract: Stepwise oxidation of the epigenetic mark 5-methylcytosine and base excision repair (BER) of the resulting 5-formylcytosine (5-fC) and 5-carboxycytosine (5-caC) may provide a mechanism for reactivation of epigenetically silenced genes; however, the functions of 5-fC and 5-caC at defined gene elements are scarcely explored. We analyzed the expression of reporter constructs containing either 2′-deoxy-(5-fC/5-caC) or their BER-resistant 2′-fluorinated analogs, asymmetrically incorporated into CG-dinucleotide of the GC box *cis*-element (5′-TGGGCGGAGC) upstream from the RNA polymerase II core promoter. In the absence of BER, 5-caC caused a strong inhibition of the promoter activity, whereas 5-fC had almost no effect, similar to 5-methylcytosine or 5-hydroxymethylcytosine. BER of 5-caC caused a transient but significant promoter reactivation, succeeded by silencing during the following hours. Both responses strictly required thymine DNA glycosylase (TDG); however, the silencing phase additionally demanded a 5′-endonuclease (likely APE1) activity and was also induced by 5-fC or an apurinic/apyrimidinic site. We propose that 5-caC may act as a repressory mark to prevent premature activation of promoters undergoing the final stages of DNA demethylation, when the symmetric CpG methylation has already been lost. Remarkably, the downstream promoter activation or repression responses are regulated by two separate BER steps, where TDG and APE1 act as potential switches.

Keywords: DNA demethylation; 5-formylcytosine; 5-carboxycytosine; thymine DNA glycosylase (TDG); base excision repair (BER); gene regulation; epigenetic marks

Citation: Müller, N.; Ponkkonen, E.; Carell, T.; Khobta, A. Direct and Base Excision Repair-Mediated Regulation of a GC-Rich *cis*-Element in Response to 5-Formylcytosine and 5-Carboxycytosine. *Int. J. Mol. Sci.* **2021**, *22*, 11025. https://doi.org/10.3390/ijms222011025

Academic Editor: Ennio Prosperi

Received: 7 September 2021
Accepted: 7 October 2021
Published: 13 October 2021

1. Introduction

Epigenetic regulation of the genome function is crucial for concerted realization of gene expression programs during development and for maintenance of the lineage-specific gene expression patterns in adulthood. During these processes, functions of the whole genome and of the individual genes are dynamically regulated by deposition and removal or maintenance of epigenetic marks, which include specific chromatin components as well as chemically modified DNA bases [1,2]. 5-methylcytosine (5-mC) at CpG dinucleotides is a major DNA modification in vertebrates, and the only inheritable one. It regulates numerous cellular processes, including tissue-specific gene expression, genomic imprinting and X-chromosome inactivation [3]. Depending on the nature of a specific gene regulatory element (and perhaps cell lineage), a methylated cytosine base can function to promote or preclude recruitment of regulatory proteins by two different mechanisms [4]. Symmetric methylation of CpG sites can be read by the methyl-CpG binding protein (MBP) family members [5,6], which mediate transcriptional repression by recruitment of chromatin-modifying enzymes [7–9]. In addition, the methyl mark sometimes acts by directly preventing transcription factor binding to their target sites [10–12].

Recent progress in the field of epigenetic reprogramming has revealed that 5-mC can undergo enzymatic oxidation by ten-eleven translocation (TET) family dioxygenases to 5-hydroxymethylcytosine (5-hmC) and further to 5-formylcytosine (5-fC) and 5-carboxycytosine (5-caC) [13–15]. The latter two modifications are recognized by the TDG DNA N-glycosylase and cleared from DNA by the base excision repair (BER) pathway [16–18]. Newly developed whole-genome sequencing techniques revealed enrichment of 5-fC and 5-caC in promoters of transcribed genes and active enhancers, indirectly suggesting their roles in transcriptional activation [19–21]. Still, investigation of the dynamics of the oxidation products of 5-mC at specific genomic loci and its significance for the regulation of transcription is hampered by low abundances of these modifications in the genome and a troublesome discrimination between different cytosine modifications by sequencing techniques [22,23]. It remains a matter of controversy whether 5-hmC should be regarded as a functionally autonomous epigenetic mark, similar to 5-mC, or barely a demethylation intermediate which does not have a dedicated function [22–24]. Potential biological functions of the downstream products of the TET-mediated oxidation pathway, 5-fC and 5-caC, are even harder to characterize, as these modifications are much scarcer in genomic DNA [23].

To directly address functional impacts of 5-hmC and its derivatives on the regulation of gene expression, these modifications can be site-specifically incorporated into a suitable reporter vector. We previously used such an approach to investigate the effects of the 5-mC oxidation products in a cyclic adenosine monophosphate response element (CRE). There, 5-mC, 5-hmC, 5-fC and 5-caC all negatively affected the promoter activity by inhibition of the cognate transcription factor CREB binding [25], thus corroborating the mechanism earlier described for 5-mC in the central CpG dinucleotide of the CRE sequence [10]. In addition, 5-fC and 5-caC initiated an indirect silencing mechanism attributed to the base excision and DNA strand cleavage [25]. Thereby, the outcomes of 5-fC and 5-caC in CRE largely recapitulated responses to common types of DNA damage processed by the BER [26–28]. It should be kept in mind, however, that binding of regulatory proteins to their cognate target elements in DNA, and consequently the outcomes of these interactions, can be highly context-specific [11,12,29]. Here, we systematically investigated the effects of synthetic 5-mC, 5-hmC, 5-fC and 5-caC incorporated into a GC-rich regulatory element (GC box), on the promoter activity. To model potential DNA demethylation stages, we placed defined cytosines modifications asymmetrically in either strand of the GC box, opposite to the non-methylated CpG dinucleotide. This led to identification of distinct functional states of the promoter conferred by defined steps of the DNA demethylation pathway.

2. Results

2.1. Gene Repression in HeLa Cells Induced by 5-fC and 5-caC in the GC Box CpG Dinucleotide

To investigate the impact of 5-mC oxidation products in a GC-rich promoter on the gene expression, we generated enhanced green fluorescence protein (EGFP) reporter constructs containing single 5-mC, 5-hmC, 5-fC or 5-caC residues at the CpG site of an artificial promoter containing the 5′-TGGGCGGAGC-3′ GC box sequence as the only *cis*-regulatory element (Figure 1a). In the pGCbox-W and pGCbox-C plasmid vectors, this GC box enhances the basal gene expression of the downstream *EGFP* gene by the factor of $\geq$2, compared to a scrambled DNA sequence [30]. Synthetic oligonucleotides containing all types of cytosine modifications were incorporated into the pGCbox-W vector, with efficiencies closely approaching 100% (Figure 1b,c). Moreover, the fractions of correctly ligated vector DNA were >92% for all modifications, which warranted that DNA topology or misalignment of the inserted synthetic oligonucleotides would not affect subsequent measurement of the promoter activity in cells.

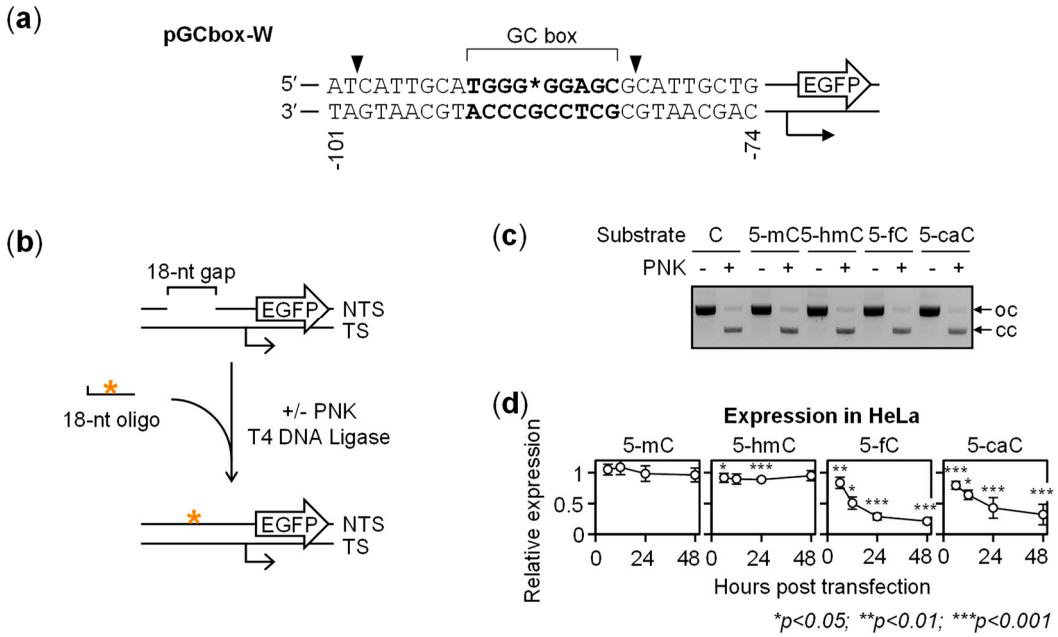


Figure 1. Effects of 5-methylcytosine (5-mC), 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC) and 5-carboxycytosine (5-caC) in the purine-rich strand on the GC box activity. (**a**) Scheme of the reporter enhanced green fluorescence protein (*EGFP*) gene under the control of a GC box as the only upstream activating element. Synthetic oligonucleotide containing cytosine modifications at the unique CpG site (position indicated with an asterisk) were incorporated into the gap generated with the Nb.BsrDI nicking endonuclease (cleavage sites indicated with arrowheads). (**b**) Procedure for the incorporation of synthetic oligonucleotides containing C/5-mC/5-hmC/5-fC/5-caC (*) into the gap generated by depletion of the purine-rich strand of the targeted GC box. Aliquots of the same annealing reactions were incubated with or without T4 polynucleotide kinase (PNK) to validate full replacement of the native DNA strand. DNA strand labeling denotes transcribed strand (TS) of the *EGFP* gene and the non-transcribed strand (NTS); broken arrow indicates the transcription start site and direction. (**c**) Agarose gel electrophoresis of the reporter constructs generated by targeted incorporation of C/5-mC/5-hmC/5-fC/5-caC into plasmid DNA. Arrows indicate the open circular (oc) and covalently closed (cc) forms. (**d**) Expression time course of constructs containing 5-mC/5-hmC/5-fC/5-caC in transfected HeLa cells. All values (mean ± SD) are calculated relative to the expression of the control construct harboring synthetic oligonucleotide containing cytosine for $n = 6$ independent experiments (the 12-h point was skipped in three of the experiments). Representative flow cytometry data is shown in Supplementary Materials Figure S1.

Time course analyses of pGCbox-W constructs containing different cytosine modifications at the CpG site in the purine-rich strand showed that 5-mC and 5-hmC did not appreciably influence the *EGFP* gene expression levels in transfected HeLa cells (Figure 1d). In contrast, constructs containing 5-fC and 5-caC showed considerably decreased EGFP expression levels. These effects were present already 6 h after transfection (the earliest time point when the expression could be reproducibly measured) and grew stronger in the course of time. At the initial time point, the decrease of the gene expression was more significant for 5-caC than for 5-fC, whereas subsequent dynamics was somewhat faster for 5-fC, leading to a stronger impairment of the gene expression at the later time points. Thus, relative to the reference construct containing unmodified cytosine, the expression levels decreased from 83.4 ± 6.5% to 21.0 ± 3.7% for 5-fC between 6 and 48 h post-transfection. For 5-caC, the promoter activity decreased from 79.4 ± 6.5% to 32.0 ± 16.7% in the same experiment series.

Also in the pyrimidine-rich DNA strand, the incorporation rates of all cytosine modifications were very close to 100%, with >88% synthetic strands fully and correctly ligated (Figure 2). As in the opposite strand, neither 5-mC nor 5-hmC caused considerable changes of the gene expression levels over the entire time interval, whereas 5-fC and 5-caC caused the gene expression to decline strongly and in a time-dependent manner. Once again, the inhibition of gene expression was initially stronger in case of 5-caC, whereas the 5-fC

construct demonstrated a faster decline of the gene expression in the course of time. The 6- and 48-h expression values decreased from 87.2 ± 8.1% to 20.3 ± 3.5% for 5-fC and from 82.0 ± 9.8% to 29.3 ± 6.3% for 5-caC. Combined, the results show that 5-hmC and 5-mC in either strand of the GC box did not tangibly affect the promoter activity over a period of at least 48 h. In contrast, 5-fC and 5-caC clearly perturbed the gene expression in a time-dependent fashion.

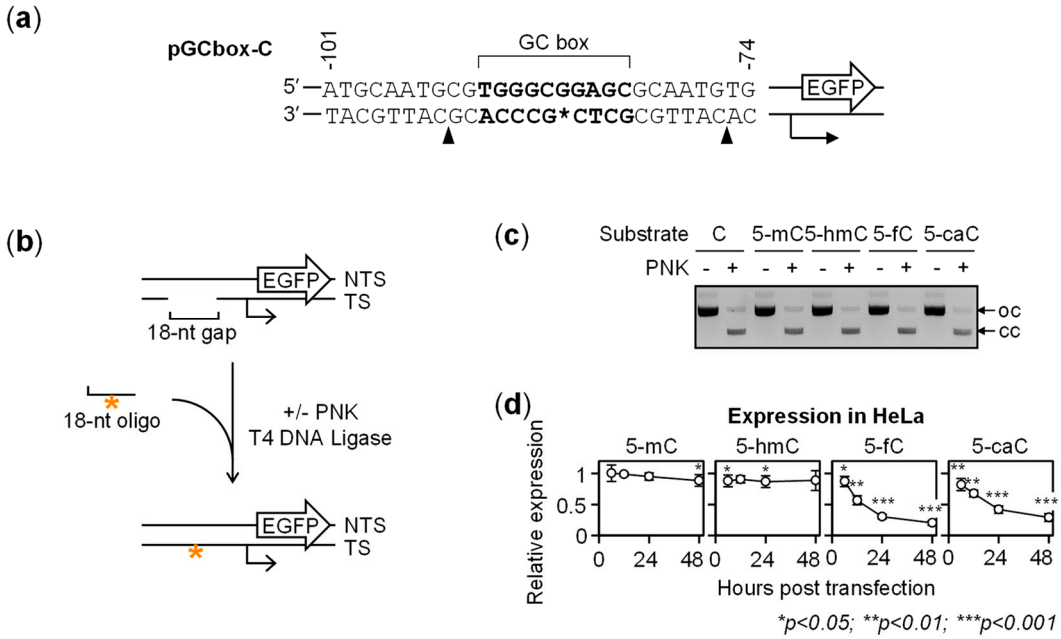


Figure 2. Effects of 5-mC, 5-hmC, 5-fC and 5-caC in the pyrimidine-rich strand of the GC box. (**a**) Promoter sequence with the modified cytosine position (asterisk) and Nb.BsrDI nicking sites (arrowheads). (**b**) Scheme of the incorporation procedure of C/5-mC/5-hmC/5-fC/5-caC (*) into the pyrimidine-rich strand of the GC box. (**c**) Agarose gel electrophoresis of the reporter constructs containing C/5-mC/5-hmC/5-fC/5-caC. The open circular (oc) and covalently closed (cc) forms are indicated by arrows. (**d**) The expression time course in transfected HeLa cells (mean ± SD) for $n = 6$ independent experiments (the 12-h point was skipped in three of them). Representative flow cytometry data are shown in Supplementary Materials Figure S1.

2.2. BER-Resistant 5-caC and, to Some Extent, 5-fC Directly Diminish the GC Box Activity

The observed dynamic changes of the expression of constructs carrying 5-fC and 5-caC, in contrast to 5-mC or 5-hmC, could likely be attributed to removal of these modifications from DNA. Therefore, to measure direct impacts of 5-fC and 5-caC on the GC box activity, it was necessary to eliminate their repair. Activities of DNA N-glycosylases towards their substrates, including 5-fC and 5-caC, can be efficiently inhibited by deoxyribose fluorination at the C2′ position [25,31]. This motivated us to generate GC box constructs carrying the 2′-(*R*)-fluorinated derivatives of 5-fC and 5-caC. Both modifications were incorporated into the purine-rich DNA strand as efficiently as the respective 2′-deoxy compounds (Figure 3a). Quantification of the EGFP expression levels in transfected HeLa cells showed that 2′-(*R*)-fluorination entirely reversed the repressory effect of 5-fC, as the difference between the 6- and 24-h points disappeared. The expression levels were 81.2 ± 9.8% (at 6 h) and 84.6 ± 8.4% (at 24 h post-transfection), relative to the construct containing cytosine. The respective values for 2′-deoxy 5-fC were 88.0 ± 5.7% and 31.0 ± 4.1% (Figure 3b). The repression of the promoter activity caused by 5-fC between 6 and 24 h post-transfection was statistically highly significant ($p = 8.6 \times 10^{-6}$, Student's two tailed heteroscedastic t-test), whereas the effect of 2′-(*R*)-fluorinated 5-fC was not ($p = 0.62$). In the case of 5-caC, the differences between the 6- and 24-h time points were, again, highly significant for deoxynucleotide ($p = 1.6 \times 10^{-4}$) but not for the 2′-fluorinated analog ($p = 0.38$). We thereby conclude that dynamic changes of the promoter activity induced by 5-fC and 5-caC are the consequence of excision of the modified bases by a DNA N-glycosylase mechanism.

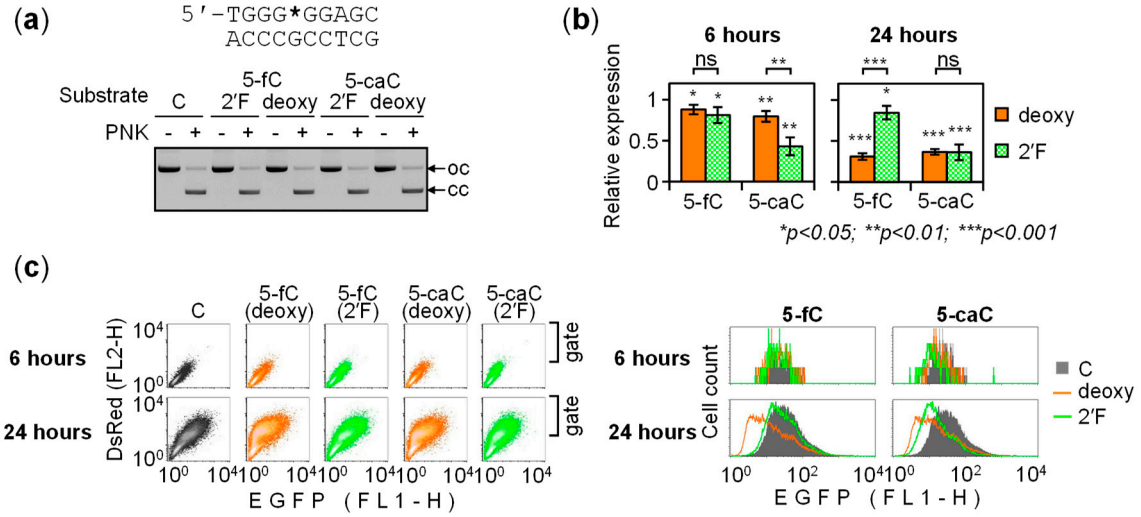


Figure 3. Effects of BER-resistant analogs of 5-fC and 5-caC in the purine-rich strand on the GC box activity. (**a**) Efficient incorporation of 5-fC and 5-caC deoxyribonuceotides (deoxy) and their 2′-(*R*)-fluorinated analogs (2′F) into the purine-rich strand of the GC box. Agarose gel electrophoresis of the fully ligated constructs and the respective "no PNK" controls. The open circular (oc) and covalently closed (cc) forms are indicated by arrows. (**b**) Quantification of the EGFP expression driven by GC box containing 5-fC and 5-caC with or without 2′-fluorination, relative to C, 6 and 24 h post-transfection (mean ± SD, $n = 4$ independent experiments). (**c**) A representative flow cytometry experiment. Cells were gated based on the expression of the transfection marker DsRed, as shown on two-dimensional fluorescence scatter-plots (left panels), and the resulting EGFP signal distributions (right panels, samples overlaid) analyzed to determine the median EGFP fluorescence.

As 2′-(*R*)-fluorinated 5-fC and 5-caC in the GC box constructs showed stable levels of the reporter gene expression, it was now possible to quantify their direct effects on the promoter activity, based on the EGFP expression levels in transfected HeLa cells. We used 24-h values to estimate these effects, since fluorescent cell counts as well as average fluorescence intensity per cell were higher at this time point (Figure 3c). Reduction of the gene expression by 2′-(*R*)-fluorinated 5-fC was minimal yet statistically significant ($p = 0.035$). The effect of 2′-(*R*)-fluorinated 5-caC was much stronger (36.2 ± 9.6% residual gene expression relative to the cytosine control) and statistically highly significant ($p = 9.3 \times 10^{-4}$). Considering that GC box enhances the gene expression by a factor of 2 to 2.5 over the basal expression level [30], the results indicate that the presence of a single 5-caC in the purine-rich strand abolished the activation attributable to GC box completely or almost completely.

We also analyzed the effects of the 2′-(*R*)-fluorinated 5-fC and 5-caC analogs in the pyrimidine-rich GC box strand (Figure 4). As in the purine-rich strand, 2′-fluorination abolished time-dependent repression by both cytosine modifications, as judged by disappeared differences between the 6- and 24-h expression values. The respective *p*-values were 8.1×10^{-5} (5-fC), 0.19 (2′-fluoro 5-fC), 9.1×10^{-4} (5-caC) and 0.30 (2′-fluoro 5-caC). The effects of BER-resistant modifications differed slightly between the strands. Thus, 2′-(*R*)-fluorinated 5-fC in the pyrimidine-rich strand did not at all inhibit the GC box activity, showing the 24-h relative expression level of 102.1 ± 4.1% ($p = 0.38$). The inhibitory effect of 2′-(*R*)-fluorinated 5-caC remained highly significant in the pyrimidine-rich strand, with relative expression of 54.7 ± 2.3% ($p = 3.5 \times 10^{-5}$).

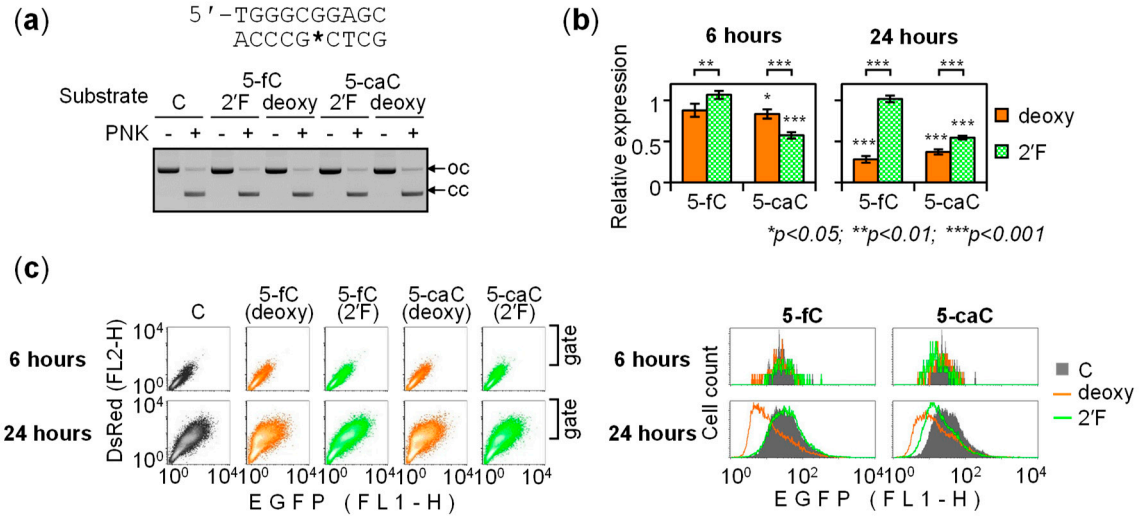


Figure 4. Effects of BER-resistant analogs of 5-fC and 5-caC in the pyrimidine-rich strand on the GC box activity. (**a**) Agarose gel electrophoresis of the fully ligated deoxy and 2′F constructs and the respective "no PNK" controls. The open circular (oc) and covalently closed (cc) forms of the vector are indicated by arrows. (**b**) Relative expression of the 5-fC and 5-caC (deoxy versus 2′F) constructs 6 and 24 h post-transfection (mean ± SD, $n = 4$ independent experiments). (**c**) A representative flow cytometry experiment: two-dimensional fluorescence scatter-plots (left) and the derived EGFP fluorescent distribution plots for the indicated transfection condition (right panels, samples overlaid).

In summary, apart from subtle quantitative differences between the purine- and pyrimidine-rich strands of the GC box, the results indicate that processing of 5-fC and 5-caC by BER plays a key role in the regulation of the gene expression. Over an extended time interval (24 h), in HeLa cells, the role of BER is manifested by dynamic repression of the promoter activity, in contrast to steady expression levels observed in the presence of the repair-resistant analogs. The potent negative effect of 2′-(*R*)-fluorinated 5-caC on the promoter activity strongly suggests that 5-caC functions as a negative regulatory mark in the GC box. The function of 5-fC seems to be different, as this modification has little or no effect on the GC box activity, depending on the DNA strand.

2.3. BER of 5-caC in the GC Box Induces a Transient Promoter Activation

The evidence implicating the base excision in the transcriptional repression overall corroborated the idea that the time-dependent decline of the gene expression observed earlier (Figures 1 and 2) was caused by gradual BER-mediated clearance of 5-fC and 5-caC from DNA. However, even though the cumulative outcomes of the excision of 5-fC and 5-caC were manifested after 24 h as repression of the reporter gene, we noticed that the initial effects of 5-caC were the opposite (Figures 3b and 4b). In particular, the 6-h EGFP expression values showed an inverse relationship between the activities of the GC box constructs containing BER-resistant versus BER-sensitive modifications, wherein the expression was significantly enhanced by 2′-deoxy 5-caC compared to its 2′-fluorinated counterpart in both the purine-rich ($p = 1.2 \times 10^{-3}$) and the pyrimidine-rich DNA strands ($p = 7.2 \times 10^{-4}$). These results indicate that removal of 5-caC reactivated the GC box promoter at least transiently, whereas subsequent gradual decrease of the EGFP expression (reported in Figures 1 and 2) suggests that activation was followed by transcriptional silencing within the next few hours. Thereby, the response to 5-caC in GC box is different from reported previously in CRE, where TDG-mediated silencing occurred without a preceding activation phase [25].

If the pulse of promoter activation during the first hours is specific to 5-caC, it would provide an explanation to the ostensibly slower silencing kinetics in comparison to the 5-fC constructs (Figures 1 and 2). Indeed, the 5-fC constructs did not display a similar early activation by the BER mechanism. Rather on the contrary, BER-sensitive 2′-deoxy 5-fC in

the pyrimidine-rich strand slightly but significantly decreased the expression with respect to its BER-resistant counterpart already after 6 h ($p = 6.4 \times 10^{-3}$), while the respective values for the purine rich strand did not differ significantly ($p = 0.28$).

2.4. The Dynamics of Transcriptional Regulation by 5-fC and 5-caC Is Entirely TDG-Dependent

We next aimed at determining whether the effects described for the 2′-(*R*)-fluorinated analogs of 5-fC and 5-caC hold true also for the modifications in the context of the 2′-deoxyribose backbone. To eliminate BER of 5-fC and 5-caC in HeLa cells, we targeted the *TDG* locus by CRISPR-Cas9 (Figure 5a). This resulted in deletions encompassing the exons 2 to 5 (Supplementary Materials Figure S2) and conferred complete elimination of the TDG protein expression in eight of the selected clones (Figure 5b).

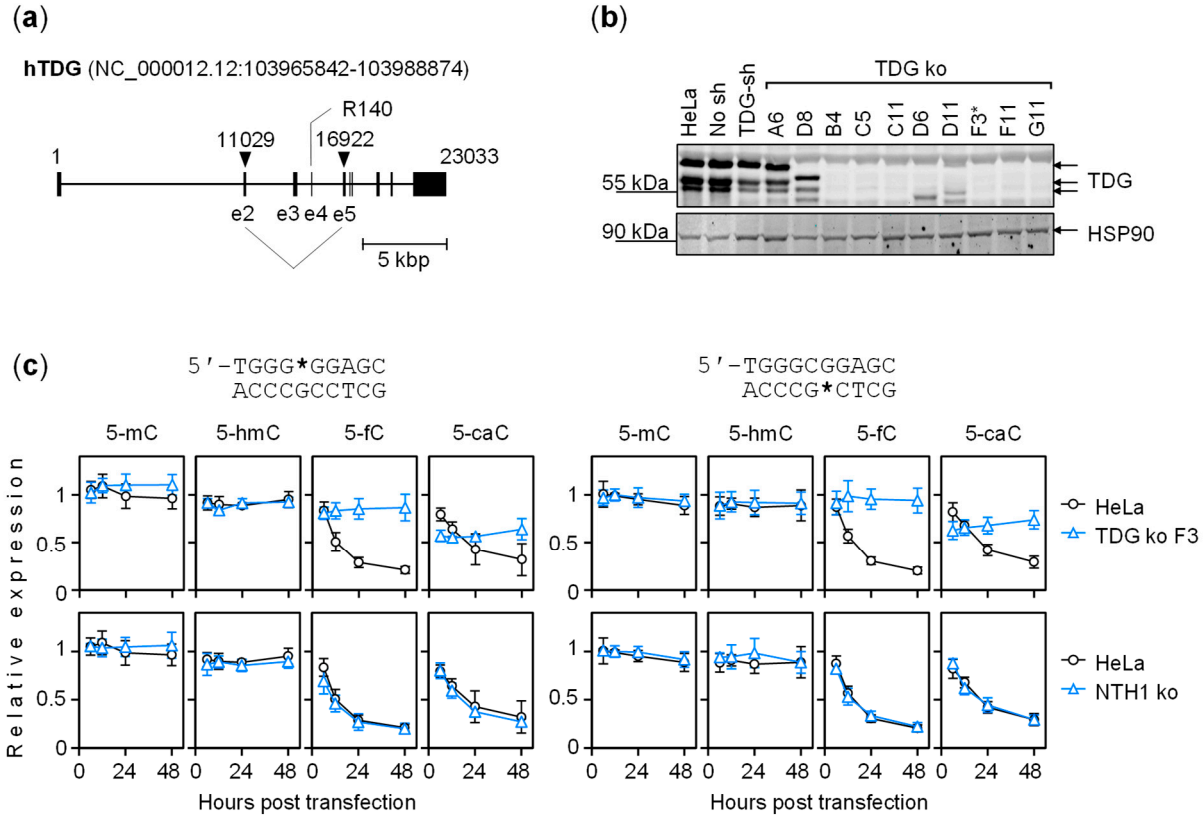


Figure 5. Effect of TDG knockout on the activity of GC box containing 5-mC/5-hmC/5-fC/5-caC. (**a**) Strategy for TDG knockout in Hela cells. The *TDG* gene was targeted simultaneously at two different sites by a pair of single guide RNAs enclosing the codon for the catalytic R140 residue. The Cas9 cut sites are indicated by arrowheads. (**b**) Validation of TDG negative clones by Western blotting. Clone F3 further used as a transfection host for the gene expression analyses is marked (*). (**c**) Expression time course of constructs containing 5-mC/5-hmC/5-fC/5-caC in the purine-rich (left group of plots) or the pyrimidine-rich GC box strand (right). Isogenic clonal TDG knockout (upper row) and NTH1 knockout (lower row) cell lines were compared with the parental HeLa cell line (overlaid in plots). All cell lines were transfected in parallel with the same sets of reporter constructs. Results of $n = 3$ independent experiments (mean ± SD).

We next performed time-course expression analyses of GC box constructs containing the whole spectrum of cytosine modifications in in a TDG knockout cell line derived from clone F3. The results showed that dynamic EGFP expression changes characteristic for the parental HeLa cells were completely abrogated by TDG knockout (Figure 5c). In the absence of TDG, 5-caC steadily inhibited promoter activity, in the same manner as its 2′-(*R*)-fluorinated analog (Figure 5c and Supplementary Materials Figure S3). As this result also closely recapitulated the effect of the 2′-(*R*)-fluorinated analog in the parental HeLa cell line (Figures 3 and 4), we conclude that 2′-fluorination does not influence biological

properties of 5-fC and 5-caC beyond stabilization of the N-glycosidic bond. In the case of 5-fC, a rather mild inhibition of transcriptional activation was observable only when the modification was present in the purine-rich strand. This outcome was, again, consistent with results obtained for the 2′-(*R*)-fluorinated analog (Figures 3 and 4; Supplementary Materials Figure S3).

In an independent TDG knockout clone C11 (not shown), responses to 5-fC and 5-caC were the same as in F3. In contrast, responses in the isogenic NTH1 knockout cell line displayed the same pattern as in the parental HeLa cell line (Figure 5c). We thereby conclude that the phenotype displayed by the TDG knockout clones was specific to the targeted locus. Finally, the expression of constructs carrying 5-mC or 5-hmC essentially did not differ between the cell lines, as expected based on the universally stable character of these modifications (Figure 5c). In summary, the results in TDG knockout cells corroborate the conclusion that unrepaired 5-caC prevents or counteracts the GC box activation in the absence of repair. Besides, the steady character of the effect of 5-caC on the gene expression in the absence of TDG strongly suggests that TDG-dependent BER is by far the most efficient pathway for 5-caC removal in the chosen cell model.

Of note, the results confirmed our previous conclusion that BER is accountable not only for long-term gene silencing induced by 5-fC/5-caC but also for transient activation by 5-caC at the beginning of the time course. Moreover, by comparison of expression levels of the same constructs between cell lines with different TDG states, we now could attribute the activation specifically to TDG, as the 6-h expression values were significantly higher in the parental HeLa cell line than in the derived TDG knockouts (Figure 5c and data not shown). For 5-caC in the purine-rich GC box strand, relative expression level was 57.0 ± 5.8% in the TDG knockout clone F3 versus 84.3 ± 5.6% in HeLa ($p = 4.2 \times 10^{-3}$). In the pyrimidine-rich strand, it was 62.6 ± 9.4% versus 84.7 ± 7.2% in HeLa ($p = 3.2 \times 10^{-2}$). The respective expression levels in clone C11 also lied significantly below the HeLa values: 66.9 ± 5.3% ($p = 1.7 \times 10^{-2}$) for 5-caC in the purine-rich strand and 65.2 ± 5.9% ($p = 2.2 \times 10^{-2}$) in the pyrimidine-rich strand.

2.5. Regulation of the GC Box Promoter by Acytosinic Sites and the Effect of Strand Cleavage

TDG is a monofunctional DNA-N glycosylase that leaves deoxyribose as a reaction product in the DNA strand [17]. Intriguingly, analogous apurinic/apyrimidinic (AP) lesions were implicated in the regulation of GC-rich DNA *cis*-elements previously [30,32]. Therefore, we constructed reporters carrying acytosinic lesions in either of the GC box strands. This was achieved by using oligonucleotides, where 2′-deoxy cytosines at the CpG sites were replaced by two types of tetrahydrofuran AP lesions (Supplementary Materials Figure S4). Being a close structural analog of the natural AP lesion, tetrahydrofuran (F) is efficiently excised by APE1; however, a combination of F with a phosphorothioate linkage on the 5′ side (SF) yields an APE1-resistant lesion [26,33].

The expression levels of GC box constructs containing APE1-resistant acytosinic sites (SF) were slightly decreased with respect to the counterparts containing unmodified C (Figure 6). In both HeLa and TDG knockout cells, they were in the range of 87–90% at the 6-h time point, regardless of the strand, indicating that AP lesion has a mild inhibitory effect on the promoter activity. The respective values for F lesions were somewhat lower (78–83%), which is probably attributable to their incision by APE1. The difference between the F and SF expression levels grew highly significant by the 24-h points: whilst expression of F constructs intensely declined, the expression levels of both SF constructs decreased only slightly. Consequently, the potent gene silencing response to F is attributable to the phosphodiester bond cleavage (Figure 6).

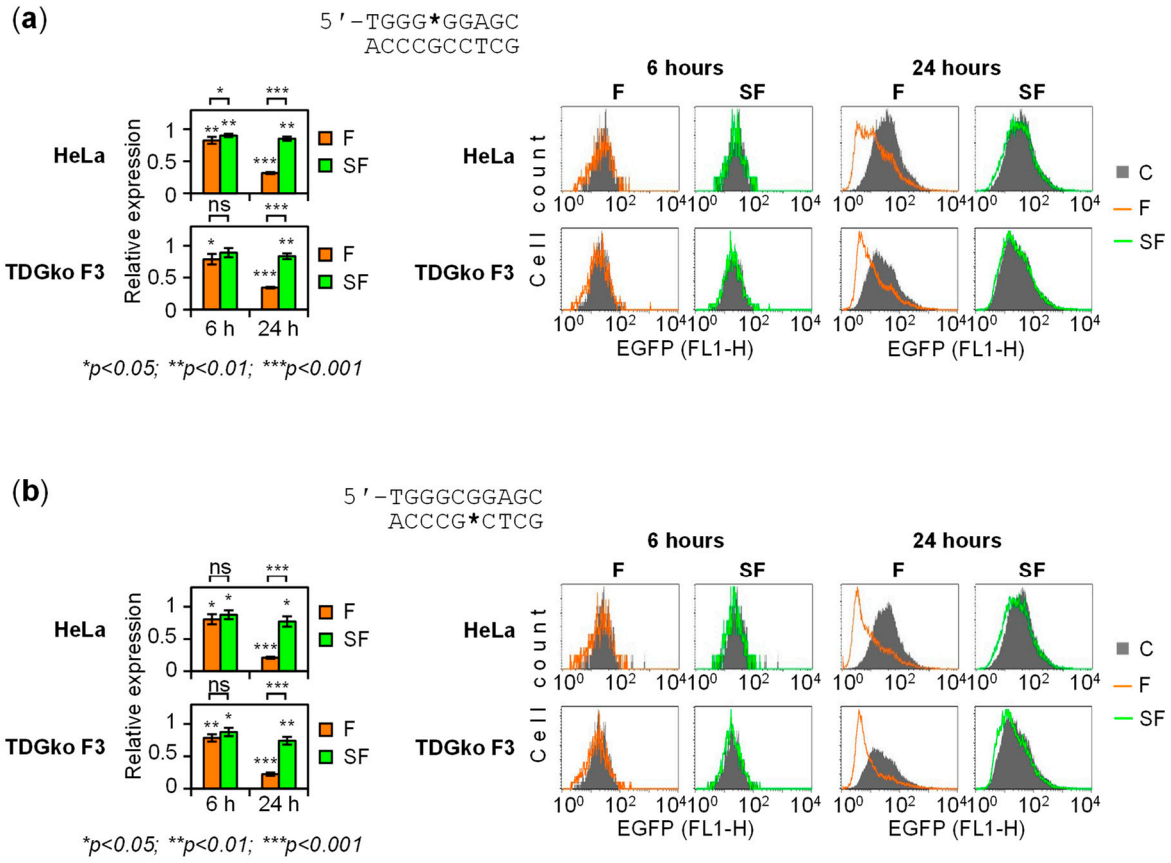


Figure 6. Effect of acytosinic sites on the GC box activity. (**a**) Expression analyses of constructs containing a synthetic abasic site in the purine-rich strand of the GC box. Relative EGFP expression (mean ± SD) for $n = 4$ independent experiments (left) and the representative fluorescent distribution plots (right). Acytosinic modifications used were tetrahydrofuran (F) or a tetrahydrofuran with nuclease resistant phosphorothioate 5′-linkage (SF). (**b**) Analogous analyses of constructs containing F or SF acytosinic sites in the pyrimidine-rich strand.

Endonucleolytic strand cleavage seems to be a rather common inducer of transcriptional repression, as previously reported for AP lesions in other contexts [28,30,32]. In contrast, endonuclease-resistant aguaninic lesions can occasionally cause promoter activation, presumably via a non-processive APE1 binding [30,32]. Based on present results, we deduce that acytosinic lesions at the GC box CpG dinucleotide do not cause a full promoter activation. Nonetheless, they enable substantially higher expression levels than observed in the presence of the unrepairable 5-caC analog (Figures 3 and 4).

3. Discussion

Both active and passive DNA demethylation mechanisms must necessarily involve a step when only one cytosine in a given double-stranded CpG dinucleotide is modified (methylated, hydroxymethylated, formylated or carboxylated) at the C5 position. Recognition of CpG sites by methyl-CpG-binding domain (MBD) proteins, including DNMT1, is strongly inhibited in the presence of 5-hmC [34–36]. This causes impaired methylation maintenance and leads to hemi-hydroxymethylated sites upon replication [37]. An even more pronounced DNMT1 inhibition was reported by 5-fC and 5-caC, implying that replication of DNA containing these modifications would generate hemi-formylated and hemi-carboxylated CpG sites [38]. Hence, we propose that the impacts of defined cytosine modifications on the GC box activity that were reported here can be extrapolated to model functional outcomes of a range of potential DNA demethylation intermediates. Our results suggested that GC box containing a hemi-methylated CpG dinucleotide (with 5-methyl group present in either DNA strand) is as active as in the absence of any modification

(Figures 1, 2 and 5). This is in agreement with biochemical evidence that methylation does not inhibit transcription factor binding to various GC box consensus motifs [39,40]. We further found that GC box remained fully active in the presence of single 5-hmC or 5-fC in the pyrimidine-rich strand, whereas the modifications in the pyrimidine-rich strand seemed to cause only a very minor decrease of the activity (Figures 1–5). In contrast, 5-caC appears to be the only cytosine modification within the TET pathway, which causes a strong direct impairment of the GC box activity (Figures 3–5). Our results thus suggest that the nature of the modification present at asymmetrically modified CpG dinucleotides could be of a critical functional importance.

The inhibitory effect of 5-caC on the GC box can be relieved by a TDG-dependent mechanism. Thus, our results show that expression of constructs carrying 5-caC is re-activated early upon their delivery to cells, but only if TDG is available (Figure 5) and base excision is unhindered (Figures 3 and 4). In HeLa cells, a short period of promoter activation was followed by the onset of a repressed state, which was TDG-dependent as well. However, this response should be regarded as separate from previous activation, since it required, besides base excision, the presence of a labile phosphodiester linkage 5′ to the target nucleotide (Figure 6). A strand cleavage reaction at the critical position is most likely catalyzed by APE1, which is by far the most important AP endonuclease in human cells, whereas the downstream gene silencing mechanism remains to be elucidated. Similar silencing responses were seen with various BER substrates previously and were attributed to adoption of a repressive chromatin structure after the completion of BER [28,41]. It is intriguing to speculate that TDG and APE1 may be subjected to regulation in particular cell lineages and genomic contexts or, perhaps, by endogenous or exogenous signals. This would allow diversification of transcriptional responses and enable plasticity of the epigenetic states in cells undergoing TET- and TDG-dependent DNA demethylation.

To understand the effects of single intermediates arising during the stepwise DNA demethylation pathway on the promoter activity, the expression levels should be related to the amount of a given cytosine modification present. For DNA modifications undergoing repair in cells, quantification of residual modifications specifically in transcription-competent DNA poses a serious technical challenge, because significant fractions of vector DNA distribute to non-nuclear compartments or undergo dynamic changes of expression due to chromatinization [41]. Nevertheless, under the assumption that the effects of BER-resistant 2′-fluorinated synthetic analogs of 5-fC and 5-caC remain steady, it was possible to derive conclusions about direct effects of these modifications on the GC box activity. Indeed, the expression levels in the presence of these modifications, relative to cytosine controls, were found constant over time (Figures 3 and 4). The same holds for natural (2′-deoxy) modifications, when delivered to repair-deficient host cells (Figure 5). Similarly, the effect of AP intermediate could be inferred based on the results obtained with APE1-resistant SF lesion, which also displayed steady expression levels (Figure 6).

4. Conclusions

In summary, our results revealed several levels of GC box regulation by cytosine modifications generated within the active DNA demethylation pathway. GC box upstream from RNA polymerase II core promoter retains its full activity when the CpG dinucleotide is hemi-methylated. Oxidation of 5-mC to 5-hmC and 5-fC is well tolerated; however, the ultimate oxidation product 5-caC behaves as a stable repressory mark in the absence of TDG or if protected from the N-glycosylase activity. TDG restores the GC box activity almost completely. It can be assumed that promoter reactivation occurs as soon as an acytosinic lesion is generated. In support of this notion, modeling of the post-excision step, with the help of AP site analogs, led to very similar effects on the promoter activity (Figure 6). Reactivation by the excision of 5-caC lasted only for a few hours in HeLa cells before getting overturned by a concurrent silencing response, which was elicited by strand cleavage at the AP lesion. Although our data provide no evidence of a more permanent promoter activation by the TDG pathway, such a scenario may take place in another cell type or

promoter context. For instance, AP lesion can be protected from endonucleolytic processing by a non-canonical DNA structure [42–45] or shielded by a specific binding protein [46,47]. Alternatively, signaling downstream from the single strand break generation could be modulated towards a different functional state of the promoter.

5. Materials and Methods

5.1. Synthetic Oligonucleotides Carrying Cytosine Modifications

Deoxyribo-oligonucleotides containing the specified cytosine modifications were 5′-CATTGCATGGG[C*]GGAGCG and 5′-CATTGCGCTC[C*]GCCCACG (where C* is C, 5-mC, 5-hmC, 5-fC or 5-caC). DNA CE-phosphoramidites Bz-dA, Bz-dC, iBu-dG, dT and Bz-mdC were obtained from Glen Research (Sterling, VA, USA) or Link Technologies (Bellshill, Scotland, UK). Syntheses of the 5-hmC, 5-fC and 5-caC phosphoramidites [48], along with the 2′-(*R*)-fluorinated derivatives of 5-fC and 5-caC [31], were performed as described previously. The solid-phase synthesis, HPLC-purification and MALDI/MS quality-control procedures of the 18-mer deoxyribo-oligonucleotides were performed by using the standard protocols described previously [31]. Synthetic apurinic/apyrimidinic (AP) lesions were tetrahydrofuran with either the phosphodiester (F) or the APE1-resistant phosphorothioate 5′-linkage (SF). Oligonucleotides 5′-CATTGCATGGG[AP]GGAGCG and 5′-CATTGCGCTC[AP]GCCCACG (where AP is F or SF) were purchased from BioSpring GmbH (Frankfurt am Main, Germany).

5.2. Generation of Reporter Constructs Containing Cytosine Modifications in the GC Box CpG Dinucleotide

Vectors pGCbox-W and pGCbox-C, allowing substitution of the selected GC box strand with synthetic oligonucleotides, were described previously [30]. Both vectors contain a common GC box motif 5′-TGGGCGGAGC as the only *cis*-regulatory element upstream from the RNA polymerase II transcription initiation site and sustain equivalent levels of the reporter *EGFP* gene expression. Defined modifications targeting cytosines of the CpG dinucleotide were introduced into the purine-rich GC box strand of pGCbox-W or into the pyrimidine-rich strand of pGCbox-C, using the available sites by the Nb.BsrDI nicking endonuclease. The procedure was described in detail previously for introducing 8-oxo-7,8-dihydro-2′-deoxyguanosine (8-oxoG) into the same vectors [30]. Plasmid DNA was cut at two tandem sites by the Nb.BsrDI nicking endonuclease, and the excised native DNA strand fragments were substituted for synthetic oligonucleotides containing the modifications of choice, as described previously for a different promoter context [25]. The efficient incorporation of synthetic DNA strands was verified by inhibition of ligation in the absence of T4 polynucleotide kinase [49]. Percentages of covalently closed DNA in the vector preparations were determined by agarose gel electrophoresis in the presence of 0.5 mg/L ethidium bromide, followed by band quantification, using a GelDoc™ EZ imager and the ImageLab™ software (Bio-Rad Laboratories, GmbH, Munich, Germany), as described previously [28]. The presence of AP lesions was verified by excision analysis, using endonuclease IV (NEB GmbH, Frankfurt am Main, Germany). Constructs were incubated with endonuclease IV (8 U/200 ng plasmid DNA) 1 h at 37 °C in 15 µL buffer composed of 10 mM HEPES (pH 7.5), 200 mM NaCl, 1 mM ethylenediaminetetraacetic acid and 0.1 mg/mL nuclease free bovine serum albumin (NEB). The enzyme was heat-inactivated for 20 min, at 85 °C.

5.3. Quantitative Analyses of EGFP Expression in Transfected Cells

HeLa cells used in experiments were clones descending from HELA cervical carcinoma cell line (German Collection of Microorganisms and Cell Cultures No. ACC 57). The derived TDG and NTH1 knockout cell lines were generated in our lab. Cells exponentially growing in 6-well plates were transfected with mixtures containing 400 ng GC box reporter vector (pGCbox-W or pGCbox-C) and 400 ng tracer pDsRed-Monomer-N1 vector (Clontech, Saint-Germain-en-Laye, France), using the Effectene reagent (QIAGEN, Hilden,

Germany), as described previously [25,30]. GC box constructs containing C or the specified modifications at the respective site were transfected in parallel. Cells were split 6 h after transfection and either fixed immediately with 1% of formaldehyde or seeded into separate wells to be fixed after the indicated time intervals, as described previously. Harvested formaldehyde-fixed cells were analyzed by using a FACSCalibur™ flow cytometer and the CellQuest™ Pro software (Beckton Dickinson GmbH, Heidelberg, Germany). The EGFP expression was quantified as median FL1-H fluorescence over the population of transfected cells, defined by the DsRed expression, as described previously [41]. Relative expression levels were calculated for each modification type in the individual experiments based on the expression of the control construct harboring cytosine.

5.4. TDG Gene Knockout in HeLa Cells

HeLa-derived clonal cell lines with deletion of a critical portion of the *TDG* gene were generated by a CRISPR-Cas9-mediated gene-editing procedure, using the pX330-sgCas9-HF1 vector (Addgene, Watertown, MA, USA) according to the supplier's instructions [50], with minor adjustments. Single guide RNA (sgRNA) sequences targeting human *TDG* locus were designed by using CHOPCHOP online tool and subcloned into the BbsI sites [51]. A pair of sgRNAs targeting the exons 2 (sg11029) and 5 (sg16922) was identified as the most efficient, based on screening of four different sgRNAs. The inserts used for cloning of these sgRNAs were obtained by pairwise annealing of synthetic oligonucleotides (Eurofins MWG Operon, Ebersberg, Germany): 5′-CACCGACGAAATATGGACGTTCAAG (sg110292, forward) with 5′-AAACCTTGAACGTCCATATTTCGTC (sg11029, reverse) and 5′-CACCGCTACCAGGGAAGTATGGTAT (sg16922, forward) with 5′-CACCGCTACCAG GGAAGTATGGTAT (sg16922, reverse). Exponentially growing HeLa cells were co-transfected in 6-well plates with the combination of both sgRNA/Cas9 expression vectors (300 ng each) and 50 ng pZAJ vector [52] as a transfection marker. Transfected cells were sorted after 40 h into two 96-well plates, based on the top 5 percentile of the EGFP expression. Single cell sorting was performed in the Flow Cytometry Core Facility of the Institute for Molecular Biology gGmbH (IMB Mainz) under supervision of scientific staff. After two weeks, growing clones were transferred to 25 cm^2 flasks and screened by PCR for the presence of a deletion spanning the sequence between the sgRNA-targeted sites in the exons 2 and 5 (including the catalytic R140 codon). The primers were 5′-TCCTCTGTAATCCACTCTAA (forward) and 5′-AGCTCAGCTTGAACTAGATA (reverse). Preselected clones were next screened for the presence of non-rearranged *TDG* alleles to eliminate the positives. The primer pair detecting the non-rearranged exon 2 was 5′-TCCTCTGTAATCCACTCTAA (forward) and 5′-ATGTCCCTACTCTGATCTTT (reverse). The remaining clones were expanded and the TDG knockout was validated by Western blotting of protein extracts with a 1:5000 dilution of the TDG rabbit polyclonal antibody #PA5-29140 (Thermo Fisher Scientific Inc.) and a 1:10,000 dilution of the IRDye® 800CW donkey anti-rabbit IgG #926-32213 (LI-COR Biosciences GmbH, Bad Homburg, Germany). Blots were analyzed by using the Odyssey 9120 infrared imaging system (LI-COR). Stripped membranes were subsequently probed with a 1:10,000 dilution of the mouse monoclonal antibody AC88 to HSP90 # ADI-SPA-830 (Enzo Life Sciences GmbH, Lörrach, Germany) and a 1:10,000 dilution of the IRDye® 800CW donkey anti-mouse secondary antibody # 926-68072 (LI-COR).

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/ijms222011025/s1.

Author Contributions: Conceptualization, A.K. and T.C.; methodology, A.K. and T.C.; validation, N.M., E.P. and A.K.; formal analysis, N.M. and A.K.; investigation, N.M.; resources, T.C. and E.P.; data curation, N.M.; writing—original draft preparation, A.K. and N.M.; writing—review and editing, A.K., N.M., T.C. and E.P.; visualization, N.M. and A.K.; supervision, A.K.; project administration, A.K.; funding acquisition, A.K. and T.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Deutsche Forschungs gemeinschaft (DFG, German Research Foundation): project-ID 393547839—SFB 1361, project-ID 321812289—GRK 2338, grant numbers KH263/2-2 and KH263/5-1. This project received funding from the European Union's Horizon 2020 Research and Innovation Programme, under the Marie Sklodowska-Curie grant agreement No. 765266.

Acknowledgments: We thank the following persons and organizations for help with experiments and reagents: IMB Flow Cytometry Core Facility for help with cell sorting; Oliver H. Krämer and Mandy Beyer at the Institute of Toxicology (UMC Mainz) for advice and shared Western blotting reagents; Martin Rossa for preparing the first batch of synthetic DNA strands; Marta Rodriguez-Alvarez for help with handling the TDG knockout clones and Nataliya Kitsera for constructive comments on the manuscript draft.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bernstein, B.E.; Meissner, A.; Lander, E.S. The mammalian epigenome. *Cell* **2007**, *128*, 669–681. [CrossRef] [PubMed]
2. Berger, S.L.; Kouzarides, T.; Shiekhattar, R.; Shilatifard, A. An operational definition of epigenetics. *Genes Dev.* **2009**, *23*, 781–783. [CrossRef] [PubMed]
3. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **2002**, *16*, 6–21. [CrossRef] [PubMed]
4. Tate, P.H.; Bird, A.P. Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr. Opin. Genet. Dev.* **1993**, *3*, 226–231. [CrossRef]
5. Boyes, J.; Bird, A. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* **1991**, *64*, 1123–1134. [CrossRef]
6. Lewis, J.D.; Meehan, R.R.; Henzel, W.J.; Maurer-Fogy, I.; Jeppesen, P.; Klein, F.; Bird, A. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* **1992**, *69*, 905–914. [CrossRef]
7. Jones, P.L.; Veenstra, G.J.; Wade, P.A.; Vermaak, D.; Kass, S.U.; Landsberger, N.; Strouboulis, J.; Wolffe, A.P. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* **1998**, *19*, 187–191. [CrossRef]
8. Nan, X.; Ng, H.H.; Johnson, C.A.; Laherty, C.D.; Turner, B.M.; Eisenman, R.N.; Bird, A. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **1998**, *393*, 386–389. [CrossRef]
9. Fuks, F.; Burgers, W.A.; Brehm, A.; Hughes-Davies, L.; Kouzarides, T. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nat. Genet.* **2000**, *24*, 88–91. [CrossRef]
10. Iguchi-Ariga, S.M.; Schaffner, W. CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes Dev.* **1989**, *3*, 612–619. [CrossRef]
11. Spruijt, C.G.; Gnerlich, F.; Smits, A.H.; Pfaffeneder, T.; Jansen, P.W.; Bauer, C.; Munzel, M.; Wagner, M.; Muller, M.; Khan, F.; et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **2013**, *152*, 1146–1159. [CrossRef]
12. Yin, Y.; Morgunova, E.; Jolma, A.; Kaasinen, E.; Sahu, B.; Khund-Sayeed, S.; Das, P.K.; Kivioja, T.; Dave, K.; Zhong, F.; et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **2017**, *356*. [CrossRef] [PubMed]
13. Tahiliani, M.; Koh, K.P.; Shen, Y.; Pastor, W.A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L.M.; Liu, D.R.; Aravind, L.; et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **2009**, *324*, 930–935. [CrossRef]
14. Ito, S.; Shen, L.; Dai, Q.; Wu, S.C.; Collins, L.B.; Swenberg, J.A.; He, C.; Zhang, Y. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **2011**, *333*, 1300–1303. [CrossRef]
15. Pfaffeneder, T.; Hackner, B.; Truss, M.; Munzel, M.; Muller, M.; Deiml, C.A.; Hagemeier, C.; Carell, T. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 7008–7012. [CrossRef] [PubMed]
16. He, Y.F.; Li, B.Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **2011**, *333*, 1303–1307. [CrossRef] [PubMed]
17. Maiti, A.; Drohat, A.C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: Potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **2011**, *286*, 35334–35338. [CrossRef] [PubMed]
18. Zhang, L.; Lu, X.; Lu, J.; Liang, H.; Dai, Q.; Xu, G.L.; Luo, C.; Jiang, H.; He, C. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat. Chem. Biol.* **2012**, *8*, 328–330. [CrossRef] [PubMed]
19. Neri, F.; Incarnato, D.; Krepelova, A.; Rapelli, S.; Anselmi, F.; Parlato, C.; Medana, C.; Dal Bello, F.; Oliviero, S. Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. *Cell Rep.* **2015**, *10*, 674–683. [CrossRef] [PubMed]
20. Iurlaro, M.; McInroy, G.R.; Burgess, H.E.; Dean, W.; Raiber, E.A.; Bachman, M.; Beraldi, D.; Balasubramanian, S.; Reik, W. In vivo genome-wide profiling reveals a tissue-specific role for 5-formylcytosine. *Genome Biol.* **2016**, *17*, 141. [CrossRef] [PubMed]
21. Licyte, J.; Gibas, P.; Skardziute, K.; Stankevicius, V.; Ruksenaite, A.; Kriukiene, E. A Bisulfite-free Approach for Base-Resolution Analysis of Genomic 5-Carboxylcytosine. *Cell Rep.* **2020**, *32*, 108155. [CrossRef]
22. Cadet, J.; Wagner, J.R. TET enzymatic oxidation of 5-methylcytosine, 5-hydroxymethylcytosine and 5-formylcytosine. *Mutat. Res. Genet. Toxicol. Environ. Mutagen* **2014**, *764–765*, 18–35. [CrossRef] [PubMed]

23. Carell, T.; Kurz, M.Q.; Muller, M.; Rossa, M.; Spada, F. Non-canonical Bases in the Genome: The Regulatory Information Layer in DNA. *Angew. Chem.-Int. Ed.* **2018**, *57*, 4296–4312. [CrossRef] [PubMed]
24. Branco, M.R.; Ficz, G.; Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* **2012**, *13*, 7–13. [CrossRef] [PubMed]
25. Kitsera, N.; Allgayer, J.; Parsa, E.; Geier, N.; Rossa, M.; Carell, T.; Khobta, A. Functional impacts of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine at a single hemi-modified CpG dinucleotide in a gene promoter. *Nucleic Acids Res.* **2017**, *45*, 11033–11042. [CrossRef]
26. Kitsera, N.; Rodriguez-Alvarez, M.; Emmert, S.; Carell, T.; Khobta, A. Nucleotide excision repair of abasic DNA lesions. *Nucleic Acids Res.* **2019**, *47*, 8537–8547. [CrossRef]
27. Luhnsdorf, B.; Epe, B.; Khobta, A. Excision of uracil from transcribed DNA negatively affects gene expression. *J. Biol. Chem.* **2014**, *289*, 22008–22018. [CrossRef]
28. Allgayer, J.; Kitsera, N.; Bartelt, S.; Epe, B.; Khobta, A. Widespread transcriptional gene inactivation initiated by a repair intermediate of 8-oxoguanine. *Nucleic Acids Res.* **2016**, *44*, 7267–7280. [CrossRef]
29. Kribelbauer, J.F.; Lu, X.J.; Rohs, R.; Mann, R.S.; Bussemaker, H.J. Toward a Mechanistic Understanding of DNA Methylation Readout by Transcription Factors. *J. Mol. Biol.* **2020**, *432*, 1801–1815. [CrossRef]
30. Muller, N.; Khobta, A. Regulation of GC box activity by 8-oxoguanine. *Redox Biol.* **2021**, *43*, 101997. [CrossRef]
31. Schroder, A.S.; Kotljarova, O.; Parsa, E.; Iwan, K.; Raddaoui, N.; Carell, T. Synthesis of (*R*)-Configured 2′-Fluorinated mC, hmC, fC, and caC Phosphoramidites and Oligonucleotides. *Org. Lett.* **2016**, *18*, 4368–4371. [CrossRef] [PubMed]
32. Fleming, A.M.; Ding, Y.; Burrows, C.J. Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 2604–2609. [CrossRef] [PubMed]
33. Wilson, D.M., 3rd; Takeshita, M.; Grollman, A.P.; Demple, B. Incision activity of human apurinic endonuclease (Ape) at abasic site analogs in DNA. *J. Biol. Chem.* **1995**, *270*, 16002–16007. [CrossRef] [PubMed]
34. Valinluck, V.; Tsai, H.H.; Rogstad, D.K.; Burdzy, A.; Bird, A.; Sowers, L.C. Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Res.* **2004**, *32*, 4100–4108. [CrossRef] [PubMed]
35. Valinluck, V.; Sowers, L.C. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res.* **2007**, *67*, 946–950. [CrossRef]
36. Hashimoto, H.; Liu, Y.; Upadhyay, A.K.; Chang, Y.; Howerton, S.B.; Vertino, P.M.; Zhang, X.; Cheng, X. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* **2012**, *40*, 4841–4849. [CrossRef]
37. Giehr, P.; Kyriakopoulos, C.; Ficz, G.; Wolf, V.; Walter, J. The Influence of Hydroxylation on Maintaining CpG Methylation Patterns: A Hidden Markov Model Approach. *PLoS Comput. Biol.* **2016**, *12*, e1004905. [CrossRef]
38. Ji, D.; Lin, K.; Song, J.; Wang, Y. Effects of Tet-induced oxidation products of 5-methylcytosine on Dnmt1- and DNMT3a-mediated cytosine methylation. *Mol. Biosyst.* **2014**, *10*, 1749–1752. [CrossRef]
39. Harrington, M.A.; Jones, P.A.; Imagawa, M.; Karin, M. Cytosine methylation does not affect binding of transcription factor Sp1. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 2066–2070. [CrossRef]
40. Holler, M.; Westin, G.; Jiricny, J.; Schaffner, W. Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. *Genes Dev.* **1988**, *2*, 1127–1135. [CrossRef]
41. Khobta, A.; Anderhub, S.; Kitsera, N.; Epe, B. Gene silencing induced by oxidative DNA base damage: Association with local decrease of histone H4 acetylation in the promoter region. *Nucleic Acids Res.* **2010**, *38*, 4285–4295. [CrossRef]
42. Broxson, C.; Hayner, J.N.; Beckett, J.; Bloom, L.B.; Tornaletti, S. Human AP endonuclease inefficiently removes abasic sites within G4 structures compared to duplex DNA. *Nucleic Acids Res.* **2014**, *42*, 7708–7719. [CrossRef]
43. Burra, S.; Marasco, D.; Malfatti, M.C.; Antoniali, G.; Virgilio, A.; Esposito, V.; Demple, B.; Galeone, A.; Tell, G. Human AP-endonuclease (Ape1) activity on telomeric G4 structures is modulated by acetylatable lysine residues in the N-terminal sequence. *DNA Repair* **2019**, *73*, 129–143. [CrossRef]
44. Fleming, A.M.; Zhu, J.; Howpay Manage, S.A.; Burrows, C.J. Human NEIL3 Gene Expression Regulated by Epigenetic-Like Oxidative DNA Modification. *J. Am. Chem. Soc.* **2019**, *141*, 11036–11049. [CrossRef]
45. Davletgildeeva, A.T.; Kuznetsova, A.A.; Fedorova, O.S.; Kuznetsov, N.A. Activity of Human Apurinic/Apyrimidinic Endonuclease APE1 toward Damaged DNA and Native RNA with Non-canonical Structures. *Front. Cell Dev. Biol.* **2020**, *8*, 590848. [CrossRef]
46. Mohni, K.N.; Wessel, S.R.; Zhao, R.; Wojciechowski, A.C.; Luzwick, J.W.; Layden, H.; Eichman, B.F.; Thompson, P.S.; Mehta, K.P.M.; Cortez, D. HMCES Maintains Genome Integrity by Shielding Abasic Sites in Single-Strand DNA. *Cell* **2019**, *176*, 144–153.e13. [CrossRef]
47. Thompson, P.S.; Amidon, K.M.; Mohni, K.N.; Cortez, D.; Eichman, B.F. Protection of abasic sites during DNA replication by a stable thiazolidine protein-DNA cross-link. *Nat. Struct. Mol. Biol.* **2019**, *26*, 613–618. [CrossRef]
48. Schroder, A.S.; Steinbacher, J.; Steigenberger, B.; Gnerlich, F.A.; Schiesser, S.; Pfaffeneder, T.; Carell, T. Synthesis of a DNA promoter segment containing all four epigenetic nucleosides: 5-methyl-, 5-hydroxymethyl-, 5-formyl-, and 5-carboxy-2′-deoxycytidine. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 315–318. [CrossRef] [PubMed]
49. Luhnsdorf, B.; Kitsera, N.; Warken, D.; Lingg, T.; Epe, B.; Khobta, A. Generation of reporter plasmids containing defined base modifications in the DNA strand of choice. *Anal. Biochem.* **2012**, *425*, 47–53. [CrossRef] [PubMed]

50. Bauer, D.E.; Canver, M.C.; Orkin, S.H. Generation of genomic deletions in mammalian cell lines via CRISPR/Cas9. *J. Vis. Exp.* **2015**, *95*, e52118. [CrossRef] [PubMed]
51. Labun, K.; Montague, T.G.; Gagnon, J.A.; Thyme, S.B.; Valen, E. CHOPCHOP v2: A web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res.* **2016**, *44*, W272–W276. [CrossRef] [PubMed]
52. Allgayer, J.; Kitsera, N.; von der Lippen, C.; Epe, B.; Khobta, A. Modulation of base excision repair of 8-oxoguanine by the nucleotide sequence. *Nucleic Acids Res.* **2013**, *41*, 8559–8571. [CrossRef] [PubMed]

4.3. Studies on ultrafast excited states of epigenetic deoxycytidine derivatives

"Unified Description of Ultrafast Excited State Decay Processes in Epigenetic Deoxycytidine Derivatives." Piotr Kabaciński#, Marco Romanelli#, Eveliina Ponkkonen, Vishal Kumar Jaiswal, Thomas Carell, Marco Garavelli*, Giulio Cerullo*, and Irene Conti*, *Phys. Chem. Lett.* **2021**, *12*, 11070−11077

#: These authors contributed equally

*: Corresponding author

Prologue

Epigenetic dC modifications play an important role in modulating gene expression and regulating cellular processes. To study the role of these modifications as a potential DNA photodamage, combination of ultrafast TA spectroscopy with sub-30-fs transient absorption spectroscopy and high-level correlated multiconfigurational CASPT2/MM computations, was used to obtain a unified picture of the photophysical decay mechanisms of mdC, hmdC, fdC and cadC. Thereby, mdC and hmdC showed slower ultrafast excited state decay times (~4ps) compared to dC (in the subpicosecond range) due to an increased energy barrier. fdC, on the other hand, was shown to have two almost isoenergetic low-lying excited energetic states and cadC was observed to display the shortest excited state lifetime among the modifications with a decay of only 130 fs back to the electronic ground state.

Contribution to this study

For this work I synthesized epigenetically relevant nucleosides hmdC, fdC and cadC in big quantities and purified the compounds via HPLC for further studies in the group of Prof. Giulio Cerullo.

License

THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCL

Letter

Unified Description of Ultrafast Excited State Decay Processes in Epigenetic Deoxycytidine Derivatives

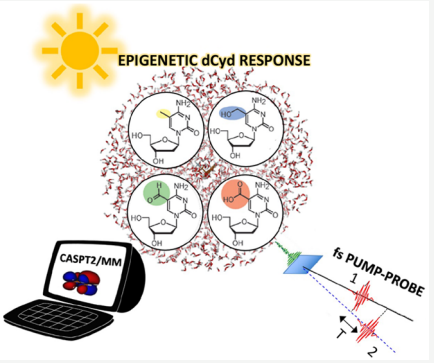
Piotr Kabaciński,[‖] Marco Romanelli,[‖] Eveliina Ponkkonen, Vishal Kumar Jaiswal, Thomas Carell, Marco Garavelli,* Giulio Cerullo,* and Irene Conti*

Cite This: *J. Phys. Chem. Lett.* 2021, 12, 11070−11077

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Epigenetic DNA modifications play a fundamental role in modulating gene expression and regulating cellular and developmental biological processes, thereby forming a second layer of information in DNA. The epigenetic 2′-deoxycytidine modification 5-methyl-2′-deoxycytidine, together with its enzymatic oxidation products (5-hydroxymethyl-2′-deoxycytidine, 5-formyl-2′-deoxycytidine, and 5-carboxyl-2′-deoxycytidine), are closely related to deactivation and reactivation of DNA transcription. Here, we combine sub-30-fs transient absorption spectroscopy with high-level correlated multiconfigurational CASPT2/MM computational methods, explicitly including the solvent, to obtain a unified picture of the photophysics of deoxycytidine-derived epigenetic DNA nucleosides. We assign all the observed time constants and identify the excited state relaxation pathways, including the competition of intersystem crossing and internal conversion for 5-formyl-2′-deoxycytidine and ballistic decay to the ground state for 5-carboxy-2′-deoxycytidine. Our work contributes to shed light on the role of epigenetic derivatives in DNA photodamage as well as on their possible therapeutic use.



Epigenetics, which is the study of heritable phenotype modifications that do not involve alterations in the genotype, is becoming a more and more important field of research, aiming to explain how living organisms adapt to external stimuli.[1] Methylation of 2′-deoxycytidine (dC) at the C_5 position of the nucleobase can lead to transcriptional silencing of the corresponding gene in certain genomic regions.[2] 5-Methyl-2′-deoxycytidine (mdC) is a prevailing epigenetic modification that plays important roles in modulating gene expression and developmental processes, and its dysregulation may cause severe diseases, including cancer.[1,3−6] Demethylation of mdC back to dC reactivates the transcription of these genes; however, the process behind this demethylation remains not yet fully understood. A decade ago, 5-hydroxymethyl-2′-deoxycytidine (hmdC),[7] 5-formyl-2′-deoxycytidine (fdC),[8] and 5-carboxyl-2′-deoxycytidine (cadC)[9] were detected as additional epigenetic elements in DNA. Furthermore, it was shown that these modified dC bases are formed from mdC via consecutive oxidation reactions catalyzed by 10−11 translocation enzymes.[9,10] These oxidized mdC derivatives are considered to form a second layer of information and to be a part of an active DNA demethylation process that potentially regulates the concentration and pattern of epigenetic markers in mammalian cells.[3−6,10]

Epigenetic dC derivatives might affect the efficient and ultrafast nonradiative excited state (ES) deactivation channels of the canonical nucleosides, which safely dissipate the absorbed light energy, possibly leading to more complex scenarios of DNA photoprotection and photodamage.[11−15] According to quantum mechanics/molecular mechanics (QM/MM)[16] calculations at the CASPT2[17,18] level, the classical cytidine dC and the most common epigenetic methylated form (mdC) show different energy barriers of ∼0.18 and ∼0.27 eV, respectively, along the same decay pathway, driving the lowest $\pi\pi^*$ ES to the "ethylene-like" conical intersection (CI)[19−22] with the ground state (S_0). This difference justifies the significantly longer lifetime of the epigenetic derivative (6.8 ps) with respect to the parent compound (1.1 ps),[23,24] observed with femtosecond transient absorption (TA) spectroscopy, which makes mdC more prone to photodamage events. In addition, dark $n\pi^*$ ESs, which are thought to play a role in the long-living component of the observed TA signal for water-solvated dC, are predicted to be destabilized in mdC and thus not to be involved in the relaxation of the lowest $\pi\pi^*$ state.[23,24] Recent experimental[25] studies, supported by CASSCF or TDDFT computations,[26,27] showed that while the photophysics of hmdC substantially resembles that of mdC, the ES relaxation pathways of fdC and cadC are remarkably different. In fdC, there is experimental evidence of an efficient ultrafast intersystem crossing (ISC) that leads to the population of the

Received: September 3, 2021
Accepted: November 1, 2021
Published: November 8, 2021

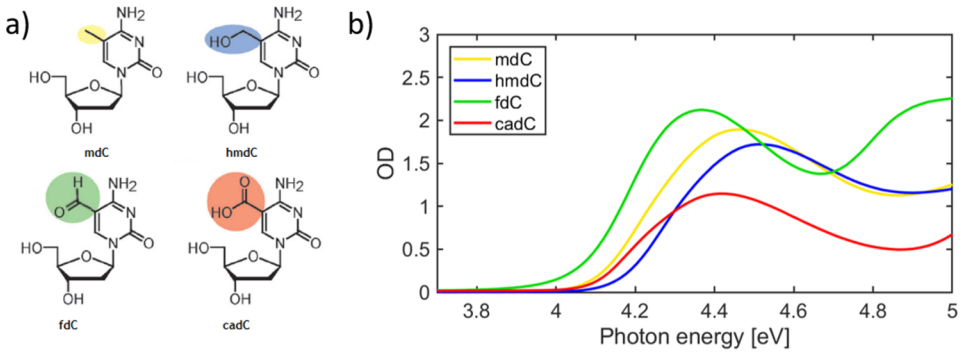


Figure 1. (a) Epigenetic derivatives of 2′-deoxycytidine studied and (b) their linear absorption spectra. The epigenetic modifications involve the C_5 position of cytidine. The cadC nucleoside is in the anionic form in our simulations, because it is the stable species at neutral pH.

lowest triplet state, accounting for a long-living component of the TA signal, whereas water-solvated cadC shows a subpicosecond (840 fs) ES decay.[25] The photophysical processes underlying these very different excited state dynamics aroused increasing interest, becoming a current matter of debate and giving rise to contradictory hypotheses on the decay mechanisms at play.[25−27]

Here, we aim to provide a unified and coherent description of the complex ES decay pathways of the epigenetic cytidines through a comprehensive experimental and theoretical investigation. On the experimental side, we perform ultrafast TA spectroscopy with state-of-the-art sub-30-fs temporal resolution to follow the rapid evolution of the photoexcited wave packet on the excited state potential energy surface (PES) and broad spectral coverage in the 1.9−3.9 eV range to identify all the photoinduced signals, including the previously unexplored UV region.[24,25] Thanks to the high sensitivity of our TA apparatus, experiments are performed at low fluences, at which no formation of solvated electrons is observed, permitting the correct assignment of the TA signals. On the computational side, we employ a hybrid SS-CASPT2/MM scheme accounting for multireference dynamically correlated energies and gradients on all the epigenetic cytidines simultaneously, including the sugar moiety, and considering explicitly the water solvent along with hydrogen bonds (instead of an implicit continuum as in the polarizable continuum model[26−28]), which is necessary for a realistic description of the ES dynamics and of the spectroscopic signals, as previously demonstrated.[11,23,29,30] We systematically map the major decay pathways (singlet and triplet, $\pi\pi^*$ and $n\pi^*$ states) based on the minimum energy paths involving all the characterized critical points and CIs driving the different photoprocesses. Eventually, this study allows new light to be shed on previously detected deactivation channels and reveals new ones falling in the so far uncharted sub-500-fs regime.

Figure 1a shows the chemical structures of the four epigenetic nucleosides, obtained from dC by substitution at the C_5 position of the nucleobase. The corresponding absorption spectra, shown in Figure 1b, are dominated by an intense band spanning 4.1−4.6 eV due to the $\pi\pi^*$ transition of the aromatic ring, similar to the canonical nucleosides.

Figure 2a plots the differential absorption (ΔA) spectrum, as a function of pump−probe delay (up to 1 ps) and probe photon energy, for mdC following photoexcitation by a sub-20-fs pulse at 4.35 eV, which populates the lowest $\pi\pi_1^*$ bright ES. At early times, we observe a negative band (Figure 2c, blue line), peaking at 3.68 eV, assigned to stimulated emission (SE) from the bright $\pi\pi_1^*$ state, together with a positive photoinduced absorption (PA) band. Both the SE and PA bands undergo a rapid partial decay on the ∼100 fs time scale, showing a subsequent SE red-shift to 3.54 eV, as illustrated in Figure 2c (purple line), which displays the evolution associated spectra (EAS) obtained by global analysis of the TA data. Subsequently, the spectrum decays further in 1 ps while still shifting to the red, followed by a longer 4.3 ps decay into a low-intensity spectrum lacking any SE signals.

Experimental TA data are consistent with the ES deactivation scenario derived from QM/MM calculations, which is summarized in Figure 2d. The calculations reveal an ultrafast relaxation from the Franck−Condon (FC) region of the $\pi\pi_1^*$ (S_1) state toward a flat region of the ES PES, characterized by low forces acting on the system (Plateau-$\pi\pi_1^*$, Figure 2d), where the computed SE band at 3.6 eV with an oscillator strength (OS) of 0.11 (SE_1 in Figure 2d) matches the experimental SE band observed immediately after excitation (empty blue circle in Figure 2c).

Calculations also reproduce the peaks of the PA spectrum (full blue circles in Figure 2c) observed in the visible at early times (PA_1), when the system came out of the Franck−Condon region and reached the Plateau-$\pi\pi_1^*$ region (vertical blue arrows in Figure 2d). The observed decrease in intensity and red-shift of the SE band seen in the ∼100 fs time scale (change from blue to purple EAS in Figure 2c) is assigned to molecules moving and oscillating across the flat Plateau-$\pi\pi_1^*$ region (see oscillating purple line in Figure 2d) until fully relaxing to the minimum (Min-$\pi\pi_1^*$) within ∼1 ps. Critical point energies are reported in Figure S6 in the Supporting Information (SI). Here, the system resides for a longer time, due to the barrier that needs to be overcome to decay to the GS (4.3 ps time constant, pink line in Figure 2d). Note that the colors of the lines and arrows in Figure 2d, indicating the different decay processes, match those of the EAS curves of Figure 2c. The computed SE (SE_2 in Figure 2d) from the Min-$\pi\pi_1^*$ structure is at 3.2 eV (OS 0.09) and is red-shifted compared to SE_1. The overlap of the SE_2 with the positive PA_2 band, predicted from Min-$\pi\pi_1^*$ at 3.0 eV (OS 0.06, pink arrow in Figure 2d), could account for the observed reduction of the intensity and the blue-shift of the experimental SE peak with respect to the computed value (SE_2, pink empty circle in Figure 2c). Moreover, starting from the ES minimum (Min-$\pi\pi_1^*$), we also calculated PA_2 signals at 3.83 eV (OS 0.03) and 4.22 eV (OS 0.08) that match with the experimental PA band observed in that spectral region (additional pink full circles in Figure 2c). The 4.3 ps decay of the SE_2 signal coming from

https://doi.org/10.1021/acs.jpclett.1c02909
J. Phys. Chem. Lett. 2021, 12, 11070−11077

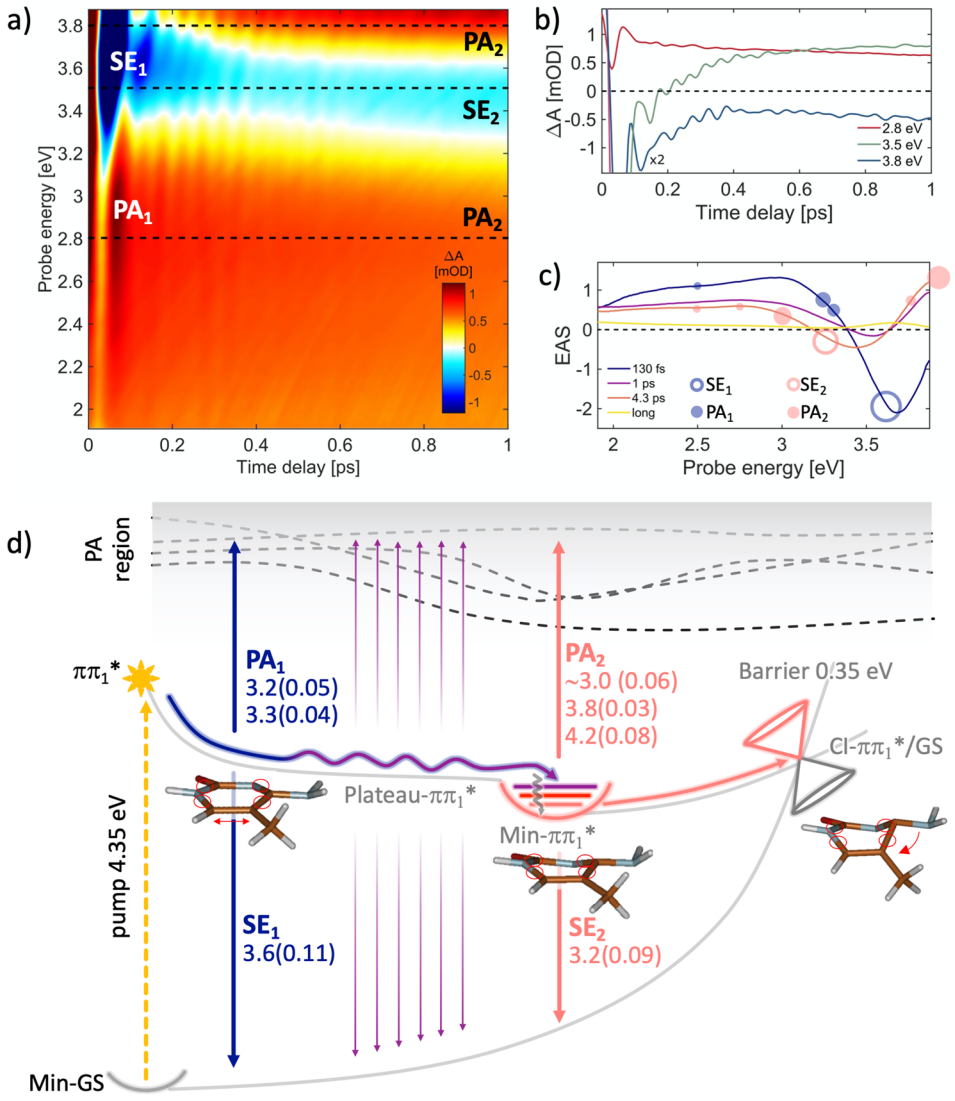


Figure 2. (a) ΔA map of mdC in water solution recorded with pulse polarizations at the magic angle. (b) Dynamics at selected probe photon energies marked with dashed lines in panel a. (c) EAS with the corresponding time constants: 130 fs (blue curve) is the time needed to relax from the FC to the Plateau-$\pi\pi_1$* flat region (see calculated paths in panel d), and 1 ps (purple curve) is the time needed to fully relax from the Plateau-$\pi\pi_1$* region to the Min-$\pi\pi_1$* (wavy purple line, panel d). The 4.3 ps time constant (pink curve in panel c) relates to the $\pi\pi_1$* → GS decay process (CI-$\pi\pi_1$*/GS, panel d), involving a 0.35 eV barrier. The yellow line corresponds to long-lived products probably due to other minor decay paths. Empty and full circles correspond to the calculated SE and PA energy values, respectively, and the circle dimensions are proportional to the computed oscillator strength values (documented in panel d). (d) Schematic decay paths of mdC, calculated at CASPT2/MM level (details in the SI section). Relaxation routes and SE/PA colors arrows are matching with the line colors of time constants in panel c. Oscillator strengths are reported in brackets. Critical point energies are in Figure S6. Molecular optimized structures refer to the QM region only.

Min-$\pi\pi_1$* together with the corresponding PA_2 (pink line, Figure 2c) is consistent with the fact that the lowest-lying CI between $\pi\pi_1$* and S_0 features an energy barrier from the Min-$\pi\pi_1$* of ∼0.35 eV. Following the internal conversion (IC) to S_0, which is the dominant decay pathway, a residual weak PA spectrum remains (yellow line) lasting longer than the probed time window (30 ps) because of other possible minor decay routes.

The calculations associate the first PA_1 and SE_1 signals (corresponding to the so far not observed shortest decay time constant on the order of 100 fs) to still planar structures just relaxed out of the FC region, beginning to distort along the ring-puckering coordinate in the flat Plateau-$\pi\pi_1$* region, leading to Min-$\pi\pi_1$* within the second time constant of ∼1 ps. The corresponding structural changes are illustrated in Figure 2d. It is worth noting that although our optimized $\pi\pi_1$*/GS CI does not exactly reproduce the structure of the "ethene-like" CI reported previously by Martinéz-Fernandéz et al.,[23] they both show comparable access energy barriers (0.35 and 0.3 eV, respectively), but the CI documented in Figure 2d should be more easily accessible, as it lies exactly along the reaction coordinate that coherently connects the planar structure to the crossing, passing through the Min-$\pi\pi_1$*. Structural details about the computed $\pi\pi_1$*/GS CI for mdC are reported in the SI section, including the Cartesian coordinates.

We also investigated the photophysics of mdC when the higher-energy bright state S_2 ($\pi\pi_2$*) is populated: a sudden decay to S_1($\pi\pi_1$*) is predicted owing to a crossing with the S_1 state nearby to the S_2 FC region, thus showing that the $\pi\pi_1$* state collects also the $\pi\pi_2$* population. In addition, dark states

https://doi.org/10.1021/acs.jpclett.1c02909
J. Phys. Chem. Lett. 2021, 12, 11070−11077

($n\pi^*$) are destabilized compared to bright states in water solution, and therefore, they are not involved in the ES decay pathway when pumping at 4.35 eV (see calculated energies for the corresponding vertical and critical points shown in Figure S6).

HmdC exhibits photophysics very similar to that of mdC upon UV photoexcitation at 4.35 eV. Both ultrafast TA spectra and the calculated decay pathways strongly resemble those of mdC (see Figure S7). Following the photoexcitation into S_1 ($\pi\pi_1^*$), the population initially decays toward the Plateau-$\pi\pi_1^*$ region with a 160 fs time constant, and moving along the plateau region, it reaches the lowest minimum (Min-$\pi\pi_1^*$) with a 735 fs time constant, again showing a red-shift of the SE spectrum. This behavior, already observed for the methylated compound, can be rationalized in the same way, including the previously undetected fast decay with a 160 fs time constant and similarly assigned to the initial planar relaxation, before population of the ring-puckering mode. A comparable energy barrier (~0.30 eV) has to be passed to reach the crossing point of $\pi\pi_1^*$ with the GS, which presents very similar molecular distortions (see molecular structures in Figure S7) to those found for mdC, through which the molecule decays in 4.6 ps. The similarity of the photophysics is confirmed by the resemblance of the TA spectra and time constants of mdC and hmdC (Figures 2 and S7, respectively). The difference between the two TA maps is mostly due to the higher intensity of the PA bands relative to the SE for hmdC. Once again, computations reveal that photoexcitation of the second bright $\pi\pi_2^*$ state (S_2) immediately leads to a crossing with $\pi\pi_1^*$ (S_1), and no dark states (including all the low-lying $n\pi^*$ states) seem to be involved in the ES decay pathway upon pumping at 4.35 eV (see the computed critical points in Figure S8).

FdC shows a very different decay scenario, compared to the other epigenetic cytidine derivatives (Figure 3). This is due to the presence of two almost isoenergetic low-lying ESs in the FC region that were not predicted before,[26,27] namely S_1 ($\pi\pi_1^*$) and S_2 ($n\pi^*$) (Figure 3d, yellow star and circle, respectively). This immediately leads to branching of the ES population. Only a CASPT2-correlated method combined with explicit solvent interactions predicts the $\pi\pi^*/n\pi^*$ degeneracy in the FC vertical region (see also Figure S4): upon pumping at 4.35 eV, the bright $\pi\pi^*$ S_1 state (Figure 3d, right part) is mainly populated, and its simulated spectral signatures (blue arrows in Figure 3d) match well with the experimentally observed SE and PA_1 signals shown as blue circles in Figure 3c. Simultaneously, the $n\pi^*$ state is partially populated, contributing with its positive PA signals in the 3.3/3.9 eV regions (see Min-$n\pi^*$ PA_2 signals in Figure 3d). For the $\pi\pi_1^*$ state, an ultrafast relaxation pathway leads to the Plateau-$\pi\pi_1^*$ (right side of Figure 3d), a planar region of the $\pi\pi_1^*$ state that spans a progressively decreasing S_1-S_0 energy gap (from ~3.30 to ~1.0 eV) due to the corresponding increase in the GS energy. Here, the structure undergoes large distortions along the "ethene-like" coordinate: starting from a quite planar geometry, a large torsion around the $C_2N_1-C_6C_5$ angle (until ~54°) and a C_5 formyl out of plane bending (until ~100°) take place (see molecular structures on top of the $\pi\pi_1^*$ decay path, Figure 3d). By continuing the optimization along S_1, one finds a low-lying CI with the GS (CI-$\pi\pi_1^*$/GS) that presents no access energy barrier from the plateau region, thus suggesting an ultrafast decay pathway (critical point energies in Figure S9). Indeed, the SE signal observed just after the excitation (blue line in Figure 3c) disappears with a 130 fs time constant: the purple line (Figure 3c) possibly represents a later stage of the evolution on the Plateau-$\pi\pi_1^*$ region where the $\pi\pi^*$-GS energy gap is reduced until the wavepacket decays to the GS in 345 fs, supporting our proposed mechanism (purple $\pi\pi_1^*$ wavy and GS arrows in Figure 3d).

While two different fdC conformers may exist in water,[27] only the one lacking an intramolecular hydrogen bond between the amino and the formyl groups was discussed in this study (*anti* isomer) while neglecting the conformer where the formyl carbonyl and the amino group are bridged through an intramolecular N−H···O bond (*syn*), which possibly could induce molecular restraints. The *anti* choice was taken, because this conformer is the one that, we believe, is more relevant for the ultrafast sub-400-fs photoinduced dynamics observed in this study (see the SI, section 4.1, for a detailed discussion). Moreover, very recent time-resolved IR experiments and TDDFT calculations[27] show that out of plane motions (described for the *anti* conformer) are indeed populated regardless of the fdC conformer.

On the other hand, the aforementioned S_2 dark state ($n\pi^*$) is almost isoenergetic with the $\pi\pi_1^*$ (S_1) at the FC point: surface crossing between these two states leads to population of the dark $n\pi^*$ singlet state (S_2) at early times through IC, already within 130 fs (blue line). In addition, our vertical calculations do not consider the vibrational degrees of freedom of the molecule, thus neglecting the plausible contribution to the S_2 OS coming from distorted molecular geometries. The simultaneous population of the $\pi\pi_1^*$ and $n\pi^*$ states could be supported by the weaker fdC steady state fluorescence spectrum[27] as compared with mdC,[22] which instead populates just the bright $\pi\pi_1^*$ state, because the $n\pi^*$ state lies at higher energies (see Figure S6). Following the dark $n\pi^*$ state, the optimization of S_2 leads to a minimum (Min-$n\pi^*$, Figure 3d), where the computed PA_2 values (3.3 and 3.9 eV in Figure 3d and dark blue circles in Figure 3c) contribute to the first three time constants (blue, purple, and pink lines, Figure 3c), because its decay via ISC processes could require picosecond time scales, as also supported by recent time-resolved mid-IR spectroscopy experiments.[27] The experimental spectrum also contains a contribution from the hot GS PA, following ultrafast decay through CI-$\pi\pi_1^*$/GS. Hot GS relaxation is a process typically falling in the picosecond time range (pink line, dashed circle in Figure 3c and pink dashed arrow in Figure 3d).

A crucial characteristic of the $n\pi^*$ relaxation path is that at the minimum geometry (Min-$n\pi^*$) the lowest $^3\pi\pi^*$ triplet excited state (gray triangle in Figure S9) is close in energy to the singlet $n\pi^*$, allowing an ISC process that results in an efficient population of the T_1 triplet excited state minimum (Min-$^3\pi\pi^*$, yellow triangle in Figure 3d). The decay of the $n\pi^*$ spectrum in 2.1 ps (pink line in Figure 3c) into the remaining long-lived spectrum (yellow line Figure 3c) is attributed to the population of this lowest triplet state, which survives for times much longer than our probing window. In support of this mechanism, the PA_3 values computed on top of the triplet minimum (yellow arrow, 2.8 and 4.0 eV) show good agreement with the experimental peak around 2.84 eV (growing in the pink and clearly recognizable in the long-living yellow line, Figure 3c) as well as with the more intense UV-shifted signal at 4.0 eV (better recognizable in the DUV probe spectrum in Figure S11), which exhibits a higher OS compared to the previous transition (0.22 vs 0.04), thus justifying the strong absorption tail on the blue edge of the spectrum. The high triplet quantum yield[25,31] could also be

https://doi.org/10.1021/acs.jpclett.1c02909
J. Phys. Chem. Lett. 2021, 12, 11070−11077

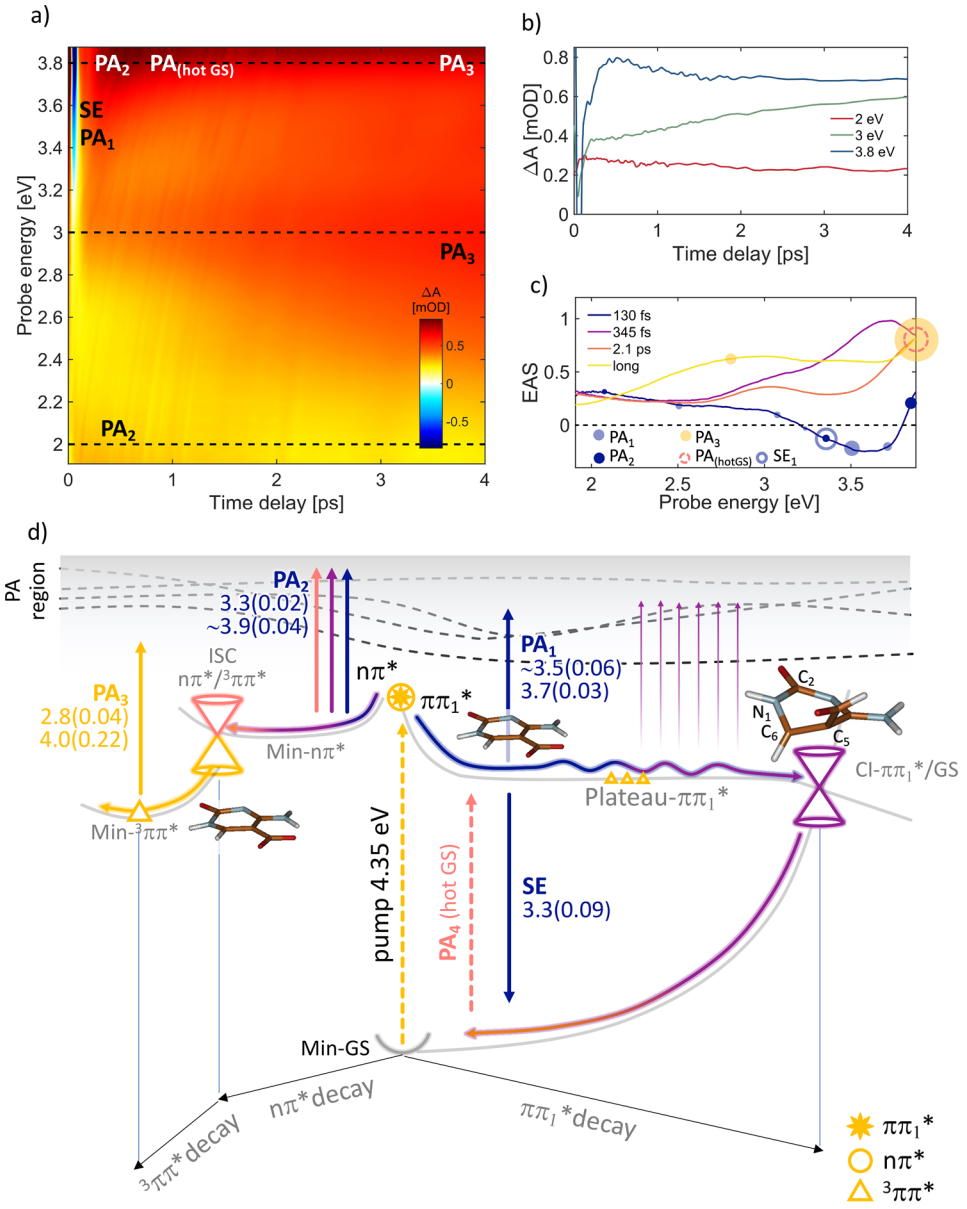


Figure 3. (a) ΔA map of fdC in water solution recorded with pulse polarizations at the magic angle. (b) Dynamics at selected probe energies (eV) marked with dashed lines on panel a. (c) EAS with the corresponding time constants: 130 fs (blue curve) is the time needed to relax from the FC to the Plateau-$\pi\pi_1$* flat region (calculated paths in d panel, right side) and simultaneously to the Min-nπ* (left side, panel d), 345 fs (purple curve) corresponds to the time to decay on the GS from the Plateau-$\pi\pi_1$* through CI-$\pi\pi_1$*-GS (wavy purple line, panel d), and the 2.1 ps time constant (pink curve, panel c) relates to the ISC ^{1}nπ*/$^3\pi\pi$* decay process (panel d, left side). The yellow line corresponds to long-living triplet state minimum (Min-$^3\pi\pi$*). Empty and full circles correspond to the calculated SE and PA energy values, respectively, and the circle dimensions are proportional to the computed oscillator strength values (documented in panel d). (d) Schematic decay paths of fdC, calculated at the CASPT2/MM level (details in the SI section). Relaxation routes and SE/PA colors arrows are matching with the line colors of time constants in panel c. Oscillator strengths are reported in brackets. Critical point energies are in Figure S9. Molecular optimized structures refer to the QM region only.

attributed to a further minor contribution coming from the Plateau-$\pi\pi$* region, in which the triplet is isoenergetic to the bright state (yellow triangles on the right side of Figure 3d).

The combination of sub-30-fs TA spectroscopy and state-of-the-art CASPT2/MM calculations thus enables one to derive a detailed picture of the different photoinduced processes in fdC, assign the observed decay time constants, and understand the pathway leading to population of the triplet state. The $\pi\pi$* ↔ nπ* IC in the FC region, later leading to the ISC process, and the ultrafast barrierless $\pi\pi_1$* → S_0 decay path are both fundamental and previously unpredicted excited state deactivation processes.[25−27]

Finally, and notably, the photophysics of cadC upon pumping at 4.35 eV is quite different from that of the previous derivative. Surprisingly, there is no evidence of triplet formation and the experimental signal shows an ultrafast relaxation that can be assigned to direct decay from the $\pi\pi_1$* ES to the GS. In this molecule, unlike the fdC derivative, computations do not identify any low-lying dark state that is isoenergetic with $\pi\pi_1$* (S_1) in the FC region. The experimental TA map (Figure 4a) and dynamics (Figure 4b) are dominated by SE (at 3.68 eV) and broad PA (at 2.33 eV) bands at early times that shift and decay on the ∼100 fs time scale to give rise to a PA band above 3.8 eV together with small remaining

https://doi.org/10.1021/acs.jpclett.1c02909
J. Phys. Chem. Lett. 2021, 12, 11070−11077

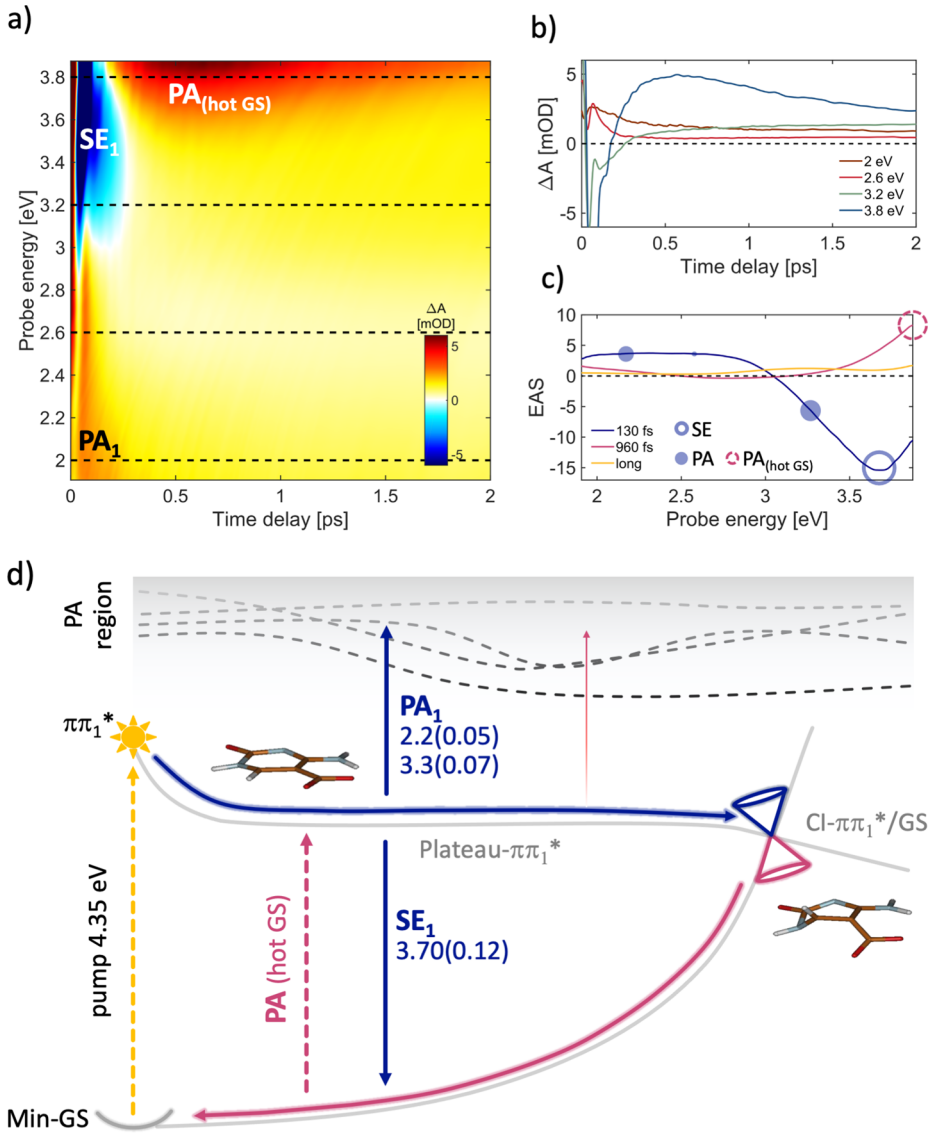


Figure 4. (a) ΔA map of cadC in water solution recorded with parallel pulse polarizations. (b) Dynamics at selected probe energies (eV) marked with dashed lines on panel a. (c) EAS with the corresponding time constants: 130 fs (blue curve) is the time needed to ballistically relax from the FC to the $\pi\pi_1$*/GS CI (see calculated blue paths in panel d), and 960 fs pink curve signals correspond to the hot GS repopulation, following the ultrafast IC (calculated decay path in panel d). The yellow line corresponds to long-lived products probably due to other minor decay paths. Empty and full circles correspond to the calculated SE and PA energy values, respectively, and the circle dimensions are proportional to the computed OS values (documented in panel d). (d) Schematic decay paths of cadC, calculated at CASPT2/MM level (details in the SI section). Relaxation routes and SE/PA colors arrows are matching with the line colors of time constants in panel c. Oscillator strengths are reported in brackets. Critical point energies are in Figure S10. Molecular optimized structures refer to the QM region only.

intensity in the shifted PA (near IR region), which in turn decays on the picosecond time scale. The corresponding EAS (Figure 4c) decays with the very fast 130 fs time constant, giving rise to a characteristic spectrum of hot GS PA decaying with a 960 fs time constant (previously incorrectly assigned to the $\pi\pi^* \rightarrow S_0$ decay[25]). By optimizing the lowest $\pi\pi_1$* state, we found once again a flat region of the PES (Plateau-$\pi\pi_1$*) where the SE signal (SE_1 at 3.7 eV in Figure 4d) matches the short-living experimental signal (130 fs, blue empty circle Figure 4c). These data indicate that the $\pi\pi_1^* \rightarrow$ GS decay process is ultrafast (with a 130 fs time constant, blue line in Figure 4c), leading straight to the CI in a ballistic fashion (CI-$\pi\pi_1$*/GS, Figure 4d), differently from the mdc and hm-dC derivatives described above, where the flatter region of the S_1 PES, leading to the Min-$\pi\pi_1$*, and the energy barrier work as a trap (Figures 2 and S7).

On the other hand, we attribute the red tail of the second EAS spectrum (pink line in Figure 4c) to the residual population remaining trapped on the Plateau-$\pi\pi_1$* (similarly to the barrierless $\pi\pi_1$* ultrafast evolution of the fdC), where the SE is almost negligible, showing only a very weak tail around 2.9 eV. While a significant part of the ES population decays on an ultrafast time scale through this $\pi\pi_1^* \rightarrow$ GS IC channel, there is also a low-intensity PA signal left for times longer than 30 ps (Figure 4c, yellow) that might be due to other minor decay pathways.

In conclusion, our joint experimental/computational study provides a comprehensive picture of the ES dynamics of all four epigenetic 2′-deoxycytidine nucleosides. By combining

https://doi.org/10.1021/acs.jpclett.1c02909
J. Phys. Chem. Lett. 2021, 12, 11070−11077

ultrafast TA spectroscopy with sub-30-fs temporal resolution with CASPT2/MM computations explicitly considering the water solvent, we have shown how the different chemical modifications dramatically affect the de-excitation pathways. By replacing the hydrogen atom at the fifth position of the pyrimidine ring with a methyl or hydroxymethyl group, the ultrafast ES decay along the S_1 PES, as compared to the parent molecule,[23] is slowed down due to an increased energy barrier to reach the $\pi\pi_1^*$/GS CI (0.35 or 0.30 eV, respectively), compared to the standard nucleoside (0.18 eV[23]). Indeed, the experimentally recorded SE signal, which provides an unambiguous spectroscopic fingerprint of the $\pi\pi_1^*$ state, decays in ∼4 ps for the methylated and hydroxymethylated derivatives in contrast with the typical subpicosecond decay of 2′-deoxycytidine. For these molecules, we also observe an initial ultrafast decay (∼130−160 fs time constant), associated with the fast relaxation out of the FC region. Moreover, the low-lying dark states that are thought to be involved in the excited state relaxation path of water-solvated 2′-deoxycytidine are destabilized in these derivatives[24] and are therefore not involved in the main relaxation pathway.

Substitution of a C_5 hydrogen of the cytosine ring by a formyl group significantly changes the ES dynamics. We first identify a dark $n\pi^*$ state, which is nearly energetically degenerate with the bright $\pi\pi^*$ state in the FC region, that can be thus immediately populated, eventually enabling an ultrafast 2 ps ISC process from the $n\pi^*$ state minimum, which gives rise to a long-lived lowest triplet state, in agreement with previous studies.[25] We also characterize a new additional and simultaneous $\pi\pi_1^* \rightarrow S_0$ ultrafast decay pathway, leading directly back to the ground state.

Finally, the carboxyl derivative displays the shortest, and previously uncharted, ES lifetime among all epigenetic dC nucleosides, dominated by the ultrafast decay of the lowest $\pi\pi_1^*$ (S_1) to the GS with a 130 fs time constant due to a ballistic wavepacket motion toward a low-lying barrierless CI. The 960 fs time constant, previously assigned to the $\pi\pi_1^* \rightarrow$ GS IC process, is now attributed to GS vibrational cooling.[25]

This work represents an important step toward a comprehensive picture of the intricate photophysical decay mechanisms of epigenetic dC derivatives in the biologically relevant aqueous environment, which display a dramatic sensitivity to C5 substitutions. Our results help to elucidate their role in the incidence of DNA photodamage, promoted by either longer excited state lifetimes or population of the triplet states, which leads to the generation of destructive singlet oxygen and makes the epigenetic derivatives more reactive or, on the other hand, possibly suitable in medical applications as phototherapeutic agents.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.1c02909.

Computational details; 5-methyl-2′-deoxycytidine: critical points and conical intersection calculations; benchmark calculations; 5-hydroxymethyl-2′-deoxycytidine: time-resolved spectra, decay paths, critical points, and conical intersection calculations; 5-formyl-2′-deoxycytidine: critical points and conical intersection calculations; 5-formyl-2′-deoxycytidine: formyl in *syn* and *anti* conformation; 5-carboxyl-2′-deoxycytidine: critical points and conical intersection calculations; chemical synthesis; sample preparation; TA setup description; DUV probe measurement of fdC; parallel polarizations for mdC and hmdC and magic angle polarization for cadC; impulsively excited vibrations; Cartesian coordinates (QM region only) (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Marco Garavelli − *Dipartimento di Chimica Industriale, Università degli Studi di Bologna, I-40136 Bologna, Italy;* orcid.org/0000-0002-0796-289X; Email: marco.garavelli@unibo.it

Giulio Cerullo − *IFN-CNR, Dipartimento di Fisica, Politecnico di Milano, I-20133 Milano, Italy;* orcid.org/0000-0002-9534-2702; Email: giulio.cerullo@polimi.it

Irene Conti − *Dipartimento di Chimica Industriale, Università degli Studi di Bologna, I-40136 Bologna, Italy;* orcid.org/0000-0001-7982-4480; Email: irene.conti@unibo.it

Authors

Piotr Kabaciński − *IFN-CNR, Dipartimento di Fisica, Politecnico di Milano, I-20133 Milano, Italy;* orcid.org/0000-0003-4591-5100

Marco Romanelli − *Dipartimento di Chimica Industriale, Università degli Studi di Bologna, I-40136 Bologna, Italy*

Eveliina Ponkkonen − *Department of Chemistry, Ludwig-Maximilians-Universität München, Munich 81377, Germany*

Vishal Kumar Jaiswal − *Dipartimento di Chimica Industriale, Università degli Studi di Bologna, I-40136 Bologna, Italy*

Thomas Carell − *Department of Chemistry, Ludwig-Maximilians-Universität München, Munich 81377, Germany;* orcid.org/0000-0001-7898-2831

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.1c02909

Author Contributions

‖P.K. and M.R. contributed equally to the work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge support from the H2020 Grant Agreement number 765266 (LightDyNAmics). G.C. and M.G. acknowledge support from the European Research Council Advanced Grant STRATUS (ERC-2011-AdG No. 291198).

■ REFERENCES

(1) Nabel, C. S.; Manning, S. A.; Kohli, R. M. The Curious Chemical Biology of Cytosine: Deamination, Methylation,and Oxidation as Modulators of Genomic Potential. *ACS Chem. Biol.* **2012**, *7*, 20−30.

(2) Yang, X.; Han, H.; DeCarvalho, D. D.; Lay, F. D.; Jones, P. A.; Liang, G. Gene Body Methylation Can Alter Gene Expression and Is a Therapeutic Target in Cancer. *Cancer Cell* **2014**, *26*, 577−590.

(3) You, C.; Ji, D.; Dai, X.; Wang, Y. Effects of Tet-Mediated Oxidation Products of 5-Methylcytosine on DNA Transcription in Vitro and in Mammalian Cells. *Sci. Rep.* **2015**, *4*, 7052.

(4) Lu, X.; Zhao, B. S.; He, C. TET Family Proteins: Oxidation Activity, Interacting Molecules, and Functions in Diseases. *Chem. Rev.* **2015**, *115*, 2225−2239.

(5) Drohat, A. C.; Coey, C. T. Role of Base Excision "Repair" Enzymes in Erasing Epigenetic Marks from DNA. *Chem. Rev.* **2016**, *116*, 12711−12729.

(6) Shen, L.; Song, C.-X.; He, C.; Zhang, Y. Mechanism and Function of Oxidative Reversal of DNA and RNA Methylation. *Annu. Rev. Biochem.* **2014**, *83*, 585−614.
(7) Kriaucionis, S.; Heintz, N. The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science (Washington, DC, U. S.)* **2009**, *324*, 929−930.
(8) Pfaffeneder, T.; Hackner, B.; Truß, M.; Münzel, M.; Müller, M.; Deiml, C. A.; Hagemeier, C.; Carell, T. The Discovery of 5-Formylcytosine in Embryonic Stem Cell DNA. *Angew. Chem., Int. Ed.* **2011**, *50*, 7008−7012.
(9) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science (Washington, DC, U. S.)* **2011**, *333*, 1300−1303.
(10) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science (Washington, DC, U. S.)* **2009**, *324*, 930−935.
(11) Martínez-Fernández, L.; Pepino, A. J.; Segarra-Martí, J.; Banyasz, A.; Garavelli, M.; Improta, R. Computing the Absorption and Emission Spectra of 5-Methylcytidine in Different Solvents: A Test-Case for Different Solvation Models. *J. Chem. Theory Comput.* **2016**, *12*, 4430−4439.
(12) Esposito, L.; Banyasz, A.; Douki, T.; Perron, M.; Markovitsi, D.; Improta, R. Effect of C5-Methylation of Cytosine on the Photoreactivity of DNA: A Joint Experimental and Computational Study of TCG Trinucleotides. *J. Am. Chem. Soc.* **2014**, *136*, 10838−10841.
(13) Tommasi, S.; Denissensko, M. F.; Pfeifer, G. P. Sunlight Induces Pyrimidine Dimers Preferentially at 5-Methylcytosine Bases. *Cancer Res* **1997**, *57*, 4727−4730.
(14) Giussani, A.; Conti, I.; Nenov, A.; Garavelli, M. Photoinduced Formation Mechanism of the Thymine−Thymine (6−4) Adduct in DNA; a QM(CASPT2//CASSCF):MM(AMBER) Study. *Faraday Discuss.* **2018**, *207*, 375−387.
(15) Conti, I.; Martínez-Fernández, L.; Esposito, L.; Hofinger, S.; Nenov, A.; Garavelli, M.; Improta, R. Multiple Electronic and Structural Factors Control Cyclobutane Pyrimidine Dimer and 6−4 Thymine−Thymine Photodimerization in a DNA Duplex. *Chem. - Eur. J.* **2017**, *23*, 15177−15188.
(16) Weingart, O.; Nenov, A.; Altoè, P.; Rivalta, I.; Segarra-Martí, J.; Dokukina, I.; Garavelli, M. COBRAMM 2.0 — A Software Interface for Tailoring Molecular Electronic Structure Calculations and Running Nanoscale (QM/MM) Simulations. *J. Mol. Model.* **2018**, *24*, 271.
(17) Aquilante, F.; Autschbach, J.; Baiardi, A.; Battaglia, S.; Borin, V. A.; Chibotaru, L. F.; Conti, I.; De Vico, L.; Delcey, M.; Fdez. Galván, I.; Ferré, N.; Freitag, L.; Garavelli, M.; Gong, X.; Knecht, S.; Larsson, E. D.; Lindh, R.; Lundberg, M.; Malmqvist, P. Å.; Nenov, A.; Norell, J.; Odelius, M.; Olivucci, M.; Pedersen, T. B.; Pedraza-González, L.; Phung, Q. M.; Pierloot, K.; Reiher, M.; Schapiro, I.; Segarra-Martí, J.; Segatta, F.; Seijo, L.; Sen, S.; Sergentu, D.-C.; Stein, C. J.; Ungur, L.; Vacher, M.; Valentini, A.; Veryazov, V. Modern Quantum Chemistry with [Open]Molcas. *J. Chem. Phys.* **2020**, *152*, 214117.
(18) Fdez. Galván, I.; Vacher, M.; Alavi, A.; Angeli, C.; Aquilante, F.; Autschbach, J.; Bao, J. J.; Bokarev, S. I.; Bogdanov, N. A.; Carlson, R. K.; Chibotaru, L. F.; Creutzberg, J.; Dattani, N.; Delcey, M. G.; Dong, S. S.; Dreuw, A.; Freitag, L.; Frutos, L. M.; Gagliardi, L.; Gendron, F.; Giussani, A.; González, L.; Grell, G.; Guo, M.; Hoyer, C. E.; Johansson, M.; Keller, S.; Knecht, S.; Kovačević, G.; Källman, E.; Li Manni, G.; Lundberg, M.; Ma, Y.; Mai, S.; Malhado, J. P.; Malmqvist, P. Å.; Marquetand, P.; Mewes, S. A.; Norell, J.; Olivucci, M.; Oppel, M.; Phung, Q. M.; Pierloot, K.; Plasser, F.; Reiher, M.; Sand, A. M.; Schapiro, I.; Sharma, P.; Stein, C. J.; Sørensen, L. K.; Truhlar, D. G.; Ugandi, M.; Ungur, L.; Valentini, A.; Vancoillie, S.; Veryazov, V.; Weser, O.; Wesołowski, T. A.; Widmark, P.-O.; Wouters, S.; Zech, A.; Zobel, J. P.; Lindh, R. OpenMolcas: From Source Code to Insight. *J. Chem. Theory Comput.* **2019**, *15*, 5925−5964.
(19) Gustavsson, T.; Bányász, Á.; Lazzarotto, E.; Markovitsi, D.; Scalmani, G.; Frisch, M. J.; Barone, V.; Improta, R. Singlet Excited-State Behavior of Uracil and Thymine in Aqueous Solution: A Combined Experimental and Computational Study of 11 Uracil Derivatives. *J. Am. Chem. Soc.* **2006**, *128*, 607−619.
(20) Onidas, D.; Markovitsi, D.; Marguet, S.; Sharonov, A.; Gustavsson, T. Fluorescence Properties of DNA Nucleosides and Nucleotides: A Refined Steady-State and Femtosecond Investigation. *J. Phys. Chem. B* **2002**, *106*, 11367−11374.
(21) Pecourt, J.-M. L.; Peon, J.; Kohler, B. DNA Excited-State Dynamics: Ultrafast Internal Conversion and Vibrational Cooling in a Series of Nucleosides. *J. Am. Chem. Soc.* **2001**, *123*, 10370−10378.
(22) Ma, C.; Cheng, C. C.-W.; Chan, C. T.-L.; Chan, R. C.-T.; Kwok, W.-M. Remarkable Effects of Solvent and Substitution on the Photo-Dynamics of Cytosine: A Femtosecond Broadband Time-Resolved Fluorescence and Transient Absorption Study. *Phys. Chem. Chem. Phys.* **2015**, *17*, 19045−19057.
(23) Martínez-Fernández, L.; Pepino, A. J.; Segarra-Martí, J.; Jovaišaitė, J.; Vaya, I.; Nenov, A.; Markovitsi, D.; Gustavsson, T.; Banyasz, A.; Garavelli, M.; Improta, R. Photophysics of Deoxycytidine and 5-Methyldeoxycytidine in Solution: A Comprehensive Picture by Quantum Mechanical Calculations and Femtosecond Fluorescence Spectroscopy. *J. Am. Chem. Soc.* **2017**, *139*, 7780−7791.
(24) Wang, X.; Zhou, Z.; Tang, Y.; Chen, J.; Zhong, D.; Xu, J. Excited State Decay Pathways of 2′-Deoxy-5-Methylcytidine and Deoxycytidine Revisited in Solution: A Comprehensive Kinetic Study by Femtosecond Transient Absorption. *J. Phys. Chem. B* **2018**, *122*, 7027−7037.
(25) Wang, X.; Yu, Y.; Zhou, Z.; Liu, Y.; Yang, Y.; Xu, J.; Chen, J. Ultrafast Intersystem Crossing in Epigenetic DNA Nucleoside 2′-Deoxy-5-Formylcytidine. *J. Phys. Chem. B* **2019**, *123*, 5782−5790.
(26) Xing, J.; Ai, Y.; Liu, Y.; Du, J.; Chen, W.; Lu, Z.; Wang, X. Theoretical Studies on the Photophysics and Photochemistry of 5-Formylcytosine and 5-Carboxylcytosine: The Oxidative Products of Epigenetic Modification of Cytosine in DNA. *J. Phys. Chem. B* **2018**, *122*, 2704−2714.
(27) Wang, X.; Martínez-Fernández, L.; Zhang, Y.; Zhang, K.; Improta, R.; Kohler, B.; Xu, J.; Chen, J. Solvent-Dependent Stabilization of a Charge Transfer State Is the Key to Ultrafast Triplet State Formation in an Epigenetic DNA Nucleoside. *Chem. - Eur. J.* **2021**, *27*, 10932−10940.
(28) Ai, Y.; Xing, J.; Zhang, A.; Zhao, C.; Liu, Y.; Xie, B.; Chen, W.; Cui, G.; Lu, Z.; Wang, X. Computational Study on the Excited-State Decay of 5-Methylcytosine and 5-Hydroxymethylcytosine: The Common Form of DNA Methylation and Its Oxidation Product. *J. Phys. Chem. B* **2018**, *122*, 10424−10434.
(29) Conti, I.; Garavelli, M. Evolution of the Excitonic State of DNA Stacked Thymines: Intrabase $\Pi\pi^*$ → S0 Decay Paths Account for Ultrafast (Subpicosecond) and Longer (>100 Ps) Deactivations. *J. Phys. Chem. Lett.* **2018**, *9*, 2373−2379.
(30) Conti, I.; Cerullo, G.; Nenov, A.; Garavelli, M. Ultrafast Spectroscopy of Photoactive Molecular Systems from First Principles: Where We Stand Today and Where We Are Going. *J. Am. Chem. Soc.* **2020**, *142*, 16117−16139.
(31) Francés-Monerris, A.; Lineros-Rosa, M.; Miranda, M. A.; Lhiaubet-Vallet, V.; Monari, A. Photoinduced Intersystem Crossing in DNA Oxidative Lesions and Epigenetic Intermediates. *Chem. Commun.* **2020**, *56*, 4404−4407.

https://doi.org/10.1021/acs.jpclett.1c02909
J. Phys. Chem. Lett. 2021, 12, 11070−11077

5. Unpublished Work

5.1. Decarboxylation studies of F-cadC

In 2018, the first *in vivo* study was published that gave evidence for direct demethylation of epigenetically modified dC-nucleosides in gDNA.[16] Therein, F-fdC was fed to mESCs and other cell types, and after gDNA isolation and digestion, UHPLC-MS/MS studies detected the C-C bond cleavage product F-dC. However, the authors were not able to answer the question whether the process occurs via oxidation of fdC to cadC and subsequent decarboxylation or via direct deformylation of fdC and whether these two processes are independent.

The aim of the present project was to investigate if decarboxylation of cadC occurs *in vivo* in gDNA by using a combination of synthetic chemistry, cell biology and analytical chemistry. Since the goal was to study a C-C bond cleavage reaction of cadC in gDNA, a probe molecule that resists the BER pathway had to be used. 2'-Fluorinated cadC (F-cadC) was synthesized modifying an already published protocol,[73] the target compound was extensively purified via HPLC and then fed to Neuro-2a cells (N2a cells). A concentration of 100 µM F-cadC was used for the cell feeding experiments and after three days of incubation, gDNA was isolated and digested enzymatically into individual nucleosides. The nucleoside mixture was then analyzed with UHPLC-MS/MS. In both biological replicates (BR1 and BR2), no F-cadC could be detected (Figure 7), however, the corresponding decarboxylation product F-dC and its subsequent remethylation product F-mdC were found to be present in quantifiable amounts. The presence of decarboxylation and remethylation products in both biological replicates could indicate that a decarboxylation process is indeed occurring inside of the cells.

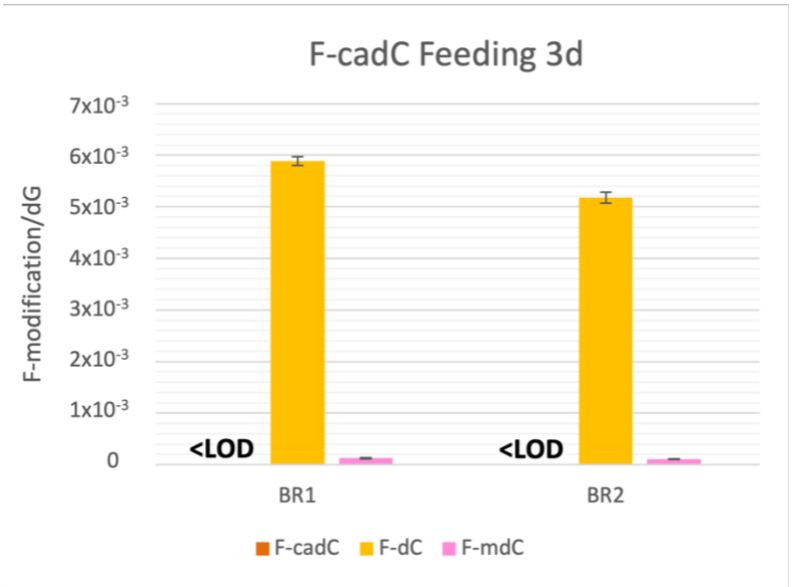


Figure 7. F-cadC nucleoside feeding of F-cadC to N2a cells for 3d. <LOD, indicates the value to be under the limit of detection and the error bars show the standard deviation of the technical triplicate.

To ensure that the observed F-dC was not a result of spontaneous decarboxylation of F-cadC in the cytoplasm, the soluble nucleotide pool of the fed cells was also analyzed. Earlier published data claims deoxycytidine kinase (DCK) to transfer a phosphate from ATP to dC, hmdC and fdC but not to cadC.[77] This would support the present data of not detecting a strong incorporation of F-cadC into the gDNA. If the detected F-dC was a product of spontaneous decarboxylation in the cytoplasm, it should, however, be also detectable in the soluble cellular nucleotide pool. To study F-cadC in the cytosol, an isolation and analysis method to distinguish between the nucleosides and nucleotides was developed. The usual nucleotide soluble pool separation protocol involves only the collection of nucleosides, while triphosphates are separated from the nucleosides beforehand. In the newly developed method, which is shown in Figure 8, the aqueous phase that is collected first, contains oligonucleotides, mono, - di, - and triphosphates that are subsequently treated with a phosphatase that dephosphorylates the nucleotides to nucleosides which are then collected according to the common protocol using MeCN/MeOH. To analyze the content of the soluble pool, these fractions were measured separately with UHPLC/MS-MS. The fractions are in the following indicated as “nucleotide” and “nucleoside” depending on the isolation step, where the sample was obtained.

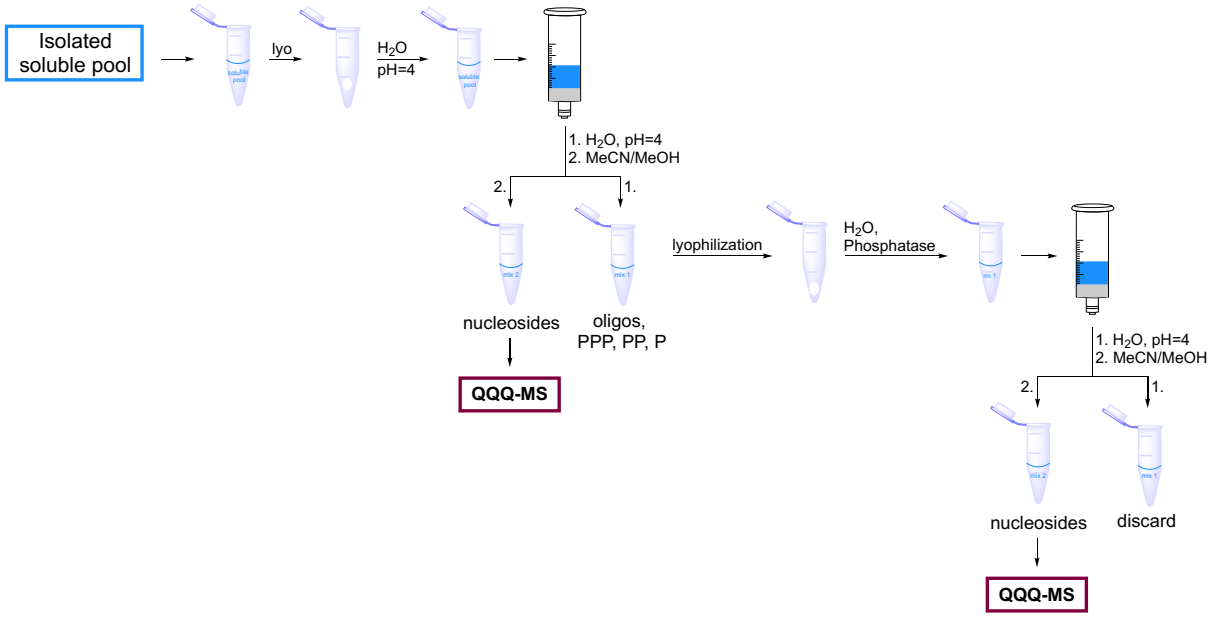


Figure 8. Nucleotide soluble pool extraction workflow to distinguish nucleosides from nucleotides.

The soluble pool analysis of BR1 showed F-cadC to be present in both, the nucleoside and the nucleotide form. Also, no F-dC nucleosides or nucleotides were detected in remarkable amounts (Figure 9). The analysis of BR2, showed similar results: F-cadC was mainly present as a nucleotide, however whether it was in the form of the mono-, di- or triphosphate was not distinguishable by the present detection method. If phosphorylation was not complete or efficient enough, the detection of its incorporation would agree with the above-shown data, where F-cadC was not detected in the gDNA. Moreover, no remarkable amounts of F-dC nucleoside/nucleotide were detected. All in all, this could indicate that F-cadC was stable towards spontaneous decarboxylation and the data could be interpreted as evidence towards decarboxylation on the genomic level. Nevertheless, it is still unclear how efficiently the F-cadC nucleoside is phosphorylated in the cytosol. Both biological replicates showed the formation of phosphorylated F-cadC to be present more than the corresponding nucleoside, while the gDNA samples unfortunately did not show quantifiable incorporation levels.

In general, the results can be interpreted to give evidence that F-cadC is able to pass through the cellular membrane due to its detection in the cellular soluble pool, but the phosphorylation to the corresponding triphosphate by a kinase in the cytosol might be on the other hand not efficient. Also, the observation of F-cadC in the soluble nucleotide pool as a

phosphorylated species is not in a full agreement with earlier published report.[77] If F-cadC is not phosphorylated efficiently enough, that would explain the unquantifiable/non detectable incorporation. On the other hand, the detected presence of F-dC in the gDNA might indicate that a decarboxylation turnover was so fast that the detection of the anyhow low abundant F-cadC was challenging. Hence, at this point further studies were required to exclude the possibility of spontaneous decarboxylation of F-cadC to F-dC in the soluble pool, as this could lead to phosphorylation of F-dC into its corresponding triphosphate and, if the uptake of F-dC nucleotide was fast, only the F-dC incorporated into the gDNA would have been detected (Scheme 2).

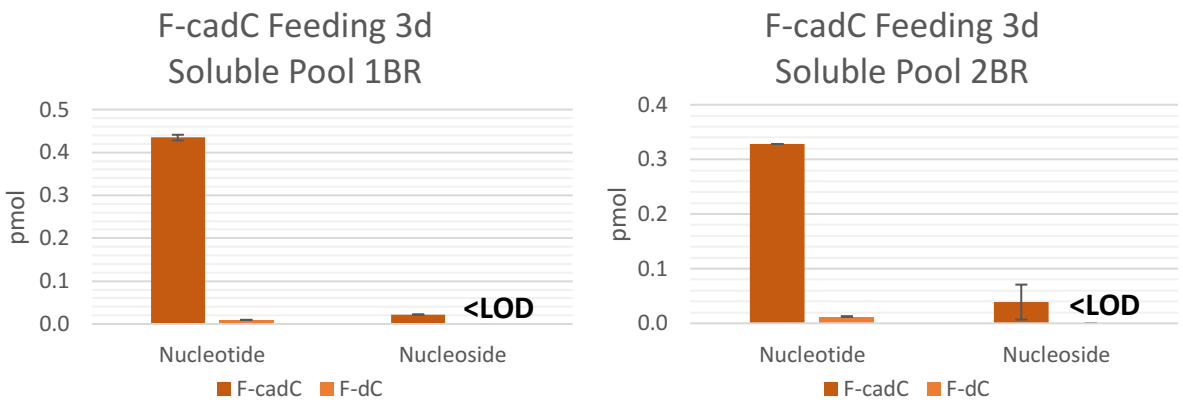
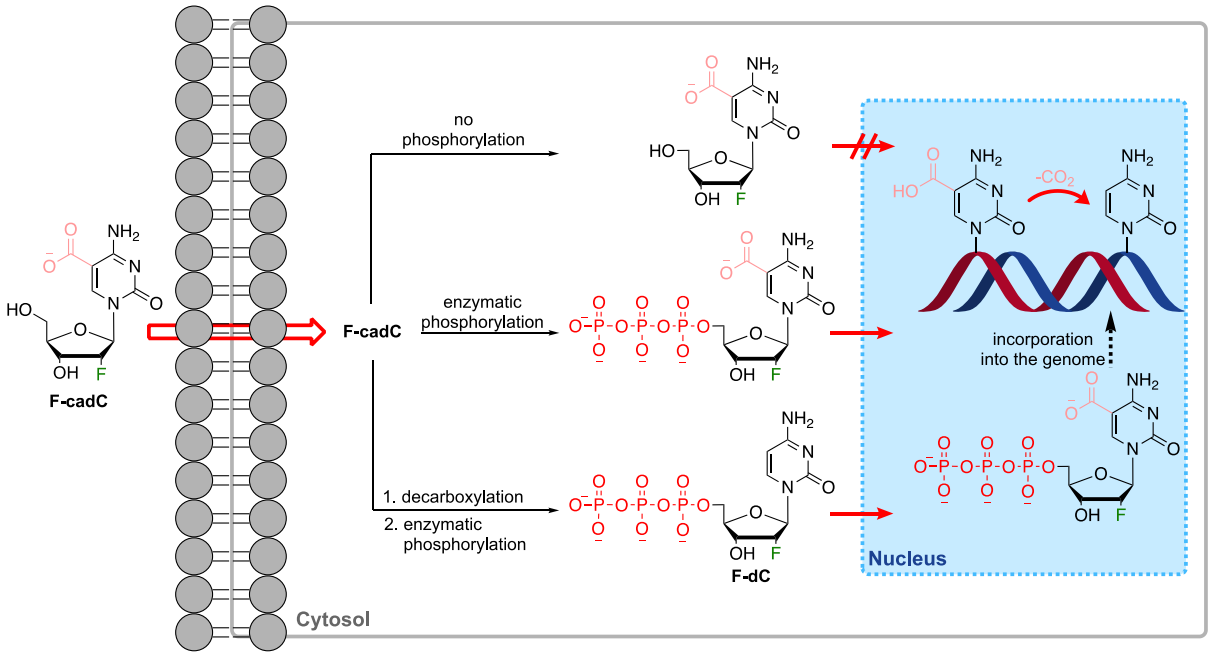
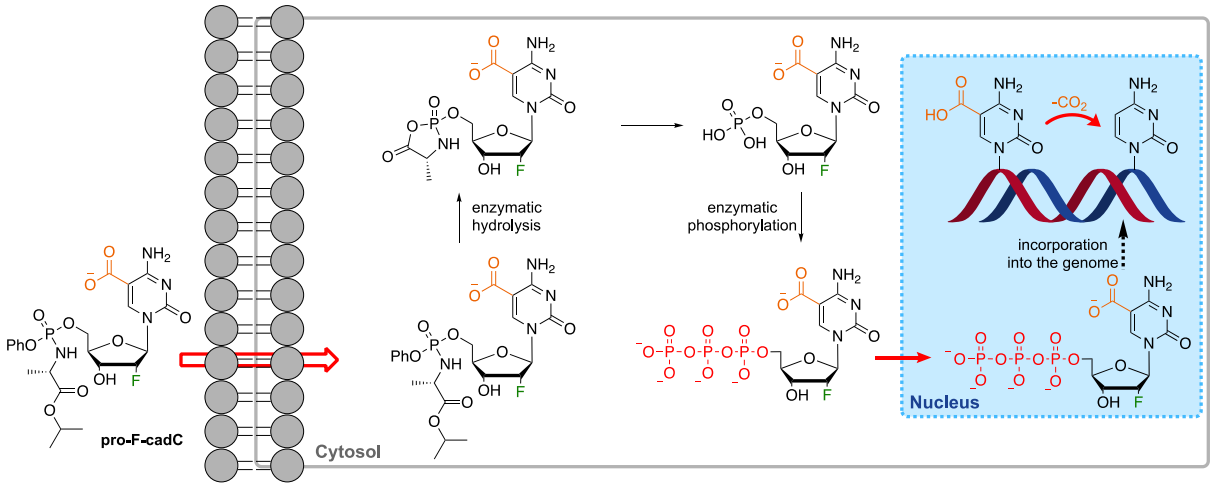


Figure 9. Soluble nucleoside pool analysis of F-cadC nucleoside feeding experiments to N2a cells.



Scheme 2. Overview over F-cadC and its potential metabolic pathways during cell feeding experiments.

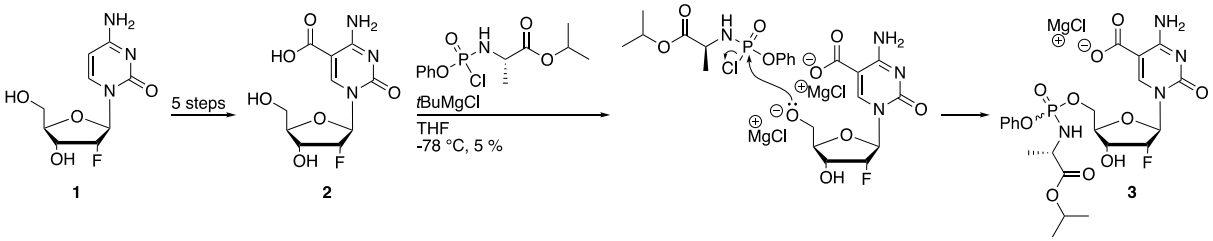
To come to a more reliable conclusion, it was of interest to achieve higher levels of incorporation of F-cadC into the genome. An approach was designed, where F-cadC was synthesized as a phosphoramidate prodrug. Since the first phosphorylation step is considered as the rate-limiting step in the triphosphate biosynthesis[78], this F-cadC prodrug would circumvent the challenging first phosphorylation step in the cytosol, as the prodrug would be able to be converted into its corresponding monophosphate via enzymatic hydrolysis without requiring phosphorylation by a kinase (see the mechanism in Scheme 3). Thereafter, the kinases catalyzing triphosphate biosynthesis were expected to recognize the resulting nucleoside monophosphate and further phosphorylate it to yield the corresponding triphosphate. Another advantage of the monophosphate prodrug is its potentially improved cell membrane permeability by masking the monophosphate group with lipophilic moieties, compared to an unmasked F-cadC bearing the charged carboxylate group under physiological pH.[79,80]



Scheme 3. Overview of cell feeding experiments using pro-F-cadC and of its potential metabolic pathway in the cell.

The synthesis of the F-cadC prodrug (pro-F-cadC) started with the known synthetic route from F-dC **1** to F-cadC **2.**[72] HPLC purified **2** was then phosphorylated at the 5'-position using the freshly synthesized corresponding aryloxyphosphorochloridate[80] and a strong organometallic base, *tert*-butylmagnesium chloride, that resulted in the formation of the corresponding nucleoside alkoxide in the first step as shown in Scheme 4. However, an excess of the Grignard reagent was required as the carboxylic acid also reacted in an acid-base reaction to form the corresponding carboxylate which competed with the deprotonation of the 5'-OH group. Due to the presence of a stereocenter at the phosphorus atom of nucleoside prodrug **3**, the final compound was obtained as a mixture of two diastereomers which were subsequently purified via HPLC. A chromatogram obtained via analytical HPLC of the crude reaction mixture before separation of the target compounds is shown in Figure 10. The absolute stereochemistry of the individual diastereomers was not elucidated, however, HPLC retention times allowed to distinguish the compounds and showed that one of the diastereomers was formed with a higher yield than the other. Figure 11 shows an analytical HPLC chromatogram of the first-eluting pro-F-cadC diastereomer after its preparative isolation.

Nevertheless, after extensive HPLC purification, the diastereomers of the target compound were combined and used as a mixture in the following cell feeding experiments due to the low overall reaction yield.



Scheme 4. Synthesis of pro-F-cadC **3**.

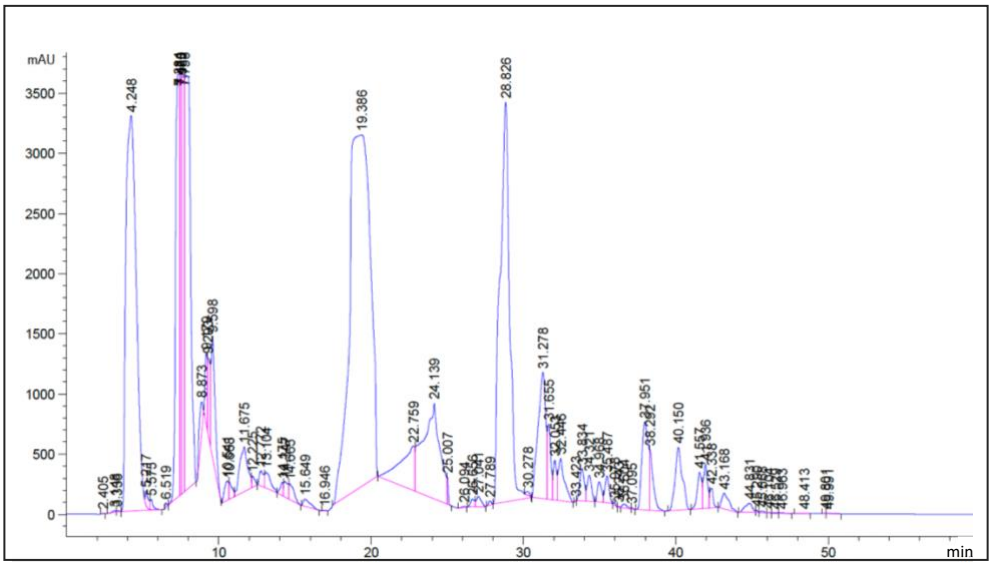


Figure 10. Analytical HPLC chromatogram of a crude reaction product of pro-F-cadC (**3**).

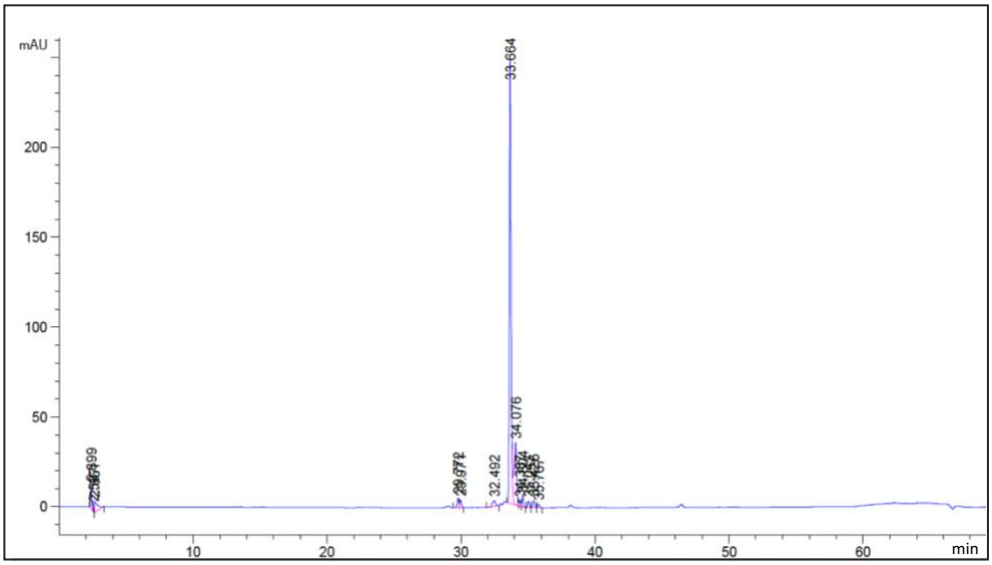


Figure 11. Analytical HPLC chromatogram of one purified diastereomer of **3**.

250 µM of pro-F-cadC was fed to N2a cells and after 3d of incubation, gDNA was isolated, digested and analyzed via UHPLC-MS/MS. F-cadC was not detected in the gDNA of the

samples, but its corresponding decarboxylation product F-dC and its remethylation product F-mdC were present in quantifiable amounts (Figure 12). Unfortunately, no soluble pool data could be obtained. Based on that, further studies are required to confirm the results.

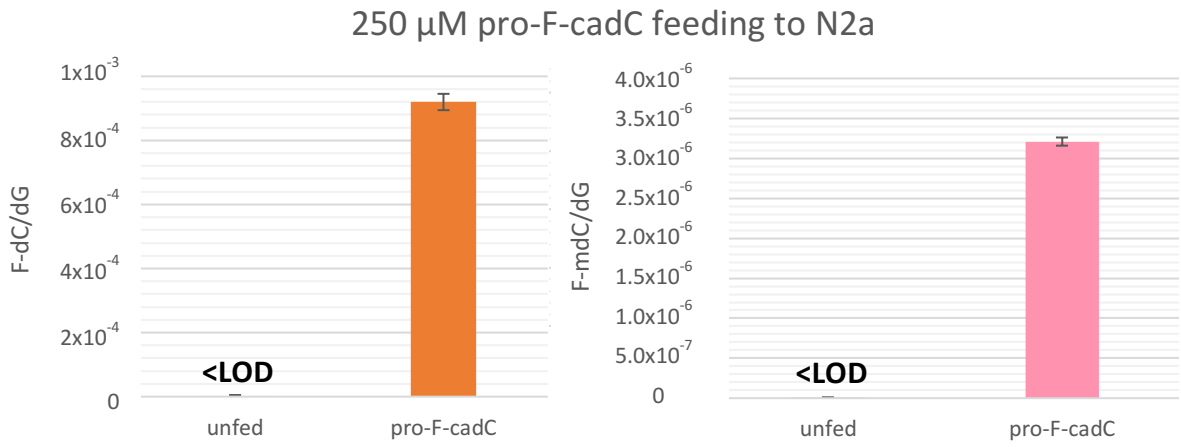
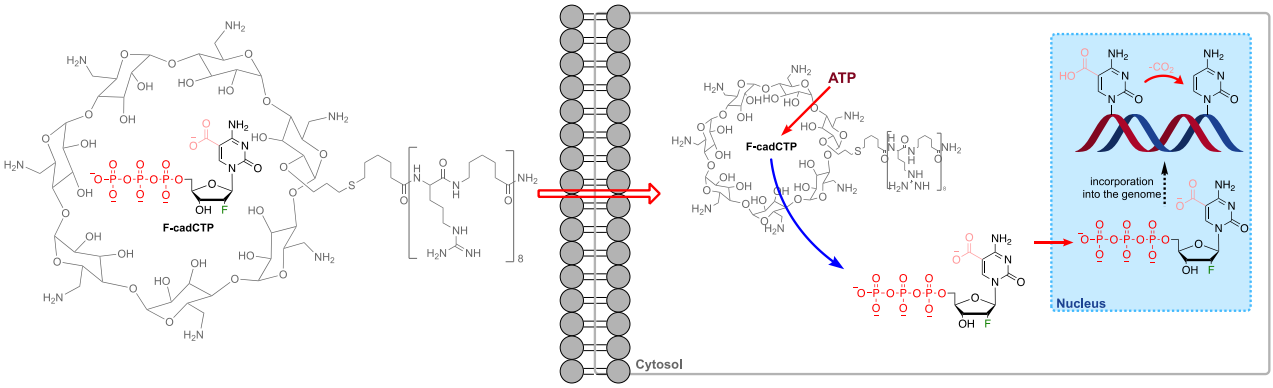


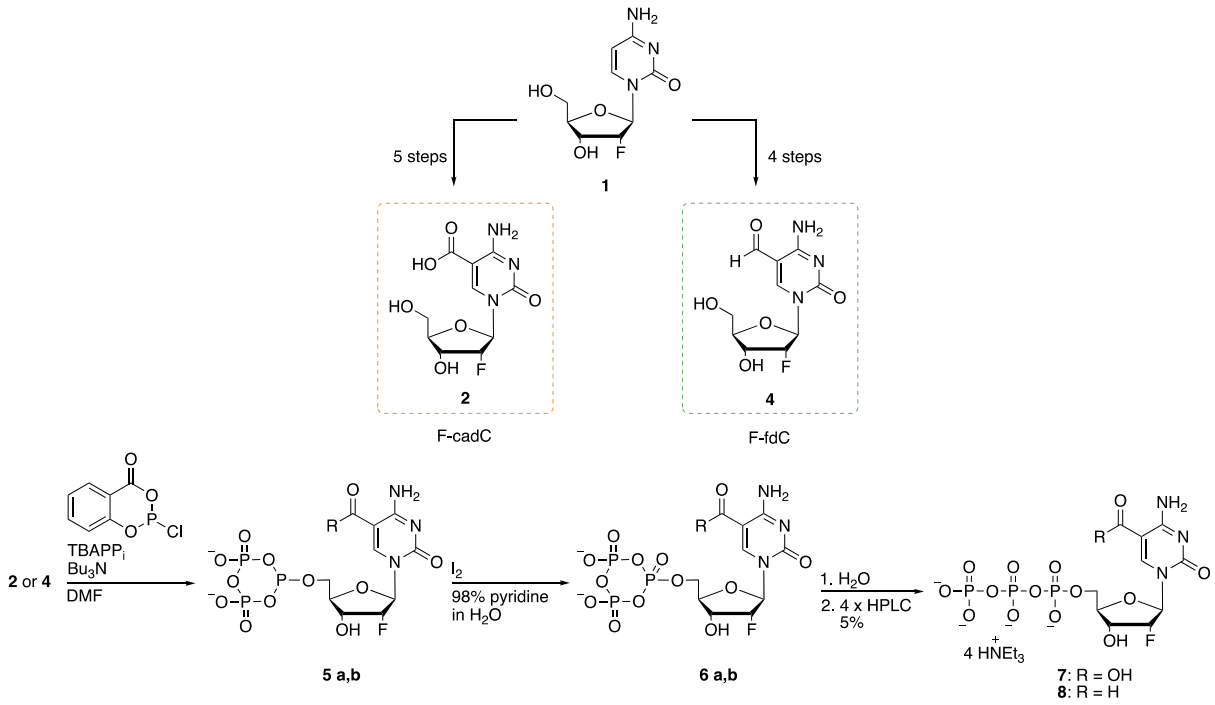
Figure 12. Results of 250 µM pro-F-cadC feeding experiments to N2a cells for 3d incubation time.

Another approach to achieve higher incorporation of F-cadC into the gDNA was based on cell feeding experiments using the triphosphate of F-cadC (F-cadCTP) which would completely circumvent the enzymatic phosphorylation steps in the cytosol and could be directly incorporated into the DNA. However, due to the in general highly charged nature of triphosphates, a transportation tool through the cellular membrane was required. Zawada *et al.*[81] presented the recent development of a nucleotide triphosphate transporter that comprises a cyclodextrin receptor which forms a non-covalent complex with the nucleotide anion, and a cell penetrating peptide derivative that allows the transportation of the target molecule across the cell membrane. Once inside of the cell, the nucleotide is released in the cytosol, and could directly enter the nucleus and be incorporated into the DNA (Scheme 5).



Scheme 5. Cell feeding experiments of F-cadC triphosphate using a transporter and its potential metabolic pathway to the nucleus.

To this aim, F-cadCTP (**7**) and F-fdC triphosphate (F-fdCTP, **8**) were synthesized according to published procedures (Scheme 6).[73] The triphosphate synthesis follows the Ludwig-Eckstein approach as its advantage is a rapid one-pot triphosphate synthesis starting from the unprotected nucleoside.[82] This synthetic pathway exploits salicyl phosphorochloridite which can undergo three nucleophilic displacement reactions. In the first step, a reaction between the nucleoside and salicyl phosphochloridite gives the corresponding nucleoside phosphite that reacts further with pyrophosphate in the reaction mixture forming a trivalent phosphorus at the 5'-O of the nucleoside (**5 a** and **b**). **5** is then further oxidized with iodine and the cyclic intermediate **6** is finally hydrolyzed to obtain the triphosphate.



Scheme 6. Synthesis of **2** and **4** and their corresponding triphosphates **7** and **8** from F-cadC **1**.

After extensive purification, the target compound was used for metabolic labelling studies together with the commercially purchased triphosphate transporter described above. All experiments involving F-cadCTP or F-fdCTP were conducted in the presence of this transporter. The selected cell line was again the Neuro-2a cell line that was previously shown to be robust enough and suitable for deformylation studies.[16,83] 100 µM F-cadCTP or F-fdC each in an 1:1 molar ratio with the triphosphate transporter were fed to N2a cells and incubated for 3 d. Subsequent mass spectrometric analysis of the isolated gDNA did not show any incorporation of F-cadC. The detected levels of F-dC and F-mdC in both BR1 and BR2 of F-cadC and F-fdC feedings are shown in Figure 13. BR1 shows slightly higher deformylation levels to F-dC than a decarboxylation of F-cadC to F-dC. BR2, however, shows F-cadC to decarboxylate more efficiently than F-fdC deformylates. To exclude spontaneous decarboxylation in the soluble cellular nucleoside pool, F-cadC, F-fdC and F-dC levels were also measured. F-cadC and F-fdC were indeed found in the cytosol and no decarboxylation or deformylation product F-dC was detected in relevant amounts, indicating the suitability of the probe molecules.

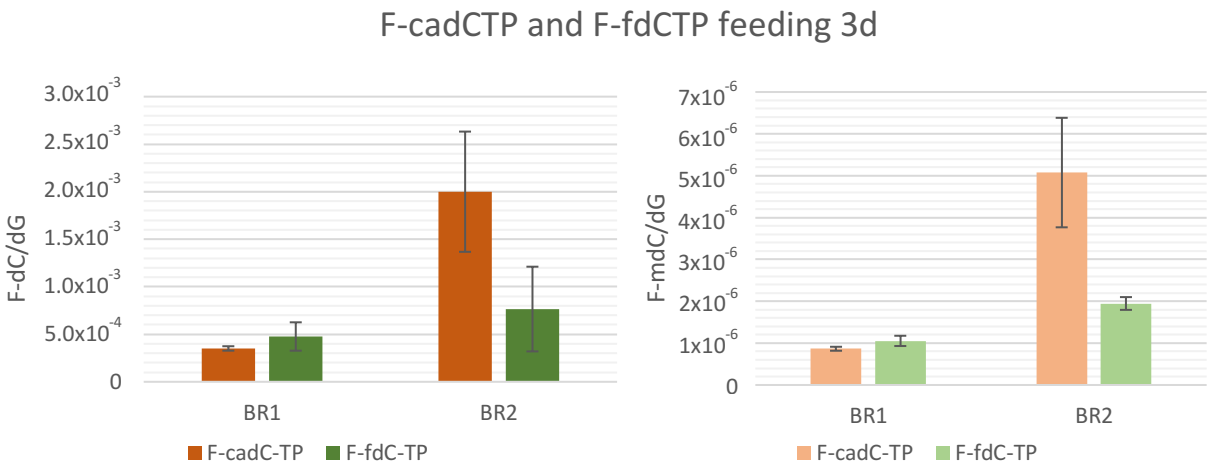


Figure 13. F-dC and F-mdC levels from 3d feeding of F-cadCTP and F-fdCTP.

It was also of interest to investigate, whether the cell fed triphosphate remained as a triphosphate in the metabolite pool after 3 d of incubation. An experimental biological duplicate revealed that F-cadCTP was the major component found, whereas F-cadC as well as F-dC were only present in almost undetectable amounts (Figure 14).

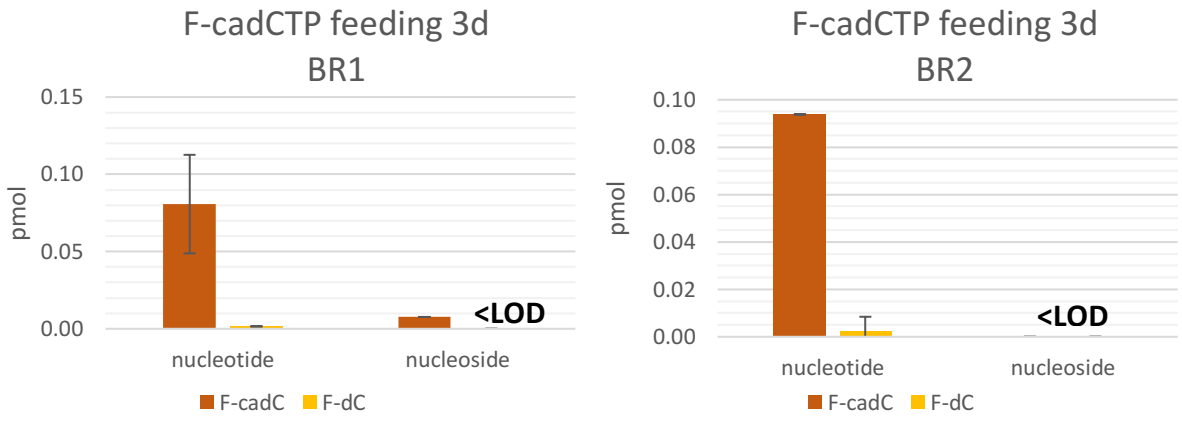


Figure 14. Soluble nucleotide pool analysis of two biological replicates of F-cadCTP cell feeding for 3d.

As a reference the soluble pool of a F-fdC feeding experiment was also analyzed, and the results showed that F-fdCTP was detected in the cytosol and that spontaneous deformylation was not remarkable (Figure 15).

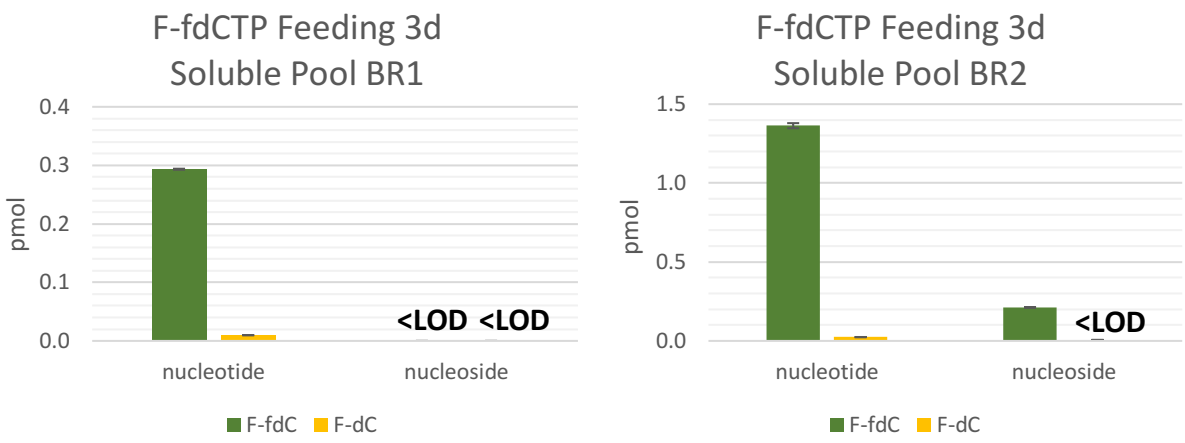


Figure 15. Two biological replicates of F-fdCTP cell feeding and 3d incubation.

In general, in all the experiments above, it was noticed that in the digested samples containing DNA, the added isotopically labelled quantification standard $^{15}N_2$-F-cadC did not show a peak as theoretically expected. Therefore, potential reasons for these detection issues of $^{15}N_2$-F-cadC were studied. First, isotopically labeled fluorinated standard molecules (F-standards) were treated with digestion enzymes in DNA-free samples and filtered with standard filters (0.2 µm). Surprisingly, only 0.6 % of the $^{15}N_2$-F-cadC signal was detected (Figure 16). The same analysis was performed on samples obtained from F-cadC feeding experiments, where only 1 % of the theoretically expected signal was detected. A similar detection effect on F-fdC was not as dramatic with a signal decrease to 40% of the expected level. In case of F-dC the detected absorption was between 15-20% of the expected level (Figure 17).

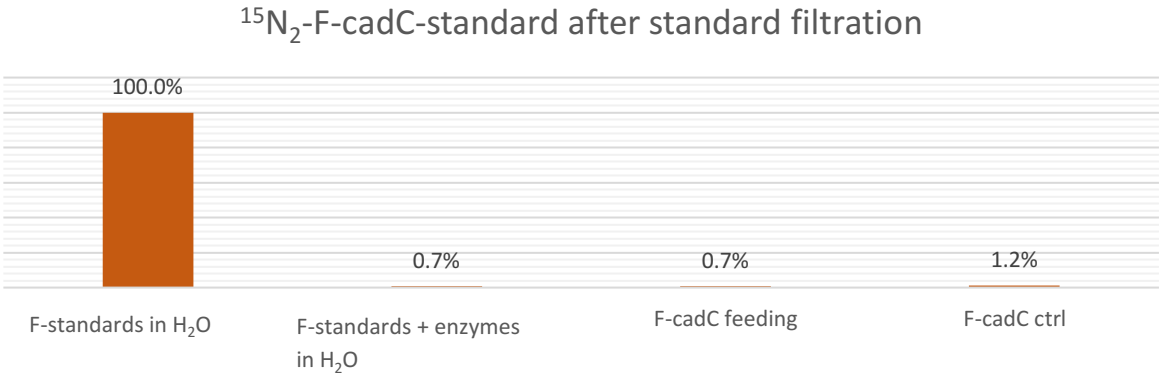


Figure 16. $^{15}N_2$-F-cadC-standard levels after standard filtration H_2O.

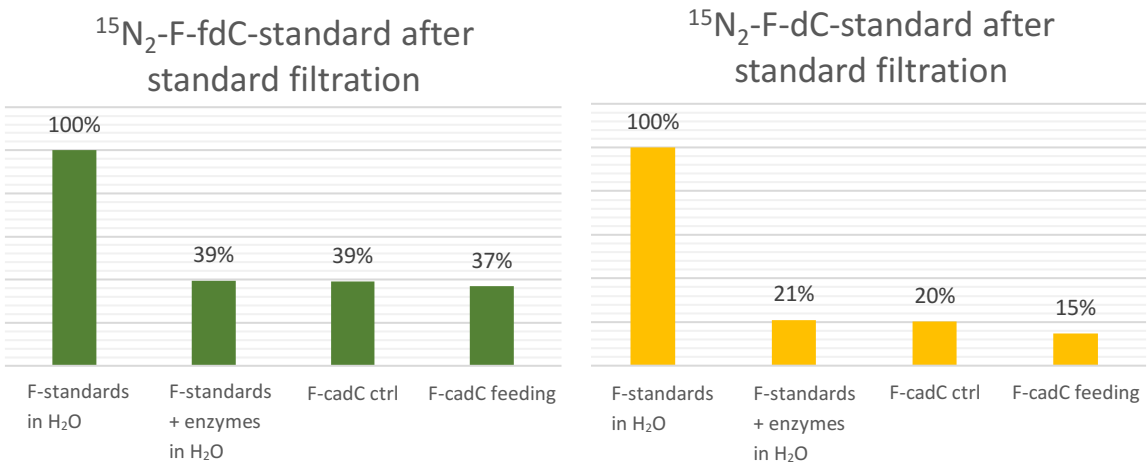


Figure 17. $^{15}N_2$-F-fdC -and $^{15}N_2$-F-dC-standard levels after standard filtration.

A possible explanation for the observed low signal was a potential loss of the compound during the filtration steps. Therefore, different filters were tested experimentally, and their impact was evaluated. Without filtration the obtained signal was the closest to the theoretically expected value, however, this approach is not sustainable for measuring DNA samples due to the resulting contaminations to be expected to influence the highly sensitive UHPLC-MS/MS. Nevertheless, the analysis of a cell feeding experiment using F-cadCTP, wherein DNA samples obtained were measured without filtration is shown and discussed further below in Figure 23 and Figure 24. From the experiments evaluating different filters available (AcroPrep Advance 96 filter plate 0.2 µm Supor, *Pall Life Sciences*) and AcroPrep Advance filter plate 10K (MWCO molecular weight cutoff, *Pall Life Sciences*), it was found that the 10K filter resulted in the best results for detection of F-cadC (Figure 18 A) and F-fdC (Figure 18 B). Nevertheless, the effect of different filters on the detected signal of F-dC was in general minimal (Figure 18C).

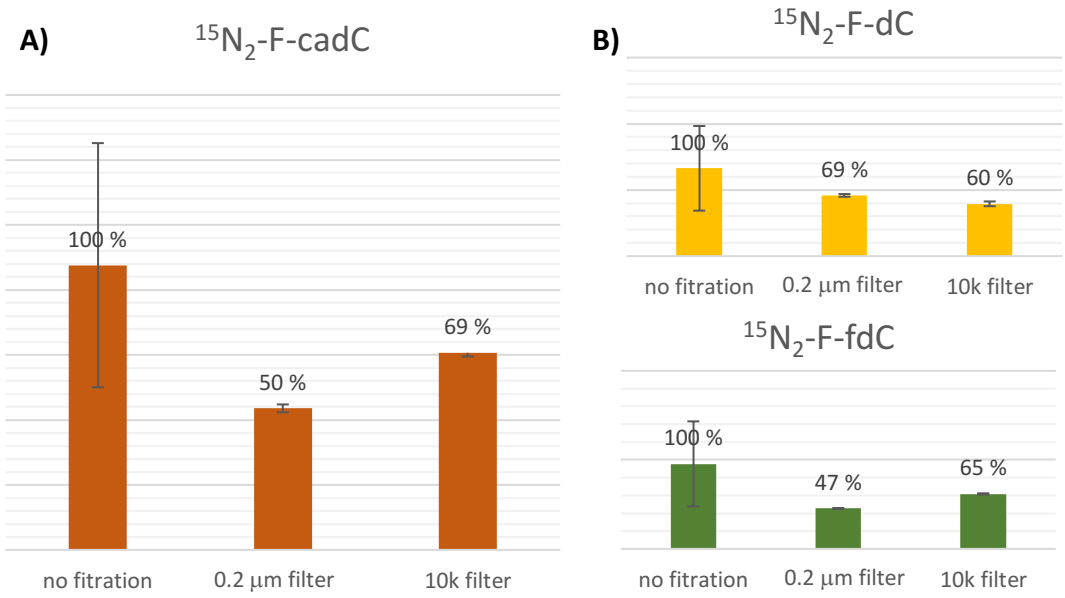


Figure 18. $^{15}N_2$-F-standards filtered with different filters.

A loss of the compound during the filtration process seems therefore unfortunately experimentally unavoidable and, thus, other variables needed to be optimized. To obtain the highest possible sample volume from the isolated DNA, the most suitable DNA digestion protocol for F-cadC analysis needed to be determined. A recently published study compared the efficiencies of common DNA hydrolysis methods to digest DNA which contained the epigenetically relevant dC modifications.[40] According to the study, the chosen DNA digestion protocol influenced the abundance of the nucleoside of interest during the analysis and quantification. When measuring mdC, the most common, one-step DNA hydrolysis methods gave the results closest to the theoretically expected value. For hmdC, the most suitable method was an enzymatic digestion using Nucleoside Digestion Mix (New England BioLabs, NEB) giving an approximately 20% higher signal intensity compared to other evaluated digestion methods. With respect to fdC, the lowest signal was detected after Degradase (Zymo) digestion, whereas other digestion methods resulted in similar signals. The biggest deviation between the digestion methods was found in case of cadC. Using the NEB to digest DNA showed the highest cadC level. Experiments using a digestion method with snake venom phosphodiesterase I (PDE1) and a PDE1/Benzonase mixture (Merck), which are widely used to hydrolyze DNA samples,[14,37] showed that only 6.2% and 6.8% of cadC, respectively, were digested completely to single nucleosides. By choosing the correct digestion protocol, this

study claims to detect an over 10-fold higher amount of cadC than reported before. Based on the published data on cadC, and the particularly special characteristics of this modification that are making it a challenging compound to study, it could be argued that the actual amount of cadC in the genomic DNA might still be underestimated.

To develop the most suitable digestion method, mixtures of 2'-fluorinated standard nucleosides were treated with Nucleoside Digestion Mix (NEB), Degradase (Zymo research)/Benzonase (Merck, *Serratia marcenscens*)[37] and Degradase-only digestion mixtures. Isotopically labelled, non-fluorinated $^{15}N_2$-cadC and isotopically labelled, fluorinated $^{15}N_2$-F-cadC nucleosides were compared and, surprisingly, the fluorinated compounds showed remarkably lower signals compared to untreated reference samples (Figure 19). Non-fluorinated $^{15}N_2$-cadC showed 33% of the theoretically expected signal when using NEB digestion, while in case of F-cadC only 2% compared to the reference were detected. The best digestion method for other compounds that were investigated was the Degradase/Benzonase treatment, which still resulted in considerably decreased signal intensity for $^{15}N_2$-F-dC at 30 % and $^{15}N_2$-F-mdC at 60 % of the theoretically expected signal at best.

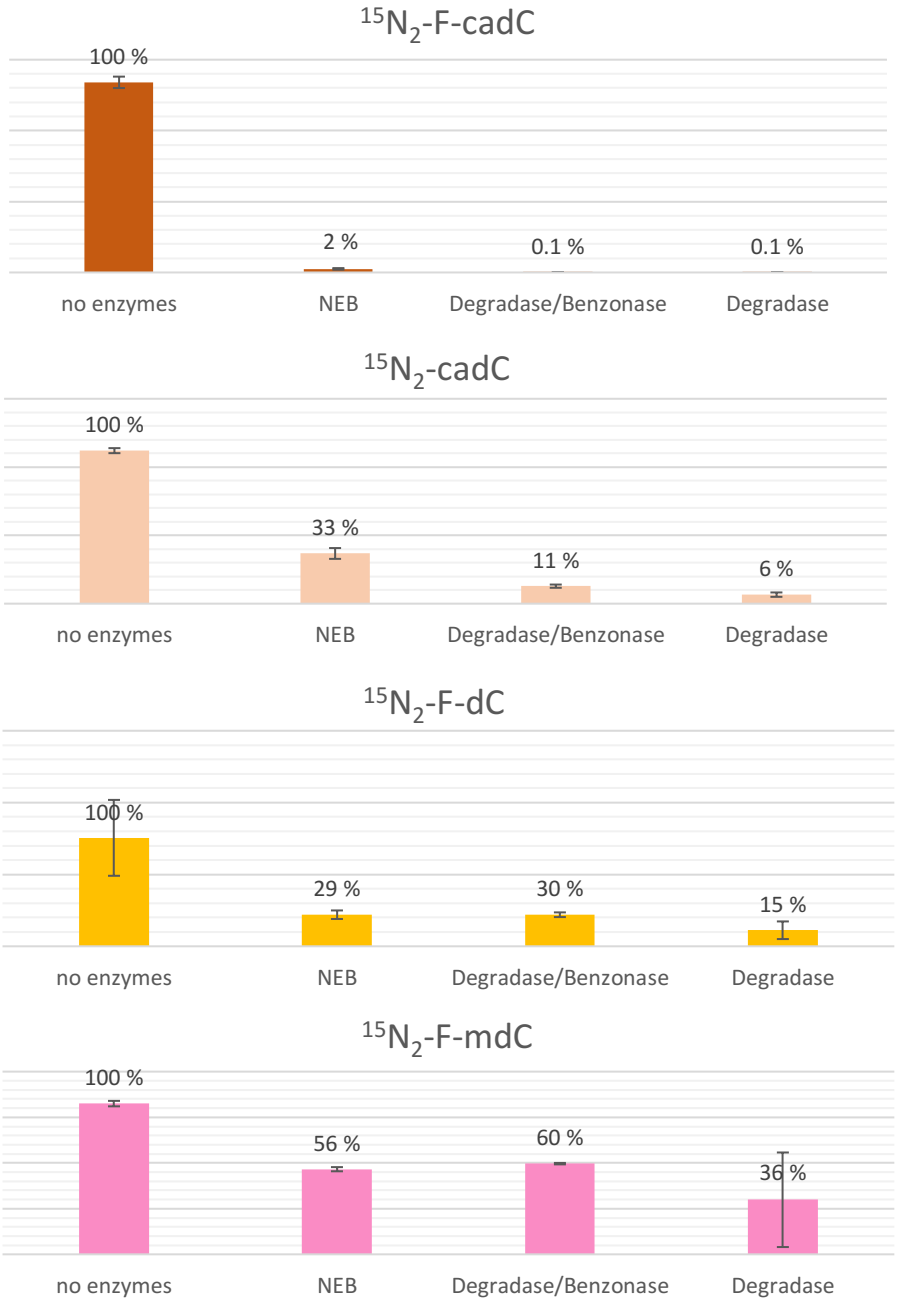


Figure 19. $^{15}N_2$-cadC and $^{15}N_2$-F-standards after different digestion methods.

Also, it was considered that due to the strong ion suppression properties of F-cadC, the treatment with digestion enzymes might influence its ionization. This could result in potentially trapping the molecule or in shifting the retention times during UHPLC and, therefore, cause loss of the signal. When a mass spectrometric UHPLC/MS-MS method to measure a certain molecule is developed, the compound to be studied is highly purified, dissolved in H_2O and analyzed. Based on that, the special nature of F-cadC and potential

changes in its ionization caused by DNA digestion treatment were not taken into account during method development. Further UHPLC studies were conducted and indeed a shift in the elution time as well as strong ion suppression of both the labelled standard compound and the compounds obtained from cell feeding experiments were found. Figure 20 A shows, in the first three chromatograms from the top, in case of pure $^{15}N_2$-F-cadC in water, a retention time of the compound of 3.9 min. The next three chromatograms show $^{15}N_2$-F-cadC treated with NEB digestion enzymes and a shift in retention time is already visible. The lowest three chromatograms show a mixture of $^{15}N_2$-F-cadC that was digested with the degradase/benzonase digestion mixture, and the retention time of F-cadC is shifted to 2.9 min. Based on an adjusted measurement time segment in UHPLC/MS-MS, the measured signal for $^{15}N_2$-F-cadC was 37% with NEB and 20 % with degradase/benzonase compared to the untreated reference (Figure 20A).

In further experiments, it could be shown that addition of 1% formic acid to the measurement buffer restored the ionization state of $^{15}N_2$-F-cadC, and that the compound eluted in the usual time frame (except for a measurement error in the 6th chromatogram from the top, Figure 20B) and gave a slightly improved signal with 50% compared to the reference after NEB digestion and 37% compared to the reference after degradase/benzonase treatment (Figure 21A), while no remarkable increase in the signals of $^{15}N_2$-F-dC (Figure 21B) or $^{15}N_2$-F-fdC was observed (Figure 21C). However, due to the sensitive nature of the DNA samples, formic acid treatment in such high amount was not a preferable option in order to avoid other potential side reactions. Nevertheless, for the further experiments the adjusted measurement time frame to detect F-cadC was used.

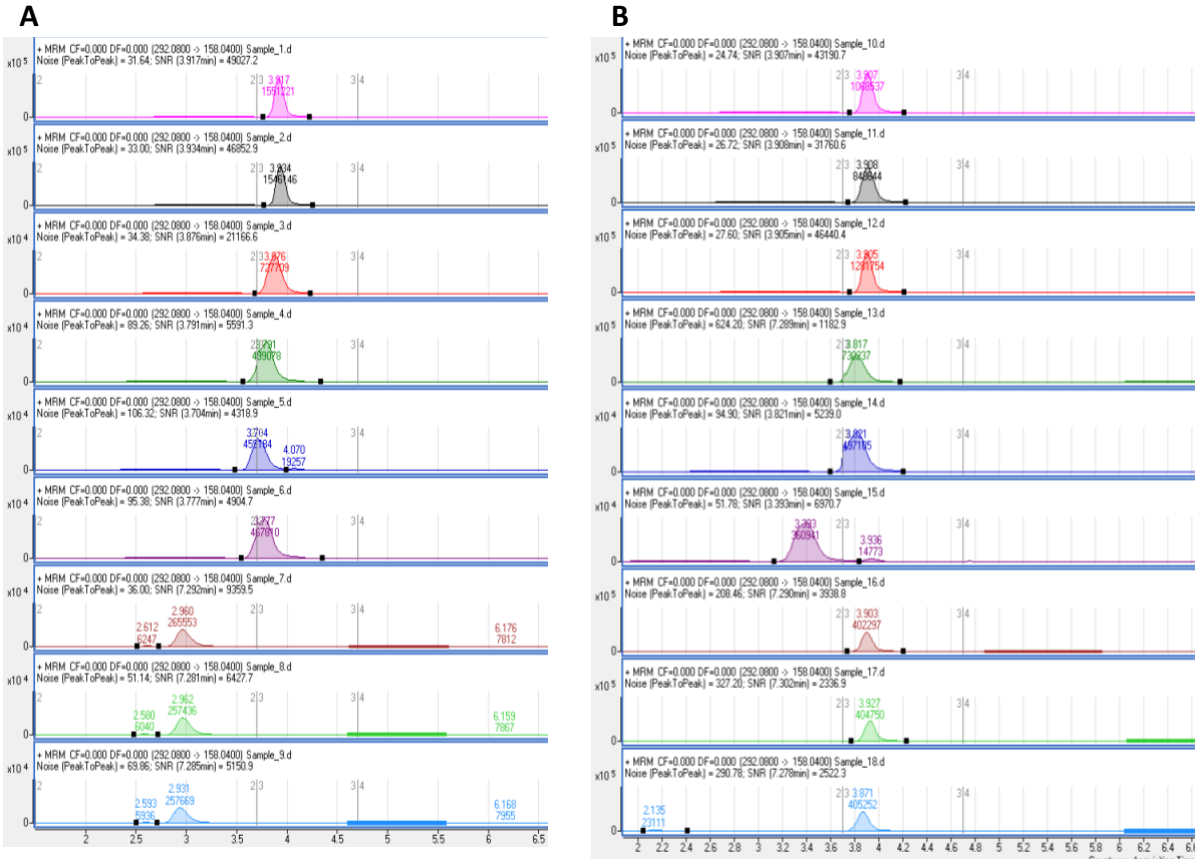


Figure 20. A) time segment for F-cadC adjusted and B) addition of FA before filtration.

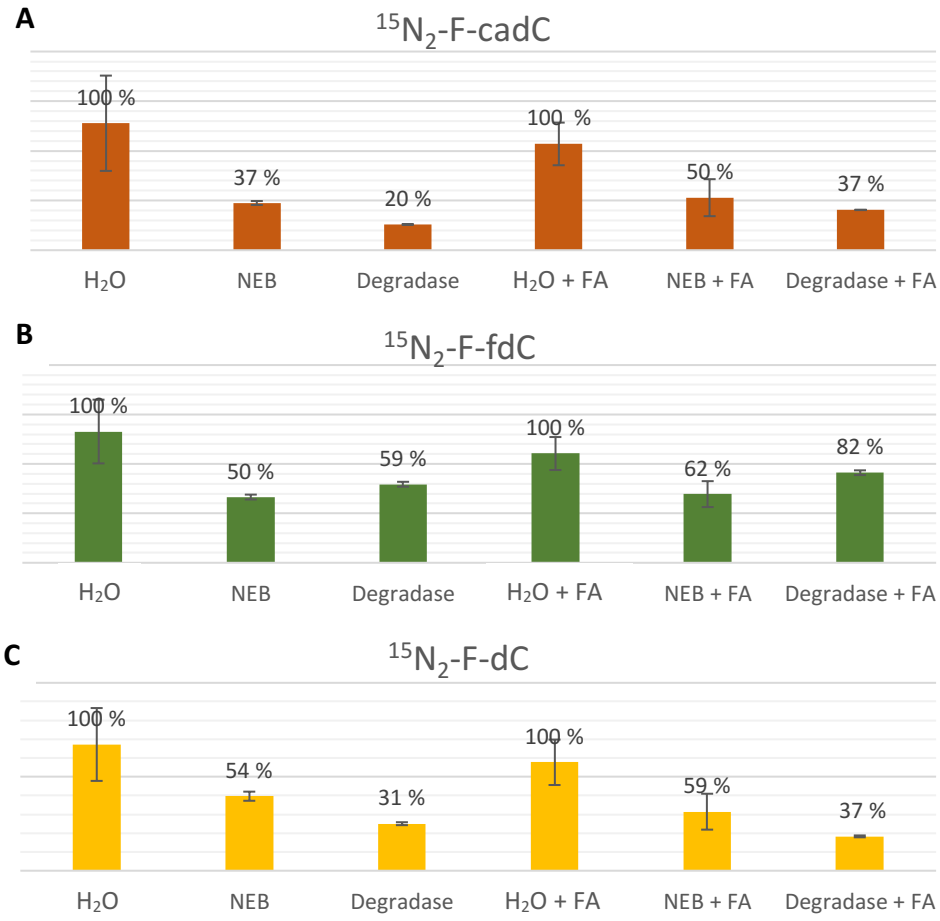


Figure 21. $^{15}N_2$-F-standards after transition shift adjustment versus after formic acid addition.

To further evaluate a preferential enzymatic digestion protocol, experiments using a synthetic F-cadC-containing 18-mer single stranded oligonucleotide were conducted. Thereby, enzymatic digestion of the strand into individual nucleosides was mimicking the reaction that occurs during treatment of the actual DNA. 10 µg of the oligonucleotide were digested with NEB or degradase/benzonase and, surprisingly, degradase/benzonase digestion gave better results compared to NEB digestion unlike in the studies investigating free nucleosides (Figure 22A). Digestion of 1 µg of the oligonucleotide decreased the performance difference, yet degradase/benzonase still gave better signals (Figure 22 B). However, a minor decarboxylation side reaction, apparently caused by the enzymatic treatment, was observed to be significantly more likely during degradase/benzonase digestion. To achieve the goal to detect F-cadC in the gDNA, degradase/benzonase digestion was chosen as the preferred method to be used based

on its superior resulting signal intensity, and the decarboxylation side reaction was taken into account in the analysis.

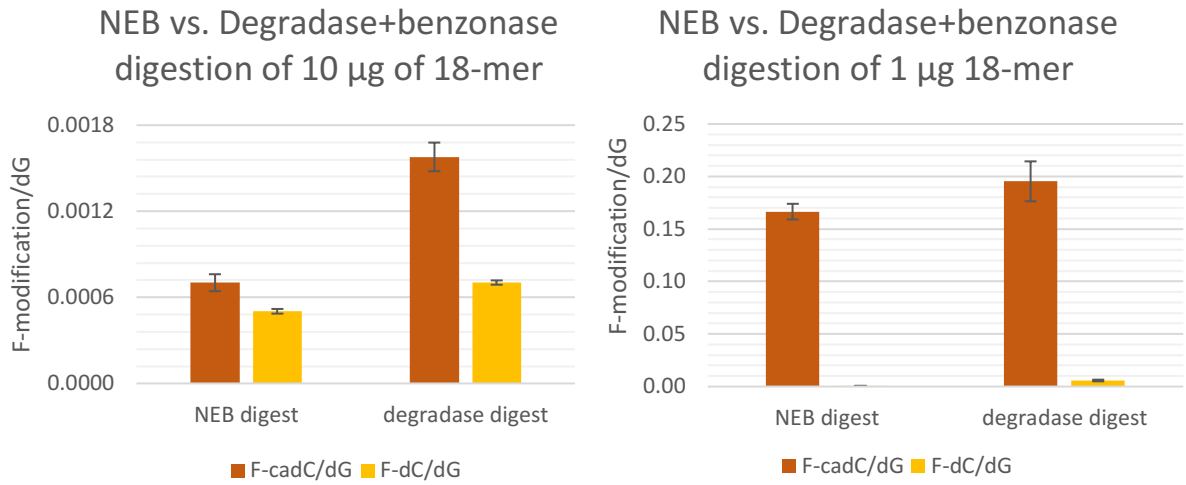


Figure 22. NEB vs Degradase/benzonase digestion of 10 µg of F-cadC containing 18mer oligonucleotide.

In the next step, 10 µg F-cadCTP together with the above-mentioned triphosphate transporter were fed to N2a cells and incubated for 3d. After gDNA isolation, the samples were digested with degradase/benzonase and measured without filtration. The same was done for a cell feeding experiment using F-fdCTP. Unfortunately, F-cadC was not detected in the DNA of the 1st biological replicate of the F-cadCTP feedings, but F-fdC was visible (Figure 23). The gDNA of a second biological replicate was digested with NEB and analyzed without prior filtration and, surprisingly, traces of F-cadC were detected, not in a quantifiable amount, but indicating its presence. However, in case of the F-fdCTP feeding experiments, the amount of F-fdC that was detected in the respective 2nd BR was under the limit of quantification (Figure 24). F-dC as well as F-mdC were detected in a quantifiable amount in both experiments. Additionally, F-cadC seemed to be decarboxylated and remethylated more efficiently than F-fdC was demethylated and remethylated.

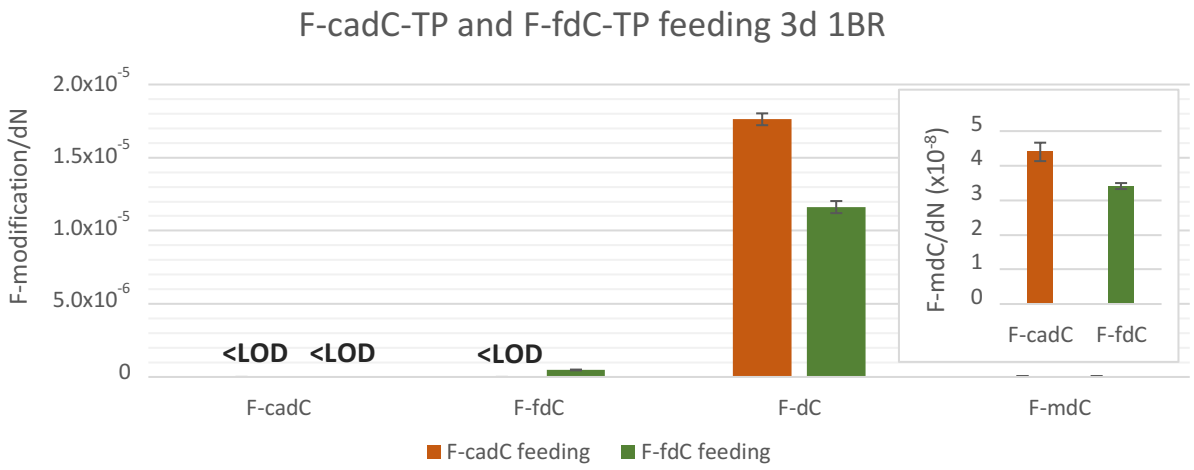


Figure 23. F-cadCTP and F-fdCTP feeding comparison 1st biological replicate.

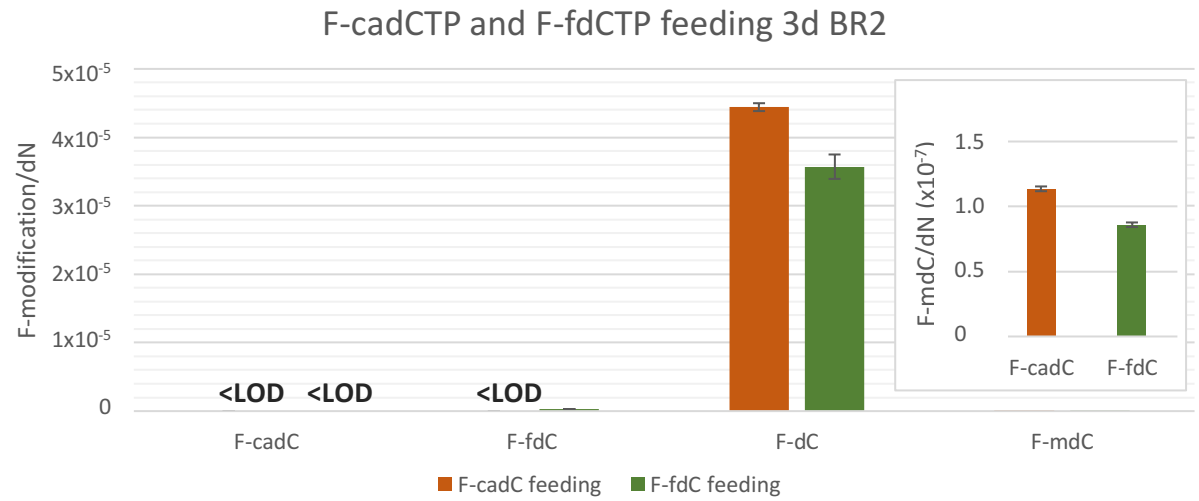
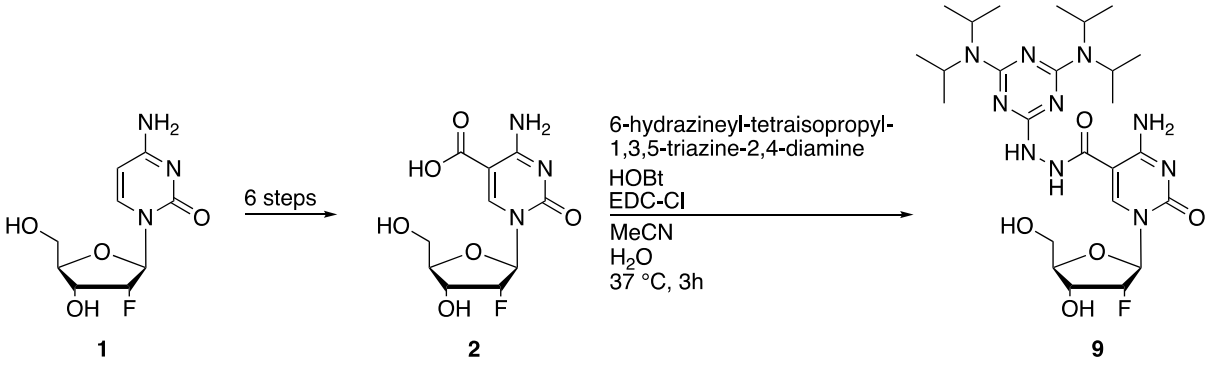


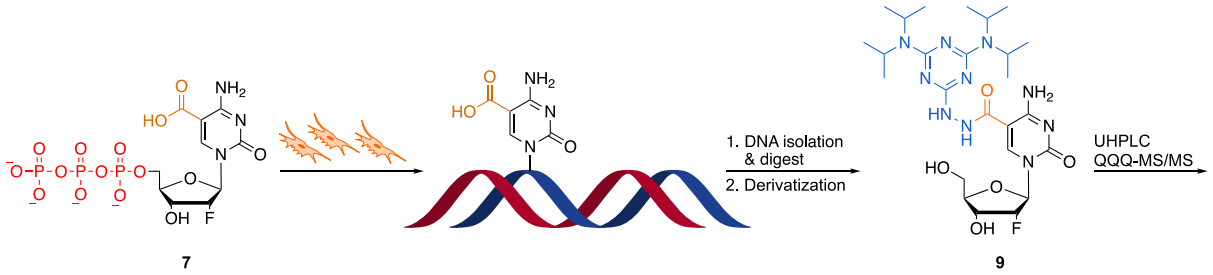
Figure 24. F-cadCTP and F-fdCTP feeding comparison 2nd biological replicate.

A recent publication showed ultrasensitive detection of modified cytidines bases by a chemical derivatization approach of fdC and cadC.[25] Mass-spectrometric detection of the derivatized compounds was claimed to enhance the detection of the compounds 125-fold and to have a limit of detection as low as 10 amol. In order to improve the detection efficiency in our experiments, the reported derivatization reagent was synthesized[84] and used for the synthesis of **9** (Scheme 7). Subsequently a method to derivatize the gDNA samples was developed.



Scheme 7. Synthesis of derivatized F-cadC for UHPLC-MS-MS method development

After F-cadCTP cell feeding experiments, the gDNA was isolated, digested, then the determined derivatization conditions were applied, and the product mixture was measured with UHPLC-MS/MS (Scheme 8). However, the lowest limit of detection achieved with this protocol was determined to be 29 pmol which was not as good as previously determined limits of detection being 0.2 fmol for F-cadC and 84 fmol for F-cadC. Experiments to derivatize $^{15}N_2$-F-cadC were performed separately in water, degradase/benzonase mix and in NEB digestion mix, but after MS-analysis no product was visible. The same was done for non-fluorinated $^{15}N_2$-cadC and only a small conversion was detected when using the NEB digestion mixture. In conclusion, the derivatization attempts of F-cadC in the digested DNA samples did not proceed to full conversion and, furthermore, even showed traces of decomposition. Thereby this method was determined to not be suitable for further studies.



Scheme 8. General workflow of F-cadCTP feeding and derivatization.

In another approach, F-cadC was exposed to N2a cells for different durations of time to determine the optimal time point where the F-cadC would be present. 100 µM F-cadCTP together with the triphosphate transporter were fed, samples were collected after 1h, 2h, 4h, 8h, 24h and 3d and investigated by mass spectrometric analysis (Figure 25). The highest levels

of F-dC were found at 24h time point for both F-cadCTP and F-fdCTP feedings (Figure 26). In a similar experimental setup, F-cadCTP and F-fdCTP were used and results obtained after 24h and 3d of incubation time showed that F-cadC indeed decarboxylates more efficiently. In Figure 26 the lighter orange and lighter green indicate the corresponding decarboxylation and deformylation product F-dC.

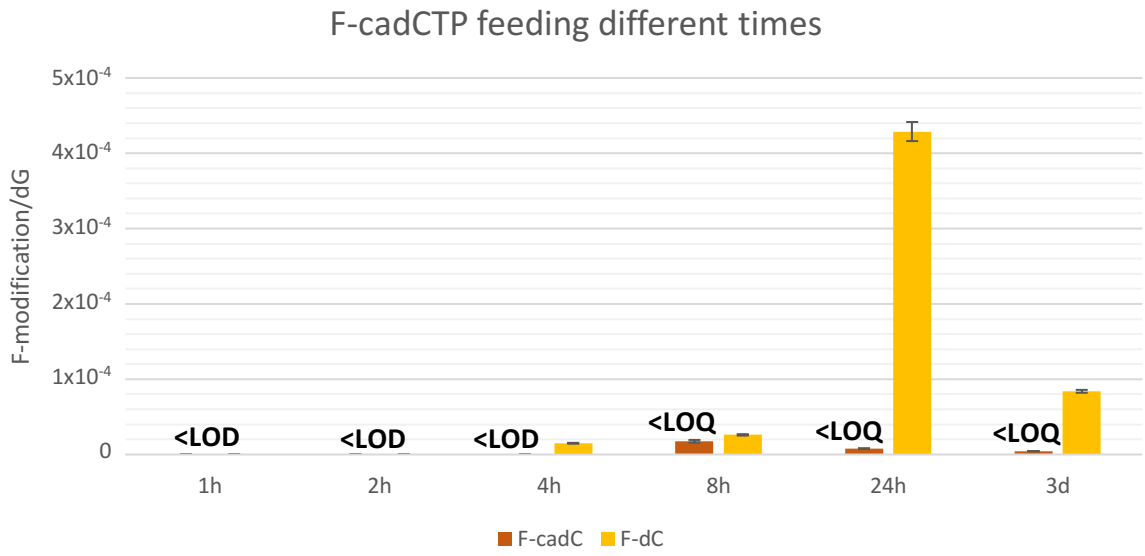


Figure 25. F-cadCTP feeding of different timepoints.

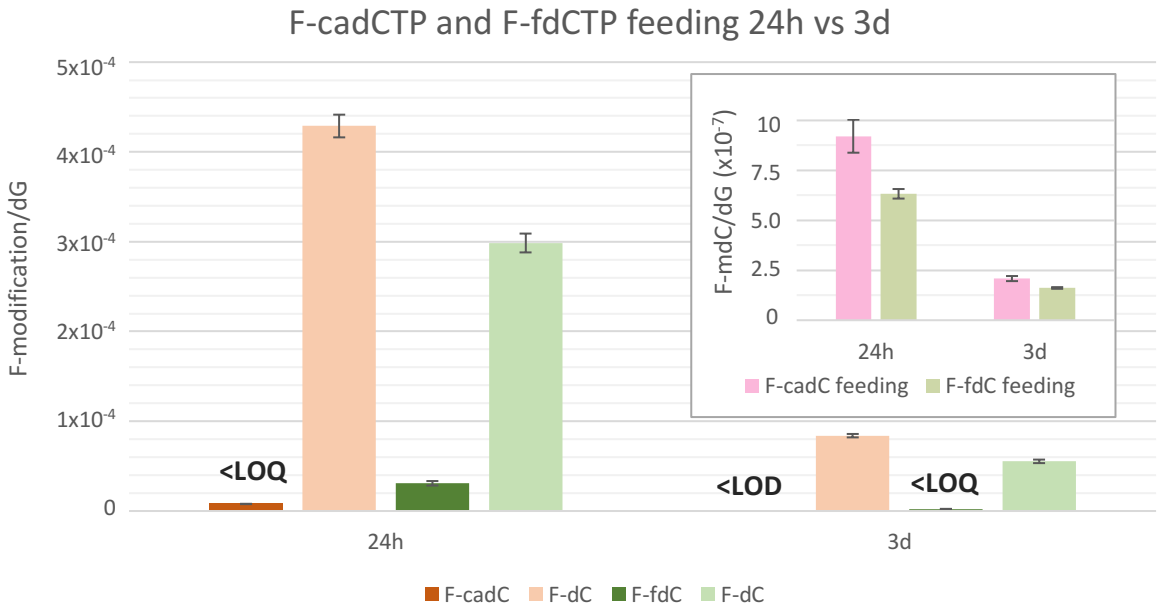


Figure 26. F-cadCTP and F-fdCTP feeding 24h versus 3d comparison.

Furthermore, it was studied whether F-cadC would incorporate better in different cell lines. The chosen cell lines were, besides already mentioned N2a cells, the CHO-K1 cell line as well as mouse embryonic stem cells J1. The robust N2a cells were already shown to be suitable for studying deformylation levels, and thus were expected to be also suitable for decarboxylation studies. However, in the recently published experiments stem cells showed ten times higher natural decarboxylation levels than somatic cells.[85] That is especially understandable for cells presenting post-implantation embryos due to the epigenetic reprogramming during cell lineage differentiation.[86]

Figure 27 shows the results of two 100 µM and 70 µM feedings of F-cadCTP for 24h incubation time to J1, CHO-K and N2a cells. In the first experiment J1 cells did not yield quantifiable amounts of either F-cadC or F-dC. CHO-K cells showed high levels of demethylation product F-dC but an unquantifiable amount of F-cadC. The results from N2a cells followed the same trend as the previous experiments. The 100 µM feeding, however, showed J1 cell to have the most efficient decarboxylation activity whereas, CHO-K cells did not result in quantifiable data at all. N2a cells behaved as previously shown. These results demonstrate the requirement of further experiments to adjust the cell stage and the feeding conditions suitable for stem cells. Irrespective thereof, N2a cells showed the most stable behavior amongst the investigated cell lines but only a yet unquantifiable amount of F-cadC was visible. Based on these experiments and previous results, N2a cells were again determined to be the most suitable cell line for cell feeding studies to investigate decarboxylation of F-cadC.

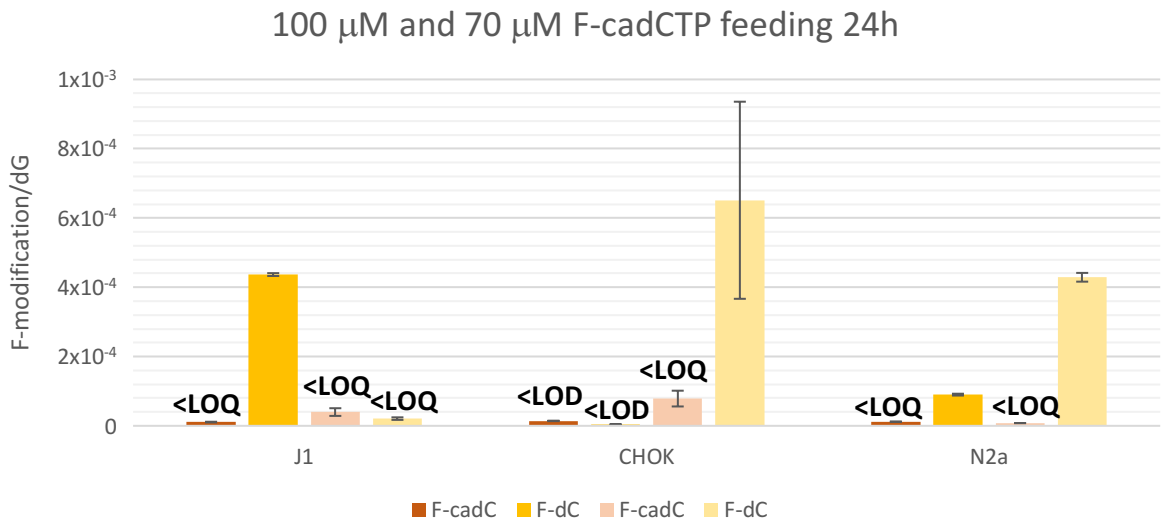


Figure 27. F-cadCTP feeding for different cell lines. Dark orange and dark yellow indicate 100 µM feeding, whereas light orange and light yellow indicate 70 µM feeding.

It was moreover of interest to investigate whether decarboxylation occurs in a dose dependent manner. N2a cells were fed with 10 µM, 50 µM, 100 µM and 400 µM F-cadCTP together with the triphosphate transporter in equimolar concentrations. The formation of decarboxylation product showed an increasing trend, however, the remethylation levels indicated by F-mdC did not increase significantly by the concentration which might be explained by the incorporation of the fed material to non-CpG sites that do not remethylate (Figure 28).

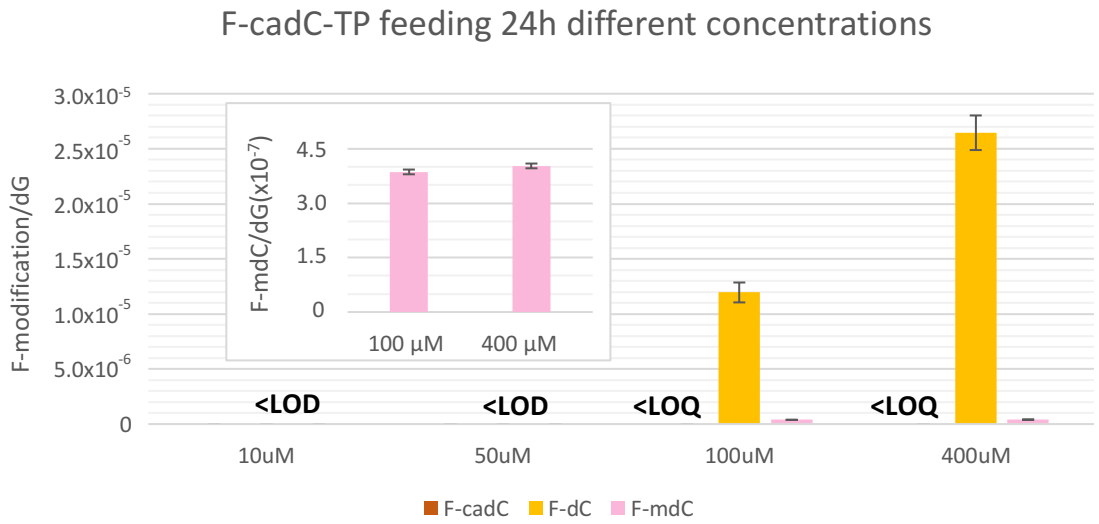


Figure 28. F-cadCTP feeding in different concentrations.

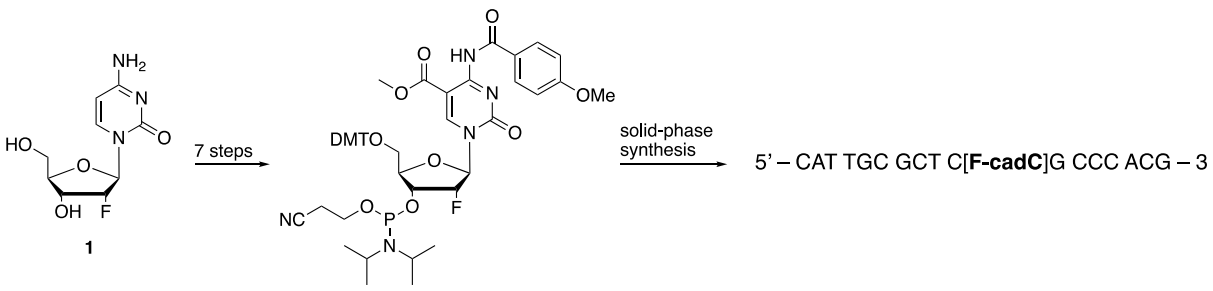
Based on this data in combination with the cellular soluble nucleotide pool analysis of nucleosides and nucleotides it seems plausible that F-cadC, pro-F-cadC and F-cadCTP are stable compounds that do not spontaneously decarboxylate in the cytosol. Furthermore, the results obtained from the cell feeding experiments and MS analysis could indicate decarboxylation of cadC to occur on the level of genomic DNA. Also, the studies show that an elaborate choice of the experimental conditions is crucial for a successful analysis.

5.1.1. Project contributions

Synthesis of the compounds used for cell feeding experiments was done by me. Cell feeding experiments were done by Ammar Ahmedani and Ewelina Kaminska. Eva Korytiakova synthesized the $^{15}N_2$-F-cadC standard and developed a method for mass-spectrometric quantification for this compound. The UHPLC-MS/MS measurements and analysis was conducted by Eva Korytiakova and me.

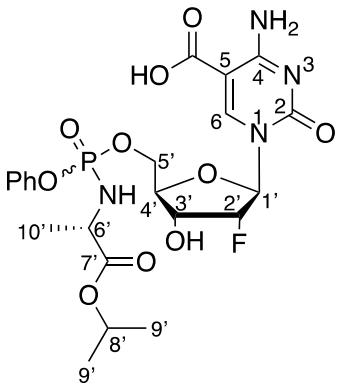
5.1.2. Synthesized compounds and analytics

Compounds **6, 7** and **8** were synthesized according to previously published procedures.[73] The phosphoramidite of F-cadC was synthesized according to published literature.[72] DNA CE-phosphoramidites Bz-dA, Bz-dC, iBu-dG, dT and Bz-mdC were obtained from Glen Research or Link Technologies. The solid-phase synthesis of the 18-mer deoxyribo-oligonucleotide, 5'-CATTGCGCTC[F-cadC]GCCCACG-3' (Scheme 9), HPLC-purification and MALDI/MS quality-control were performed according to previously described protocols.[72]



Scheme 9. Synthesis of F-cadC phosphoramidite and its incorporation into a 2'-deoxyribo-oligonucleotide via solid-phase synthesis.[72]

4-Amino-1-((2*R*,3*R*,4*R*,5*R*)-3-fluoro-4-hydroxy-5-((((((*S*)-1-isopropoxy-1-oxopropan-2-yl) amino)(phenoxy)phosphoryl)oxy)methyl)tetrahydrofuran-2-yl)-2-oxo-1,2-dihydropyrimidine-5-carboxylate (**3**)



1.06 ml *t*BuMgCl (4.4 eq, 1.06 mmol, 124 mg) in THF (1.0 M) was added dropwise to 68 mg F-cadC (1.0 eq, 0.24 mmol) in THF (0.1 M) at -78 °C. The mixture was slowly warmed to 0°C over 30 min and then cooled back to -78 °C. 0.58 ml of Isopropyl (Chloro(phenoxy)phosphoryl)-L-alaninate[80] (2.4 eq. 0.58 mmol, 177 mg) in THF (1.0 M) was added slowly and the reaction mixture was allowed to come to room temperature overnight. The reaction mixture was poured into 20 ml H_2O, THF was evaporated *in vacuo* and the residue was lyophilized. The crude mixture was purified via HPLC (20 to 60 % 0.1 % TFA H_2O/MeCN in 45 min) to obtain **3** as a white powder (5%).

^{1}H NMR (400 MHz, CD_3CN) δ 9.17 (2x s, 1H, H-6), 7.38-7.28 (m, 2H, OPh), 7.18 – 7.10 (m, 3H, OPh), 5.88 (d, *J* = 16.1 Hz, 1H, H-1´), 5.15 (dd, 1H, H-2´), 4.98 – 4.79 (m, 2H, H-3´and H-8´), 4.37 – 4.16 (m, 2H, H-6´), 3.97 – 3.79 (m, 2H, H-4´and H-5´), 3.75 (d, J = 12.6 Hz, 1H, H-5´), 1.23 (dd, *J* = 7.1, 3.3 Hz, 3H, H-10´), 1.12 (t, *J* = 5.8 Hz, 6H, H-9´).

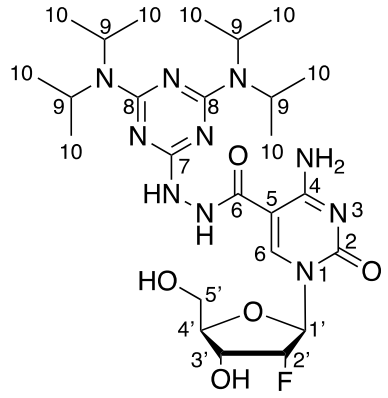
^{13}C NMR (101 MHz, CD_3CN) δ 171.56 (C-7´), 149.60 (O-C_{Ph}), 147.54 (C-4), 144.44 (C-2), 128.46 (C-6), 124.00 (C_{Ph}), 123.60 (C_{Ph}) 119.11 (C-5), 117.00 (C_{Ph}), 111.58 (C_{Ph}), 92.6 (d, $^{1}J_{C\text{-}F}$=184.6 Hz, C-2'), 138.22 (d, $^{2}J_{C\text{-}F}$ = 34.2 Hz, C-1´), 80.65 (C-4´), 70.47 (d, $^{2}J_{C\text{-}F}$ = 15.4 Hz, C-3´), 67.56 (C-8´), 56.64 (C-5´), 50.20 (C-6´), 21.30 (C-10´), 20.20 (C-9´);

^{31}P NMR (162 MHz, CD_3CN) δ 2.44, 1.68.

HRMS (ESI$^+$): m/z value calc. for ion $C_{22}H_{29}FN_4O_{10}P$ $[M+H]^+$=559.1600; found m/z=559.1598, D=0.2 mDa;

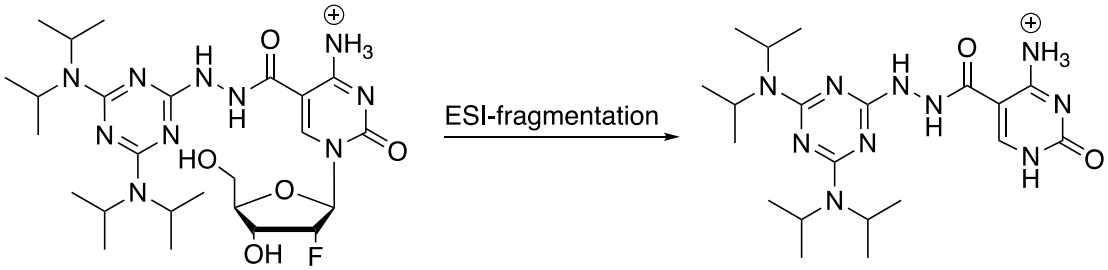
HRMS (ESI$^-$): m/z value calc. for ion $C_{22}H_{27}FN_4O_{10}P$ $[M-H]^-$ =557.1454; found m/z=557.1453, D=0.1mDa

4-Amino-*N*'-(4,6-bis(diisopropylamino)-1,3,5-triazin-2-yl)-1-((2*R*,3*R*,4*R*,5*R*)-3-fluoro-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-2-oxo-1,2-dihydropyrimidine-5-carbohydrazide (**9**)



A solution of 16 mg 6-Hydrazinyl-*N2,N2,N4,N4*-tetraisopropyl-1,3,5-triazin-2,4-diamine (2.0 eq, 51.2 μmol) in 600 μl of MeCN was added to a solution of 2'-F-5-cadC (1.0 eq, 25.6 μmol, 10 mg) in H_2O (300 μl). 7.8 mg HOBt (51.2 μmol, 2.0 eq) in 1:1 H_2O/MeCN (1.8 ml) was added to the nucleoside solution and the mixture was sonicated for 1 min. 25 mg EDC-Cl (128 μmol, 5.0 eq.) in H_2O (300 μl) was added and reaction mixture was incubated at 37 °C for 3h. Reaction mixture was lyophilized and purified via HPLC (50:50 H_2O/MeCN in 45 min) giving **9** 37 % yield.

^{1}H NMR (400 MHz, CD_3CN) δ 8.86 (s, 1H, H-6), 5.86 (d, *J* = 16.2 Hz, 1H, H-1´), 4.98 (dd, *J* = 52.5, 4.1 Hz, 1H, H-2´), 4.24 (ddd, *J* = 25.2, 9.2, 4.1 Hz, 2H, H-3´), 4.06 (dt, *J* = 9.2, 2.4 Hz, 1H, H-4´), 3.98 (dd, *J* = 13.0, 2.1 Hz, 1H, H-5´), 3.77 (dd, *J* = 12.8, 2.6 Hz, 1H, H-5´), 1.21 (d, *J* = 7.8 Hz, 24H, H-10);



QQQ-MS/MS: product ion scan for $C_{25}H_{42}FN_{10}O_5$ $[M+H]^+$ =581 found, after ESI fragmentation precursor ion for 447 found. CE=27, standard nucleoside method[87] with H_2O/MeCN, C8 column, retention time 10 min.

5.2. Design and Synthesis of a cadC analogue to study C-C bond cleavage *in vivo*

Since the enzymatic entity behind the direct DNA demethylation via a C-C-bond cleavage reaction is still unknown, efforts to design probe molecules that would have influence on decarboxylation *in vivo* are being made. In nature, enzymes that catalyze decarboxylation reactions are known. Iso-orotate decarboxylase, for example, has been shown to decarboxylate iso-orotate to uracil.[88] On the other hand, it is also known that 5-NO_2-uracil inhibits decarboxylation of iso-orotate in bacteria.[89]

In the context of this project, 5-NO_2-F-dC that mimics the structure of cadC, was designed as a potentially BER resistant probe molecule to investigate the inhibition of a potential decarboxylation reaction in gDNA (Figure 29). Due to the electron withdrawing nature of the nitro-substituent, the probe molecule should have a strongly activated C6 position and, furthermore, the nitro group is considered to be stable towards potential denitration reactions. In addition, the NO_2-group is part of the sp^2-system of the pyrimidine heterocycle, which could potentially lead the probe molecule to trap the enzyme that attacks at C6 position (Scheme 10).

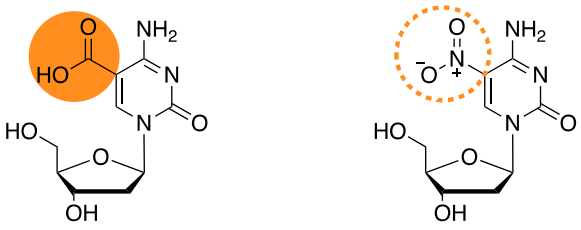
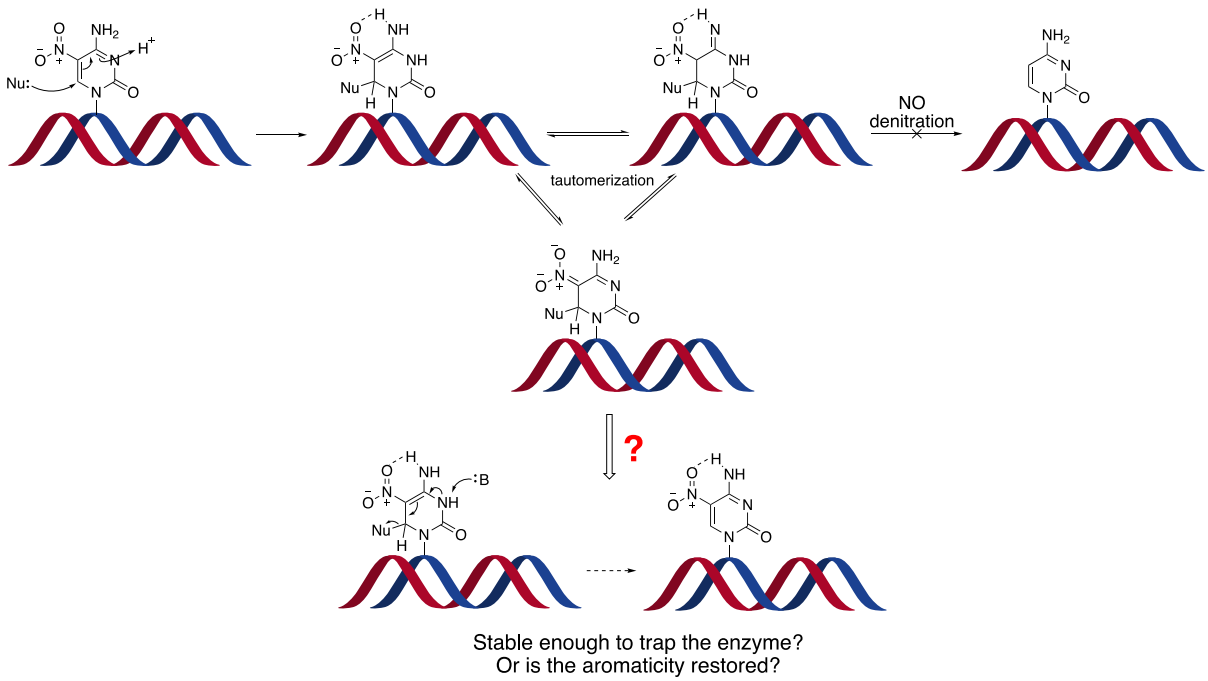
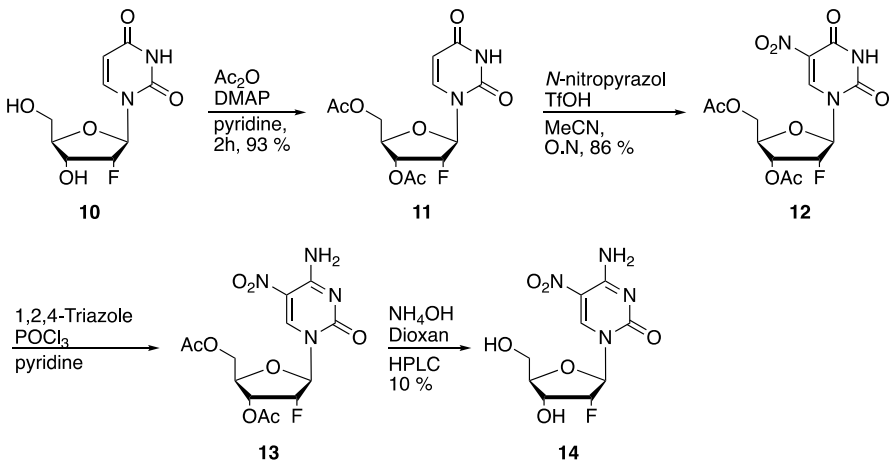


Figure 29. CadC and the potential decarboxylation inhibitor 5-NO_2-dC



Scheme 10. Suggested pathway to trap the potential decarboxylase

The synthesis of 5-NO_2-F-dC (**14**) was carried out as depicted in Scheme 11. The synthesis started with acetyl protection of the 3'- and 5'-hydroxyl groups of **10** followed by an electrophilic aromatic substitution reaction to introduce a nitro group to the C-5 of the pyrimidine ring of **11** using *N*-nitropyrazole in the presence of triflic acid. 5-NO_2-F-dU (**12**) was converted into the corresponding cytidine derivative by a standard amination protocol which in the first step activates the C4-position with 1,2,4-triazole and subsequently uses ammonium hydroxide to form the cytidine derivative. Simultaneously, deprotection from the acetyl protecting groups in the 3'- and 5'-position was achieved. After HPLC purification, the final compound **14** was obtained with a yield of 10 %.



Scheme 11. Synthesis of 5-NO_2-F-dC.

For subsequent metabolic labelling experiments, N2a cells were cultured and fed in separate experiments with 100 μM, 250 μM, 400 μM aqueous solutions of 5-NO_2-F-dC and incubated for 3d. After DNA isolation and DNA digestion, UHPLC-MS/MS was measured and analyzed. However, incorporation of the probe molecule was not detected. On the other hand, in the 400 μM feedings F-dC was surprisingly detected, which might be a product of denitration of 5-NO_2-F-dC, an impurity in the chromatography column or a contamination during the feeding process. Based on the amount of detected unnatural nucleoside, F-dC, in the unfed control sample the results rather indicate the presence of an experimental impurity rather than the detection of a denitration product. To understand whether the feeding of 5-NO_2-F-dC would have an influence on natural cadC and fdC levels, fdC and cadC were also analyzed mass spectrometrically and their levels were compared between fed and unfed cells. Figure 30 shows that in all of the different concentrations, no significant difference in natural cadC levels between the fed and the unfed samples could be detected. However, except at 400 μM concentration, natural fdC values show a slight increase correlating with the increase of 5-NO_2-F-dC concentration.

These results, nevertheless, could indicate an activity of 5-NO_2-F-dC towards inhibition of deformylation, while the detection of cadC requires improving to be able to draw conclusions about decarboxylation.

The analysis of the corresponding cellular nucleotide soluble pool at 400 µM 5-NO_2-F-dC concentration showed that the probe molecule passed through the cellular membrane and also no potential denitration product F-dC was detected (Figure 31). In general, the levels of natural nucleotides in the nucleotide pool range from 1 nmol to 8 nmoles per 10^6 cells,[90] the high amount of 16 nmol of the 5-NO_2-dC found in the soluble nucleotide pool could indicate that the kinases in the cytosol have low affinity towards the probe molecule and, thus, the phosphorylation level of the corresponding triphosphate would be insufficient to detect stable incorporation into the DNA. This might explain the low response of the probe molecule to the cadC and fdC levels. Further studies are, therefore, required especially investigate better transportation methods.

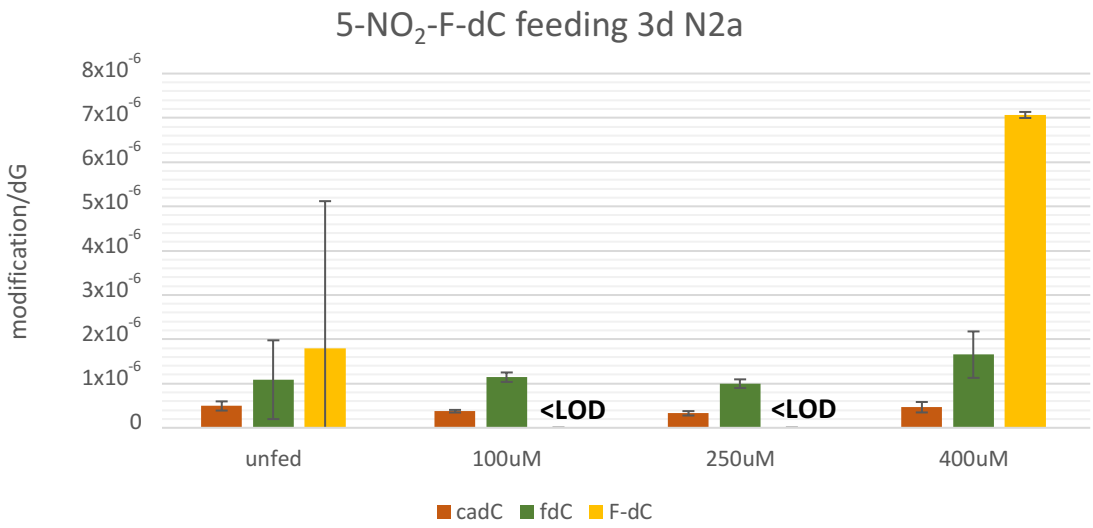


Figure 30. Cell feeding experiments with different concentrations of the probe molecule 5-NO_2-F-dC.

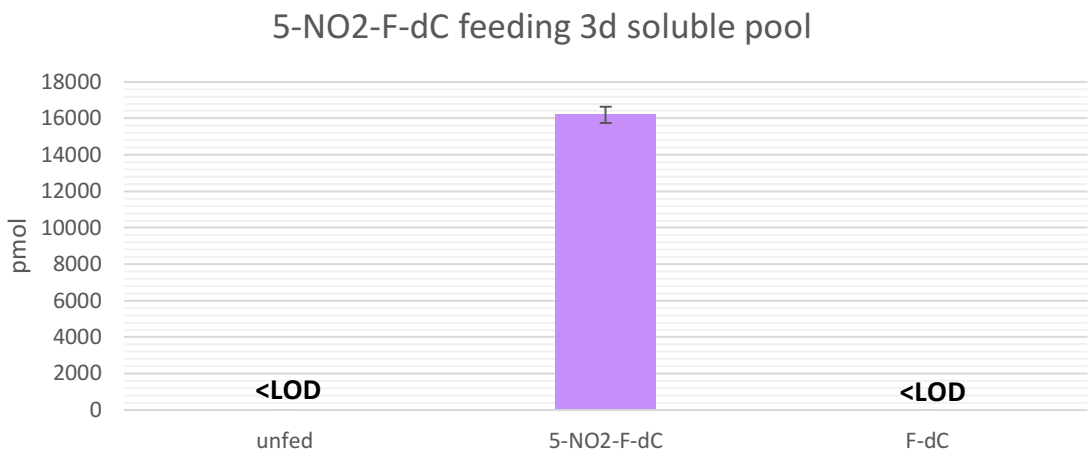
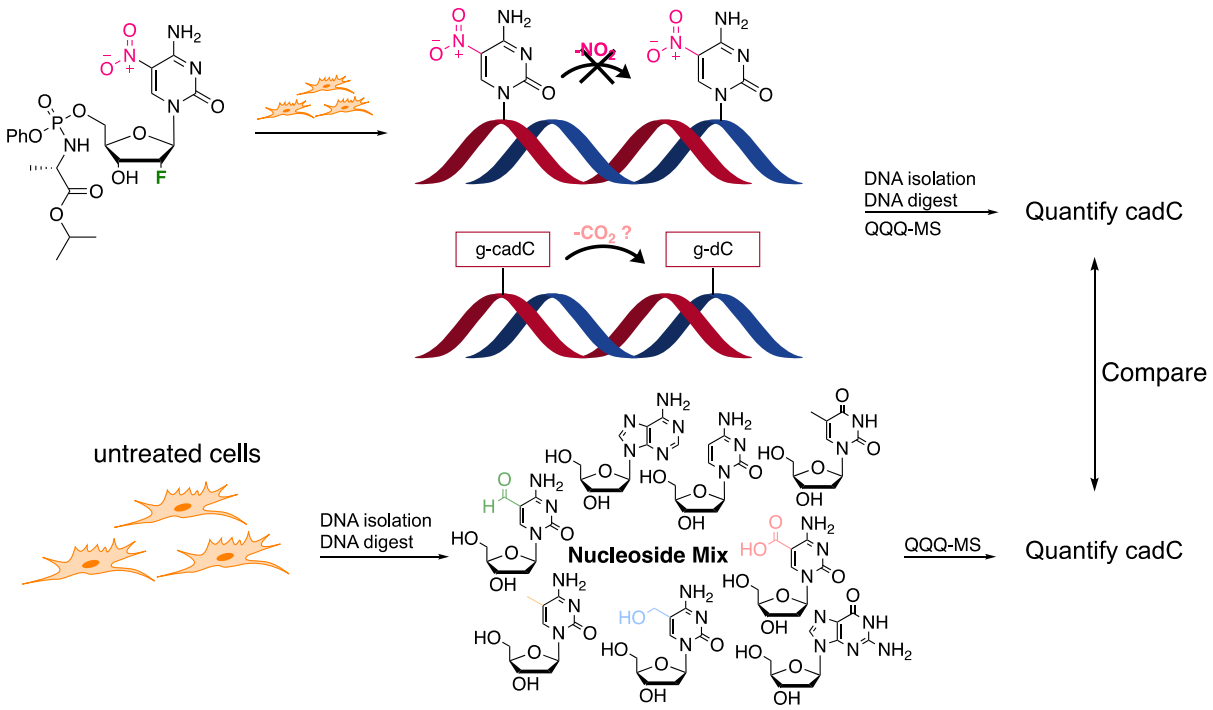


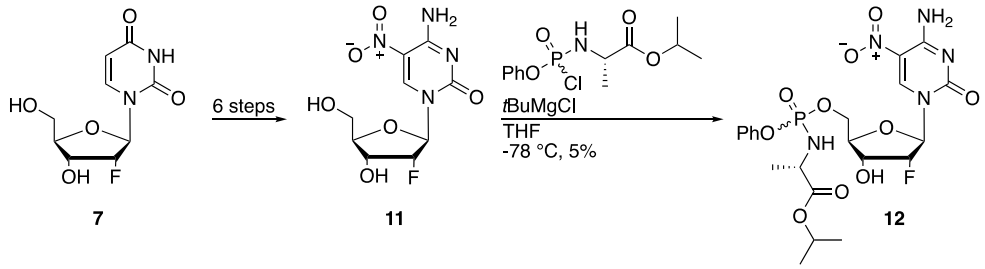
Figure 31. Soluble pool data of 5-NO_2-F-dC feeding.

If the enzymatic phosphorylation of the probe molecule could be circumvented synthetically, potentially more prominent data on cadC and fdC could be obtained. Since the first phosphorylation step of the nucleosides in the cytosol is considered as the most challenging step,[78] another approach utilizing a phosphoroxyarylamidate prodrug (pro-5-NO_2-F-dC) was designed, where 5-NO_2-F-dC was introduced with a masked monophosphate at its 5'-position. The aim was to feed pro-5-NO_2-F-dC and while studying its incorporation, also to quantify the natural genomic cadC levels in fed and unfed samples (Scheme 12). If pro-5-NO_2-F-dC had an eventual inhibitory effect on the decarboxylation reaction of genomic cadC, the observed g-cadC levels should be higher in the pro-5-NO_2-F-dC fed samples than in the unfed samples.



Scheme 12. Pro-5-NO_2-F-dC feeding workflow.

The synthesis of pro-5-NO_2-F-dC started with the synthesis of **11**. HPLC purified **11** was phosphorylated with isopropyl (chloro(phenoxy)phosphoryl)-L-alaninate, (Scheme 13) that was synthesized according to a published protocol.[80] For the phosphorylation, the 5'-hydroxy group of 5-NO_2-F-dC was selectively deprotonated using *tert*-butyl magnesium chloride, and subsequently *iso*-propyl (chloro(phenoxy)phoshoryl)-L-alaninate was added to form the corresponding phosphoroxyarylamidate prodrug (**12**) in a one-pot reaction via a nucleophilic substitution on the phosphorous center.



Scheme 13. Synthesis of pro-NO_2-F-dC.

The HPLC purification of the crude product of **12** was challenging due to the numerous side products and the two diastereomers in the product (Figure 32). Also, the isolated amount of one of the isomers was found to be extremely low. Based on the collected fractions, their ESI-

MS analysis and based on the NMR spectra of the final compound, the final fraction collected was purely one diastereomer with 5 % yield (Figure 33).

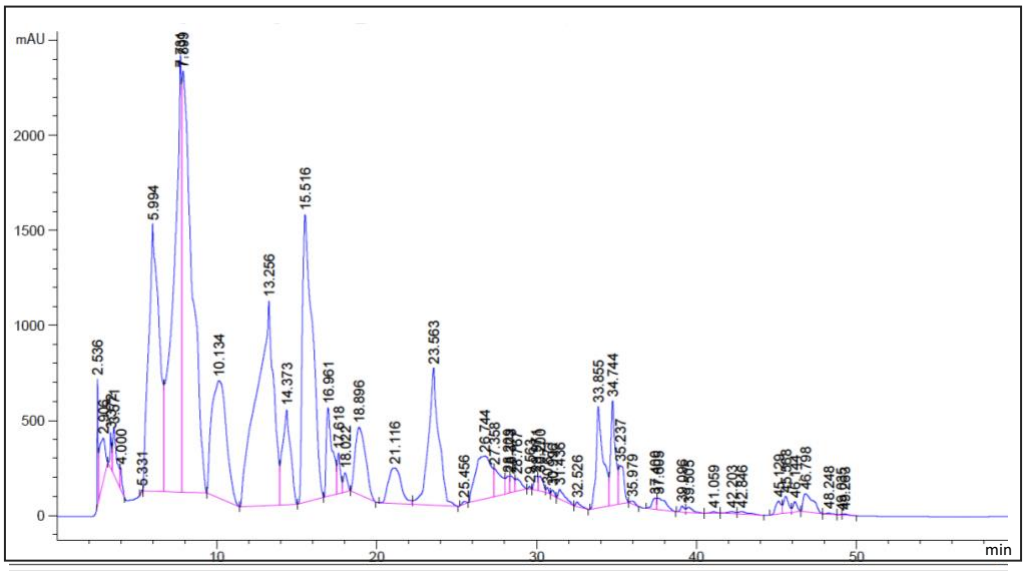


Figure 32. HPLC chromatogram of crude product of **12**.

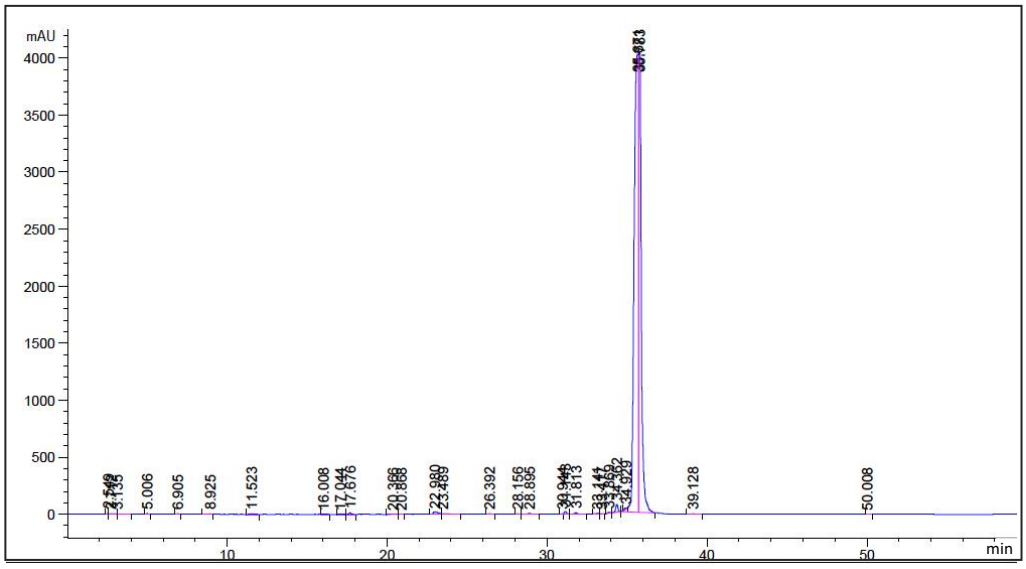
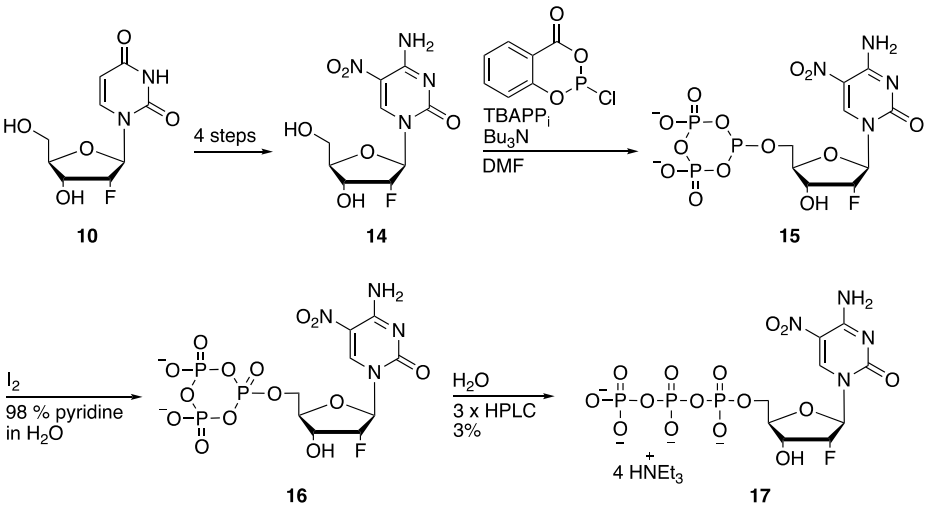


Figure 33. HPLC chromatogram of pure product of **12**.

Another transportation method to achieve incorporation of 5-NO_2-F-dC into the gDNA that was studied, was based on cell feeding experiments using the triphosphate of 5-NO_2-F-dC

which would circumvent the enzymatic phosphorylation steps in the cytosol and could be directly incorporated into the DNA as described in *chapter 5.1.*

To this aim, 5-NO_2-F-dC triphosphate was synthesized modifying the published procedure (Scheme 14).[73] The triphosphate synthesis follows the Ludwig-Eckstein approach as described in *chapter 5.1*. Free nucleoside **14** is phosphorylated at 5´-OH giving a phoshorus III (**15)** species that is oxidized to phosphorus VI (**16**) with I_2. The cyclic structure of **16** is hydrolyzed by addition of water giving a triphosphate **17**.



Scheme 14. Synthesis of 5-NO_2-F-dC-triphosphate.

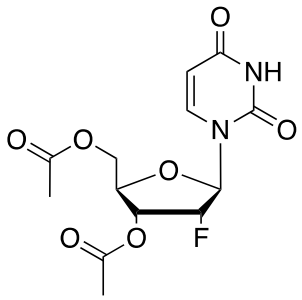
The further cell feeding experiments were delayed due to COVID-19 pandemic and future experiments will be conducted by the Andreas Reichle who will investigate how this compound affects *in vivo* decarboxylation of genomic cadC.

5.2.1. Project contributions

The synthesis of the compounds was done by me and cell feeding experiments were conducted by Ewelina Kaminska. Method development for 5-NO_2-F-dC with UHPLC-MS/MS was done by Dr. Katharina Iwan and the UHPLC-QQQ-MS-MS measurements of the cell feeding experiments were performed by me and Florian Schelter.

5.2.2. Synthesized compounds and analytics

3',5'-Di-*O*-acetyl-2'-fluoro-2'-deoxyuridine (**8**)



To a solution of 2'-deoxy-2'-fluoro-uridine (2.0 g, 8.10 mmol, 1.0 eq.) and DMAP (440 mg, 3.60 mmol, 0.45 eq.) in pyridine (40 ml) Ac_2O (4.4 g, 4.10 ml, 43.0 mmol, 5.3 eq.) was added and the reaction mixture was stirred two hours at room temperature. Volatiles were evaporated and the residue was diluted in H_2O (40 ml) and extracted with EtOAc (3 x 40 ml). Combined organic layers were washed with sat. aq. $NaHCO_3$ (120 ml) and dried over Na_2SO_4. After evaporation of all volatiles, the crude mixture was purified with flash chromatography (8% MeOH/DCM). The acetylated product was obtained as white foam (2.5 g, 7.60 mmol, 93%).

R_f (EtOAc/iHex 1:1)=0.35

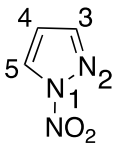
^{1}H NMR ($DMSOd_6$): d 11.49 (s, 1H, N-H), 7.72 (d, 1H, *J*=8.08 Hz, H-6), 5.87 (dd, 1H, ^1J=22.47 Hz, ^2J=1.96 Hz, H-1'), 5.68 (dd, 1H, ^1J=8.03, ^2J=2.20, H-5), 5.53 (ddd, 1H, ^1J=54.58 Hz, ^2J=5.28, ^3J=2.00 Hz, H-2'), 5.26 (ddd, 1H, ^1J=13.28 Hz, ^2J=7.96 Hz, ^3J=2.60 Hz, H-3'), 4.34 (dd, 1H, ^1J=11.95, ^2J=2.86, H-4'), 4.29–4.25 (m, 1H, H-5'), 4.16 (dd, 1H, ^1J=12.08 Hz, ^2J=5.71 Hz, H-5'), 2.11 (s, 3H,CO-CH_3), 2.04 (s, 3H, CO-CH_3)

^{13}C ($CDCl_3$, 77.16 ppm) d 170.33 ($OCOCH_3$), 169.91 ($OCOCH_3$), 162.70 (C-4), 149.65 (C-2), 140.70 (C-6), 103.07 (C-5), 91.33 (d, C-2'), 90.38 (C-1'), 78.79 (C-4'), 69.82 (C-3'), 62.38 (C-5'), 20.87 ($OCOCH_3$), 20.54 ($OCOCH_3$)

HRMS (ESI^+): m/z calc. for ion $C_{13}H_{16}FN_2O_7$ $[M+H]^+$=331.0942; found m/z=331.0936, D=0.6 mDa

HRMS (ESI^-): m/z calc. for ion $C_{13}H_{14}FN_2O_7$ $[M-H]^-$ =329.0785; found m/z=329.0794, D=0.9 mDa.

N-Nitropyrazole[91]



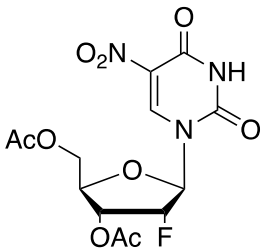
First, fuming HNO_3 (0.90 ml, 22.0 mmol) was added dropwise to vigorously stirred Ac_2O (2.10 ml, 22.0 mmol) at 0 °C. The mixture was allowed to warm to room temperature and after 30 min, the freshly prepared acetyl nitrate (3.00 ml) was added to a solution of pyrazole (500 mg, 7.5 mmol, 1.0 eq.) in glacial AcOH (1.40 ml) and stirred for further 1 h. The reaction mixture was poured into ice cold H_2O (20 ml), filtered and washed with H_2O (30 ml) to yield *N*-nitropyrazole as white crystals (400 mg, 3.57 mmol, 48 %).

R_f (EtOAc/iHex 2:1)=0.77;

^1H NMR (DMSOd6): d 8.79 (dd, 1H, J1=3.00 Hz, J2=0.76 Hz, H-5), 7.87 (d, 1H, J=0.80 Hz, H-4), 6.70 (q, 1H, J=4.68 Hz, H-3);

EIGC: m/z calc. for $C_3H_3N_2O$=83.0245; found m/z=83.0240, D=0.5 mDa

5-nitro-3',5'-di-*O*-acetyl-2'-fluoro-2'-deoxyuridine[92] (**9**)



TfOH (1.0 g, 610 µl, 6.90 mmol, 2.3 eq.) was added to a solution of **8** (1.0 g, 3.00 mmol, 1.0 eq.) and *N*-nitropyrazole (510 mg, 4.50 mmol, 1.5 eq.) in MeCN (30 ml) and stirred at room temperature for 15 h. After evaporation of the volatiles, the residue was diluted in H_2O (30 ml) and extracted with EtOAc (3x40 ml). Combined organic layers were washed with sat. aq. NaCl (120 ml) and dried over Na_2SO_4. After evaporation of volatiles, the crude mixture was

purified with flash chromatography (EtOAc/iHex 1:1 to 2:1) yielding **9** (1.05 g, 2.80 mmol, 93%) as white foam.

R_f (EtOAc/iHex 2:1)=0.59

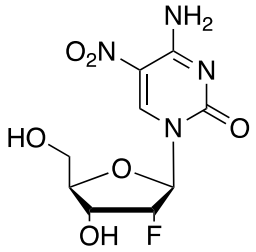
^{1}H NMR (DMSOd_6): d 12.26 (br, 1H, NH), 9.13 (s, 1H, H-6), 6.02 (d, 1H, *J*=20.4 Hz, H-1'), 5.52 (dd, 1H, ^1J=51.9 Hz, ^2J=5.1 Hz, H-2'), 5.24 (ddd, 1H, ^1J=19.6 Hz, ^2J=8.4 Hz, ^3J=5.1 Hz, H-3'), 4.42–4.35 (m, 2H, H-5' + H-4'),4.25 (dd, 1H, ^1J=12.4 Hz, ^2J=4.4 Hz, H-5'), 2.11 (s, 3H, C-OCO-CH_3), 2.04 (s, 3H, C-OCO-CH_3)

^{13}C ($CDCl_3$, 77.16 ppm) d 177.95 ($OCOCH_3$), 170.09 ($OCOCH_3$), 154.73 (C-4), 148.15 (C-2), 144.58 (C-6), 126.10 (C-5), 91.66 (d, C2'), 79.37 (C-1'), 68.71 (C-4'), 61.09 (C-3'), 60.60 (C-5'), 20.69 ($OCOCH_3$), 20.49 97 ($OCOCH_3$)

HRMS (ESI$^+$): m/z value calc. for ion $C_{13}H_{18}FN_4O_9$ $[M+NH4]^+$ =393.3024; found m/z=393.1052, D=0.197 mDa.

HRMS (ESI$^-$): m/z calc. for $C_{13}H_{13}FN_3O_9$ $[M-H]^-$ =374.0636; found m/z=374.0645, D=0.9 mDa

5-nitro-2'-fluoro-2'-deoxycytidine (**11**)



184 mg of 1,2,4-triazole (2.66 mmol, 10.0 eq.) in pyridine (10 ml) was suspended with 83 mg of $POCl_3$ (0.5 mmol, 2.0 eq) giving a white suspension. Subsequently, 100 mg of **9** (0.27 mmol, 1.0 eq.) was added to the suspension giving a bright yellow solution. The reaction mixture was stirred at room temperature for 26h giving an orange solution. 1 ml of conc. NH_4OH (26.0 mmol, 0.91 g, 100 eq.) was added and reaction mixture was stirred at room temperature for 16h giving a red solution. Then, the reaction mixture was poured into sat. aq. NH_4Cl (12 ml), extracted with DCM (2 x 12 ml) and the aqueous layer was lyophilized. The crude mixture was purified via HPLC (0 to 7 % MeCN/H_2O in 60 min) giving **11** (38.5 mg, 0.13 mmol, 10%) as a white powder.

R_f (MeOH/DCM 1:9)=0.2;

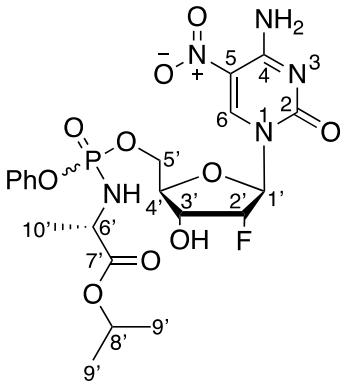
^{1}H NMR (D_2O): δ 9.59 (s, 1H, H-6), 5.96 (d, 1H, *J*=16.4 Hz, H-1'), 5.04 (dd, 1H, ^1J=52.1 Hz, ^2J=4.0 Hz, H-2'), 4.24 (ddd, 1H, ^1J=25.1 Hz, ^2J=9.5 Hz, J3=4.0 Hz, H-3'), 4.15–4.10 (m, 1H, H-5'), 4.02 (dd, 1H, ^1J=13.2 Hz, ^2J=2.5 Hz, H-5'), 3.78 (dd, 1H, ^1J=13.2, ^2J=2.5 Hz, H-4')

^{13}C NMR (101 MHz, D_2O): δ 158.1 (C-4), 153.9 (C2), 147.3 (C-6), 120.5 (C-5), 93.6 (d, $^1J_{C\text{-}F}$=185.7 Hz, C-2'), 89.5 (d, $^2J_{C\text{-}F}$=34.7 Hz, C-1'), 82.4 (C-4'), 66.6 (d, $^2J_{C\text{-}F}$ =16.6 Hz, C3'), 58.1 (C-5')

HRMS (ESI$^+$): m/z value calc. for ion $C_9H_{12}FN_4O_6$ $[M+H]^+$ =291.0741; found m/z=291.0734, D=0.7 mDa

HRMS (ESI$^-$): m/z value calc. for ion $C_9H_{10}FN_4O_6$ $[M-H]^-$ =289.0584; found m/z=289.0592, D=0.8mDa

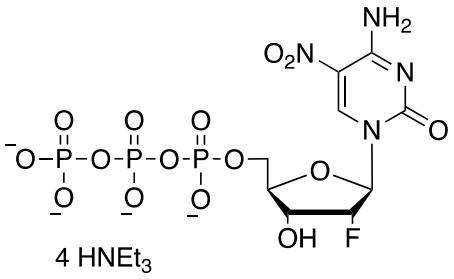
Isopropyl((((2*R*,3*R*,4*R*,5*R*)-5-(4-amino-5-nitro-2-oxopyrimidin-1(2*H*)-yl)-4-fluoro-3-hydroxytetrahydrofuran-2-yl)methoxy)(phenoxy)phosphoryl)-*L*-alaninate (**12**)



62 mg of the nitrated nucleoside derivative **11** (1.0 eq, 0.21 mmol, 0.1 M) were dissolved in THF and cooled to -78 °C. 0.46 ml of *t*BuMgCl (2.2 eq, 1.0 M in THF) were added dropwise, the mixture was warmed to 0°C for 30 min and then again cooled to -78 °C. 0.2 ml of phosphorus reagent isopropyl (chloro(phenoxy)phosphoryl)-L-alaninate (2.2 eq. x mmol, as a 1.0 M solution in THF)[80] were added slowly and the mixture was allowed to come to room temperature overnight. Subsequently, the reaction mixture was poured in 20 ml H_2O, THF was evaporated, and water was lyophilized. The residue was purified via HPLC (20 to 60 % 0.1 % TFA H_2O/MeCN in 45 min) to yield 1.7 mg (3.0 μmol, 1.5%) of **12**.

^{1}H NMR (400 MHz, CD_3CN) δ 9.68 (s, 1H, H-6), 7.97 (s, 1H, NH_2), 7.71 (s, 1H, NH_2), 7.41 – 7.30 (m, 2H, -Ph), 7.27 – 7.16 (m, 3H, -Ph) 5.96 (d, *J* = 16.3 Hz, 1H, H-H1´), 5.21 (dd, *J* = 51.7, 4.0 Hz, 1H, H-2´), 5.07 – 4.88 (m, 2H, H-3´and H-8´), 4.46 – 4.26 (m, 2H, H-6´), 4.06 (dd, *J* = 12.8, 2.2 Hz, 1H, H-4´), 3.96 – 3.82 (m, 2H, H-5´), 1.33 – 1.28 (m, 3H, H-10´), 1.20 (t, *J* = 6.0 Hz, 6H, H-9´)
^{13}C NMR (101 MHz, CD_3CN) δ 173.65 (C-7´), 151.60 (O-C_{Ph}), 151.54 (C-4), 148.43 (C-2), 130.66 (C-6), 126.01 (C_{Ph}), 125.80 (C_{Ph}) 121.13 (C-5), 121.08 (C_{Ph}), 111.78 (C_{Ph}), 94.8 (d, $^{1}J_{C\text{-}F}$=184.6 Hz, C-2'), 90.42 (d, $^{2}J_{C\text{-}F}$ = 34.2 Hz, C-1´), 82.85 (C-4´), 70.67 (d, $^{2}J_{C\text{-}F}$ = 15.4 Hz, C-3´), 69.76 (C-8´), 58.84 (C-5´), 51.40 (C-6´), 21.73 (C-10´), 20.64 (C-9´)
^{31}P NMR (162 MHz, CD_3CN) δ 3.74, 2.88;
HRMS (ESI^+): m/z value calc. for ion $C_{21}H_{28}FN_5O_{10}P$ $[M+H]^+$ =560.1552; found m/z=560.1552
HRMS (ESI^-): m/z value calc. for ion $C_{21}H_{26}FN_5O_{10}P$ $[M-H]^-$ =558.1407; found m/z=558.1413, D=0.6mDa

((2*R*,3*R*,4*R*,5*R*)-5-(4-amino-5-nitro-2-oxopyrimidin-1(2*H*)-yl)-4-fluoro-3-hydroxy-tetrahydrofuran-2-yl)methyl tetrahydrogen triphosphate (**17**)



91.6 mg Tributylammonium pyrophosphate $(HNBu_3)_2PPi$ (55.8 µmol, 1.0 eq) was dried under vacuum overnight and subsequently dissolved in 282 µl DMF. 245 mg of tributylamine (1.32 mmol, 315 µl, 23.7 eq) was added to the solution. The decomposed white layer was scraped off of the surface of 2-chloro-1,3,2-benzodioxaphosphorin-4-one prior the use. 33.9 mg of 2-chloro-1,3,2-benzodioxaphosphorin-4-one (167 µmol, 3.0 eq) in 282 µl DMF was added to $(HNBu_3)_2PPi$-solution and stirred 30 min at room temperature. Solution was cooled to 0 °C and 16.2 mg of nitrated nucleoside derivative **11** (55.8 µmol, 1.0 eq) was added. Reaction mixture turned yellow and stirring was continued at 0 °C for 4h. 80 mM I_2 in pyridine/H_2O (98:2) was added to the reaction mixture until the color change to dark orange was permanent and stirred for additional 1h at room temperature. 1.7 ml ddH_2O (31 ml/mmol) was added and the

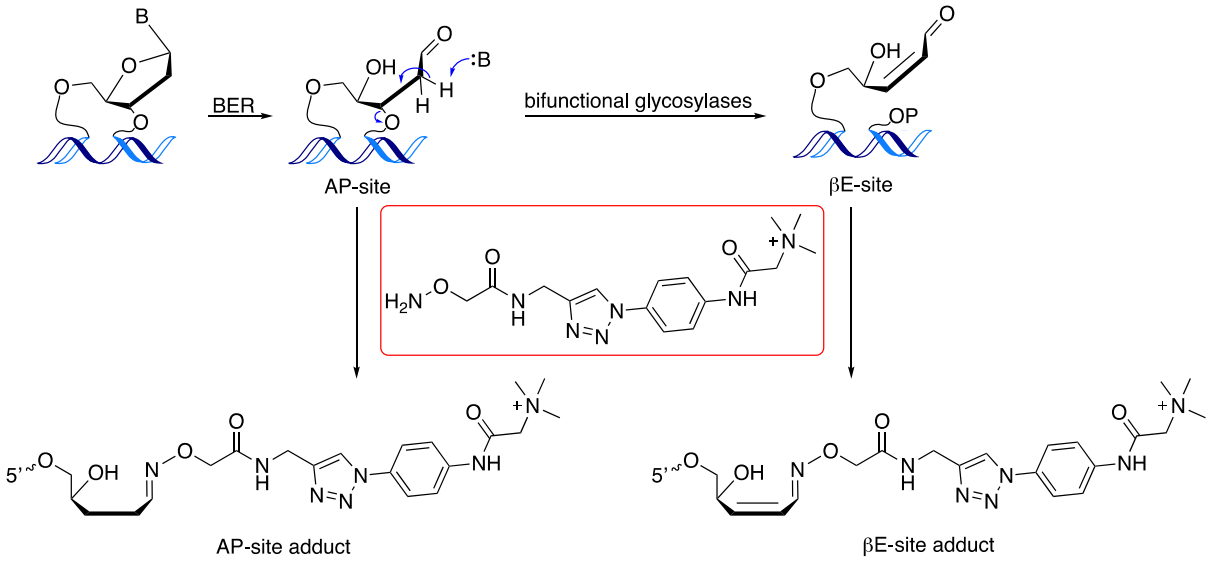
reaction mixture was stirred further 1h at room temperature, while the color changed back to light yellow. 1.04 ml of aq. NaCl (3M, 18.6 ml/mmol) was added, and the mixture was vortexted for 10 seconds. 11.6 ml abs. EtOH (208 ml/mmol) was added and the mixture was placed in -80 °C overnight. Thawed sample was centrifuged 6000 rpm for 10 min and supernatant was removed. The residue was purified via HPLC (0 to 13 % $HNEt_3OAc$ in H_2O/MeCN in 45 min) to yield 1 mg (1.9 μmol, 3%) of **13**.

HRMS (ESI^+): m/z value calc. for ion $C_9H_{15}FN_4O_{15}P_3$ $[M+H]^+$ =530.9725; found m/z=530.2525, D=0.72Da

HRMS (ESI^-): m/z value calc. for ion $C_9H_{13}FN_4O_{15}P_3$ $[M-H]^-$ =528.9580; found m/z=528.9579, D=0.3mDa

5.3. Design and Synthesis of a Novel Reagent for MS-based studies of epigenetically relevant nucleosides

To quantify genomic abasic sites, a hydroxylamine based derivatization reagent[83] has been reported which was designed to specifically react with the abasic sites in the genome of cells. Then UHPLC-QQQ-MS was used to quantify the resulting derivatized abasic sites. The known reagent contains a carbonyl-specific reactive hydroxylamine moiety that is able to form oximes by a reaction with aldehydes or ketones. Thus the reagent can react with the open-chain aldehydic form of both, abasic sites and their β-elimination intermediates, and enables their mass spectrometric quantification (Scheme 15).[83] Due to their different molecular weights, mass-spectrometric analysis is able to distinguish and quantify these reaction products.



Scheme 15. Derivatization of BER products with a known abasic site derivatization reagent.

It was shown that the derivatized BER products did not influence the hydrolytic enzymes acting on gDNA which resulted in quantitative excision of the reaction products from the genome. When utilizing this reagent, not only the signal of the charged derivatized compound is detected, but also a specific mass transition by the fragmentation of the parent ion is observed, where the collision induced dissociation (CID) causes a loss of N_2 of the triazole moiety. In general, the permanent positive charge in the betaine moiety of the derivatization

reagent enhances the detection efficiency during MS. In addition, the positive charge was also found to improve the reaction kinetics of the reagent with AP-sites and β-elimination intermediates by interacting with the negatively charged DNA backbone.

Nevertheless, it was of interest to develop an even improved reagent wherein the sensitivity during the LS-MS analysis was comparably higher. Therefore, a new reagent bearing a tetra-fluorinated π-stacking core was designed. It was considered to provide additional benefits by potentially enhancing the stability of the aryl-triazole bond and by increasing DNA affinity due to the properties of its electronic structure (Figure 34).

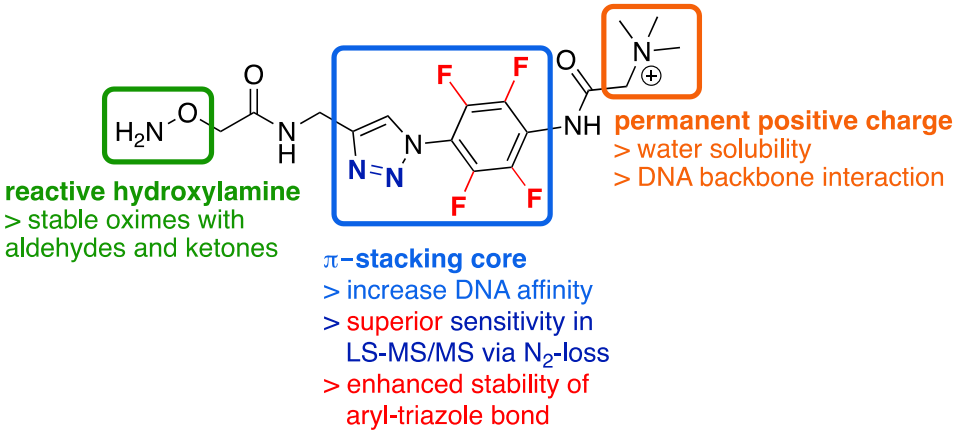
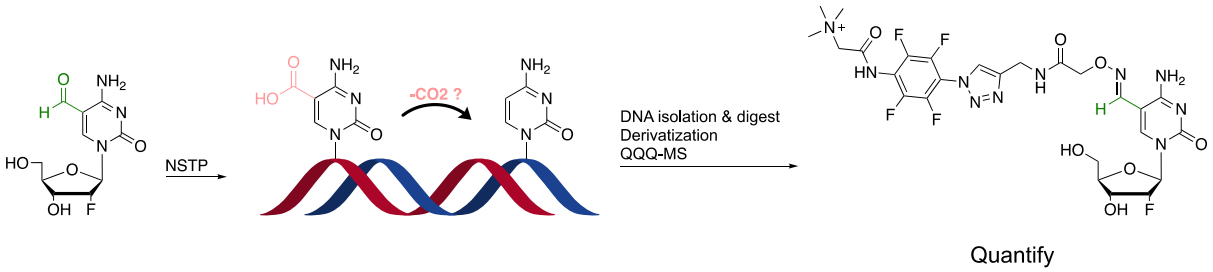


Figure 34. Modified hydroxylamine-based derivatization reagent target molecule.

In the course of this project, this new reagent should be chemically synthesized and it should be investigated whether it could be used as a tool to improve the sensitivity of the LC-MS/MS analysis when quantifying the exemplary target compound F-fdC in gDNA samples obtained from cell feeding experiments using F-fdC (Scheme 16).

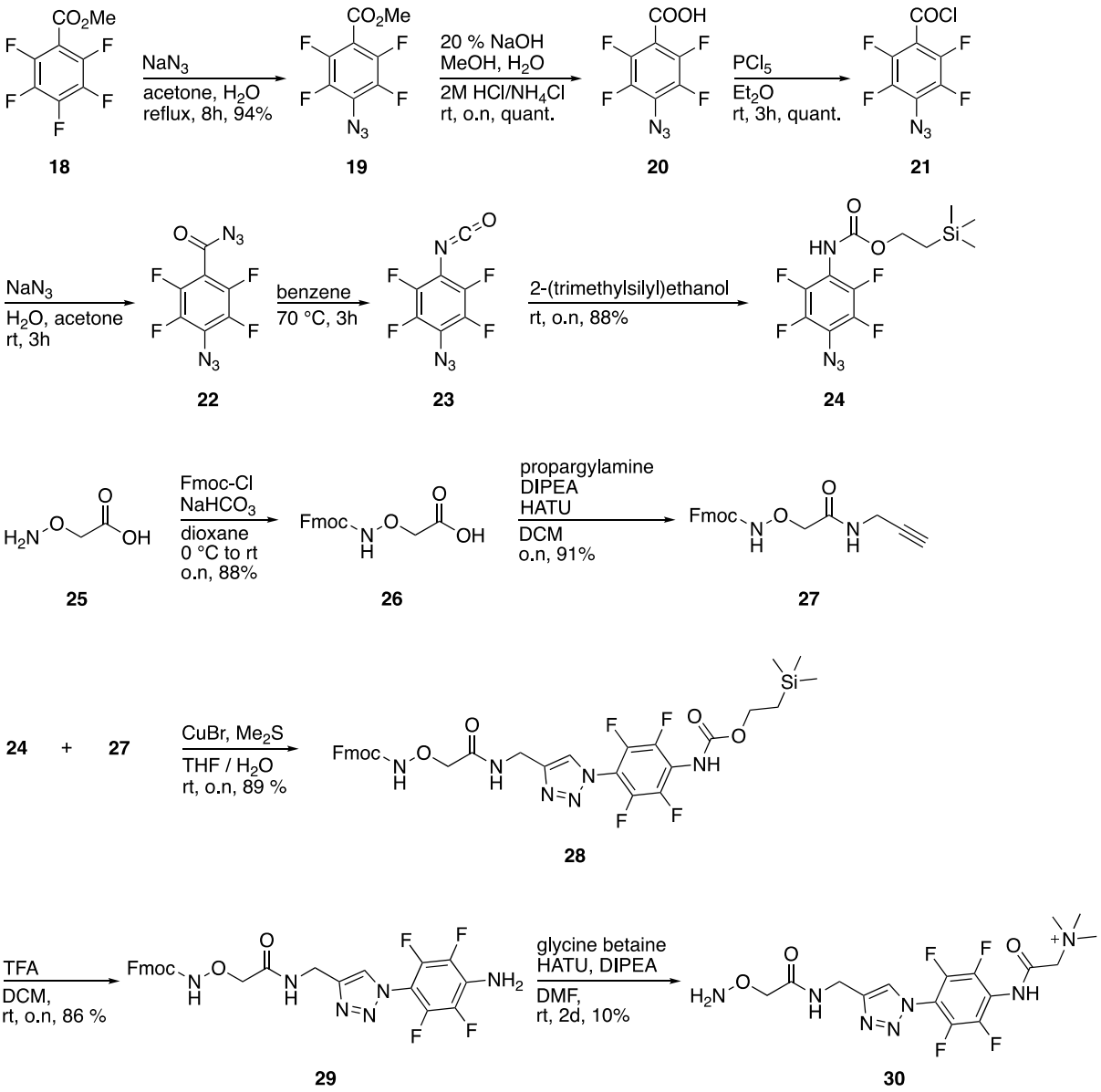


Scheme 16. F-fdC cell feeding and quantification of genomic F-fdC after derivatization.

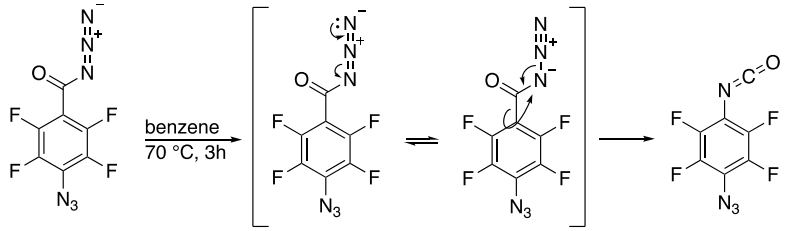
The synthesis of the new derivatization and quantification reagent is depicted in Scheme 17. First, 4-(N-((2-trimethylsilyl)ethoxy)carbonyl)amino)tetrafluorophenyl azide (**19**) was synthesized according to a previously published protocol.[93] The synthesis started with nucleophilic aromatic substitution in the para-position of **18** using NaN_3 to give azide **19** with 94 % yield. Saponification of the ester of **19** using NaOH in aqueous methanol gave **20** with almost quantitative yield. Subsequently, the carboxylic acid **20** was converted to its acid chloride (**21**) with PCl_5 followed by the formation of the acyl azide **22** using NaN_3 in acetone/water. Acyl azide **22** was then further converted into the intermediate isocyanate (**23**) via a modified Curtius rearrangement (mechanism depicted in Scheme 18) followed by an addition reaction using 2-(trimethylsilyl)ethanol resulting in stable carbamate **24**.

The other building block to be used in an alkyne-azide cycloaddition with **24** was synthesized starting from commercially available hydroxylamino acetic acid **25**, which was protected with Fmoc[94] in the first step yielding the Fmoc-protected hydroxylamino acetic acid **26.** The amine-protected compound **26** was subsequently used in an amide-bond forming coupling reaction with propargyl amine in the presence of by HATU to give amide **27**.

Finally, alkyne **24** and azide **27** were reacted via a Cu(I)-catalyzed alkyne-azide-1,3-dipolar cycloaddition to give triazole **28** in 89% yield. Subsequently, the carbamate moiety of **28** was cleaved using TFA in DCM to obtain aryl amine **29** which was then reacted with glycine betaine in a coupling reaction using HATU as the coupling reagent. Simultaneously in the last step, the Fmoc-protecting group was cleaved by addition of DIPEA to the reaction mixture and the final target compound was purified via HPLC using a formic acid buffer system resulting in a 10 % yield of **30**.



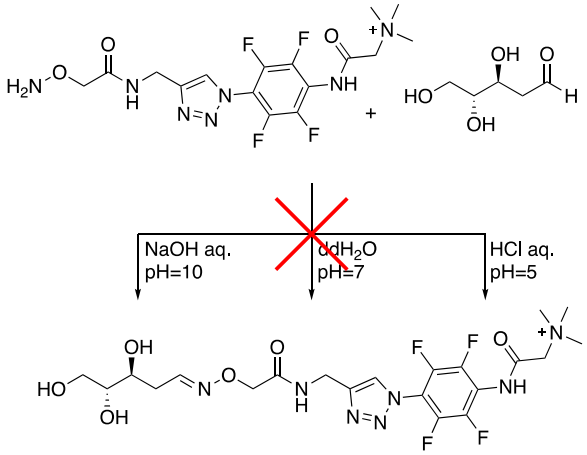
Scheme 17. Synthesis scheme of **29**.



Scheme 18. Reaction mechanism of Curtius rearrangement.

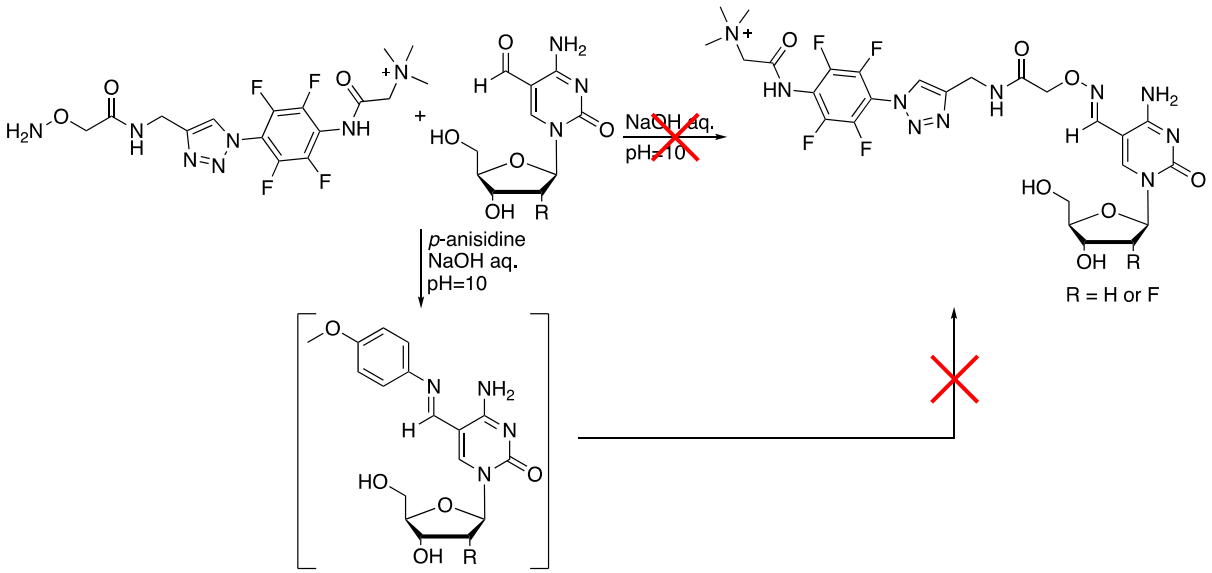
Next, possible reaction conditions to derivatize 2'-deoxyribose with the new reagent were investigated (Scheme 19). Initial reaction conditions applied were an aqueous reaction

solution with basic, neutral, or acidic pH at room temperature. Under none of the applied conditions any conversion was observed. Consequently, also reaction temperatures of 40 °C, 50 °C and 60 °C were applied. However, no reaction occurred also under these conditions.



Scheme 19. Derivatization with studies of 2-deoxyribose with derivatization reagent **30**.

Analogously, experiments to derivatize fdC and F-fdC were performed under basic aqueous conditions and also in presence of *p*-anisidine which was considered to work as a nucleophilic catalyst to form an aromatic Schiff base intermediate. This intermediate would be more reactive than the parent aldehyde towards the nitrogen nucleophile of the hydroxylamine (Scheme 20).[95] Figure 35 shows the formation of anisidine activated fdC and the starting materials fdC and **30**. However again, no conversion was observed, and also reaction temperatures of 40 °C, 50 °C and 60 °C were investigated. Nevertheless, no reaction could be observed in any case. Apparently, the four fluorine substituents of derivatization agent **30** unexpectedly decreased the reactivity of the hydroxylamine moiety of the compound due to their strong electron withdrawing effect so drastically that no reaction with any of the investigated probe molecules occurs. Due to its low reactivity, no further studies with derivatization agent **30** were conducted.



Scheme 20. Derivatization studies of fdC and F-fdC with derivatization reagent **30**.

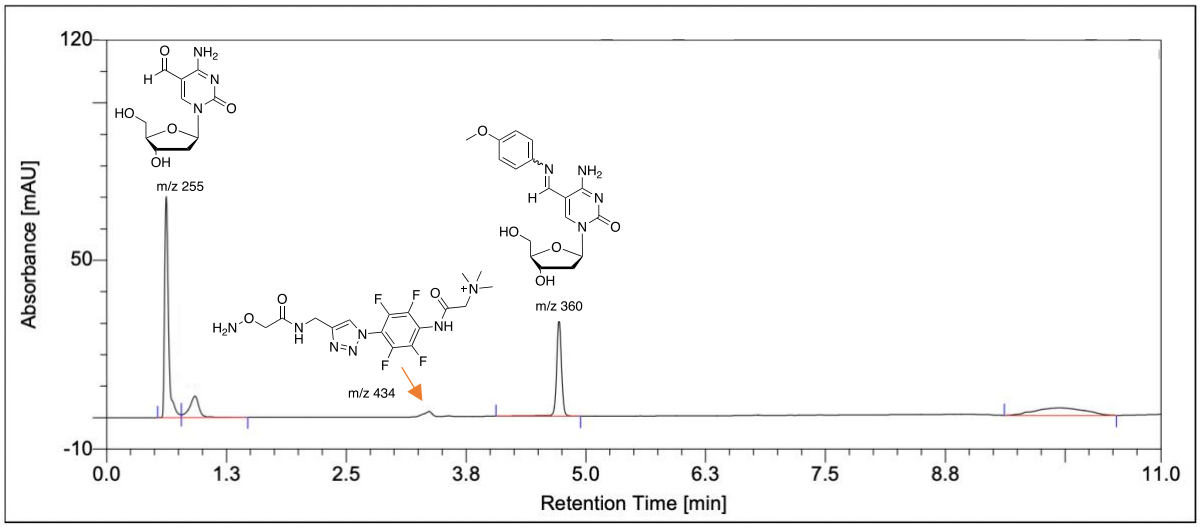


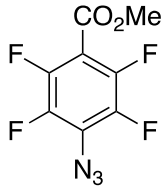
Figure 35. UV-Vis liquid chromatogram at 260 nm of derivatization reaction of fdC and **30**.

5.3.1. Project contributions

I was responsible for the synthesis of the target molecule and the screening of the derivatization reaction conditions.

5.3.2. Synthesized compounds and analytics

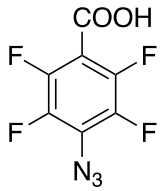
4-Azido-tetrafluoro methylbenzoate[93] (**19**)



0.9 ml methyl pentafluoro benzoate (6.10 mmol, 1.4 g, 1.0 eq.) were added to a solution of NaN_3 (9.48 mmol, 615mg, 1.6 eq.) in 8 ml acetone and 3 ml H_2O. The reaction mixture was refluxed for 8h and simultaneously covered from light. After cooling the reaction mixture to room temperature, it was further diluted with H_2O (10 ml) and extracted with DCM (3 x 20 ml). Combined organic layers were dried over $MgSO_4$ and solvents were evaporated resulting in yellowish needles. Due to the unstable character of the compound no further analysis was measured and the product was used without further purification in the next step.

Retention time same with the starting material, Seebach stain (Phosphomolybdic acid solution) stains azides dark blue which indicated the full conversion on TLC.
R_f(1:2 EtOAc/*i*Hex)=0.8

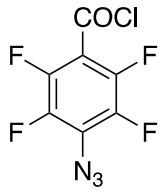
4-Azido-tetrafluoro benzoic acid[93] (**20**)



1.5 g 4-azido-tetrafluorobenzoate (6.02 mmol, 1.0 eq) were dissolved in 26 ml MeOH and 2.8 ml H_2O. 20 % aq. NaOH (2 ml) were added, and the reaction mixture was stirred at room temperature overnight and covered from light. After full conversion, the reaction mixture was cooled to 0 °C and acidified with 2 M HCl to pH 1. The mixture was extracted with DCM (3 x 40 ml), organic layers were dried over $MgSO_4$ and solvents were evaporated resulting in yellowish crystals (1.41 g, 98 %). Due to the unstable character of the compound no further analysis was measured and the product was used without further purification in the next step.

Seebach staining gives the compound as light blue spot
R_f(5% MeOH/DCM)=0.10, Seebach gives blue spot

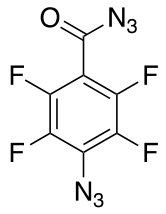
4-Azido-tetrafluoro benzoic acid chloride[93] (**21**)



1.32 g PCl_5 (6.36 mmol, 1.06 eq) were added to a mixture of 1.4 g 4-azido-tetrafluoro benzoic acid (6.00 mmol, 1.0 eq) in Et_2O (21 ml). The reaction mixture was stirred at room temperature for 1 h, volatiles were subsequently evaporated *in vacuo* and the residue was dried under high vacuum for 30 min in the dark. Due to the unstable character of the compound no further analysis was measured and the resulting yellow oil was used without further purification in the next step.

R_f(20 % MeOH/DCM)=0.70, Seebach staining gives the product as a red spot.

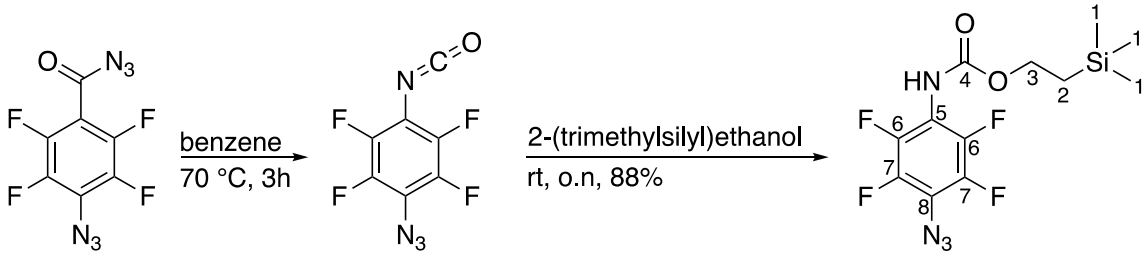
4-Azido-tetrafluoro benzoic acid azide[93] (**22**)



515 mg 4-azido-tetrafluoro benzoic acid chloride (2.19 mmol, 1.0 eq.) were dissolved in acetone (12 ml) and cooled to 0 °C. NaN_3 was dissolved in H_2O (1.7 ml) and cooled to 0 °C. The acid chloride solution was added dropwise to the rapidly stirred NaN_3 solution at 0 °C and the reaction mixture was stirred at 0 °C for 15 min. Then, the mixture was poured into a mixture of *i*-Hex/H_2O (50 ml/ 40 ml). The organic layer was separated, and the aqueous layer was further extracted with *i*-Hex (50 ml). Organic layers were combined, dried over $MgSO_4$ and approximately 80% of the volatiles were evaporated in vacuo giving yellow needles.
Due to the extremely unstable and explosive character of the compound no further analysis was measured and the product was used without further purification in the next step.

R_f (30 % EtOAc/iHex)=0.46, Seebach yellow spot.

4-(N-((2-Trimethylsilyl)ethoxy)carbonyl)amino)tetrafluorophenyl azide[93] (**23** and **24**)

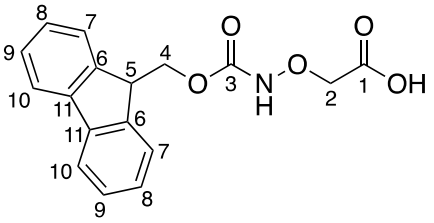


569 mg 4-azido-tetrafluoro acylazide (2.19 mmol, 1.0 eq.) were dissolved in benzene (5.3 ml) and the solution was stirred at 70 °C for 2 h. After full conversion of the starting material (R_f (30 % EtOAc/i-Hex): 0.68, Seebach stain gives red spot), 0.65 ml 2-(trimethylsilyl)ethanol (4.5 mmol, 528 mg, 2.04 eq.) were added and the reaction mixture was stirred at 70 °C for 6h. Volatiles were evaporated in vacuo and the residue was purified via column chromatography (EtOAc/i-Hex 5% -> 10% -> 30%) giving an orange oil in 72 % yield (1.57 mmol, 550 mg).
R_f (30 % EtOAc/i-Hex): 0.79, Seebach gives purple spot.

^{1}H NMR (400 MHz, CD_2Cl_2) δ 6.26 (s, 1H, -NH), 4.32 – 4.20 (m, 2H, H-3), 1.09 – 0.99 (m, 2H, H-2), 0.05 (s, 9H, H-1).

^{13}C NMR (101 MHz, CD_2Cl_2) δ 154.40 (C-4), 148.80 (m, C2, C-6), 146.42 (m, 2C, C-7), 120.24 (C-5), 106.26 (C-8), 68.22 (C-3), 20.00 (C-2), 1.60 (3C, C-1).

N-(9-Fluorenylmethyloxycarbonyl)-2-aminooxyacetic acid[94] (**26**)



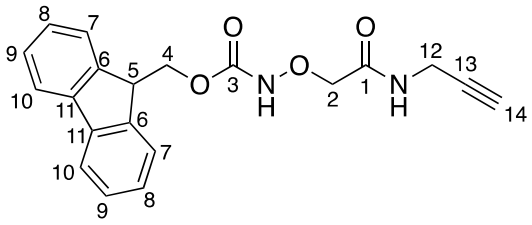
1.0 g hydroxylamino acetic acid hemihydrochloride (9.45 mmol, 1.0 eq.) were dissolved in aq. 10 % $NaHCO_3$ (29 ml) and cooled to 0 °C. A solution of 2.36 g Fmoc-Cl (9.45 mmol, 1.0 eq.) in dioxane (24 ml) was added dropwise and the reaction mixture was allowed to warm from 0 °C to room temperature overnight. Volatile dioxane was evaporated, the residue was diluted with H_2O (40 ml) and washed with Et_2O (20 ml) and EtOAc (20 ml). Then, the aqueous layer was acidified with 1N HCl to pH 3 and extracted with DCM (2 x 200 ml). Organic layers were combined, dried over Na_2SO_4 and volatiles were evaporated giving the Fmoc protected compound in 85 % yield.

R_f (12% MeOH/DCM)=0.56

^{1}H NMR (400 MHz, CD_2Cl_2) δ 8.40 (s, 1H, -OH), 7.79 (d, *J* = 7.6 Hz, 2H, H-10), 7.59 (dd, *J* = 7.5 Hz, 1.0 Hz, 2H, H-7), 7.46-7.29 (m, 4H, H-8 and H-9), 4.56 (d, *J* = 6.6 Hz, 2H, H-4), 4.45 (s, 2H, H-2), 4.26 (t, *J* = 6.5 Hz, 1H, H-5)

^{13}C NMR (101 MHz, CD_2Cl_2) δ 171.60 (C-1), 159.66 (C-3), 143.70 (2C, m, C-6), 141.86 (2C, m, C-11), 128.47 (2C, m, C-9), 127.72 (2C, m, C-10), 125.44 (2C, m, C-8), 120.61 (2C, m, C-7), 68.95 (C-2), 60.90 (C-4) 47.35 (C-5)

N-(Prop-2-ene-1-yl)-2-((9-flueronylmethoxycarbonyl amino)oxy)acetamide[94] (**27**)



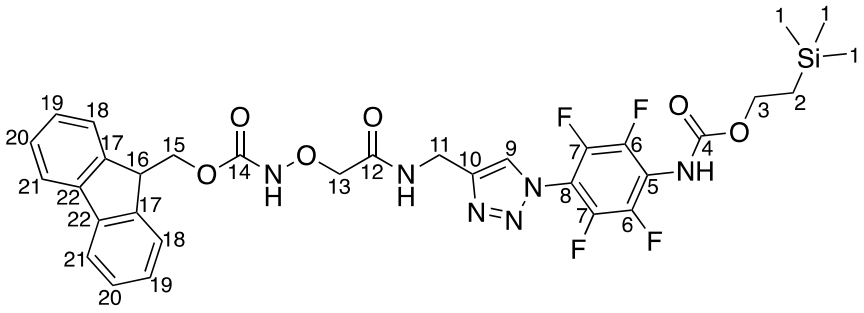
2.43 g N-(9-fluorenylmethyloxycarbonyl)-2-aminooxyacetic acid (7.76 mmol, 1.0 eq.) were dissolved in DCM (43 ml). 1.49 ml propargylamine (1.28 g, 23.30 mmol, 3.0 eq.) were added followed by addition of 3.54 g of HATU (9.32 mmol, 1.2 eq.) and 1.62 ml of DIPEA (9.32 mmol, 1.20 g, 1.20 eq.). Then, the reaction mixture was stirred at room temperature for 1 h. Subsequently the mixture was poured into aq. sat. NaCl (50 ml) and extracted with EtOAc (3 x 50 ml). Combined organic layers were dried over Na_2SO_4 and volatiles were evaporated. The crude mixture was purified via column chromatography (5% -> 10% DCM/MeOH) giving a white powder with 57% yield (1.55 g, 4.43 mmol).

R_f (12% MeOH/DCM)=0.64

^{1}H NMR (400 MHz, CD_2Cl_2) δ 7.97 (s, 2H, -NH), 7.79 (dt, *J* = 7.7, 1.0 Hz, 2H, H-10), 7.58 (dq, *J* = 7.4, 0.9 Hz, 2H, H-7), 7.46 – 7.29 (m, 4H, H-8 and H-9), 4.54 (d, *J* = 6.6 Hz, 2H, H-4), 4.31 (s, 2H, H-2), 4.26 (t, *J* = 6.6 Hz, 1H, H-5), 4.04 (dd, *J* = 5.5, 2.6 Hz, 2H, H-12), 2.23 (t, *J* = 2.6 Hz, 1H, H-14)

^{13}C NMR (101 MHz, CD_2Cl_2) δ 168.01 (C-1), 158.40 (C-3), 143.28 (m, 2C, C-6), 141.31 (m, 2C, C-11), 127.90 (m, 2C, C-9), 127.16 (m, 2C, C-10), 124.91 (m, 2C, C-8), 120.06 (m, 2C, C-7), 79.45 (C-13), 76.26 (C-14), 70.88 (C-2), 68.03 (C-4), 46.86 (C-5), 28.43 (C-12)

(9H-Fluoren-9-yl)methyl(2-oxo-2-(((1-(2,3,5,6-tetrafluoro-4-(((2-(trimethylsilyl)ethoxy)-carbonyl)amino)phenyl)-1*H*-1,2,3-triazol-4-yl)methyl)amino)ethoxy)carbamate (**28**)



350 mg 4-(N-((2-trimethylsilyl)ethoxy)carbonyl)amino)tetrafluorophenyl azide (1 mmol, 1 eq.) (**24**) were dissolved in THF/H_2O (1:1, 16 ml) and 385 mg N-(prop-2-ene-1-yl)-2-((9-flueronylmethoxycarbonyl amino)oxy)acetamide (1.1 mmol, 1.1 eq.) **27** were added. The mixture was degassed three times (freeze-pump-thaw), thereafter 62 mg CuBr•Me_2S (0.3 mmol, 0.3 eq.) were added, and reaction mixture was stirred at room temperature overnight. Then, the mixture was evaporated to dryness, the residue was dissolved in DCM (10 ml) and H_2O (10 ml), the aqueous layer was extracted with DCM (3 x 20 ml) and dried over Na_2SO_4. Volatiles were evaporated in vacuo and the residue was purified via column chromatography (EtOAc/i-Hex 30 % -> 50 %) giving the cyloaddition product **28** with 77 % yield.

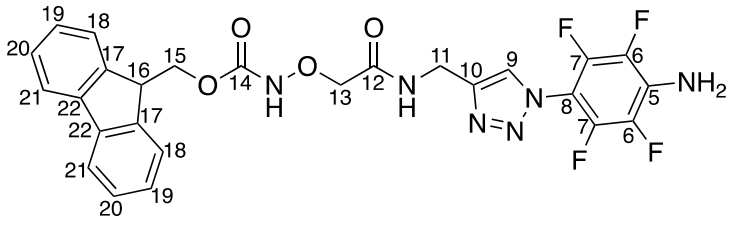
R_f (50 % EtOAc/i-Hex) = 0.3

^{1}H NMR (400 MHz, CD_2Cl_2) δ 8.20 (s, 1H, -NH), 7.82 (t, *J* = 1.0 Hz, 1H, H-9), 7.77 (dt, *J* = 7.6, 1.0 Hz, 2H, H-21), 7.56 (dq, *J* = 7.5, 1.0 Hz, 2H, H-18), 7.44 – 7.27 (m, 4H, H-19 and H-20), 6.50 (s, 1H, -NH), 4.65 (d, *J* = 5.8 Hz, 2H, H-15), 4.49 (d, *J* = 6.7 Hz, 2H, H-13), 4.36 (s, 2H, H-11), 4.34 – 4.28 (m, 2H, H-3), 4.23 (t, *J* = 6.7 Hz, 1H, H-16), 1.11 – 1.05 (m, 2H, H-2), 0.07 (s, 9H, H-1).

^{13}C NMR (101 MHz, CD_2Cl_2) δ 169.74 (C-12), 153.21 (C-14), 151.21 (C-4), 144.38 (m, 4C, C-6 and 7), 129.34 (m, 4C, C-17, 22), 128.48 (m, 9C, C-18-21, C-10), 127.67 (C-9), 125.21 (C-5), 101.30 (C-8), 74.71 (C-13), 68.74 (C-15), 65.97 (C-3), 46.88 (C-16), 34.66 (C-11), 18.04 (C-2), -1.38 (3C, C-1).

LC-MS: H_2O/MeCN 5% to 80% in 10 min, UV-Vis 260 nm, R_t: 8.1 min, LRMS (ESI^+): value calc. for $C_{32}H_{33}F_4N_6O_6Si$ $[M+H]^+$ =701.2161; found 701.4500. LRMS (ESI^-): value calc. for $C_{32}H_{31}F_4N_6O_4Si$ $[M-H]^-$ =699.2016; found 699.0300

(9H-Fluoren-9-yl)methyl(2-(((1-(4-amino-2,3,5,6-tetrafluorophenyl)-1H-1,2,3-triazol-4-yl)-methyl)amino)-2-oxoethoxy)carbamate (**29**)



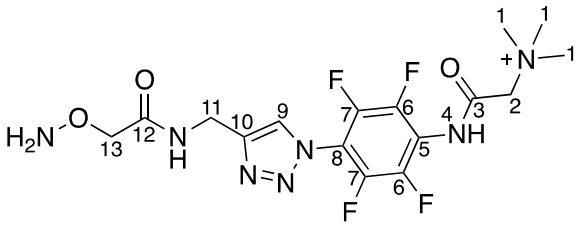
0.282 g of **28** (0.402 mmol, 1 eq.) were dissolved in DCM (2.5 ml), 0.21 ml TFA (2.82 mmol, 0.32 g, 7 eq.) were added and the reaction mixture was stirred at room temperature overnight. Subsequently sat. aq. $NaHCO_3$ (5 ml) was added, the mixture was extracted with DCM (3 x 5 ml) and the combined organic layers were dried over Na_2SO_4. Volatiles were evaporated in vacuo and the crude product was purified via column chromatography (DCM/MeOH 100% -> 2% -> 4%)

R_f (2% MeOH/DCM) = 0.35, Ninhydrin stain gives dark orange spot

^{1}H NMR (400 MHz, CD_2Cl_2) δ 7.76 (dt, *J* = 7.6, 1.0 Hz, 2H, H-21), 7.72 (s, 1H, H-9), 7.55 (dq, *J* = 7.5, 0.9 Hz, 2H, H-18), 7.43 – 7.26 (m, 4H, H-19 and H-20), 5.33 (s, 1H, -NH), 4.63 (d, *J* = 5.8 Hz, 2H, H-15), 4.46 (s, 2H, H-13), 4.36 (s, 2H, H-11), 4.21 (t, *J* = 6.8 Hz, 1H, H-16).

LC-MS: H_2O/MeCN 5% to 80% in 10 min, UV-Vis 260 nm, R_t: 6.6 min, LRMS (ESI^+): value calc. for $C_{26}H_{21}F_4N_6O_4$ $[M+H]^+$ =557.4851; found 557.3700. LRMS (ESI^-): value calc. for $C_{26}H_{19}F_4N_6O_4$ $[M-H]^-$ =555.4702; found 555.0900, $C_{27}H_{21}F_4N_6O_6$ $[M+FA-H]^-$ = 601.1464; found 601.2300

2-((4-(4-((2-(Aminooxy)acetamido)methyl)-1*H*-1,2,3-triazol-1-yl)-2,3,5,6-tetrafluorophenyl)-amino)-*N*,*N*,*N*-trimethyl-2-oxoethan-1-aminium (**30**)



Before starting, betaine was dried under high vacuum at 100 °C for 6h. Subsequently, 42 mg betaine (0.36 mmol, 2 eq.) were dissolved at room temperature together with 137 mg HATU (0.36 mmol, 2 eq.) in DMF (1ml) and the mixture was stirred at room temperature overnight. 100 mg of a solution of **29** (0.18 mmol, 1 eq.) in DMF (0.8 ml) were added and the mixture was

stirred at room temperature another time overnight. 50 µl DIPEA (0.29 mmol, 37 mg, 1.6 eq.) were added and the reaction mixture was stirred further 3,5 h at room temperature. Subsequently, the reaction mixture was poured into aq. NH_4Cl (20 ml) and the mixture was washed with DCM (3 x 20 ml). The aqueous layer was collected, lyophillized, redissolved in aq. ammonium formiate buffer and purified via HPLC using the HCO_2NH_4 buffer system (0% -> 60% 45 min).

^{1}H NMR (400 MHz, D_2O) δ 8.40 (s, 1H, H-4), 8.19 (s, 1H, H-9), 4.59 (s, 2H, H-13), 4.35 (s, 2H, H-2), 3.56 (s, 2H, H-11), 3.07 (s, 9H, H-1).
^{13}C NMR (101 MHz, D_2O) δ 173.24 (C-12), 170.97 (C-3), 160.21 (m, 2C, C-6), 160.19 (m, 2C, C-7), 144.31 (m, 2C, C-10) 127.12 (m, 2C, C-9 and C-5), 71.05 (C-13), 67.38 (C-2), 53.76 (C-1), 33.59 (C-11).
IR (ATR): (cm^{-1}) = 3323, 3193, 1659, 1601, 1539, 1516, 1488, 1377, 1334, 1285, 1228, 1176, 1069, 1040, 946, 836.
LC-MS: H_2O/MeCN 5% to 80% in 10 min, UV-Vis 260 nm, R_t: 3 min, LRMS (ESI^+): value calc. for $C_{16}H_{20}F_4N_7O_3$=434.3751; found 434.3700.

6. Materials and methods

Chemical Synthesis

Unless otherwise noted, all chemical transformations were performed using glassware that was heat-dried under high vacuum, utilizing the Schlenk technique under an atmosphere of nitrogen or argon. Chemicals were purchased from *Sigma- Aldrich*, *TCI*, *Fluka*, *ABCR*, *Carbosynth* or *Acros Organics* and used without further purification. Technical grade solvents were distilled prior to extraction. Solvents that were dried over molecular sieve or stored under argon atmosphere were purchased from Acros Organics. The removal of solvents *in vacuo* was performed with a rotary evaporator, Laborota 4000 from Heidolph, where the water bath temperature was set to 40 °C unless otherwise noted.

Reactions under carbon monoxide over pressure

Reactions performed under carbon monoxide over pressure (3.5 bar) utilized a pressure resistant autoclave vessel. Addition of liquids to the reaction mixture were administered via syringe pump from KD Scientific. The autoclave vessel was oven dried and cooled under argon flow prior to the experiment and subsequently flushed three times with 3.5 bar of carbon monoxide for 15 min.

Thin layer chromatography (TLC)

Reactions and column chromatography fractions were monitored by thin-layer chromatography (TLC) on silica gel F_{254}TLC plates from Merck KGaA. Visualization of the developed TLC plates was achieved through UV-absorption or fluorescence under UV-lamp (254 nm and/or 366 nm) and/or through staining with Hanessian's stain (ammonium molybdate tetrahydrate 2.5 g, cerium ammonium sulfate dihydrate 1.0 g, H_2SO_4 10 ml, ddH_2O 90 ml), $KMnO_4$ stain ($KMnO_4$ 3 g, K_2CO_3 10 g, ddH_2O 300 ml), Ninhydrin stain (Ninhydrin 0.1 g, acetic acid 0.5 ml, acetone 100 ml) or Seebach's stain (phosphomolybdic acid 2.5 g, $CeSO_4$ x H_2O 1.0g, conc. H_2SO_4 6.0 ml, ddH_2O 94 ml)[96].

Column chromatography

Column chromatographic purifications of obtained crude products were performed under slight nitrogen pressure via the Flash-method. The stationary phase was Geduran®Si60 (0.063

– 0.200 nm) silica gel from Merck KgaA. Suitable eluents were determined via thin layer chromatography of the crude product.

High-Performance Liquid-Chromatography (HPLC)

Analytical RP-HPLC was performed on an analytical HPLC Waters Alliance with Waters e2695 Separation Module in combination with 2487 UV/Vis Detector (Waters) or on an Agilent 1260 Infinity II 800 bar-Flexible Prime-Pump system with 1260 infinity II MWD at 260 nm. As a stationary phase 120/3 Nucleosil C18 columns from Macherey Nagel applying an eluent flow of 0.5 mL/min were used. HPLC grade acetonitrile was purchased from VWR and double distilled water (ddH_2O) was obtained from Arium® Pro manufactured by Sartorius Stedim Biotech. In case of utilizing buffered eluent systems, the pH-value was determined with a MP 220-pH-meter from Mettler Toledo.

Semi-preparative purification of synthesized compounds was performed on a HPLC Waters Breeze system consisting of a 1525 Binary HPLC pump and a 2487 UV/Vis detector or on an Agilent 1260 Infinity II 400 bar pump system with an Agilent 1260 Infinity II VWD at 260 nm. As a stationary phase VP 250/10 Nucleosil C18 columns or Nucleodur (250/10 mm C18, 5 µm) reverse phase chromatography columns from Macherey Nagel were used. A flowrate of 5mL/min was applied.

Lyophilization

Solvent removal via lyophilization was performed with Alpha 2-4 LD plus-Lyophilizers from Christ. Solvents were frozen with liquid nitrogen prior the lyophilization.

Nuclear magnetic resonance spectroscopy (NMR-Spectroscopy)

NMR spectra were recorded on *Bruker Avance III HD 400 (*400 MHz), *Varian NMR-System 600* (600 MHz) and *Bruker Avance III HD* with *Cryo-Head 800* (800 MHz) spectrometers. ^{1}H-NMR shifts were calibrated to the residual solvent resonances: $CDCl_3$ (7.26 ppm), DMSO-d_6 (2.50ppm) D_2O (4.79 ppm), or D_3CN (1.94 ppm). ^{13}C-NMR shifts were calibrated to the residual solvent: $CDCl_3$ (77.16 ppm), DMSO-d_6 (39.52 ppm), or D_3CN (1.32 ppm). All NMR spectra were analysed using the program MestReNova 10.0.1 from Mestrelab and NMR data are reported as follows: chemical shift (multiplicity, coupling constants where applicable, number of

hydrogens, assignment). Splitting is reported with the following symbols: s = singlet, d = doublet, t = triplet, dd = doublet of doublets, dt = doublet of triplets, ddd = doublet of doublets of doublets, m = multiplet.

IR spectra were recorded on a *PerkinElmer* Spectrum *BX II FT-IR* system.

Mass spectrometry

Low resolution mass spectra were measured on a *MSQ* Plus Mass spectrometer by *Thermo Scientific* in combination with *Dionex Ultimate3000* liquid chromatography and injection. High resolution mass spectra were measured by the analytical section of the Department of Chemistry of Ludwigs-Maximilians-Universität München on *LTQ FTI-CR* by *Thermo Finnigan GmbH*. MALDI-TOF mass spectrometry was measured with *Bruker Autoflex II* after desalting the samples with 0.025 µm VSWP filters (*Millipore*) in ddH_2O and cocrystallization of the sample with HPA-matrix (3-Hydroxypicolin acid).

Quantitative UHPLC-MS/MS analysis of digested DNA samples was performed using an Agilent 1290 UHPLC system equipped with a UV detector and an Agilent 6490 triple quadrupole mass spectrometer.[87] Nucleosides of interest were quantified using the stable isotope dilution technique.[16] The source-dependent parameters were as follows: gas temperature 80 °C, gas flow 15 L/min (N_2), nebulizer 30 psi, sheath gas heater 275 °C, sheath gas flow 11 L/min (N_2), capillary voltage 2.500 V in the positive ion mode, capillary voltage −2.250 V in the negative ion mode and nozzle voltage 500 V. The fragmentor voltage was 380 V/ 250 V. Delta EMV was set to 500 (positive mode) and 800 (negative mode). Compound- dependent parameters are summarized in the tables below. Chromatography was performed by a Poroshell 120 SB-C8 column (Agilent, 2.7 µm, 2.1 mm × 150 mm) at 35 °C using a gradient of water and MeCN, each containing 0.0085% (v/v) formic acid, at a flow rate of 0.35 mL/min: 0 →4 min; 0 →3.5% (v/v) MeCN; 4 →7.9 min; 3.5 →5% MeCN; 7.9 →8.2 min; 5 →80% MeCN; 8.2 →11.5 min; 80% MeCN; 11.5 →12 min; 80 →0% MeCN; 12 →14 min; 0% MeCN. The autosampler was cooled to 4 °C. The injection volume was amounted to 39 µL.

Quantification of nucleosides in DNA samples

Nucleosides were quantified using internal calibration curves and the stable isotope dilution technique as described in the literature.[87]

Table 1. Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA fed with fluorinated nucleosides. CE: collision energy, CAV: collision cell accelerator voltage. The nucleosides were analyzed in the positive mode ($(M+H)^+$ species) ion selected reaction monitoring mode (SRM).

Compound	Precursor ion (m/z)	MS1 Resolution	Product Ion (m/z)	MS2 Resolution	Dwell time {ms}	CE (V)	CAV (C)	Polarity
Time segment 1.5-3.7 min								
F-dC	246.09	Wide	112.06	Wide	70	15	3	Positive
($^{15}N_2$)-F-dC	248.08	Wide	114.04	Wide	70	15	3	Positive
($^{15}N_2$)-cadC	274.08	Wide	158.03	Wide	40	5	5	Positive
cadC	272.09	Wide	156.04	Wide	40	5	5	Positive
F-hmdC	276.10	Wide	142.06	Wide	50	10	3	Positive
($^{15}N_2$)-F-hmdC	278.09	Wide	144.06	Wide	50	10	3	Positive
F-cadC	290.08	Wide	156.04	Wide	80	5	5	Positive
($^{15}N_2$)-F-cadC	292.08	Wide	158.04	Wide	80	5	5	Positive
Time segment 3.7-4.7 min								
(D_3)-F-mdC	263.12	Wide	129.09	Wide	80	15	3	Positive
F-mdC	260.10	Wide	126.07	Wide	80	15	3	Positive
F-cadC	290.08	Wide	156.04	Wide	80	5	5	Positive
($^{15}N_2$)-F-cadC	292.08	Wide	158.04	Wide	80	5	5	Positive
Time segment 4.7-10 min								
F-fdC	274.08	Wide	140.05	Wide	90	15	3	Positive
($^{15}N_2$)-F-fdC	276.08	Wide	142.04	Wide	90	15	3	Positive

Table 2. Compound-dependent LC-MS/MS ranges of the corresponding linear equations.

Compound	**n (<LOQ)**	**n (LLOQ)**	**A/A* (<LOQ)**	**A/A* (LLOQ)**
F-dC	800 fmol	3.13 fmol	4.074	0.01512
F-mdC	49.9 fmol	0.78 fmol	0.04174	0.0005833
F-fdC	25.1 fmol	0.390 fmol	0.2267	0.003345
F-cadC	12.7 fmol	0.2 fmol	0.240576	0.0217614
F-cadC	202.5 fmol	6.3 fmol	3.92816	0.1271398

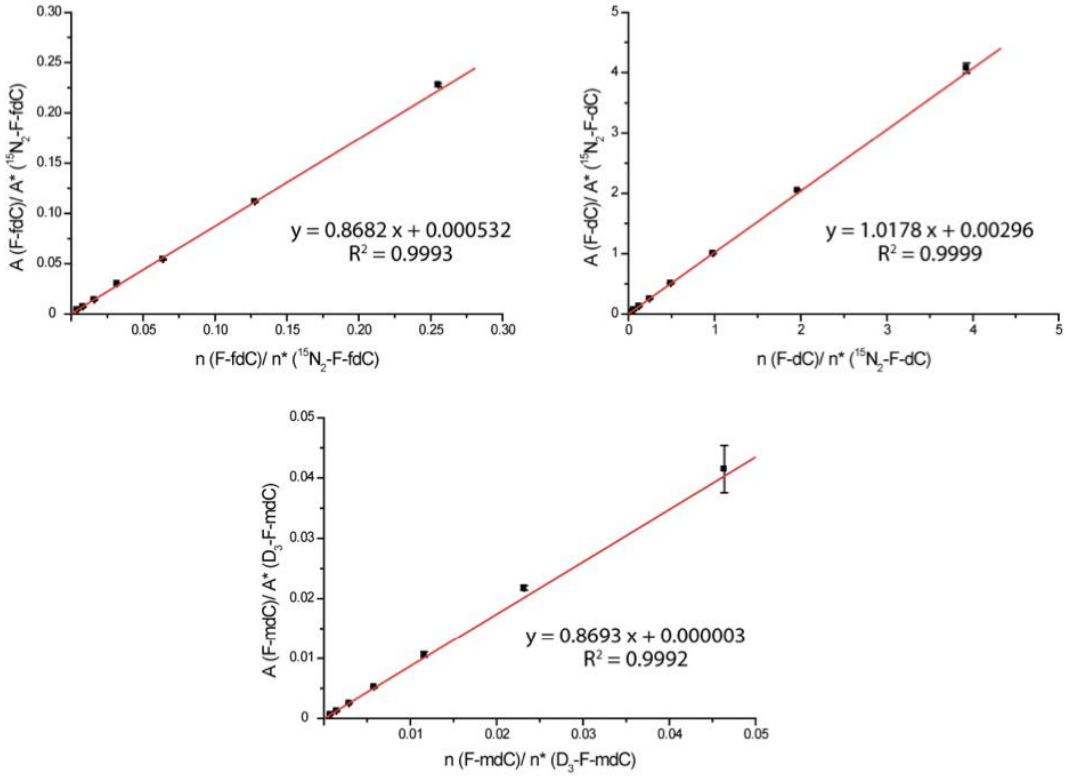


Figure 36. Internal calibration curves with corresbonding linear equation and coefficient of determination for the exact quantification of F-fdC, F-dC and F-mdC.

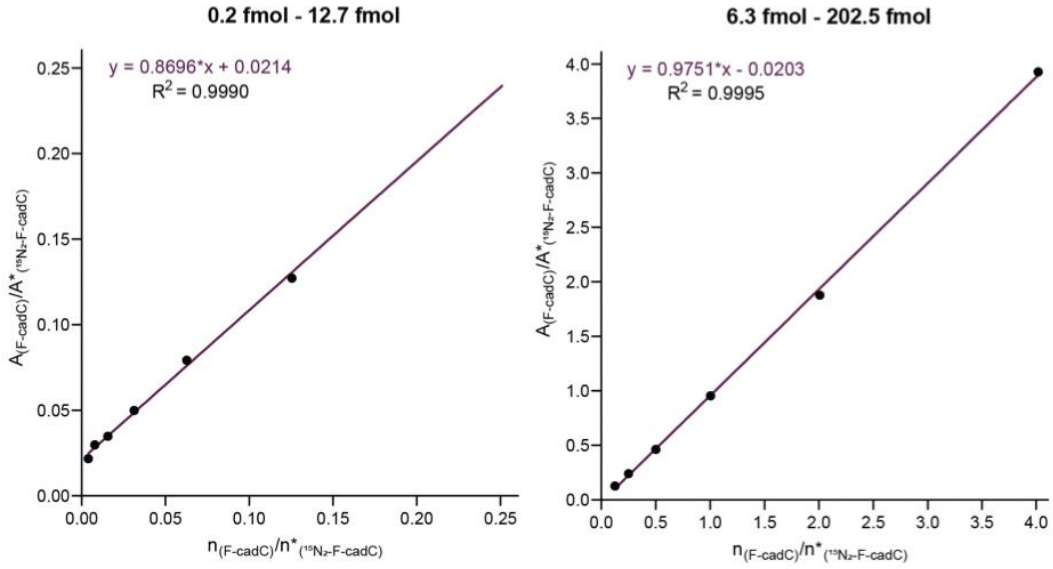


Figure 37. Internal calibration curves for the exact quantification of F-cadC with the linear equation and coefficient of determination.

Quantification of 5-NO_2-F-dC

To determine the exact amount of the detected molecule in the MS/MS experiments, an extinction coefficient has to be determined. That is determined from the measured absorptions that are plotted in function of concentration (Figure 38.) The absorption of 5-NO_2-F-dC was measured at three different wave lengths, l_{max} = 227 nm, 260 nm and l_{min} 280 nm, from 10 different concentrations. The extinction coefficient was determined from Beer-Lambert Law. The plotted values formed a straight line and for which an equation of form y = kx + b and the slope was determined as arithmetic mean of two calculated slopes. The calculated values for wave lengths 227 nm, 260 nm and 280 nm were 7718.05 $mmol^{-1}$ x l x cm^{-1}, 3085.95 $mmol^{-1}$ x l x cm^{-1}, 1369.9 $mmol^{-1}$ x l x cm^{-1} respectively.

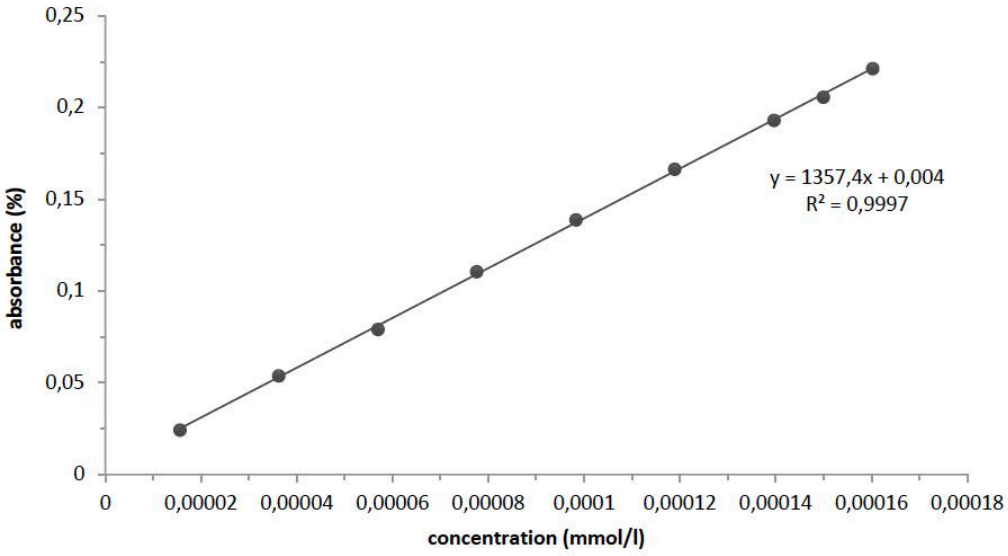


Figure 38. Calculated equation as function of concentration at 280 nm.

Part 2:

Synthesis of oxazole-containing building blocks for therapeutical peptidomimetics

7. Peptides and peptidomimetics

Generally, peptides are defined to consist of a chain of 50 or less amino acids that are linked to each other via amide bonds. Polypeptides on the other hand are understood as longer linear chains of amino acids and a polypeptide with a molecular mass of 10 000 Da or more is called a protein. Protein-protein interactions are essential for regulating a vast array of biological processes *in vivo* and it has been estimated that from the molecular overall number of interactions in a cell, around 650 000 molecular interactions are protein-protein interactions.[97] Therefore, the research how to discover active compounds, particularly small organic molecules which are able to modify such interactions is believed to exhibit a huge therapeutical potential. Traditionally, small molecules have been studied in their application to regulate the function and/or activity of enzymes. However, typically binding surfaces of proteins are relatively large on a molecular level (1500 – 3000 $Å^2$) involving several polar and hydrophobic interaction sites and leaving no well-defined binding pockets for small-molecule drugs. It has been estimated that 15 to 40 % of all protein-protein interactions are indeed mediated by short linear peptides.[98] These bioactive peptides are produced in animals, humans and plants and many of these peptides are known to exhibit a strong bio-specificity and efficacy in essential biological processes, making them another group of desirable target molecules for therapeutical applications.

Research interest in therapeutic peptides is believed to have started with insulin, oxytocin, vasopressin and gonadotropin releasing hormone that are all natural human hormones. In fact, the first endogenously present peptide based therapeutic, insulin isolated from bovine pancreas, was used to treat a child with diabetes in early 1920s.[99] Since then, more than 80 peptide drugs have been approved worldwide and over 170 peptides are in under clinical development.[100] However, due to their chemical properties, their membrane permeability is typically poor and they tend to show a short plasma half-life typically in the range of minutes caused by rapid enzymatic degradation. Nevertheless, also peptides exhibiting an *in vivo* plasma half-life in the order of magnitude of years have been discovered, however the total average is approximated to be around 1-2 days in mammalian cells. These limitations can be overcome or at least improved i.e by modifying the existing peptide sequence so as to create peptidomimetics. Exemplary modifications can comprise a substitution of a natural amino acid

in the peptide chain by a chemically altered amino acid. Other approaches are based on other chemical or enzymatical modifications, such as *L* to *D* isomerization, cyclization or linkage to cell penetrating peptides resulting in an own promising class of therapeutics. These intended changes in the peptide aim to improve the therapeutical potential of the compounds and simultaneously to reduce potential side effects. For example, vasopressin, a peptide containing 9 amino acids is known to have very short plasma half-life of 16 to 24 min and, furthermore, the undesirable side effect of causing a raise in the blood pressure of the patient to whom the vasopressin is administered. By alternating the amino acid sequence by deaminating the first of the nine amino acids and by isomerizing the arginine in position 8 from *L* to *D* (Figure 39), desmopressin can be obtained which has been demonstrated to degrade slower and to additionally merely have a little effect on the blood pressure of the patients.

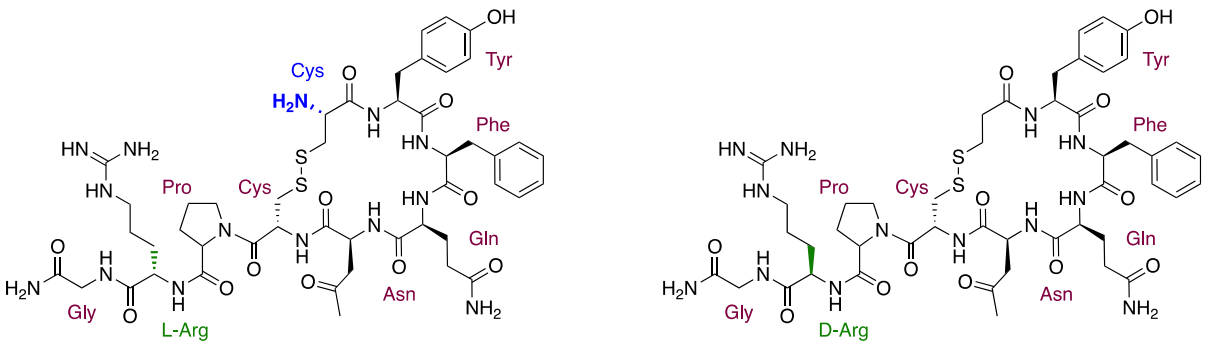


Figure 39. Natural peptide vasopressin and a peptidomimetic thereof.

7.1. Oxazole containing peptides and peptidomimetics

Another approach is the modification of the peptide sequence with a planar aromatic heterocycle, such as for example an oxazole heterocycle which is believed to lack nucleophilicity towards carbon and sulfur centers and to simultaneously have only a low Bronsted acidity and basicity. Oxazole-containing peptidomimetics are thus considered to be an effective strategy to improve the bioavailability of the peptidomimetics as well as to increase their resistance towards hydrolysis and their stability against proteases.

Moreover, it has been reported that the oxazole moiety can form an important pharmacophore that may directly influence the peptide structure by creating a conformational strain[101–103] on the molecular level and by modifying the electronic distribution within the molecule. Introduction of an oxazole heterocycle can thus give the opportunity to selectively modify the electronic properties of the peptidomimetics. It has also been reported that such modifications may even affect the hydrogen bond formation[104] or can result in an improved interaction with the nucleobases of DNA/RNA.[105]

To summarize, it has been reported that oxazole heterocycles are structurally more diverse bioisosteres than other previously reported amide bond isosteres, such as ketomethylenes, (*E*)-alkenes or aminomethylenes[106] and the improved characteristics of these compounds are fundamental importance for the design of new drugs, particularly in view of their beneficial effects with respect to the recognition of target proteins and their stability.

With respect to the direct impact of an oxazole moiety in the peptide chain of a peptidomimetic on the molecular level, Siodłak *et al*.[107] have reported a study focusing on conformational properties of amino acid residues containing an oxazole ring. Theoretical calculations supported by analytical methods including NMR and IR spectroscopy as well as single-crystal X-ray diffraction were used to investigate selected model compounds of naturally occurring oxazole-amino acids, that comprise oxazole-alanine, oxazole-dehydroalanine and oxazole-dehydrobutyrine which were then detected in four different conformations: β2, C5, β and αD and αL (Figure 40). The most stable conformation among the studied oxazole-containing amino acids was found to be conformation β2. It was further shown that this conformation is stabilized by an intramolecular N-H---N hydrogen bond and it is predominantly present in the low polar environment.[107]

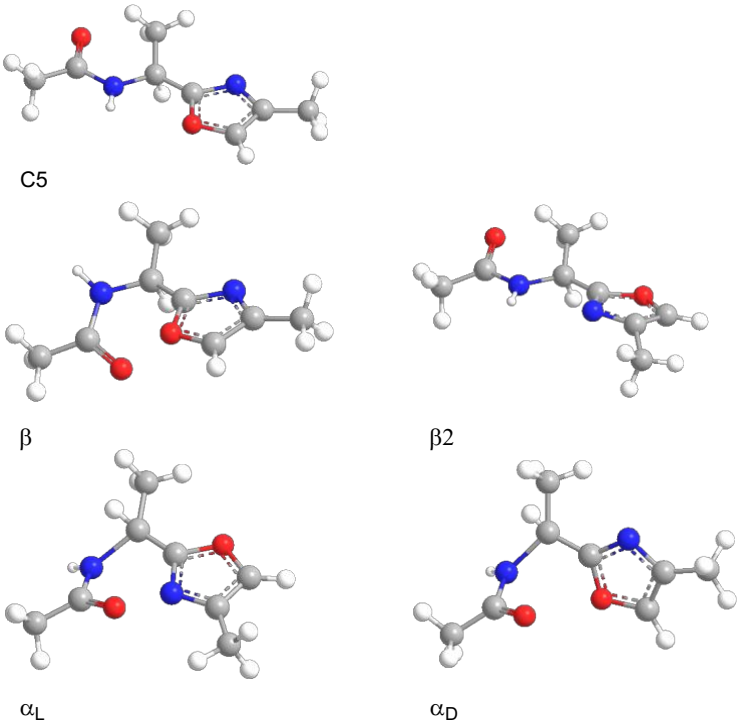


Figure 40. Four different conformations calculated with DFT methods for oxazole-alanine.[107]

Based on this conformational information of the oxazole moiety, it was reported that the possibility exists to design a peptidomimetic, wherein the side chains are spatially directed in specific directions, for instance a molecular conformation projecting two antiparallel loops perpendicularly from the surface of the main cycle.[108] Indeed, some studies suggest that oxazole-constrained macrocycles can form various supramolecular structures, including cylinders, throughs, helix bundles, and multi-loop structures (Figure 41).[103] Due to the provided additional effects on the molecular configuration, it has been reported that even in the late-stage functionalization of peptides and macrocycles, oxazole fragments can play an important role, i.e. in directing side chains, resulting in high-regioselective and site-selective reactions.[109]

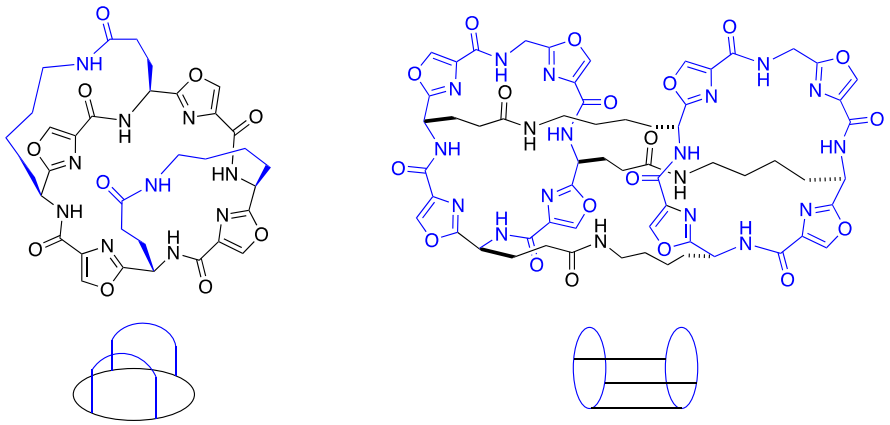


Figure 41. Supramolecular structures of oxazole containing macrocycles.[109] Left: Multi-loop structure of tetraoxazole macrocyclic scaffold attached to two loops. Right: Trough structure of octaoxazole macrocyclic scaffold consisting of two tetraoxazole cycles connected by three linkers.

Generally, the secondary structure of a polypeptide is determined by the hydrogen bonds between the atoms of its backbone. A study from D. Kaur and S. Khanna[110] compared intermolecular hydrogen bonding interactions between water and the aromatic heterocycles furan, isoxazole and oxazole. Besides the heteroatoms, the investigated heterocycles also have an aromatic π-electron system as an attractive site to form hydrogen bond. The studies on the hydrogen bonding character of different five-membered heterocycles concluded that the investigated heterocycles could take part in the interaction both as hydrogen bond donors as well as hydrogen bond acceptors and that the number and type of heteroatoms can influence the type of hydrogen bonds.

With respect to furan, a single heteroatom containing heterocycle, it was shown that hydrogen bonding with water through the ring oxygen acting as a hydrogen bond acceptor is relatively weak and when compared to the same hydrogen bond acceptor properties of the non-aromatic tetrahydrofuran with water, the hydrogen bonding is even weaker. On the other hand, the hydrogen bonding interactions between isoxazole and oxazole with water only occurred through heteroatoms and no adduct with aromatic π-electron system of the ring as a hydrogen bond acceptor was formed (Figure 42).

Additionally, it was found that the isoxazole heterocycle, where oxygen and nitrogen atoms are situated in ring positions 1 and 2, can form two different possible hydrogen bonding

adducts with water (Figure 42).[110] The first adduct forms not only a bond between H_{water} and $O_{isoxazole}$, but also a bond between $H_{isoxazole}$ in ring position 5 and O_{water}. In the second adduct where the hydrogen bonding interactions were formed only between the $N_{isoxazole}$ and H_{water} were relatively stronger compared to the first adduct. Nevertheless, two adducts were also observed between oxazole and water where nitrogen and oxygen are found to be not directly adjacent to each other, i.e. when they are positioned in ring positions 1 and 3 in the heterocycle.[110] Again, the N-H electron donation was found stronger in the second adduct of oxazole and overall, as the strongest hydrogen bonding interaction of the studied molecules. However, the oxygen atom in oxazole is a weaker H-bond acceptor than in isoxazole.

Furthermore, and as already anticipated by the authors[110], a competition between different heteroatoms with respect to hydrogen bonding was observed. The hydrogen bonding of the in-ring nitrogen was found to be comparably stronger than the hydrogen bonding properties of the in-ring oxygen which can be explained by the contribution of the ring oxygen to the conjugated aromatic system, whereas the in-ring nitrogen comprises an electron pair which does not contribute to the aromatic π-electron system making it more electron-rich and resulting in improved properties as a better hydrogen bond acceptor.[107]

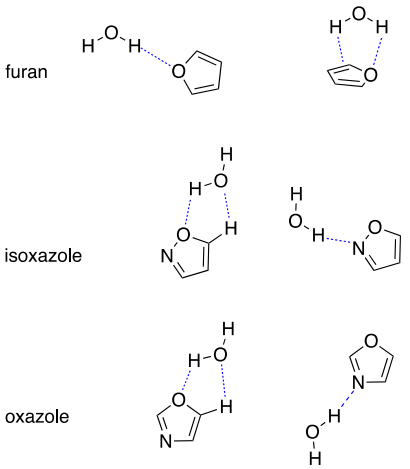


Figure 42. Hydrogen bonding interactions of furan, isoxazole and oxazole with water.

In fact, oxazole heterocycles can be found in the backbone of natural bioactive peptides and it has been discovered that several oxazole-containing peptides might exhibit improved stability and activity and are considered promising candidates

for medicinal applications.[111–113] Prominent examples are the naturally occurring cyclic pentapeptides, Nazumazole D to F (Figure 43), which can be isolated from the marine sponge *Theonella swinhoei* and which have already been shown to inhibit chymotrypsin.[114] Another highly interesting example is Beminamicyn A (Figure 43) which can be isolated from S.bernensis, and which is a cyclic peptide that inhibits protein biosynthesis in Gram-positive bacteria functioning as an antibiotic.[115] Additionally, an oxazole containing dipeptide, Almazole D (Figure 43), which was found in delesseriacean seaweed shows activity as an anti-bacterial agent, but the same compound can also act as a potent anti-viral, in particular, anti-malarian.[111,116–119]

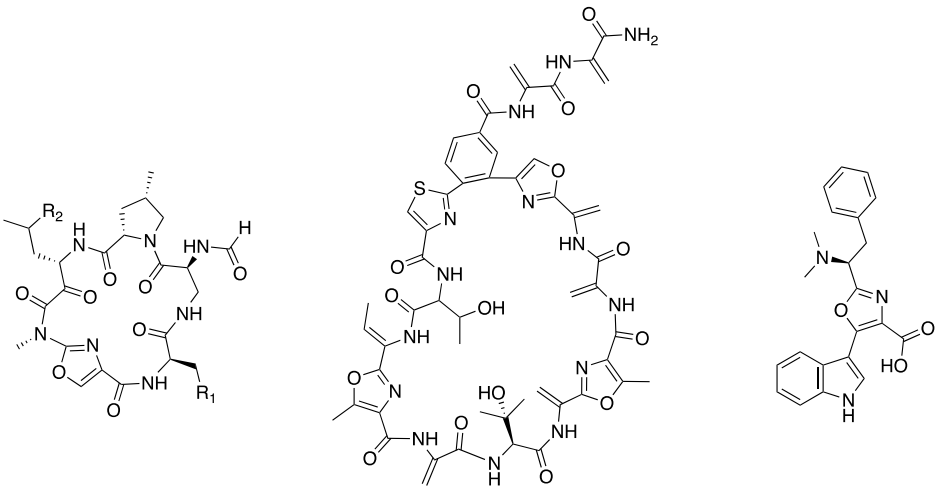


Figure 43. Naturally occurring bioactive oxazole containing peptides.

Another study demonstrated that the oxazole-containing cyclic peptide, Leucamide A (Figure 44), exhibits cytotoxic activity against various types of cancer cells[120] and certain oxazole-modified glycopeptides have been found to exhibit a significant inhibitory activity towards the immune response in case of rheumatoid arthritis.[121] In the field of treatment of bacterial strains having developed antibiotic resistance, some macrocycles, for example Methylsulfomycin I (Figure 44), have shown activity against a wide range of Gram-positive bacteria, including vancomycin- and teicoplanin-resistant organisms.[116]

An additional important class of oxazoles are methyloxazoles that are found as structural elements in bacterial peptides. They are formed via posttranslational modifications of amino acid residues, wherein instead of Ser residues being modified to above-mentioned oxazoles, Thr is used to form methyloxazoles. These methyloxazoles are known to be present in many bioactive cyclic peptides like aforementioned Beminamycin A.[115]

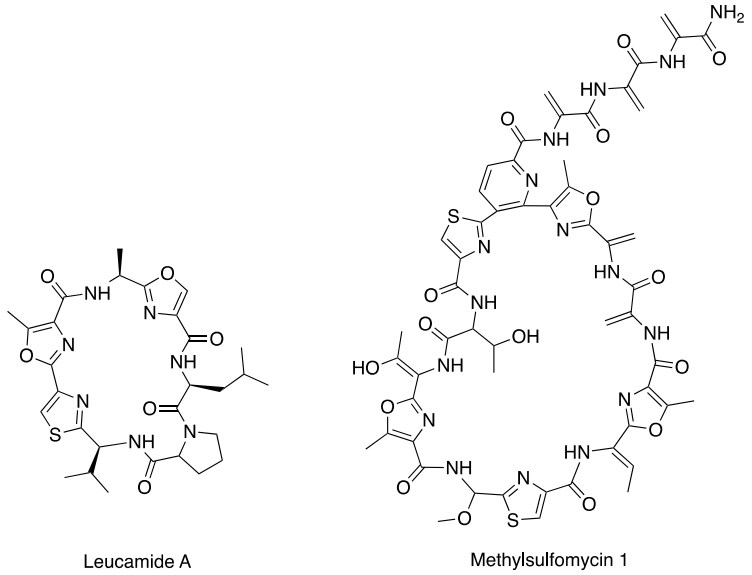


Figure 44. Naturally occurring bioactive cyclic peptides.

In addition to naturally occurring oxazole-containing peptides, synthetic derivatives have been developed which show promising activities against various types of viruses. In this regard bisheterocycle-containing compounds such as open chain derivatives of Leucamide A (Figure 45) have been evaluated as potential anti-influenza virus agents, as they provide bioactivity through specific interaction with DNA and RNA.[105] Also synthetic oxazole-based macrocycles (Figure 45), have been found to be effective against SARS-CoV-2 main protease.[122]

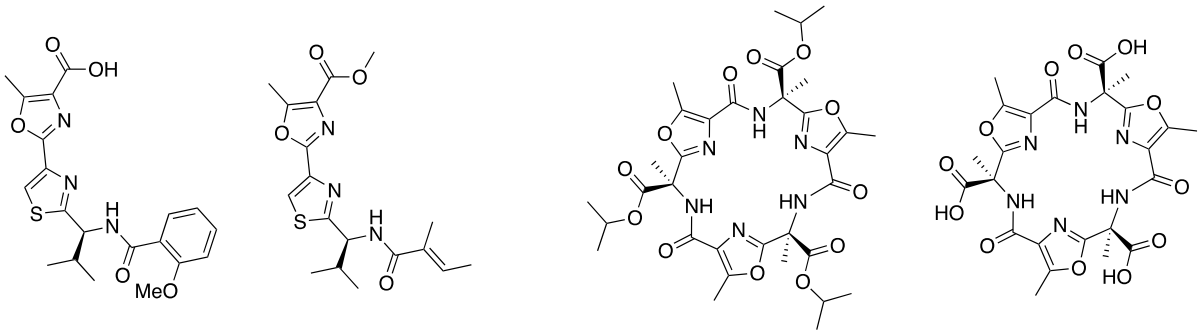


Figure 45. Synthetical derivatives of oxazole-based peptides.

Furthermore, oxazole-containing peptides have been described to potentially target different types of cancer cells. Microcyclamide (Figure 46) has been shown to possess moderate cytotoxicity against P388 murine leukemia cells.[123] and Diazonamide A has been found to inhibit the growth not only of B-16 murine melanoma cancer cells but also of HCT-116 human colon carcinoma.[124] Leucamide A (Figure 46) has shown cytotoxicity against HM02 stomach carcinoma and different types of liver carcinoma, while the macrocyclic Urukthapelstatin A (Figure 46) inhibited the growth of A549 human lung cancer.[120,125]

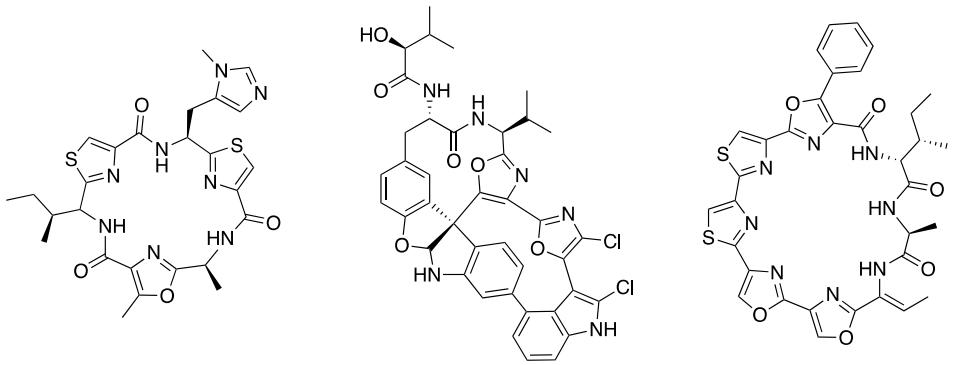


Figure 46. Cancer targeting oxazole containing peptides.

8. Aim of the work

As the demand in pharmaceutically active heterocycle-containing peptidomimetics is increasing, five highly functionalized oxazole containing building blocks are synthesized. These compounds can serve as a versatile basis for the preparation of peptidomimetics comprising the oxazole structural unit that can enhance the proteolytic stability and bring structural rigidity. To this end, the compounds reported herein are designed to be selectively and orthogonally deprotectable which allows simple and efficient introduction of the oxazole structural unit and further peptide chain couplings as well as mild and selective final deprotection of the peptidomimetic to be synthesized, thus, enabling the fast and efficient development of new synthetic methodologies to oxazole-containing peptidomimetics.

9. Unpublished Work

9.1. Design and synthesis of oxazole containing peptide building blocks

The oxazole-based building blocks reported herein were designed to have functional groups at positions 2, 4 and 5 of the heterocycle in order to be able to couple an amino acid in any position of interest (Figure 47). In general, the building blocks are synthesized as *O*-linked oxazole isosteres comprising an enol ether-like structural unit, since it is known that β–γ- enol ethers may have properties as enzyme inhibitors. A hydroxyl group introduced by a serine residue in the C5 side chain allows hetero cyclization into an additional second oxazole that can give further development possibilities to create a mimic of a heterocyclic tandem pair, similar to the oxazole-thiazole moiety in Leucamide A (Figure 44). Another potential advantage thereof may be that heterocyclic tandem pairs are reported to play an important role in their bioactivity via specific interactions with DNA.[103]

Besides to oxazoles, also access to a methyloxazole structures is provided by creating the corresponding building blocks thereof which form a new tool for creating novel bioactive methyloxazole-based peptidomimetics. Multi functionalized oxazoles on the other hand for which also respective building blocks have been developed may give access to the

development of even more complex and branched peptidomimetic structures and can be used to mimic certain demanding bioactive structures like Almazole D.

For the target compounds synthesized in this work an orthogonal protecting group strategy was designed to be suitable for subsequent standard peptide chemistry protocols (Figure 47), wherein generally the side chain hydroxyl group of the L-serine basis was protected with the bulky *tert*-butyldiphenylsilyl (TBDPS) group which has been reported to offer a unique stability under acidic conditions compared to related silyl analogues.[126] This stability against acids is considered to be required during the deprotection of the *tert*-butyl protected carboxyl group at the C-terminus. On the other hand, TBDPS can be cleaved selectively under mild conditions using a fluoride source like TBAF (tetra-n-butylammonium fluoride) or HF in pyridine. Benzyloxycarbonyl (Cbz) was chosen as the protecting group of the N-terminus as it can be selectively cleaved via reductive hydrogenation on Pd/C. For branched peptidomimetics a phtalimide (Pht) protecting group was introduced on the amide in the R-residue at the C2 side chain of the oxazole moiety which can be selectively deprotected by using hydrazine.

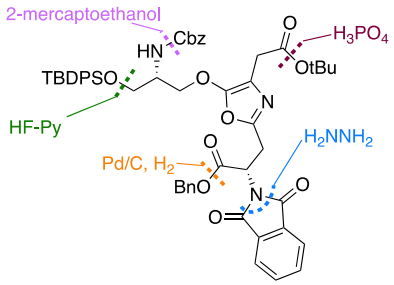


Figure 47. Oxazole containing target building block.

Additionally, in the course of the studies and during the development of aforementioned orthogonal protecting group strategy, two different protecting groups, tert-Butyldiphenylsilyl (TBDPS) and Acetyl (Ac), were investigated for the protection of the hydroxyl group of methyl((benzyloxy)carbonyl)serinate (**2** and **3** in Scheme 21) to study their respective influence on the oxazole synthesis . Furthermore, to add a potential functionality to C4 of the oxazole and to investigate the impact of a longer carbon chain on the oxazole-forming cyclization step as well as on the whole peptidomimetic structure, two amino acids, L-aspartic acid and L-glutamic acid, were coupled to TBDPS-protected compound **4**. To functionalize the

side chain at C2 of the oxazole, aspartic acid, glutamic acid and γ-butyric acid derivatives were coupled with the serine derivative **9** to give highly functionalized oxazole containing building blocks.

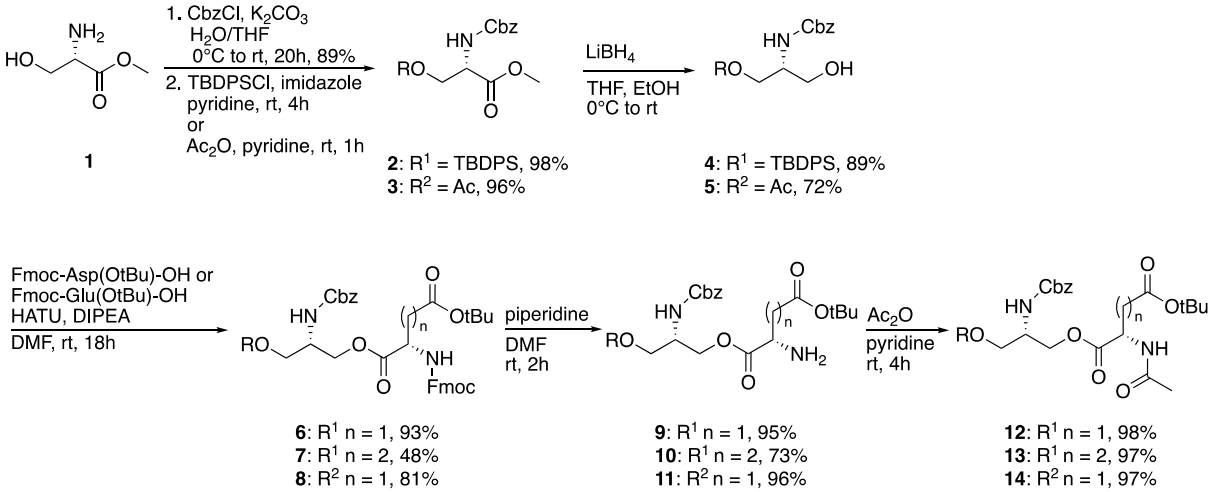
The general synthesis started with a Cbz-protection of the amine group of L-serine methyl ester (**1**) using benzylchloroformiate (Cbz-Cl) providing the protected compound with 89% yield as has been already described earlier (Scheme 21).[127] Subsequently, the remaining hydroxyl group was silylated using TBDPS-Cl giving **2** in quantitative yield. The respective acetyl protected derivative **3** was obtained in quantitative yield as well. These two protecting groups were chosen because of their different characteristics such as geometry, stability, and reactivity. Acetyl (Ac) is a small and planar group and therefore, it creates a minimal steric hindrance. On the other hand, the TBDPS group is a very bulky protecting group and due to the tetrahedral configuration of the silicon, its steric effect in comparison to the planar acetyl group is much more prominent. Moreover, due to its bulky substituents, a tert-butyl group (tBu) and two phenyl groups, TBDPS can also impact the reactivity of neighboring functional groups, in comparison, whereas Ac has almost no influence on the functional groups in its proximity. Nevertheless, *tert*-butyldiphenylsilyl is more stable against basic conditions and one of the most stable silyl-based protecting groups under acidic conditions, while acetyl groups are typically stable under acidic conditions, but are known to be more sensitive under basic/nucleophilic conditions.

In the next step, the methyl esters of **2** and **3** were reduced to a hydroxyl group (Scheme 21). TBDPS-protected compound **2** was successfully reduced with 2.0 equivalents of $LiBH_4$ under cooling to 0 °C for the first 30 min, and subsequent slow warming to room temperature over 18 h giving **4** with yield of 89%. Under the same conditions, ^{1}H-NMR analysis of the reduction product of compound **3**, surprisingly, revealed a symmetrical molecule with two hydroxyl groups indicating a double-reduced product. The room temperature presumably provides the necessary energy to overcome the activation barrier for the reduction of the acetyl group, which has a similar structure to the methyl ester in the same molecule. Moreover, the particular electronic environment given by neighboring amide and ester groups could also influence the electron density of the acetyl group, resulting in an easier reduction compared

to the methyl ester. After screening the reaction conditions, the best conversion was obtained with 1.5 eq of $LiBH_4$ at constant 0 °C for 22 h giving **5** in 72 % yield.

In the next step, an amide coupling was performed between the resulting alcohols **4** and **5,** respectively, and commercially available Fmoc-L-aspartic acid-4-*tert*-butylester using hexafluorophosphate azabenzotriazole tetramethyl uronium (HATU) as a coupling agent and *N,N*-diisopropylenethylamine (DIPEA) as a base to give the corresponding coupling products **6** with 93 % yield and **8** with 81 % yield. Another analogue was synthesized by coupling the alcohol **4** with commercially available Fmoc-L-glutamic acid-5-*tert*-butylester under similar conditions to obtain **7** in 48% yield**.** By elongating the carbon chain of the molecule, the influence of the increasing distance between the tert-butyl group and the space for intramolecular cyclization to be performed later should be investigated.

Thereafter, the Fmoc protecting group of **6**-**8** was removed with piperidine using a standard deprotection protocol to give **9**-**11** with yields of 95%, 73% and 96% respectively. In order to obtain the corresponding methyloxazole precursors, deprotected compounds **9-11** were acetylated with acetic anhydride to give **12-14** with quantitative yields.

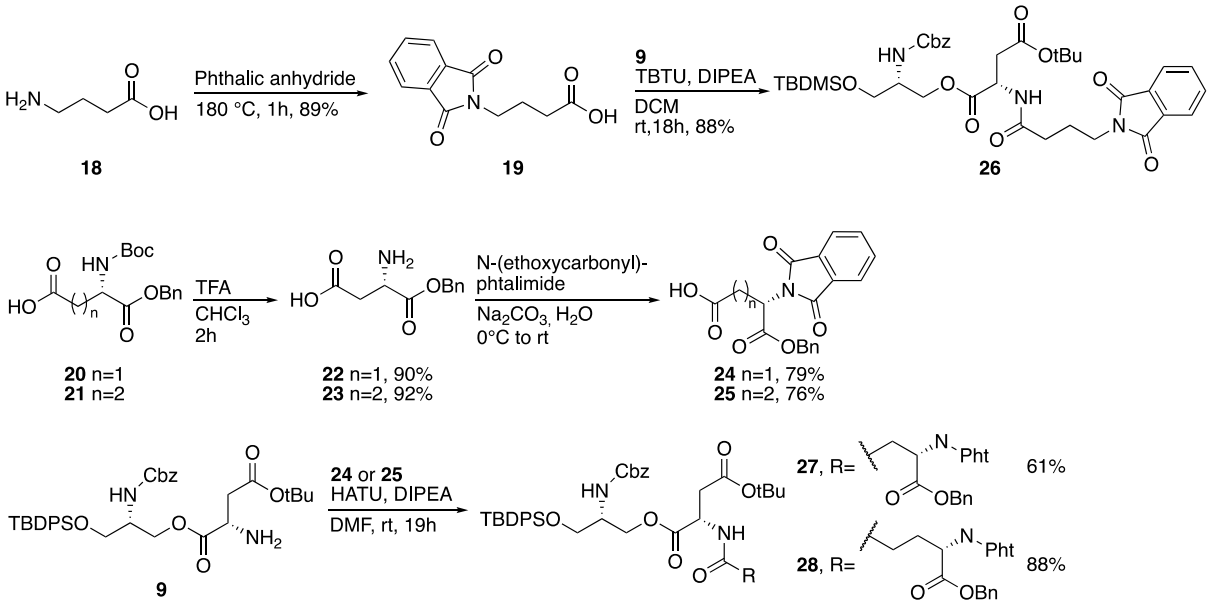


Scheme 21. Synthesis of precursor **9** for the further functionalization and precursors **12**-**14** for the synthesis of methyloxazoles.

To investigate the influence of the carbon chain length at C2-position of the oxazole on the cyclization yield, three precursors were synthesized. Additionally, it was thought that due to the rigidity of the oxazole moiety, the length of the carbon chain at C2 might be able to

introduce more structural flexibility for the desired peptidomimetic. For instance, lysin residues are believed to provide a comparably strong nucleophilic site and to improve the surface exposure of the final compounds in aqueous environment.

To synthesize the final precursors for the oxazole-forming cyclization, the amine group of γ-amino butyric acid (**18**) was first protected with phthalic anhydride giving **19** with 89% yield (Scheme 22). **19** was then further coupled with **9** giving lysin-like residue in **26** with 88% yield. On the other hand and besides to aforementioned lysin-mimicking structure, Asp and Glu residues should introduce an additional amide functionality allowing further coupling possibilities to amino acids of interest. Therefore, the amide of Boc-Asp-OBzl (**20**) and Boc-Glu-OBzl (**21**) were deprotected with TFA and subsequently reprotected with phtalimide using *N*-(ethoxycarbonyl)-phtalimide giving **24** and **25** with 79% and 76% yields, respectively (Scheme 22). The amide coupling between **9** and **24** afforded **27** with 61 % yield while the amide coupling between **9** and **23** gave **28** with 88% yield.



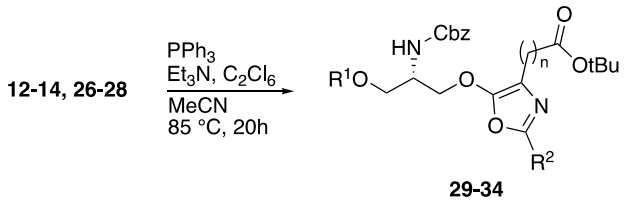
Scheme 22. Synthesis of precursors for oxazole C2 functionalization and to amide coupling with **9**.

In the next step the cyclization to introduce the oxazole moiety was performed. Oxazole synthesis is typically conducted under rather harsh conditions utilizing strong acids like concentrated sulfuric acid in the Robinson-Gabriel synthesis[128] or using triflic acid anhydride to mediate a cyclodehydration of *N*-acyl amino acid esters[129]. However, these conditions are limited to cyclization precursors with acid stable protecting groups. On the other hand,

structural elements containing β-hydroxy amides were shown to give C2 and C4 functionalized oxazolines using a fluoride source from diethylaminosulfur trifluoride (DAST) or Deoxo-Fluor in good yields.[130] However, subsequent oxidation of the functionalized oxazolines to oxazoles with 1,8-diazabicyclo(5.4.0)undec-7-ene (DBU) and $BrCCl_3$ provided the corresponding 2,4-difunctionalized oxazoles only in moderate yields. Another approach to obtain C2 and C4 functionalized oxazolines from β-hydroxy amides utilizes Burgess reagent[131] The oxazole moiety is obtained in via a dehydration and cyclization mechanism and also requires subsequent oxidation with DBU and CCl_4 to give 2, 4-functionalized oxazoles in moderate yields. The disadvantage of this method lies in its limitation to certain structures of β-hydroxy amides and, thus, is not generally applicable when aiming for multifunctionalized oxazoles.
Besides to the approaches mentioned above, mild Appel conditions were previously reported to form glycosylated oxazole derivatives in good yields.[132] These conditions were found to be tolerated by the sensitive *O*-glycosyl bond and thereby expected to be also suitable for the hydroxy like functionality at C5 of the target molecules (**29**-**34**) as well as the protecting groups used in this project. Methyloxazole-precursors **12-14** were subjected to cyclization under analogous Appel conditions using triphenyl phosphine (PPh_3) and hexachloroethane (C_2Cl_6) affording **29-31** with 42% to 57% yield (Scheme 23). Cyclization of the Ac-protected compound **14** yielded 57% of the cyclized product, while the TBDPS-protected peptide **12** yielded only 42%. The potential cause of this difference is the steric hindrance of the tert-butyldiphenylsilyl group compared to the acetyl group. Yamada *et al.*[133] reported for example that bulky protecting groups such as tert-butyldimethylsilyl (TBDMS) and tert-butyldiphenylsilyl (TBDPS) can affect the structure of some sugars and cause a conformational flip.

Mechanistically, the Appel reaction is believed to proceed via the activation of PPh_3 with a halogen source like C_2Cl_6. Chlorine-containing halogen sources are more reactive and the produced side products are typically more easily removable in comparison to iodine-based cyclization. Additionally, chlorine based activation has been reported to give better yields than cyclization with bromides.[134] Deprotonation of the tautomerized iminol promotes the nucleophilicity of α-hydroxyl group which attacks the triphenylphosphonium chloride introducing a good leaving group (Scheme 24). Intramolecular ring closure releases triphenylphosphine oxide and deprotonation of the oxazoline gives the final oxazole.

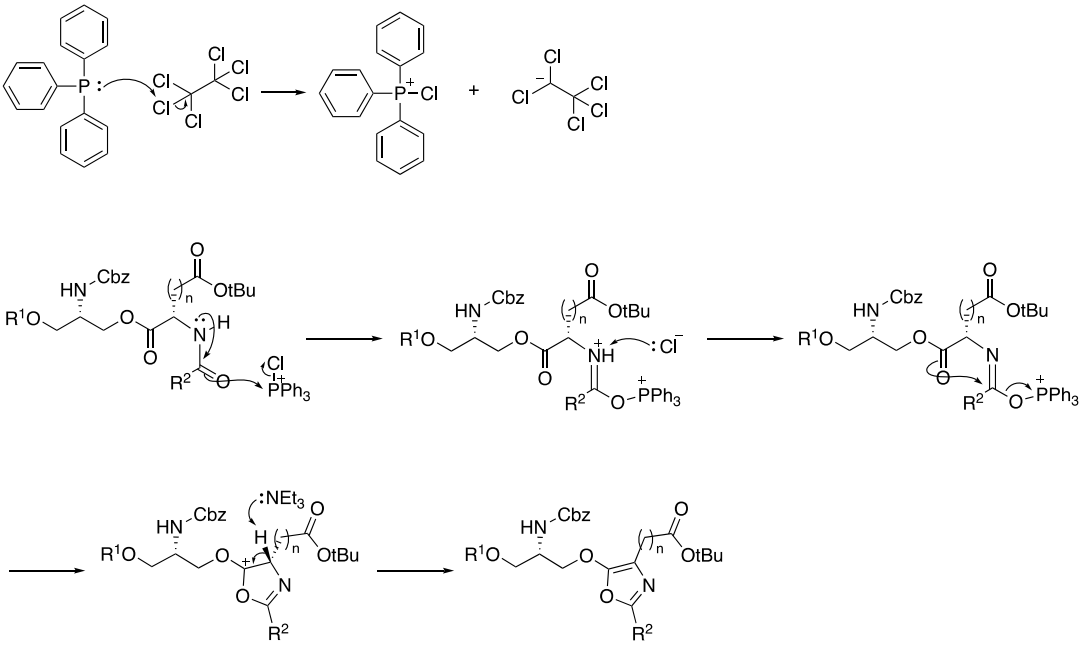
The additional CH_2-group (n=2) at precursor **13** seems to confer more flexibility to the molecule and to increase the distance of the bulky tert-butyl group from the cyclization site resulting in a higher yield (57%) of the cyclized product than cyclization of **12** to **29** (42%). The cyclization of **26**-**28** resulted in **32-34** with yields between 33-62% (Table 3). Unexpectedly, cyclization of **28** to **34** gave the lowest yield.



Scheme 23. Cyclization of the precursors to give methyloxazole- and C2 functionalized oxazole building blocks.

Table 3. Synthesized oxazoles **29-34**, sm refers to starting material.

Sm	n	R^1	R^2	Yield, %	Product
12	1	TBDPS	Me	42	**29**
13	2	TBDPS	Me	57	**30**
14	1	Ac	Me	57	**31**
26	1	TBDMS	$CH_2CH_2CH_2NHPht$	56	**32**
27	1	TBDPS	$CH_2CH(NHPht)CO_2Bn$	62	**33**
28	1	TBDPS	$CH_2CH_2CH(NHPht)CO_2Bn$	33	**34**

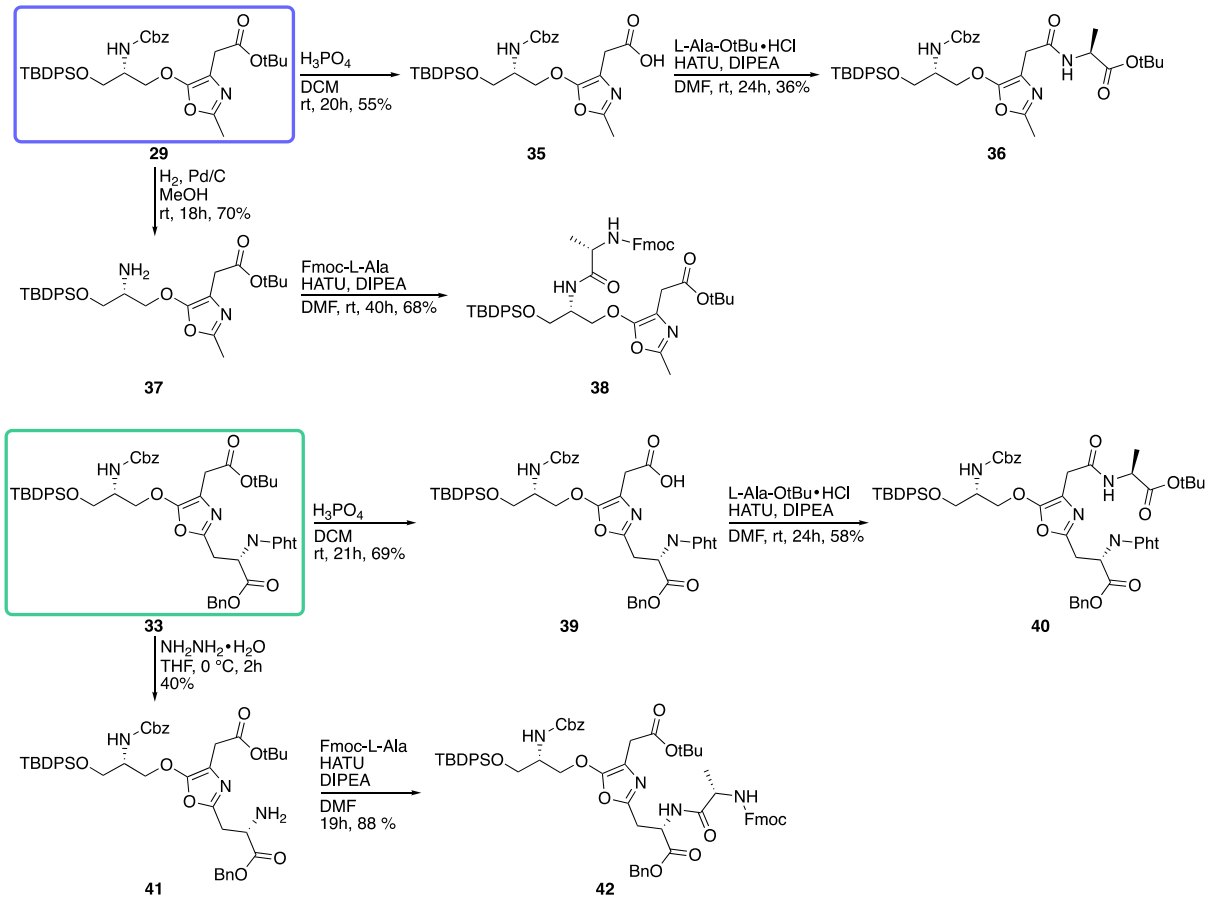


Scheme 24. Proposed reaction mechanism for the oxazole formation under Appel conditions.

The oxazoles **29** and **33** were tested for their applicability to orthogonal deprotection (Scheme 25). The *tert*-butyl ester was cleaved with 85% aq. H_3PO_4 giving carboxylates **35** and **39** with 55% and 69% yield accordingly. Additionally, the *tert*-butyl ester of **31** was deprotected under the same conditions giving 85 % yield. The Cbz group of **29** was cleaved via reductive hydrogenation on Pd/C giving amine **37** with 70% yield. To obtain a free amine in the C2 side chain of the oxazole **33**, the phtalimide group was cleaved using hydrazine monohydrate giving 40% of compound **41**.

Compounds **35**, **37**, **39** and **41** were then subjected to amide coupling reactions under common coupling conditions using HATU as coupling reagent (Scheme 25). Commercially available *L*-alanine-O*tBu* was coupled with **35** and **39** to give coupling products **36** and **40** with 36% and 58% yields accordingly. Also, Fmoc-L-Ala was used and coupled under the analogous conditions to **37** and **41** giving 68% of **38** and 88% of **42**.

These studies demonstrate that the building blocks **29** and **33** are well suited for further peptide synthesis and applications thereof. To improve the cyclization yields, more screening of different halogen sources could, however, be tested. Another potential modification to the reaction conditions could be the application of previously reported modified Appel conditions using PPh_3 in the presence of I_2 at -78 °C, which were reported to result in yields up to 73% which is slightly better than reported in this work.[131,135] Additionally, more optimization in the deprotection steps of the ready building blocks **29-34** and coupling conditions is needed to gain more satisfactory yields.

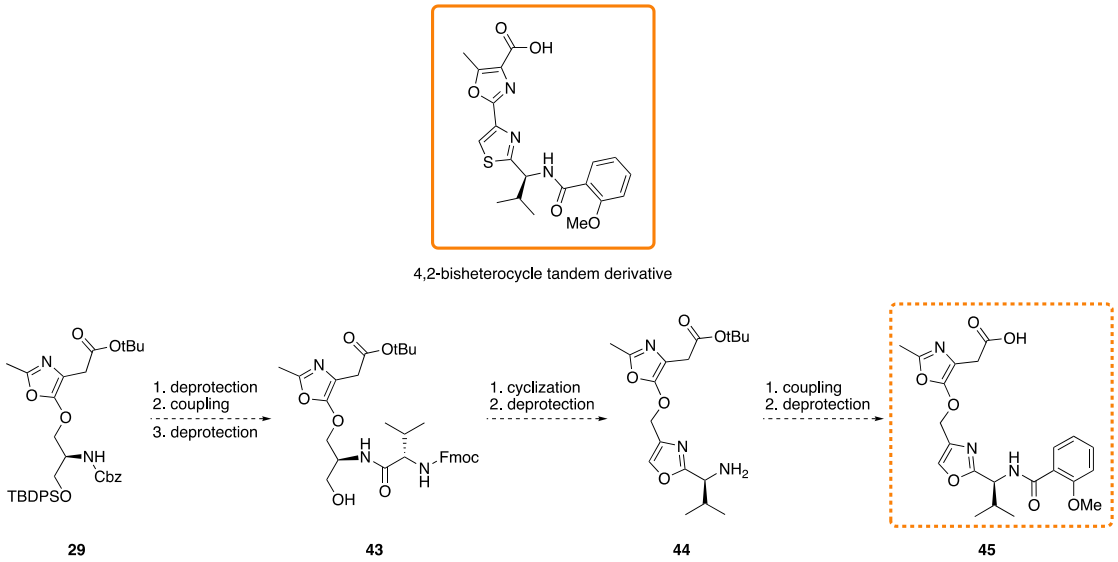


Scheme 25. Deprotection studies of the oxazole containing building block.

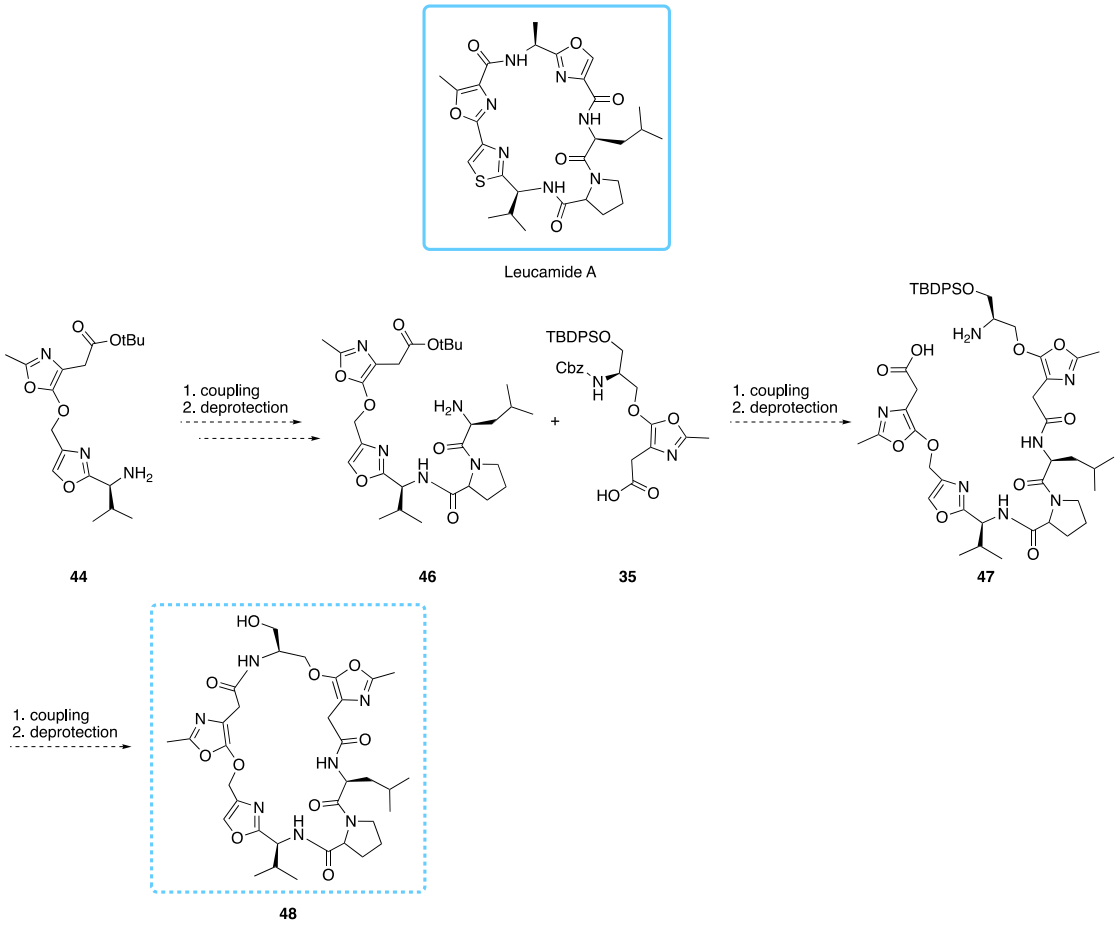
As regards to the next steps, this building block could be further modified to mimic structures of synthetic or naturally occurring oxazole containing peptides. Leucamide A as a structure of interest has shown its potential as an anti-influenza agent through specific interaction with DNA and RNA due to its bisheterocyclic structure. Compound **29** of this work could be synthetically modified to mimic this structure with simple deprotection and coupling steps to access **43** (Scheme 26). Thereby, side chain at C5 of the oxazole building block **29** would be further modified to incorporate an additional oxazole by first deprotecting the TBDPS-protected alcohol and to subsequently create a β-hydroxy amide **43** that can be cyclized, for example, via the DAST method as described earlier to obtain a bisheterocyclic structure (**44**). A standard amide coupling reaction and deprotection of the *tert*-butyl ester would afford an open chain mimic of Leucamide A **45**.

To access a cyclic peptide, the side chain of compound **44** could be elongated by two consecutive amide coupling steps to obtain **46** which is then coupled to a building block **35**

(Scheme 27). After cleaving the carbamate and *tert*-butyl ester, **47** could be subjected to amide coupling conditions and provide a cyclic structure **48**. These simple deprotection and coupling steps to modify synthesized building blocks **29-34** could give an access to variety of different peptidomimetics.



Scheme 26. Synthetic strategy to mimic an open chain derivative of Leucamide A.



Scheme 27. Synthetic strategy to mimic an oxazole containing cyclic peptide Leucamide A.

9.1.1. Project contributions

I was responsible for the synthesis and design of the target molecules.

9.1.2. Synthesized compounds and analytics

Materials and methods

Unless otherwise noted, all chemical transformations were performed using glassware that was heat-dried under high vacuum, utilizing the Schlenk technique under an atmosphere of nitrogen. Chemicals were purchased from Sigma-Aldrich, TCI, Thermo Scientific, OPC Orpegen, Novabiochem, ABCR, Carbosynth, or Acros Organics and used without further purification. Solvents that were dried over molecular sieve or stored under argon atmosphere were purchased from Acros Organics. The removal of solvents in vacuo was performed with a rotary evaporator, Laborota 4000 from Heidolph, where the water bath temperature was set to 50 °C unless otherwise noted.

Column chromatography

Column chromatographic purifications of obtained crude products were performed under slight nitrogen pressure on silica (25-70 µm particle size) from *Acros* Organics. Suitable eluents were determined via thin layer chromatography of the crude product.

Thin layer chromatography (TLC)

Reactions and column chromatography fractions were monitored by TLC on silica gel F_{254}TLC plates from Merck KGaA. Visualization of the developed TLC plates was achieved through UV-absorption or fluorescence under UV-lamp (254 nm and/or 366 nm) and/or through staining with $KMnO_4$ stain ($KMnO_4$ 3 g, K_2CO_3 10 g, ddH_2O 300 ml), Ninhydrin stain (Ninhydrin 0.1 g, acetic acid 0.5 ml, acetone 100 ml) or Seebach's stain (phosphomolybdic acid 2.5 g, $CeSO_4$ x H_2O 1.0g, conc. H_2SO_4 6.0 ml, ddH_2O 94 ml)[96]. Suitable eluents were determined via thin layer chromatography of the crude product.

Nuclear magnetic resonance spectroscopy (NMR-Spectroscopy)

NMR spectra were recorded on Bruker Avance III HD 400 (400 MHz) spectrometer. ^{1}H-NMR shifts were calibrated to the residual solvent resonances: $CDCl_3$ (7.26 ppm), CD_2Cl_2 (5.32 ppm)

or MeOD (3.31 ppm). ^{13}C-NMR shifts were calibrated to the residual solvent: $CDCl_3$ (77.16 ppm), CD_2Cl_2 (53.84 ppm), or MeOD (49.00 ppm). All NMR spectra were analyzed using the program MestReNova 14.3.1 from Mestrelab and NMR data are reported as follows: chemical shift (multiplicity, coupling constants where applicable, number of hydrogens, assignment). Splitting is reported with the following symbols: s = singlet, d = doublet, t = triplet, q = quartet, h = heptet, dd = doublet of doublets, dt = doublet of triplets, dq = doublet of quartets, ddd = doublet of doublets of doublets, m = multiplet. First-order chemical shifts and coupling constants were obtained from one-dimensional spectra; carbon and proton resonances were assigned from COSY, HSQC and HMBC experiments.

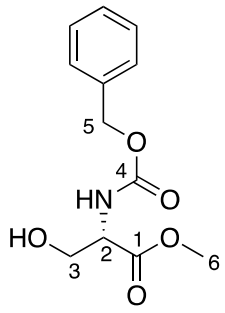
Mass spectrometry

High resolution mass spectra were measured by the analytical section of the Department of Chemistry of Ludwig-Maximilians-Universität München on *LTQ FTI-CR* by *Thermo Finnigan GmbH*.

Optical rotation

Optical rotation was measured on Krüss Optronic GmbH polarimeter and values were determined at the Sodium-D-line (589.5 nm). The temperature for each value is indicated in °C, the concentration in g/dl and the solvent in brackets.

Methyl ((benzyloxy)carbonyl)-L-serinate (**1**)[127]



11.3 g of $NaHCO_3$ (134 mmol, 3.2 eq.) was added to a suspension of 5.00 g of L-serine methylester in water (62 ml) and DCM (85 ml). Cbz-Cl was added slowly and reaction mixture

was stirred at room temperature for 16 h. Subsequently the reaction mixture was poured in ice water (100 ml) and organic layer was collected. Organic layer was washed with 0.5 M HCl (80 ml) and brine (2 x 80 ml). Organic layer was dried over $MgSO_4$ and volatiles were evaporated. The crude mixture was purified via column chromatography (EtOAc/cyclohexane 1:1) giving 7.86 g of **1** (31.0 mmol, 74%).

R_f (EtOAc/cyclohexane = 1:1) = 0.26;

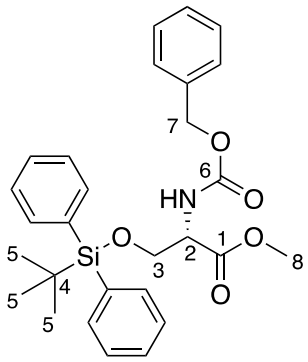
^{1}H NMR (400 MHz, CD_2Cl_2) δ 7.37 – 7.30 (s, 5H, ar), 5.10 (s, 2H, H5), 4.40 (dt, *J* = 7.8, 3.7 Hz, 1H, H2), 4.00 – 3.81 (m, 2H, H3), 3.74 (s, 3H, H6).

Values correspond to literature values J. N. Hernandez, V. S. Martin, *J. Org. Chem.* 2004, *69*, 3590–3592.

^{13}C NMR (101 MHz, CD_2Cl_2): δ = 171.59 (C1), 156.72 (C4), 137.00 (ar.), 129.00 (ar.), 128.65 (ar.), 128.48 (ar.), 67.49.12 (C5), 63.57 (C3), 56.62 (C2), 53.09 (C6).

HRMS (ESI$^+$): m/z calc. for ion $C_{12}H_{15}NNaO_5$ $[M+Na]^+$ = 276.0842; found m/z = 276.0844, D = 0.2 mDa.

Methyl *N*-((benzyloxy)carbonyl)-*O*-(*tert*-butyldiphenylsilyl)-L-serinate (**2**)[136]



2.5 ml of TBDPS-Cl (2.5 mL, 9.64 mmol, 1.3 eq.) and 753 mg of imidazole (11.0 mmol, 1.4 eq.) were added to 2.00 g of **1** (7.90 mmol, 1.0 eq.) in pyridine (32 ml). The reaction mixture was stirred at room temperature for 4 h. The solvent was removed under reduced pressure and the crude mixture was purified via column chromatography (EtOAc/cyclohexane = 1:3, then 1:2). The reaction yielded 4.00 g of **2** as a clear oil (8.14 mmol, quant.).

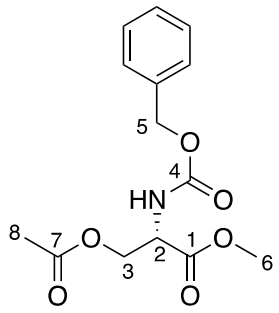
R_f (EtOAc/cyclohexane = 1:1) = 0.88;

^{1}H NMR (400 MHz, CD_2Cl_2): δ = 7.61 – 7.31 (m, 15H, ar), 5.69 (d, *J* = 8.9 Hz, 1H, NH), 5.10 (s, 2H, H7), 4.42 (dt, *J* = 8.7, 3.1 Hz, 1H, H2), 4.00 – 3.82 (m, 2H, H3), 3.73 (s, 3H, H8), 1.02 (s, 9H, H5);

^{13}C NMR (101 MHz, CD_2Cl_2): δ = 171.29 (C1), 156.24 (C6), 136.07 (ar.), 136.02 (ar.), 135.27 (ar.), 135.86 (ar.), 135.12 (ar.), 133.27 (ar.), 133.15 (ar.), 130.28 (ar.), 130.25 (ar.), 129.97 (ar.), 128.86 (ar.), 128.45 (ar.), 128.31 (ar.), 128.16 (ar.), 128.14 (ar.), 128.07 (ar.), 67.17 (C7), 64.84 (C3), 56.35 (C2), 52.70 (C8), 26.98 (C5), 19.48 (C4);

HRMS (ESI$^+$): m/z calc. for ion $C_{28}H_{33}NNaO_5Si$ $[M+Na]^+$ = 514.2026; found m/z = 514.2014, D = 1.2 mDa.

Methyl *O*-acetyl-*N*-((benzyloxy)carbonyl)-L-serinate (**3**)[137]



1.08 g of **1** (4.26 mmol, 1.0 eq.) was dissolved in pyridine (20 mL), DMAP (29.0 mg, 0.24 mmol, 0.06 eq.) and 0.50 ml acetic anhydride (5.29 mmol, 1.2 eq.) were added and the reaction mixture was stirred at room temperature for 1 h. The solvent was removed under reduced pressure and the crude mixture was purified through column chromatography (EtOAc/cyclohexane = 1:1). The reaction yielded **3** as clear oil (1.23 g, 4.15 mmol, 97%).

R_f (EtOAc/cyclohexane = 1:1) = 0.63;

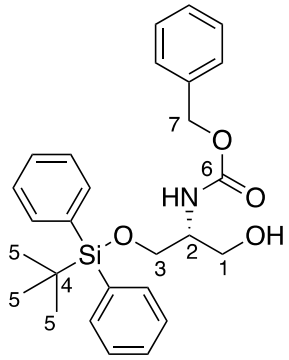
^{1}H NMR (400 MHz, CD_2Cl_2): δ = 7.39-7.30 (m, 5H, ar), 5.61 (d, *J* = 8.3 Hz, 1H, NH), 5.11 (s, 2H, H5), 4.60 (dt, *J* = 8.1, 3.8 Hz, 1H, H2), 4.41 (dd, *J* = 11.4, 4.0 Hz, 1H, H3), 4.33 (dd, *J* = 11.4, 3.7 Hz, 1H, H3), 3.75 (s, 3H, H6), 2.01 (s, 3H, H8);

^{13}C NMR (101 MHz, CD_2Cl_2): δ = 170.78 (C1), 170.29 (C7), 156.03 (C4), 136.83 (ar), 128.86 (ar), 128.53 (ar), 128.36 (ar), 67.37 (C5), 64.31 (C3), 53.69 (C2), 53.09 (C6), 20.79 (C8);

HRMS (ESI$^+$): m/z calc. for ion $C_{14}H_{17}NO_6Na$ $[M+Na]^+$ = 318.0954; found m/z = 318.0947, D = 0.7 mDa;

HRMS (ESI$^-$): m/z calc. for ion $C_{15}H_{17}NO_8$ $[M+COO]^-$ = 339.0954; found m/z = 339.2326, D = 137.2 mDa.

Benzyl (*R*)-(1-((*tert*-butyldiphenylsilyl)oxy)-3-hydroxypropan-2-yl)carbamate (**4**)[136]



4.00 g of **2** (8.14 mmol, 1.0 eq.) was dissolved in THF (30 mL) and absolute EtOH (45 mL). The solution was cooled to 0 °C and 354 mg of $LiBH_4$ (16.3 mmol, 2.0 eq.) was added. The reaction mixture was stirred at 0 °C for 30 min and slowly brought to room temperature and stirred for 20 h. The reaction was quenched with 0.5 M citric acid until no further gas evolution was observed. The crude mixture was concentrated under reduced pressure, subsequently suspended with water, and extracted with DCM (3 x 100 mL). Organic layer was dried over $MgSO_4$, and volatiles were removed under reduced pressure. The crude mixture was purified via column chromatography (EtOAc/cyclohexane = 1:3). The reaction yielded 3.37 g of **4** as a clear oil (7.28 mmol, 89%).

R_f (EtOAc/cyclohexane = 1:1) = 0.66;

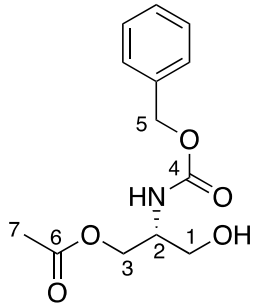
^{1}H NMR (400 MHz, CD_2Cl_2): δ = 7.67 – 7.30 (m, 15H, ar), 5.07 (s, 2H, H7), 3.84 – 3.67 (m, 5H, H1-3), 2.15 (bs, 1H, OH), 1.06 (s, 9H, H5);

^{13}C NMR (101 MHz, CD_2Cl_2): δ = 171.24 (C6), 137.23 (ar.), 135.95 (ar.), 135.94 (ar.), 133.35 (ar.), 133.34 (ar.), 130.32 (ar.), 130.30 (ar.), 128.86 (ar.), 128.41 (ar.), 128.31 (ar.), 128.23 (ar.), 67.01 (C7), 64.18 (C3), 63.31 (C1), 60.65 (C2), 27.03 (C5), 19.47 (C4);

HRMS (ESI$^+$): m/z calc. for ion $C_{27}H_{33}NNaO_4Si$ $[M+Na]^+$ = 486.2077; found m/z = 486.2065, D = 1.2 mDa;

HRMS (ESI$^-$): m/z calc. for ion $C_{27}H_{32}NO_4Si$ $[M-H]^-$ = 462.2101; found m/z = 462.2108, D = 0.7 mDa.

(*R*)-2-(((Benzyloxy)carbonyl)amino)-3-hydroxypropyl acetate (**5**)[138]



100 mg of **3** (0.34 mmol, 1.0 eq.) was dissolved in THF (1.25 mL), absolute EtOH (1.90 mL) was added, and mixture was cooled to 0 °C. 7.4 mg of $LiBH_4$ (0.34 mmol, 1.0 eq.) was added and the solution was stirred at 0 °C for 2 h. Then 3.7 mg of $LiBH_4$ (0.17 mmol, 0.5 eq.) was added and the reaction mixture was stirred at 0 °C for 20 h. 0.5 M citric acid was added to the reaction mixture until no further gas evolution was observed. The crude mixture was concentrated under reduced pressure, subsequently suspended with water and extracted with DCM (3 × 20 mL). Organic layer was dried over $MgSO_4$, and volatiles were removed under reduced pressure. The crude mixture was purified through column chromatography (EtOAc/cyclohexane = 1:1, then 2:1). The reaction yielded **5** as clear oil (65 mg, 0.24 mmol, 72%).

R_f (EtOAc/cyclohexane = 1:1) = 0.24;

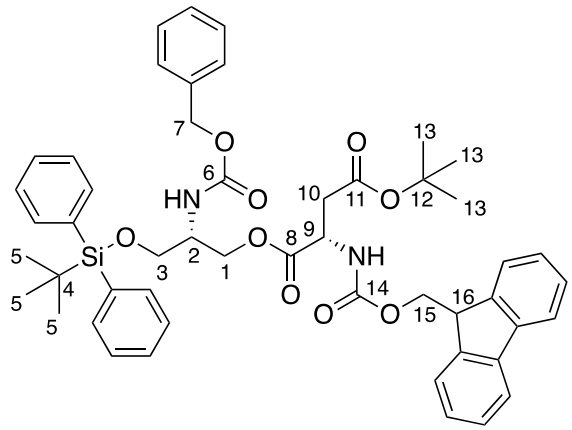
^{1}H NMR (400 MHz, CD_2Cl_2): δ = 7.39-7.28 (m, 5H, ar), 5.27 (d, 1H, NH), 5.09 (s, *J* = 7.0 Hz, 2H, H5), 4.18 (h, *J* = 5.7 Hz, 2H, H1), 3.95-3.88 (m, 1H, H2), 3.70-3.58 (m, 2H, H3), 2.35 (t, *J* = 6.1 Hz, 1H, OH), 2.04 (s, 3H, H7);

^{13}C NMR (101 MHz, CD_2Cl_2): δ = 171.22 (C6), 156.13 (C4), 136.63 (ar), 128.48 (ar), 128.09 (ar), 127.97 (ar), 66.77 (ar), 62.80 (C2), 61.65 (C1), 51.62 (C3), 20.58 (C7);

HRMS (ESI$^+$):m/z calc. for ion $C_{13}H_{17}NO_5Na$ $[M+Na]^+$ = 290.1004; found m/z = 290.0998, D = 0.6 mDa;

HRMS (ESI$^-$): m/z calc. for ion $C_{13}H_{17}NO_5Cl$ $[M+Cl]^-$ = 302.0795; found m/z = 302.0798, D = 0.3 mDa.

1-((R)-2-(((benzyloxy)carbonyl)amino)-3-((tert-butyldiphenylsilyl)oxy)propyl) 4-(tert-butyl) (((9H-fluoren-9-yl)methoxy)carbonyl)-L-aspartate (**6**)[139]



3.59 g of (*R*)-2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-4-(*tert*-butoxy)-4-oxobutanoic acid (8.73 mmol, 1.2 eq.) and 4.15 g of HATU (10.9 mmol, 1.5 eq.) were dissolved in DMF (10 mL) and the solution was stirred at room temperature for 10 min. 2.50 ml of DIPEA (14.6 mmol, 2.0 equiv.) was added and color change from white to yellow was observed. The flask was covered from light and the solution was stirred at room temperature for 20 min. 3.37 g of **4** (7.28 mmol, 1.0 eq.) in DMF (13 mL) was slowly added to the first solution and the reaction mixture was stirred at room temperature for 20 h. Reaction mixture was poured in sat. aq. NH_4Cl (5 mL) and the mixture was concentrated under reduced pressure. The residue was suspended with water and extracted with DCM (3 x 70 mL). Organic layer was dried over $MgSO_4$, and volatiles were removed under reduced pressure. The crude mixture was purified through column chromatography (EtOAc/cyclohexane = 1:3). The reaction yielded 5.78 g of **6** as a white foam (6.75 mmol, 93%).

R_f (EtOAc/cyclohexane = 1:3) = 0.55;

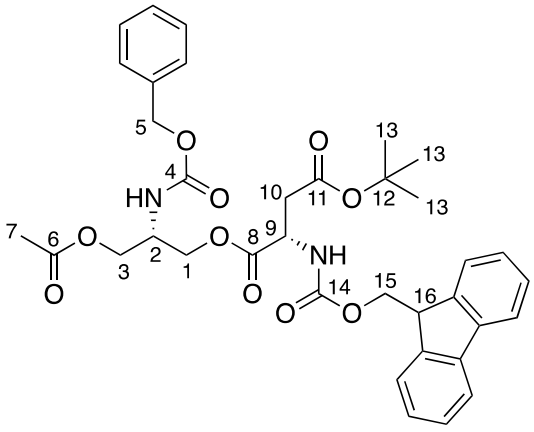
^{1}H NMR (400 MHz, CD_2Cl_2): δ = 7.78-7.27 (m, 23H, ar), 5.69 (d, *J* = 8.4 Hz, 1H, NH), 5.25 (d, *J* = 8.9 Hz, 1H, NH), 5.09-5.02 (m, 2H, H7), 4.49 (dd, *J* = 9.5, 4.4 Hz, 1H, H9), 4.39-4.28 (m, 4H, H1, H15), 4.23-4.19 (m, 1H, H16), 4.05-3.99 (m, 1H, H2), 3.76-3.68 (m, 2H, H3), 2.79-2.68 (m, 2H, H10), 1.39 (s, 9H, H13), 1.05 (s, 9H, H5);

^{13}C NMR (101 MHz, CD_2Cl_2): δ = 171.06 (C8), 170.29 (C11), 156.26 (C14), 156.18 (C6), 144.39 (ar.), 144.25 (ar.), 141.67 (ar.), 137.14 (ar.), 135.96 (ar.), 135.95 (ar.), 135.94 (ar.), 133.35 (ar.), 133.31 (ar.), 133.26 (ar.), 130.34 (ar.), 130.28 (ar.), 128.83 (ar.), 128.81 (ar.), 128.40 (ar.), 128.37 (ar.), 128.24 (ar.), 128.22 (ar.), 128.08 (ar.), 127.46 (ar.), 125.47 (ar.), 120.33 (ar.), 82.17 (C12), 67.44 (C15), 67.05 (C7) 64.41 (C1), 62.94 (C3), 51.64 (C2), 51.09 (C16), 47.54 (C9), 38.09 (C10), 28.14 (C13), 27.00 (C5), 19.49 (C4);

HRMS (ESI^+): m/z calc. for ion $C_{50}H_{56}N_2O_9NaSi$ $[M+Na]^+$ = 879.3653; found m/z = 879.3648, D = 0.5 mDa;

HRMS (ESI^-): m/z calc. for ion $C_{50}H_{56}ClN_2O_9Si$ $[M+Cl]^-$ = 891.3444; found m/z = 891.3461, D = 1.7 mDa.

1-((*S*)-3-Acetoxy-2-(((benzyloxy)carbonyl)amino)propyl) 4-(*tert*-butyl) (((9H-fluoren-9-yl) methoxy)carbonyl)-L-aspartate (**8**)



194 mg of (S)-2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-4-(tert-butoxy)-4-oxobutanoic acid (0.47 mmol, 1.2 eq.) and 224 mg of HATU (0.59 mmol, 1.5 eq.) were dissolved in DMF (0.8 mL) and the solution was stirred at room temperature for 10 min. DIPEA (0.14 mL, 0.80 mmol, 2.0 eq.) was added and color change from white to yellow was observed. The flask was covered from light and the solution was stirred at room temperature for 50 min. 105 mg of **5** (0.39 mmol, 1.0 eq.) in DMF (1 mL) was slowly added to the first solution and the reaction

mixture was stirred at room temperature for 20 h. The reaction mixture was poured aq. sat. NH_4Cl (0.2 mL) and concentrated under reduced pressure, subsequently suspended with water and extracted with DCM (3 × 30 mL). Organic layer was dried over $MgSO_4$, and volatiles were removed under reduced pressure. The crude mixture was purified through column chromatography (EtOAc/cyclohexane = 1:3, then 1:2). The reaction yielded **8** as a beige foam (211 mg, 0.32 mmol, 81%).

R_f (EtOAc/cyclohexane = 1:1) = 0.60;

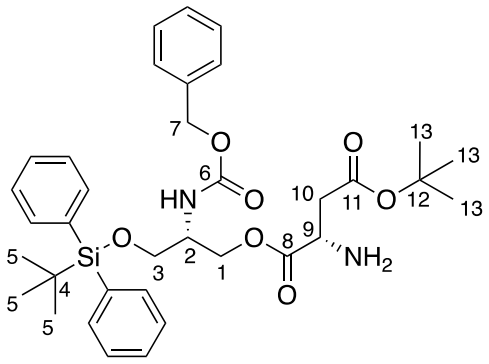
^{1}H NMR (400 MHz, CD_2Cl_2): δ = 7.79-7.29 (m, 13H, ar) 5.75 (dd, *J* = 8.8, 3.3 Hz, 1H, NH), 5.35 (m, 1H, NH), 5.07 (td, *J* = 11.5, 6.0 Hz, 2H, H5), 4.57-4.53 (m, 1H, H9), 4.38 (d, *J* = 7.2 Hz, 3H, H45, H2), 4.26-4.13 (m, 5H, H1, H3, H16), 2.93-2.75 (dd, 2H, H10), 2.01 (s, 3H, H7), 1.41 (s, 9H, H13);

^{13}C NMR (101 MHz, CD_2Cl_2): δ = 171.09 (C8), 170.86 (C14), 170.74 (C11), 156.28 (C6), 155.07 (C4), 144.93 (ar), 144.39 (ar), 144.26 (ar), 141.68 (ar), 128.85 (ar), 128.46 (ar), 128.42 (ar), 128.37 (ar), 128.10 (ar), 127.46 (ar), 125.48 (ar), 120.34 (ar), 82.52 (C12), 67.53(C5), 67.22 (C15), 64.45 (C1), 63.12 (C3), 51.10 (C2), 49.5 (C16), 47.54 (C9), 38.04 (C10), 28.15 (C13), 20.89 (C7);

HRMS (ESI$^+$): m/z calc. for ion $C_{36}H_{40}N_2O_{10}Na$ $[M+Na]^+$ = 683.2581; found m/z = 683.2566, D = 1.5 mDa;

HRMS (ESI$^-$): m/z calc. for ion $C_{36}H_{40}N_2O_{10}Cl$ $[M+Cl]^-$ = 695.2371; found m/z = 695.2370, D = 0.1 mDa.

1-((R)-2-(((benzyloxy)carbonyl)amino)-3-((tert-butyldiphenylsilyl)oxy)propyl) 4-(tert-butyl) L-aspartate (**9**)[140]



16 ml of piperidine (162 mmol, 24 eq.) was added to 5.78 g of **6** (6.75 mmol, 1.0 eq.) in DMF (70 ml). Reaction mixture was stirred at room temperature for 2 h. Volatiles were removed under reduced pressure and the crude mixture was purified via column chromatography (EtOAc/cyclohexane = 1:2 -> 1:1 -> 2:1 -> 3:1). The reaction yielded 4.21 g of **9** as a yellow oil (6.64 mmol, 95%).

R_f (EtOAc/cyclohexane = 1:3) = 0.28;

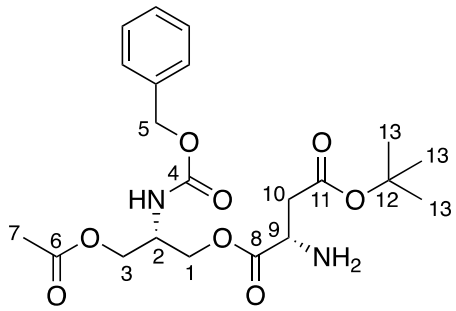
^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.63-7.28 (m, 15H, ar), 5.25-5.18 (m, 1H, NH), 5.08 (s, 2H, H7), 4.32-4.22 (m, 2H, H1), 4.08-4.02 (m, 1H, H2), 3.78-3.59 (m, 3H, H3, H9), 2.64-2.51 (m, 2H, H10), 1.40 (s, 9H, H13), 1.06 (s, 9H, H5);

^{13}C NMR (101 MHz, $CDCl_3$): 174.38 (C8), 170.54 (C11), 156.09 (C6), 136.47 (ar.), 135.65 (ar.), 135.52 (ar), 132.85 (ar.), 130.10 (ar.), 130.07 (ar.), 128.65 (ar.), 128.62 (ar.), 128.32 (ar.), 128.30 (ar.), 128.25 (ar.), 127.97 (ar.), 81.50 (C12), 66.99 (C7), 63.83 (C1), 62.64 (C3), 51.43 (C2), 51.28 (C9), 40.04 (C10), 28.14 (C13), 26.96 (C5), 19.36 (C4);

HRMS (ESI^+): m/z calc. for ion $C_{35}H_{47}N_2O_7Si$ $[M+H]^+$ = 635.3153; found m/z = 635.3142, D = 0.9 mDa;

HRMS (ESI^-): m/z calc. for ion $C_{35}H_{46}N_2O_7SiCl$ $[M+Cl]^-$ = 669.2763; found m/z = 669.2765, D = 0.2 mDa.

1-((*S*)-3-Acetoxy-2-(((benzyloxy)carbonyl)amino)propyl) 4-(*tert*-butyl) L-aspartate (**11**)



2.38g of **8** (3.60 mmol, 1.0 eq.) was dissolved in DMF (42 mL), 8.50 ml of piperidine (86.4 mmol, 24 eq.) was added and the reaction mixture was stirred at room temperature for 2 h. The solvent was coevaporated with toluene under reduced pressure till dryness and the crude mixture was purified through column chromatography (EtOAc/cyclohexane = 3:1, then pure EtOAc). The reaction yielded 1.52 g of **11** as a clear oil (3.47 mmol, 96%).

R_f (EtOAc/cyclohexane = 1:1) = 0.55;

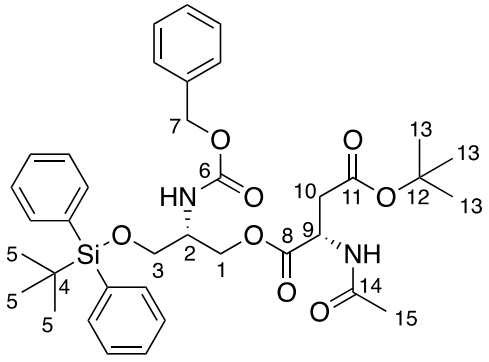
^{1}H NMR (400 MHz, CD_2Cl_2): δ = 7.38-7.29 (m, 5H, ar), 5.48 (d, *J* = 25.8 Hz, 1H, NH), 5.08 (q, 2H, H5), 4.35-4.07 (m, 5H, H1, H2, H3), 3.69 (q, *J* = 5.6 Hz, 1H, H9), 2.69-2.60 (m, 2H, H10), 2.02 (s, 3H, H7), 1.76 (s, 2H, NH), 1.41 (s, 9H, H13);

^{13}C NMR (101 MHz, CD_2Cl_2): δ = 174.67 (C6), 171.04 (C11), 170.90 (C8), 156.23 (C4), 137.05 (ar.), 128.84 (ar.), 128.49 (ar.), 128.39 (ar.), 81.59 (C12), 67.15 (C5), 63.94 (C1), 63.21 (C3), 51.72 (C2), 49.57 (C9), 40.52 (C10), 28.16 (C13), 20.89 (C7);

HRMS (ESI^+): m/z calc. for ion $C_{21}H_{30}N_2O_8Na$ $[M+Na]^+$ = 461.1900; found m/z = 461.1889, D = 1.1 mDa;

HRMS (ESI^-): m/z calc. for ion $C_{36}H_{40}N_2O_{10}Cl$ $[M+Cl]^-$ = 437.1924; found m/z = 437.1930, D = 0.6 mDa.

1-((*R*)-2-(((Benzyloxy)carbonyl)amino)-3-((*tert*-butyldiphenylsilyl)oxy)propyl)4-(*tert*-butyl) acetyl-L-aspartate (**12**)[141]



0.05 ml of acetic anhydride (0.56 mmol, 1.2 eq.) was slowly added to 298 mg of **9** (0.47 mmol, 1.0 eq.) in 2.4 ml of pyridine and. The reaction mixture was stirred at room temperature for 4 h. Subsequently the reaction mixture was poured in aq. sat. $NaHCO_3$ (0.4 mL). Solvents were removed under reduced pressure, the residue was suspended in water and extracted with DCM (3 x 10 mL). Organic layer was dried over $MgSO_4$, and volatiles were removed under reduced pressure. The reaction yielded 312 mg of **12** as a white foam (0.46 mmol, 96 %).

R_f (EtOAc/cyclohexane = 1:1) = 0.45;

^{1}H NMR (400 MHz, CD_2Cl_2): δ = 7.69-7.30 (m, 15H, ar), 6.46 (d, *J* = 8.1 Hz, 1H, H36), 5.13-5.05 (m, 2H, H11), 4.70 (dt, *J* = 8.0, 4.8 Hz, 1H, H9), 4.09-4.02 (m, 2H, H1), 4.07-3.98 (m, 1H, H2),

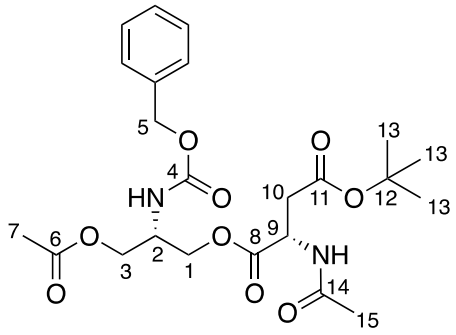
3.78-3.62 (m, 2H, H3), 2.81-2.68 (m, 2H, H10), 1.92 (s, 3H, H15), 1.39 (s, 9H, H13), 1.07 (s, 9H, H5);

^{13}C NMR (101 MHz, CD_2Cl_2): δ = 170.67 (C8), 170.14 (C11), 169.43 (C6), 155.85 (C14), 136.70 (ar), 135.50 (ar), 135.48 (ar), 132.90 (ar), 129.89 (ar), 129.84 (ar), 128.38 (ar), 128.35 (ar), 127.95 (ar), 127.90 (ar.), 127.82 (ar.), 127.78 (ar.), 81.61 (C12), 66.57 (C7), 63.87 (C1), 62.54 (C3), 51.22 (C2), 48.83 (C9), 37.31 (C10), 27.66 (C13), 26.55 (C5), 22.71 (C15), 19.04 (C4);

HRMS (ESI$^+$): m/z calc. for ion $C_{37}H_{49}N_2O_8Si$ $[M+H]^+$ = 677.3258; found m/z = 677.3250, D = 0.8 mDa;

HRMS (ESI$^-$): m/z calc. for ion $C_{37}H_{48}N_2O_8SiCl$ $[M+Cl]^-$ = 711.2868; found m/z = 711.2849, D = 1.9 mDa.

1-((*S*)-3-Acetoxy-2-(((benzyloxy)carbonyl)amino)propyl) 4-(*tert*-butyl) acetyl-L-aspartate (**14**)



0.08 ml of acetic anhydride (0.85 mmol, 1.2 eq.) was slowly added to 305 mg of **11** (0.70 mmol, 1.0 eq.) in pyridine (3.5 mL). The reaction mixture was stirred at room temperature for 1.5 h. The reaction mixture was poured into aq. sat. $NaHCO_3$ (0.3 mL). Solvents were removed under reduced pressure, subsequently the residue was resuspended with water and extracted with DCM (3 × 30 mL). Organic layer was dried over $MgSO_4$, and volatiles were removed under reduced pressure. The reaction yielded **13** as a yellowish oil (328 mg, 0.68 mmol, 98%).

R_f (EtOAc/cyclohexane = 3:1) = 0.34;

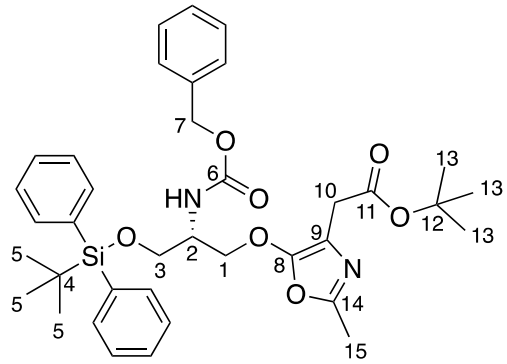
^{1}H NMR (400 MHz, CD_2Cl_2): δ = 7.38-7.28 (m, 5H, ar), 6.59-6.56 (m, 1H, NH), 5.55-5.52 (m, 1H, NH), 5.13-5.04 (m, 2H, H5), 4.75 (dt, *J* = 8.1, 4.8 Hz, 1H, H9), 4.41-4.06 (m, 5H, H1, H2, H3), 2.92-2.71 (m, 2H, H10), 2.02 (s, 3H, H7), 1.97 (s, 3H, H15), 1.40 (s, 9H, H13);

^{13}C NMR (101 MHz, CD_2Cl_2): δ = 171.17 (C11), 171.14 (C8), 171.08 (C6), 170.46 (C14), 156.27 (C4), 137.06 (ar), 128.87 (ar), 128.44 (ar), 128.34 (ar), 82.22 (C12), 67.16 (C5), 64.37 (C1), 63.10 (C3), 49.52 (C2), 49.28 (C9), 37.69 (C10), 28.12 (C13), 23.12 (C15), 20.88 (C7);

HRMS (ESI^+): m/z calc. for ion $C_{23}H_{32}N_2O_9Na$ $[M+Na]^+$ = 503.2006; found m/z = 503.1993, D = 1.3 mDa;

HRMS (ESI^-): m/z calc. for ion $C_{23}H_{32}N_2O_9Cl$ $[M+Cl]^-$ = 515.1796; found m/z = 515.1804, D = 0.8 mDa.

tert-Butyl (R)-2-(5-(2-(((benzyloxy)carbonyl)amino)-3-((tert-butyldiphenylsilyl)oxy)propoxy)-2-methyloxazol-4-yl)acetate (**29**)[134]



204 mg of **12** (0.30 mmol, 1.0 eq.) was dissolved in acetonitrile (2 mL). Solution of triphenylphosphine (412 mg, 1.57 mmol, 5.2 eq.) and hexachloroethane (199 mg, 0.84 mmol, 2.8 eq.) in acetonitrile (3 mL) was added to the first solution and it was refluxed at 85 °C for 30 min. Upon addition of triethylamine (0.38 mL, 2.78 mmol, 9.2 eq.) gas formation and colour change from light to darker yellow were observed. The reaction mixture was refluxed at 85 °C for 23 h. Reaction mixture was filtered through celite, and volatiles were removed under reduced pressure. The crude mixture was purified through column chromatography (EtOAc/cyclohexane = 1:2). The reaction yielded 84 mg of **29** as a yellow oil (0.13 mmol, 42%).

R_f (EtOAc/cyclohexane = 1:1) = 0.67;

^{1}H NMR (400 MHz, CD_2Cl_2): δ = 7.66-7.28 (m, 15 H, ar), 5.41 (d, *J* = 8.3 Hz, 1H, NH), 5.06 (s, 2H, H7), 4.36-4.24 (m, 2H, H1), 4.11-4.09 (m, 1H, H2), 3.83-3.73 (m, 2H, H3), 3.21 (s, 2H, H10), 2.26 (s, 3H, H15), 1.38 (s, 9H, H13), 1.05 (s, 9H, H5);

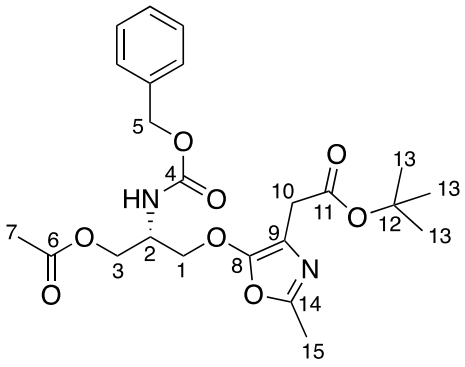
^{13}C NMR (101 MHz, CD_2Cl_2): δ = 170.18 (C11), 156.15 (C6), 155.20 (C8), 152.68 (C14), 137.13 (ar.), 135.89 (ar.), 133.40 (ar.), 133.31 (ar.), 130.24 (ar.), 130.22 (ar.), 128.82 (ar.), 128.38 (ar.), 128.29 (ar.), 128.17 (ar.), 110.52 (C9), 81.34 (C12), 72.90 (C1), 66.97 (C7), 62.53 (C3), 52.29 (C2), 32.28 (C10), 28.09 (C13), 26.95 (C5), 19.48 (C4), 14.35 (C15);

HRMS (ESI^+): m/z calc. for ion $C_{37}H_{46}NaN_2O_7Si$ $[M+Na]^+$ = 681.2972; found m/z = 681.2962, D = 1 mDa;

HRMS (ESI^-): m/z calc. for ion $C_{37}H_{45}N_2O_7Si$ $[M-H]^-$ = 657.2996; found m/z = 657.3004, D = 0.8 mDa;

Optical rotation: α_D^{23}= 0.0014 (c = 12.80, $CHCl_3$).

tert-Butyl (*R*)-2-(5-(3-acetoxy-2-(((benzyloxy)carbonyl)amino)propoxy)-2-methyloxazol-4-yl)-acetate (**31**)



151 mg of **14** (0.31 mmol, 1.0 eq.) was dissolved in acetonitrile (3.2 mL). 165 mg of triphenylphosphine (0.63 mmol, 2.0 eq.) and 89.0 mg of hexachloroethane (0.38 mmol, 1.2 eq.) in acetonitrile (1.5 mL) were added to the first solution and the mixture was stirred at room temperature for 25 min. Upon addition of triethylamine (0.18 mL, 1.29 mmol, 4.1 equiv.) color change from light to dark yellow and then to orange when heated was observed. The reaction mixture was refluxed at 85 °C for 19 h. Triethylamine (0.04 mL, 0.29 mmol, 0.9 equiv.) was added and the reaction mixture was refluxed at 85 °C for 4.5 h. The solution was filtered through celite and volatiles were removed under reduced pressure. The crude mixture was purified through column chromatography (EtOAc/cyclohexane = 2:1). The reaction yielded 83 mg of **31** as a yellow oil (0.18 mmol, 57%).

R_f (EtOAc/cyclohexane = 3:1) = 0.61;

^{1}H NMR (400 MHz, CD_2Cl_2): δ = 7.39-7.28 (m, 5H, ar), 5.54-5.53 (m, 1H, NH), 5.10 (d, *J* = 2.9 Hz, 2H, H5), 4.30-4.10 (m, 5H, H1-H3), 3.26 (s, 2H, H10), 2.27 (s, 3H, H15), 2.03 (s, 3H, H7), 1.42 (s, 9H, H13);

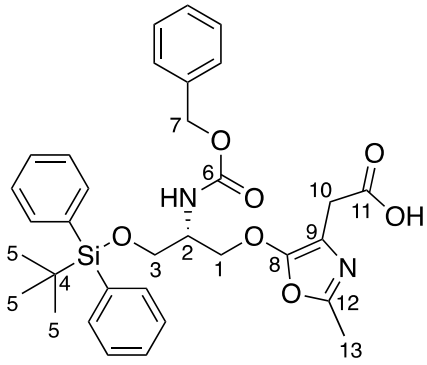
^{13}C NMR (101 MHz, CD_2Cl_2): δ = 170.89 (C11), 170.21 (C6), 156.15 (C4), 155.01 (C6), 152.82 (C14), 136.99 (ar), 128.85 (ar), 128.47 (ar), 128.35 (ar), 110.81 (C9), 81.48 (C12), 73.22 (C1), 67.15 (C5), 62.89 (C3), 50.06 (C2), 32.33 (C10), 28.11 (C13), 20.90 (C7), 14.38 (C15);

HRMS (ESI^+): m/z calc. for ion $C_{23}H_{30}N_2O_8Na$ $[M+Na]^+$ = 485.1900; found m/z = 485.1888, D = 0.2 mDa;

HRMS (ESI^-): m/z calc. for ion $C_{23}H_{30}N_2O_8Cl$ $[M+Cl]^-$ = 497.1691; found m/z = 497.1699, D = 0.8 mDa;

Optical rotation: α_D^{23}= - 0.0015 (c = 0.67, $CHCl_3$).

(*R*)-2-(5-(2-(((Benzyloxy)carbonyl)amino)-3-((tert-butyldiphenylsilyl)oxy)propoxy)-2-methyloxazol-4-yl)acetic acid (**35**)



29 (127 mg, 0.19 mmol, 1.0 equiv.) was dissolved in DCM (0.4 mL) and H_3PO_4 (0.07 mL, 1.04 mmol, 5.4 equiv.) was added. The reaction mixture was stirred at room temperature for 30 min and TLC control (EtOAc/cyclohexane = 1:1) showed full conversion of the starting material. The reaction was quenched with water (3 mL) and extracted with DCM (3 × 15 mL). Organic layer was dried over $MgSO_4$, and volatiles were removed under reduced pressure. The reaction yielded **35** as a yellow oil (64 mg, 0.11 mmol, 55%).

R_f (EtOAc/cyclohexane = 1:1) = 0.14;

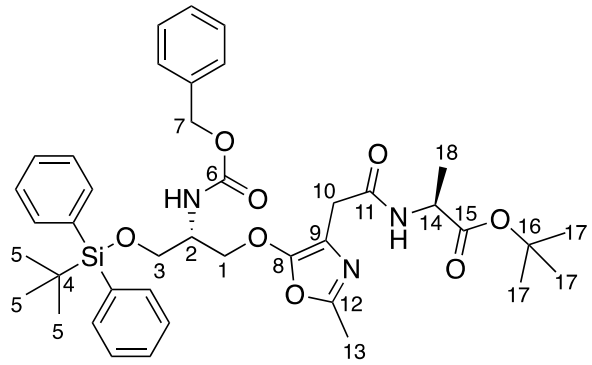
^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.65-7.31 (m, 15H), 5.26 (d, *J* = 8.8 Hz, 1H, NH), 5.09 (s, 2H, H7), 4.34-4.23 (m, 2H, H1), 4.10-4.08 (m, 1H, H2), 3.87-3.67 (m, 2H, H3), 3.38 (s, 2H, H10), 2.30 (s, 3H, H13), 1.06 (s, 9H, H5);

^{13}C NMR (101 MHz, $CDCl_3$): δ = 173.19 (C11), 156.04 (C6), 154.70 (C8), 153.25 (C12), 136.31 (ar.), 135.59 (ar.), 134.92 (ar.), 132.91 (ar.), 132.79 (ar.), 130.07 (ar.), 128.65 (ar.), 128.32 (ar.), 128.26 (ar.), 127.95 (ar.), 127.79 (ar.), 109.29 (C9), 72.46 (C1), 67.09 (C7), 62.11 (C3), 51.70 (C2), 30.15 (C10), 26.82 (C5), 19.36 (C4), 14.09 (C13);

HRMS (ESI$^+$): m/z calc. for ion $C_{33}H_{39}N_2O_7Si$ $[M+H]^+$ = 603.2527; found m/z = 603.2521, D = 0.6 mDa;

HRMS (ESI$^-$): m/z calc. for ion $C_{33}H_{37}N_2O_7Si$ $[M-H]^-$ = 601.2370; found m/z = 601.2356, D = 1.4 mDa.

tert-Butyl (2-(5-((*R*)-2-(((benzyloxy)carbonyl)amino)-3-((tert-butyldiphenylsilyl)oxy)propoxy)-2-methyloxazol-4-yl)acetyl)-L-alaninate (**36**)



35 (50 mg, 0.08 mmol, 1.0 equiv.) was dissolved in DMF (0.5 mL), HATU (95 mg, 0.25 mmol, 3.0 equiv.) and DIPEA (0.06 mL, 0.33 mmol, 4.0 equiv.) were added. Colour change from orange to brown was observed and the solution was stirred at room temperature for 20 min. *tert*-Butyl L-alaninate (36 mg, 0.20 mmol, 2.4 equiv.) was added and the reaction mixture was stirred at room temperature for 19 h. TLC control (EtOAc/cyclohexane = 1:1) showed full conversion of the starting material. The reaction was quenched with aqueous saturated NH_4Cl (0.5 mL), subsequently suspended in water, and extracted with DCM (3 × 30 mL). Organic layer was dried over $MgSO_4$, and volatiles were removed under reduced pressure. The crude mixture was purified through column chromatography (EtOAc/cyclohexane = 1:2, then stepwise to pure EtOAc). The reaction yielded **36** as an orange oil (22 mg, 0.03 mmol, 36%).

R_f (EtOAc/cyclohexane = 1:1) = 0.34;

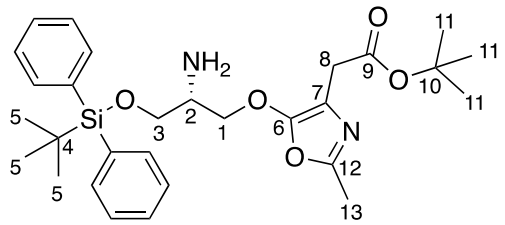
^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.64-7.28 (m, 15H, ar) 6.88 (d, *J* = 7.3 Hz, 1H, NH), 5.50 (d, *J* = 8.6 Hz, 1H, NH), 5.08 (s, 2H, H7), 4.50-4.24 (m, 3H, H1, H14), 4.09-4.07 (m, 1H, H2), 3.86-3.70 (m, 2H, H3), 3.24 (s, 2H, H10), 2.32 (s, 3H, H13), 1.42 (s, 9H, H17), 1.32 (d, 3H, H18), 1.05 (s, 9H, H5);

^{13}C NMR (101 MHz, $CDCl_3$): δ = 172.11 (C15), 168.77 (C11), 155.97 (6), 154.58 (C8), 152.77 (C12), 136.38 (ar), 135.54 (ar), 135.52 (ar), 132.92 (ar), 132.79 (ar), 129.94 (ar), 129.92 (ar), 128.53 (ar), 128.14 (ar), 128.11 (ar), 127.85 (ar), 127.84 (ar), 109.73 (C9), 81.87 (C16), 72.38 (C1), 66.84 (C7), 62.06 (C3), 51.79 (C2), 48.88 (C14), 32.40 (C10), 27.95 (C17), 26.84 (C5), 19.29 (4), 18.61 (C18), 14.30 (C13);

HRMS (ESI$^+$): m/z calc. for ion $C_{40}H_{52}N_3O_8Si$ $[M+H]^+$ = 730.3518; found m/z = 730.3320, D = 19.8 mDa;

HRMS (ESI$^-$): m/z calc. for ion $C_{40}H_{51}N_3O_8SiCl$ $[M+Cl]^-$ = 764.3; found m/z = 764.1, D = 0.2 Da.

tert-Butyl (R)-2-(5-(2-amino-3-((tert-butyldiphenylsilyl)oxy)propoxy)-2 methyloxazol-4-yl)acetate (**37**)



29 (164 mg, 0.25 mmol, 1.0 equiv.) was dissolved in methanol (7.5 mL) and Pd/C (8.00 mg, 0.08 mmol, 5 wt%) was added. Reaction was set under H_2 atmosphere and stirred at room temperature for 23 h. TLC control (EtOAc/cyclohexane = 1:1) showed only little conversion of the starting material. Pd/C (9.00 mg, 0.08 mmol, 5 wt%) was added and reaction was stirred at room temperature for 23.5 h. TLC control (EtOAc/cyclohexane = 1:1) showed full conversion of the starting material. The solution was filtered through celite, and volatiles were evaporated. The crude mixture was purified through column chromatography (EtOAc/cyclohexane = 2:1, then stepwise to pure EtOAc). The reaction yielded **37** as a yellow oil (92 mg, 0.18 mmol, 70%).

R_f (EtOAc/cyclohexane = 1:1) = 0.11;

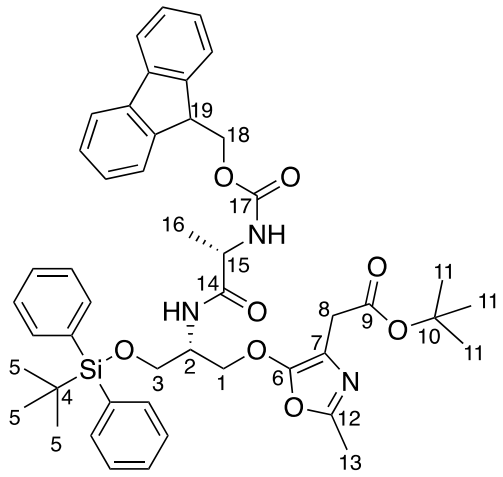
^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.68-7.35 (m, 10H, ar), 4.48 (s, 1H, NH), 4.28-4.14 (m, 2H, H1), 3.80-3.72 (m, 2H, H3), 3.37-3.31 (m, 3H, H2, H8), 2.31 (s, 3H, H13), 1.42 (s, 9H, H11), 1.06 (s, 9H, H5);

^{13}C NMR (101 MHz, $CDCl_3$): δ = 170.34 (C9), 154.83 (C6), 152.60 (C12), 135.68 (ar.), 135.65 (ar.), 132.92 (ar.), 130.07 (ar.), 128.15 (ar.), 127.99 (ar.), 110.15 (C7), 81.57 (C10), 74.64 (C1), 63.79 (C3), 52.46 (C2), 32.09 (C8), 28.19 (C11), 26.97 (C5), 19.40 (C4), 14.44 (C13);

HRMS (ESI$^+$): m/z calc. for ion $C_{29}H_{41}N_2O_5Si$ $[M+H]^+$ = 525.2779; found m/z = 525.2759, D = 2 mDa;

Optical rotation: α_D^{23}= + 0.0010 (c = 1.59, $CHCl_3$).

tert-Butyl 2-(5-((*R*)-2-((*S*)-2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)propanamido)-3-((tert-butyldiphenylsilyl)oxy)propoxy)-2-methyloxazol-4-yl)acetate (**38**)



(((9H-fluoren-9-yl)methoxy)carbonyl)-L-alanine (138 mg, 0.42 mmol, 2.4 equiv.) and HATU (282 mg, 0.74 mmol, 4.2 equiv.) were dissolved in DMF (0.6 mL) and DIPEA (0.24 mL, 1.38 mmol, 7.9 equiv.) was added. Upon addition of DIPEA colour change from white to yellow was observed. The solution was stirred at room temperature for 20 minutes. **37** (92 mg, 0.18 mmol, 1.0 equiv.) was dissolved in DMF (0.4 mL) and the first solution (0.3 mL) was added. The reaction mixture was covered from light and stirred at room temperature for 19.5 h. TLC control (EtOAc/cyclohexane = 1:1) showed only partial conversion of the starting material. A second solution of (((9H-fluoren-9-yl)methoxy)carbonyl)-L-alanine (69 mg,

0.21 mmol, 1.2 equiv.), HATU (138 mg, 0.36 mmol, 2.1 equiv.) and DIPEA (0.12 mL, 0.69 mmol, 3.9 equiv.) dissolved in DMF (0.4 mL) was stirred at room temperature for 20 minutes and 0.2 mL were added to the reaction mixture. The reaction mixture was stirred at room temperature for 20.5 h and TLC control (EtOAc/cyclohexane = 1:1) showed full conversion of the starting material. The reaction was quenched with aqueous saturated NH_4Cl (0.5 mL), and solvent was removed under reduced pressure, subsequently suspended with water and extracted with DCM (3 × 30 mL). Organic layer was dried over MgSO4, and volatiles were removed under reduced pressure. The crude mixture was purified through column chromatography (EtOAc/cyclohexane = 1:3, then 1:2). The reaction yielded **38** as a white foam (97 mg, 0.12 mmol, 68%).

R_f (EtOAc/cyclohexane = 1:1) = 0.52 ;

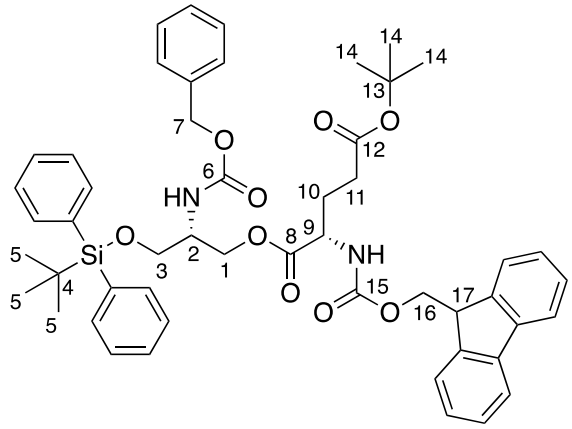
^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.66-7.27 (m, 18H, ar), 6.88 (d, *J* = 8.3 Hz, 1H, NH), 5.44 (d, *J* = 6.8 Hz, 1H, NH), 4.46-4.27 (m, 5H, H2, H1, H18) , 4.20-4.10 (m, 2H, H15, H19), 3.82-3.68 (m, 2H, H3), 3.24 (s, 2H, H8), 2.29 (s, 3H, H13), 1.38 (s, 9H, H411), 1.23 (d, 3H, H16) 1.06 (s, 9H, H5);

^{13}C NMR (101 MHz, $CDCl_3$): δ = 172.16 (C9), 170.85 (C14), 155.79 (C17), 155.17 (C6), 152.27 (C12), 143.92 (ar.) , 141.42 (ar.), 135.69 (ar.), 135.65 (ar.), 133.04 (ar.), 130.10 (ar.), 128.00 (ar.), 127.98 (ar.), 127.85 (ar.), 127.20 (ar.), 125.25 (ar.), 120.12 (ar.), 109.29 (C7), 81.90 (C10), 71.62 (C1), 67.07 (C18) 61.33 (C3), 50.58 (C2), 50.42 (C15), 47.25 (C19), 32.02 (C8), 28.11 (C11), 26.92 (C5), 19.37 (C4), 14.37 (C16), 14.35 (C13);

HRMS (ESI$^+$): m/z calc. for ion $C_{47}H_{56}N_3O_8Si$ $[M+H]^+$ = 818.3837; found m/z = 818.3809, D = 2.8 mDa;

HRMS (ESI$^+$): m/z calc. for ion $C_{48}H_{56}N_3O_{10}Si$ $[M+HCOO^-]^-$ = 862.4; found m/z = 862.9, D = 0.5 Da.

1-((*R*)-2-(((Benzyloxy)carbonyl)amino)-3-((*tert*-butyldiphenylsilyl)oxy)propyl) 5-(*tert*-butyl) (((9H-fluoren-9-yl)methoxy)carbonyl)-L-glutamate (**7**)[139]



(*S*)-2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-5-(*tert*-butoxy)-5-oxopentanoic acid (1.00 g, 2.35 mmol, 1.2 equiv.) and HATU (1.23 g, 3.23 mmol, 1.7 equiv.) were dissolved in DMF (1 mL) and the solution was stirred at room temperature for 10 min. DIPEA (0.75 mL, 4.31 mmol, 2.2 equiv.) was added to the solution and color change from white to yellow was observed. The solution was stirred at room temperature for 20 min. **4** (888 mg, 1.92 mmol, 1.0 equiv.) was dissolved in DMF (6.9 mL) and added dropwise to the previous solution. The reaction mixture was covered from light and stirred at room temperature for 8.3 h. TLC control (EtOAc/cyclohexane = 1:1) showed full conversion of the starting material and the reaction was quenched with saturated aqueous NH_4Cl (3 mL). The crude mixture was concentrated under reduced pressure, subsequently suspended with water and extracted with DCM (3 × 60 mL). Organic layer was dried over $MgSO_4$, and volatiles were removed under reduced pressure. The crude mixture was purified through column chromatography (EtOAc/cyclohexane = 1:3). The reaction yielded **7** as a white foam (796 mg, 0.91 mmol, 48%).

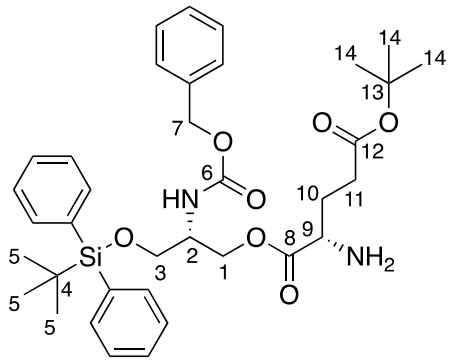
R_f (EtOAc/cyclohexane = 1:1) = 0.87 ;
^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.76-7.28 (m, 23H, ar), 5.47 (d, *J* = 7.9 Hz, 1H, NH), 5.25 (d, *J* = 9.7 Hz, 1H, NH), 5.07 (s, 2H, H7), 4.37-4.17 (m, 6H, H1, H9, H16, H17), 4.05-3.99 (m, 1H, H2), 3.75-3.67 (m, 2H, H3), 2.34-2.17 (m, 2H, H11), 2.10-1.87 (m, 2H, H10), 1.43 (s, 9H, H14), 1.05 (s, 9H, H5) ;
^{13}C NMR (101 MHz, $CDCl_3$): δ = 172.09 (C12), 171.81 (C8, C15), 156.02 (C6), 143.89 (ar.), 143.72 (ar.), 141.32 (ar.), 141.30 (ar.), 136.35 (ar.), 135.57 (ar.), 135.55 (ar.), 135.19 (ar.), 134.82 (ar.),

132.64 (ar.), 130.01 (ar.), 129.96 (ar.), 129.68 (ar.), 128.52 (ar.), 128.15 (ar.), 127.89 (ar.), 127.87 (ar.), 127.74 (ar.), 127.10 (ar.) , 125.10 (ar.), 119.99 (ar.), 80.93 (C13), 67.02 (C16), 66.89 (C7), 64.02 (C1), 62.49 (C3), 53.54 (C9), 51.17 (C2), 47.14 (C17), 31.36 (C11), 28.08 (C14), 26.87 (C10), 26.57 (C5), 19.26 (C4) ;

HRMS (ESI^+): m/z calc. for ion $C_{51}H_{62}N_3O_9Si$ $[M+NH_4]^+$ = 888.4255; found m/z = 888.4192, D = 63 mDa;

HRMS (ESI^-): m/z calc. for ion $C_{52}H_{59}N_2O_{11}Si$ $[M+HCOO^-]^-$ = 915.4; found m/z = 915.0, D = 0.4 Da.

1-((*R*)-2-(((Benzyloxy)carbonyl)amino)-3-((*tert*-butyldiphenylsilyl)oxy)propyl) 5-(*tert*-butyl)-L-glutamate (**10**)[140]



7 (796 mg, 0.91 mmol, 1.0 equiv.) was dissolved in DMF (10.5 mL), piperidine (2.2 mL, 22.3 mmol, 24 equiv.) was slowly added and the reaction mixture was stirred at room temperature for 18.5 h. TLC control (EtOAc/cyclohexane = 1:1) showed full conversion of the starting material. Volatiles were removed under reduced pressure and the crude mixture was purified through column chromatography (EtOAc/cyclohexane = 1:1, then 2:1). The reaction yielded **10** as a clear oil (435 mg, 0.67 mmol, 73%).

R_f (EtOAc/cyclohexane = 1:1) = 0.39;

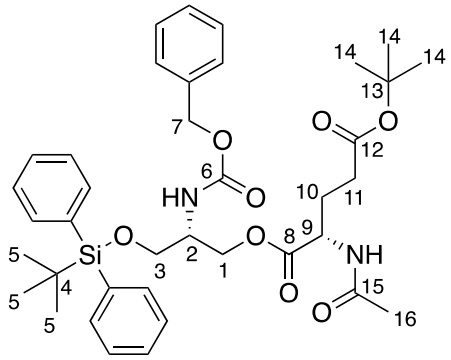
^1H NMR (400 MHz, $CDCl_3$): δ = 7.64-7.30 (m, 15H, ar), 5.16(d, *J* = 9.3 Hz, 1H, NH), 5.09 (s, 2H, H7), 4.30-4.22 (m, 2H, H1), 4.06 (m, 1H, H2), 3.72 (qd, *J* = 10.3, 4.2 Hz, 2H, H1), 3.32 (dd, *J* = 8.2, 5.2 Hz, 1H, H9), 2.31 (t, *J* = 7.5 Hz, 2H, H11), 1.93-1.74 (m, 2H, H10), 1.42 (s, 9H, H14), 1.05 (s, 9H, H5);

^{13}C NMR (101 MHz, $CDCl_3$): δ = 175.69 (C8), 172.47 (C12), 156.09 (C6), 136.49 (ar.), 135.68 (ar.), 135.65 (ar.), 132.89 (ar.), 132.80 (ar.), 130.11 (ar.), 130.09 (ar.), 128.68 (ar.), 128.65 (ar.), 128.34 (ar.), 128.32 (ar.), 127.99 (ar.), 80.56 (C13), 67.02 (C7), 63.79 (C1), 62.77 (C3), 53.86 (C9), 51.22 (C2), 31.84 (C11), 29.70 (C10), 28.21 (C14), 26.99 (C5), 19.39 (C4);

HRMS (ESI^+): m/z calc. for ion $C_{36}H_{49}N_2O_7Si$ $[M+H]^+$ = 649.3304; found m/z = 649.3296, D = 0.8 mDa;

HRMS (ESI^-): m/z calc. for ion $C_{36}H_{48}N_2O_7SiCl$ $[M+Cl]^-$ = 683.2919; found m/z = 683.2910, D = 0.9 mDa.

1-((*R*)-2-(((Benzyloxy)carbonyl)amino)-3-((*tert*-butyldiphenylsilyl)oxy)propyl) 5-(*tert*-butyl) acetyl-L-glutamate (**13**)[141]



10 (415 mg, 0.64 mmol, 1.0 equiv.) was dissolved in pyridine (3.2 mL) and acetic anhydride (0.07 mL, 0.74 mmol, 1.2 mmol) was added. The reaction mixture was stirred for 19.5 h and TLC control (EtOAc/cyclohexane = 1:1) showed full conversion of the starting material. The reaction was quenched with aqueous saturated $NaHCO_3$ (0.8 mL), and the solvent was removed through coevaporation with toluene under reduced pressure. Subsequently the crude mixture was suspended in water and extracted with DCM (3 × 30 mL), volatiles were removed under reduced pressure. The reaction yielded **13** as a yellow foam (430 mg, 0.62 mmol, 97%).

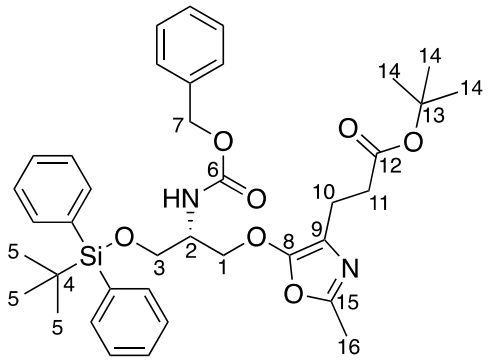
R_f (EtOAc/cyclohexane = 1:1) = 0.45;

^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.63-7.29 (m, 15H, ar), 6.29 (d, *J* = 7.7 Hz, 1H, NH), 5.26 (d, *J* = 8.9 Hz, 1H, NH), 5.09 (s, 2H, H7), 4.48 (td, *J* = 7.9, 5.0 Hz, 1H, H9), 4.35-4.30 (m, 2H, H1), 4.08-

4.02 (m, 1H, H2), 3.80-3.70 (m, 2H, H1), 2.35-2.17 (m, 2H, H11), 2.09-1.87 (m, 2H, H10), 1.43 (s, 9H, H14), 1.05 (s, 9H, H5);

^{13}C NMR (101 MHz, $CDCl_3$): δ = 172.39 (C8), 171.80 (C12), 170.08 (C15), 156.02 (C6), 136.37 (ar), 135.56 (ar), 135.54 (ar), 132.81 (ar), 132.65 (ar), 130.03 (ar), 129.96 (ar), 128.54 (ar), 128.16 (ar), 128.13 (ar), 127.90 (ar), 127.87 (ar), 80.99 (C13), 66.86 (7), 64.07 (C1), 62.57 (C3), 52.08 (C9), 51.21 (C2), 31.43 (C11), 29.73 (C10), 28.06 (C14), 26.87 (C5), 23.08 (C16), 19.27 (C4).

tert-Butyl (*R*)-3-(5-(2-(((benzyloxy)carbonyl)amino)-3-((*tert*-butyldiphenylsilyl)oxy)propoxy)-2-methyloxazol-4-yl)propanoate (**27**)[134]



13 (157 mg, 0.23 mmol, 1.0 equiv.) was dissolved in acetonitrile (3 mL). Solution of triphenylphosphine (155 mg, 0.59 mmol, 2.6 equiv.) and hexachloroethane (76 mg, 0.32 mmol, 1.4 equiv.) in acetonitrile (1 mL) were added to the first solution and the mixture was refluxed at 85 °C for 20 min. Upon addition of triethylamine (0.15 mL, 1.08 mmol, 4.7 equiv.) gas formation and colour change from yellow to light brown were observed. The reaction mixture was refluxed at 85 °C for 3 h and TLC control (EtOAc/cyclohexane = 1:1) showed partial conversion of the starting material. A solution of triphenylphosphine (78 mg, 0.30 mmol, 1.3 equiv.) and hexachloroethane (38.0 mg, 0.16 mmol, 0.7 equiv.) in acetonitrile (0.5 mL) and triethylamine (0.10 mL, 0.72 mmol, 3.16 equiv.) were added to the reaction mixture. The reaction mixture was refluxed at 85 °C for 21 h and TLC control (EtOAc/cyclohexane = 1:1) showed partial conversion of the starting material. Triphenylphosphine (156 mg, 0.59 mmol, 2.6 equiv.) and hexachloroethane (74 mg, 0.31 mmol, 1.4 equiv.) were added. The reaction mixture was refluxed at 85 °C for 1 h and triethylamine (0.15 mL, 1.08 mmol, 4.7 equiv.) was added. The reaction mixture was refluxed

at 85 °C for 2 h and TLC control (EtOAc/cyclohexane = 1:1) showed full conversion of the starting material. Volatiles were removed under reduced pressure. The crude mixture was dissolved in DCM and purified through column chromatography (EtOAc/cyclohexane = 1:3). The reaction yielded **27** as a yellow oil (88 mg, 0.13 mmol, 57%).

R_f (EtOAc/cyclohexane 1:1) = 0.75;

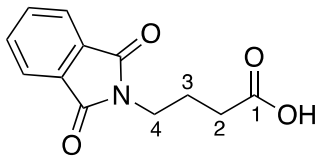
^{1}H NMR (400 MHz, $CDCl_3$): δ 7.65-7.29 (m, 15H, ar), 5.40 (d, *J* = 8.7 Hz, 1H, NH), 5.09 (s, 2H, H7), 4.29 (dd, *J* = 9.7, 4.7 Hz, 1H, H1a), 4.19 (dd, *J* = 9.6, 5.5 Hz, 1H, H1b), 4.12-4.08 (m, 1H, H2), 3.83-3.78 (m, 2H, H3) 2.60-2.44 (m, 4H, H10, H11), 2.28 (s, 3H, H16), 1.38 (s, 9H, H14), 1.06 (s, 9H, H5);

^{13}C NMR (101 MHz, $CDCl_3$): δ = 172.31 (C12), 156.04 (C6), 153.53 (C8), 152.38 (C15), 136.51 (ar.), 135.63 (ar.), 135.62 (ar.), 133.08 (ar.), 132.95 (ar.), 130.00 (ar.), 129.98 (ar.), 128.62 (ar.), 128.24 (ar.), 128.21 (ar.), 127.93 (ar.), 127.91 (ar.), 114.79 (C9), 80.37 (C13), 72.35 (C1), 66.92 (C7), 62.21 (C3), 51.86 (C2), 34.00 (C11), 28.16 (C14), 26.94 (C5), 20.25 (C10), 19.39 (C4), 14.35 (C16);

HRMS (ESI$^+$): m/z calc. for ion $C_{38}H_{49}N_2O_7Si$ $[M+H]^+$ = 673.3304; found m/z = 673.3295, D = 0.9 mDa;

HRMS (ESI$^-$): m/z calc. for ion $C_{36}H_{48}N_2O_7SiCl$ $[M+Cl]^-$ = 707.2919; found m/z = 707.2912, D = 0.7 mDa.

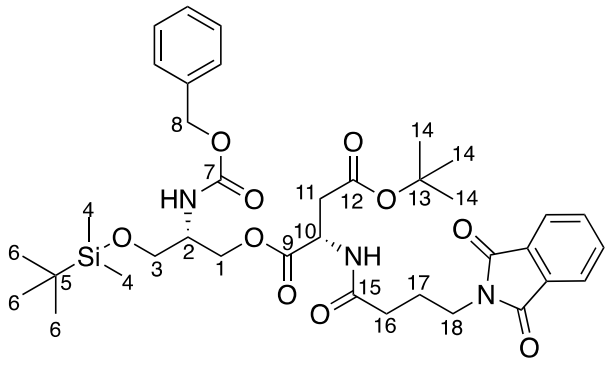
4-(1,3-dioxoisoindolin-2-yl)butanoic acid (**19**)



3.00 g of g-aminobutyric acid (29.0 mmol, 1 eq) and 4.50 g phtalic anhydride (31.0 mmol, 1.05 eq) were dried under high vacuum for 15 h and subsequently heated to 180 °C for 30 min (at 130 °C phtalic anhydride started melting). The mixture was poured in 36 ml of H_2O and the product precipitated as white solid. The solid was filtered and washed with 1 M HCl and H_2O. The product was dried under high vacuum at 85 °C for 5 h and at room temperature for 15 h.

^{1}H NMR (400 MHz, DMSO-d_6) δ 7.91 – 7.76 (m, 4H, ar), 3.60 (t, *J* = 6.8 Hz, 2H, H3), 2.27 (t, *J* = 7.2 Hz, 2H, H2), 1.81 (p, *J* = 7.0 Hz, 2H, H3);
^{13}C NMR (101 MHz, DMSO) δ 173.91 (C1), 168.05 (pht C=O), 134.34 (Pht), 131.72 (Pht), 123.02 (Pht), 36.94 (C4), 30.99 (C2), 23.34 (C3).

1-((*R*)-2-(((benzyloxy)carbonyl)amino)-3-((tert-butyldiphenylsilyl)oxy)propyl) 4-(tert-butyl) (4-(1,3-dioxoisoindolin-2-yl)butanoyl)-L-aspartate (**26**)

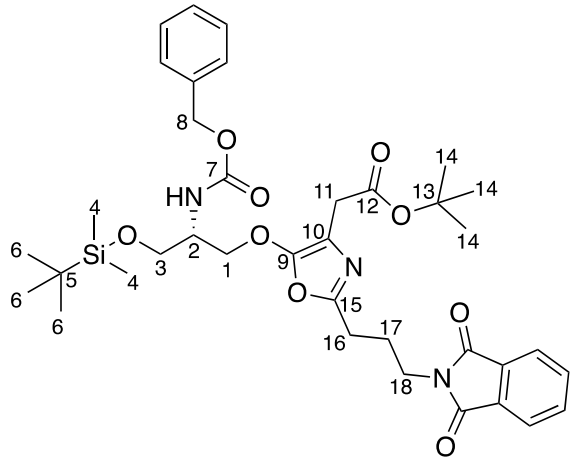


67.0 mg of TBTU (0.21 mmol, 1.5 eq) and 0.05 ml of DIPEA (35.0 mg, 0.27 mmol, 2.0 eq) were added to a solution of **9** in DCM (1.2 ml) and mixture was stirred at room temperature for 20 min. 4-(1,3-dioxoisoindolin-2-yl)butanoic acid (**19**) in DCM (0.5 ml) was added and reaction mixture was stirred at room temperature covered from light for 18h. Subsequently, reaction mixture was diluted with DCM (4 ml) and poured in 1 M HCl (4 ml) and organic phase was collected and washed with aq. sat. $NaHCO_3$ (2 x 5 ml) and H_2O (5 ml). Organic phase was dried over $MgSO_4$, and volatiles were evaporated. The crude product was purified via column chromatography (EtOAc/cyclohex 1:3 -> 1:2 -> 1:1) giving 88 % of **26** (90 mg, 1.24 mmol).

Rf (EtOAc/iHex 1:1) = 0.56;
^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.88-7.65 (m, 4H, Pht), 7.38-7.27 (m, 5H, ar), 6.85 (d, *J* = 6.4 Hz, 1H, NH), 5.37 (d, *J* = 8.8 Hz, 1H, NH), 5.08 (s, 2H, H8), 4.82 (dt, *J* = 7.9, 4.7 Hz, 1H, H10), 4.40-4.17 (m, 2H, H1), 4.05-3.90 (m, 1H, H2), 3.81-3.57 (m, 4H, H3, H18), 2.95-2.67 (m, 2H, H11), 2.29-1.94 (m, 4H, H16, H17), 1.41 (s, 9H, H14), 0.86 (s, 9H, H6), 0.03 (s, 6H, H4);
HRMS (ESI$^+$): m/z calc. for ion $C_{37}H_{51}N_3NaO_{10}Si$ $[M+Na]^+$ = 748.3241; found m/z = 748.3226, D = 1.5 mDa;

HRMS (ESI$^-$): m/z calc. for ion $C_{37}H_{51}N_3O_{10}ClSi$ $[M+Cl]^-$ = 760.3032; found m/z = 760.3016, D = 1.6 mDa.

tert-butyl (*R*)-2-(5-(2-(((benzyloxy)carbonyl)amino)-3-((*tert*-butyldiphenylsilyl)oxy)propoxy)-2-(3-(1,3-dioxoisoindolin-2-yl)propyl)oxazol-4-yl)acetate (**32**)



72 mg of PPh_3 (0.28 mmol, 2.3 eq) and 33 mg of C_2Cl_6 (0.14 mmol, 1.2 eq) were added to 90 mg of **26** (0.12 mmol, 1.0 eq) in DCM (2 ml). 0.08 ml of Et_3N (0.56 mmol, 57 mg, 4.6 eq) was added and reaction mixture was heated to 40 °C for 22 h. Subsequently the reaction mixture was diluted with DCM and filetered through celite and volatiles were evaporated. The crude mixture was purified via column chromatography (3:1 -> 2:1 -> 1:1) giving 50 mg of **32** (0.071 mmol, 56%).

Rf (EtOAc/iHex 1:1) = 0.67

^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.89-7.63 (m, 4H, Pht), 7.35-7.30 (m, 5H, ar), 5.42 (d, *J* = 8.7 Hz, 1H, NH), 5.08 (s, 2H, H8), 4.31-4.09 (m, 2H, H1), 4.08-3.93 (m, 1H, H2), 3.85-3.64 (m, 4H, H3, H18), 3.25 (s, 2H, H11), 2.68 (t, *J* = 7.5 Hz, 2H, H16), 2.10 (t, *J* = 7.6 Hz, 2H, H17), 1.41 (s, 9H, H14), 0.87 (s, 9H, H6), 0.05 (s, 6H, H4);

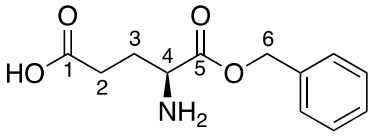
^{13}C NMR (101 MHz, CD_3OD): δ = 171.46 (C12), 168.24 (C=OPht), 155.94 (C7), 154.92 (C9), 136.33 (ar), 133.99 (Pht), 132.05 (Pht), 128.54 (ar), 128.18 (ar), 126.02 (ar), 123.27 (Pht), 109.86 (C10), 72.33 (C1), 66.86, 64.95, 61.19 (C3), 51.75 (C2), 37.14 (C18), 31.82 (C11), 28.03 (C14), 25.97 (C14), 25.93 (C16), 25.84 (C6), 25.60 (C17), 18.24, 1.04, -5.51 (C4);

HRMS (ESI$^+$): m/z calc. for ion $C_{37}H_{49}N_3NaO_9Si$ $[M+Na]^+$ = 730.3136; found m/z = 730.3125, D = 1.5 mDa;

HRMS (ESI$^-$): m/z calc. for ion $C_{37}H_{48}N_3O_9Si$ $[M-H]^-$ = 706.3160; found m/z = 706.31598, D = 0.02 mDa;

Optical rotation: α_D^{23}= - 0.0012 (c = 1.86, $CHCl_3$).

(*S*)-4-Amino-5-(benzyloxy)-5-oxopentanoic acid (**23**)[142]



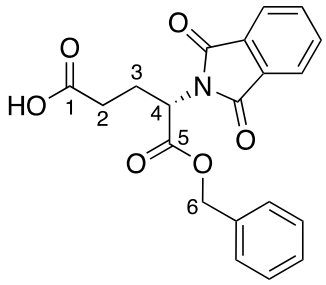
2.00 g of (*S*)-5-(benzyloxy)-4-((*tert*-butoxycarbonyl)amino)-5-oxopentanoic acid (5.93 mmol, 1.0 eq.) was dissolved in 9 ml of chloroform and 9 ml of TFA was added. The reaction mixture was stirred at room temperature for 1.5 h. 0.5 ml of triethylamine was added, resulting in the gas evolution. Solvents were removed under reduced pressure and the residue was dried under high vacuum for 24 h. The reaction yielded **23** as a clear oil and was used as obtained for the next step. Based on TLC control (DCM:MeOH = 9:1) that showed full conversion of the starting material, quantitative yield was expected.

R_f (DCM/MeOH = 9:1) = 0.44;

^{1}H NMR (400 MHz, CD_3OD): δ = 7.44-7.33 (m, 5H, ar), 5.28 (s, 2H, H6), 4.15 (t, *J* = 6.7 Hz,1H, H4), 2.56-2.42 (m, 2H, H2), 2.25-2.09 (m, 2H, H3);

^{13}C NMR (101 MHz, CD_3OD): δ = 173.99 (C1), 168.77 (C5), 134.94 (ar.), 128.43 (ar.), 128.33 (ar.), 67.85 (C6), 51.89 (C4), 28.71 (C2), 25.29 (C3).

(*S*)-5-(Benzyloxy)-4-(1,3-dioxoisoindolin-2-yl)-5-oxopentanoic acid (**25**)[143]



1.41 g of **23** (5.93 mmol, 1.0 equiv.) was dissolved in water (7.3 mL) and 1.57 g of Na_2CO_3 (14.8 mmol, 2.5 equiv.) was added. Upon addition of Na_2CO_3 gas formation was observed. 2 ml of water was added, and pH 8.3 was measured. The solution was stirred 30 min at room temperature until Na_2CO_3 was completely dissolved, the solution became transparent, and pH (8.7) was measured. Upon addition of 1.56 g pf *N*(ethoxycarbonyl)phthalimide (7.12 mmol, 1.2 equiv.) colour change to white was observed. The reaction mixture was stirred at room temperature for 1.5 h and subsequently quenched with 6M HCl until no gas formation was observed and white precipitate formed. The mixture was stirred for 3 h, and subsequently was extracted with DCM (3 × 60 mL). Organic layer was dried over $MgSO_4$, and volatiles were removed under reduced pressure. The reaction yielded **25** as a clear oil (1.69 g, 4.52 mmol, 76% over two steps).

R_f (DCM/MeOH = 9:1) = 0.71;

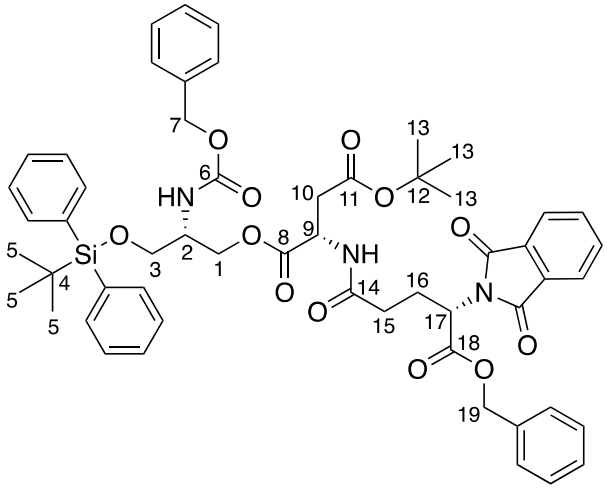
^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.91-7.73 (m, 4H, ar), 7.40-7.27 (m, 5H, ar), 5.22 (s, 2H, H6), 5.04 (dd, *J* = 10.3, 5.0 Hz, 1H, H4), 2.72-2.50 (m, 2H, H3), 2.44 (t, *J* = 7.6 Hz, 2H, H2);

^{13}C NMR (101 MHz, $CDCl_3$): δ = 176.68 (C1), 168.74 (C5), 167.66 (pht), 135.13 (ar.), 134.36 (ar.), 131.66 (ar.), 128.57 (ar.), 128.38 (ar.), 128.14 (ar.), 123.66 (ar.), 67.65 (C6), 51.40 (C4), 30.53 (C2), 24.08 (C3);

HRMS (ESI^+): m/z calc. for ion $C_{20}H_{17}NO_6Na$ $[M+Na]^+$ = 390.0954; found m/z = 390.0944, D = 1 mDa;

HRMS (ESI^-): m/z calc. for ion $C_{20}H_{16}NO_6$ $[M-H]^-$ = 366.0983; found m/z = 366.0979, D = 0.4 mDa.

1-((*R*)-2-(((Benzyloxy)carbonyl)amino)-3-((tert-butyldiphenylsilyl)oxy)propyl) 4-(tert-butyl) ((*S*)-4-(benzyloxy)-3-(1,3-dioxoisoindolin-2-yl)-4-oxobutanoyl)-L-aspartate (**28**)[139]



126 mg of HATU (0.33 mmol, 2.1 eq.) and 0.11 ml of DIPEA (0.63 mmol, 4.0 eq.) were added in 69 mg of **25** (0.19 mmol, 1.2 eq.) in DMF (0.4 mL). Upon addition of DIPEA color change from white to yellow was observed. The solution was covered from light and stirred at room temperature for 25 min. 100 mg of **9** (0.16 mmol, 1.0 eq.) in DMF (0.5 mL) was slowly added to the first solution and the reaction mixture was stirred at room temperature for 4.5 h. The reaction mixture was poured in aq. sat. NH_4Cl (0.3 mL). The mixture was concentrated under reduced pressure, subsequently suspended with water and extracted with DCM (3 x 30 mL). Organic layer was dried over $MgSO_4$, and volatiles were removed under reduced pressure. The crude mixture was purified through column chromatography (EtOAc/cyclohexane = 1:2, then 1:1). The reaction yielded 98% of **28** as a yellowish foam (153 mg, 0.16 mmol).

R_f (EtOAc/cyclohexane = 1:1) = 0.78;

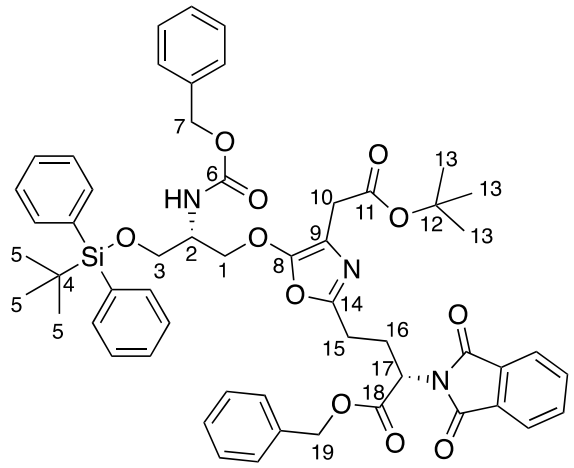
^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.86-7.28 (m, 24H, ar), 6.48-6.43 (m, 1H,NH), 5.37-5.26 (m, 1H, NH), 5.18 (s, 2H, H19), 5.06 (s, 2H, H7), 4.92 (dd, *J* = 10.5, 4.7 Hz, 1H, H17), 4.70-4.61 (m, 1H, H9), 4.37-4.21 (m, 2H, H1), 4.07-3.97 (m, 1H, H2), 3.75-3.66 (m, 2H, H3), 2.80-2.12 (m, 6H, H10, H15, H16), 1.36 (s, 9H, H13), 1.04 (s, 9H, H5);

^{13}C NMR (101 MHz, $CDCl_3$): δ = 170.94 (C11), 170.71 (C8), 170.36 (C14), 168.79 (C18), 167.72 (C-Pht), 156.06 (C6), 136.48 (ar.), 135.65 (ar.), 135.61 (ar.), 135.24 (ar.), 134.35 (ar.), 132.89 (ar.), 132.76 (ar.), 131.83 (ar.), 130.13 (ar.), 130.06 (ar.), 128.65 (ar.), 128.63 (ar.), 128.56 (ar.), 128.45 (ar.), 128.27 (ar.), 128.25 (ar.), 128.12 (ar.), 127.98 (ar.) , 127.96 (ar.), 123.72 (ar.),

81.93 (C12), 67.71 (C-19), 66.93 (C-7), 64.28 (C1), 62.59 (C3), 51.64 (C17), 51.23 (C2), 48.82 (C9), 37.44 (C15), 32.53 (C10), 28.08 (C13), 26.94 (C5), 24.79 (C16), 19.35 (C4);

HRMS (ESI^-): m/z calc. for ion $C_{56}H_{52}N_3O_{14}Si$ $[M+HCOO]^-$ = 1028.4; found m/z = 1028.2, D = 0.2 Da.

Benzyl (*S*)-3-(5-((*R*)-2-(((benzyloxy)carbonyl)amino)-3-((tert-butyldiphenylsilyl)oxy)propoxy)-4-(2-(tert-butoxy)-2-oxoethyl)oxazol-2-yl)-2-(1,3-dioxoisoindolin-2-yl)butanoate (**34**)[134]



214 mg of Triphenylphosphine (0.81 mmol, 6.0 equiv.) and 96 mg of hexachloroethane (0.41 mmol, 3.0 equiv.) were added to 133 mf of **28** (0.14 mmol, 1.0 eq.) in acetonitrile (2 ml). Reaction mixture was refluxed at 85 °C for 40 min. Upon addition of triethylamine (0.13 ml, 0.93 mmol, 6.9 equiv.) gas formation and colour change from yellow to brown and then to dark brown were observed. The reaction mixture was refluxed at 85 °C for 18 h. Volatiles were removed under reduced pressure and the crude mixture was purified through column chromatography (EtOAc/cyclohexane = 1:3). The reaction yielded 33% of **34** as an orange oil (45 mg, 0.05 mmol).

R_f (EtOAc/cyclohexane = 1:1) = 0.52;

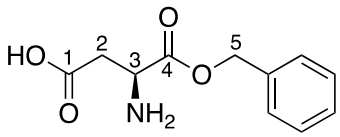
^1H NMR (400 MHz, $CDCl_3$): δ = 7.63-7.27 (m, 24H, ar), 5.41 (d, *J* = 8.9 Hz, 1H, NH), 5.23-5.14 (m, 2H, H19), 5.09-5.02 (m, 2H, H7), 4.93-4.84 (m, 1H, H17), 4.28-4.12 (m, 2H, H1), 4.13-4.06 (m, 1H, H2), 3.77-3.60 (m, 2H, H3), 3.25 (s, 2H, H10), 2.81-2.22 (m, 4H, H15, H16), 1.37 (s, 9H, H13), 1.04 (s, 9H, H5);

^{13}C NMR (101 MHz, $CDCl_3$): δ = 169.84 (C11), 168.76 (C18), 167.77 (pht), 156.02 (C6), 155.00 (C8), 153.79 (C14), 136.52 (ar), 135.63 (ar), 135.22 (ar), 134.36 (ar), 133.08 (ar), 132.94 (ar), 131.85 (ar), 130.02 (ar), 129.99 (ar), 128.66 (ar), 128.62 (ar), 128.48 (ar), 128.30 (ar), 128.27 (ar), 128.22 (ar), 128.19 (ar), 127.99 (ar), 127.95 (ar), 127.93 (ar), 123.69 (ar), 109.98 (C9), 81.33 (C12), 72.38 (C1), 67.75 (C19), 66.87 (C7), 62.17 (C3), 51.93 (C2), 51.60 (C17), 37.21 (C15), 31.86 (C10), 28.09 (C13), 26.92 (C5), 25.55 (C16), 19.36 (C4).

HRMS (ESI$^+$): m/z calc. for ion $C_{55}H_{60}N_3O_{11}Si$ $[M+H]^+$ = 966.3992; found m/z = 966.3997, D = 0.5 mDa;

Optical rotation: α_D^{23}= - 0.0034 (c = 0.88, $CHCl_3$).

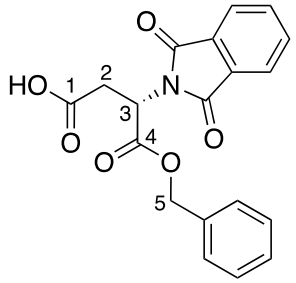
(*S*)-3-amino-4-(benzyloxy)-4-oxobutanoic acid (**22**)[142]



2.00 g of (*S*)-3-amino-4-(benzyloxy)-4-oxobutanoic acid (6.20 mmol, 1.0 eq.) was dissolved in 9.3 ml of chloroform and 9.3 ml of TFA was added. The reaction mixture was stirred at room temperature for 1.5 h. 0.5 ml of triethylamine was added, resulting in the gas evolution. Solvents were removed under reduced pressure and the residue was dried under high vacuum for 24 h. The reaction yielded **22** as a clear oil and was used as obtained for the next step. Based on full conversion on TLC control (DCM:MeOH = 9:1) quantitative yield was expected.

R_f (DCM/MeOH = 9:1) = 0.15;

(*S*)-4-(benzyloxy)-3-(1,3-dioxoisoindolin-2-yl)-4-oxobutanoic acid (**24**)[143]



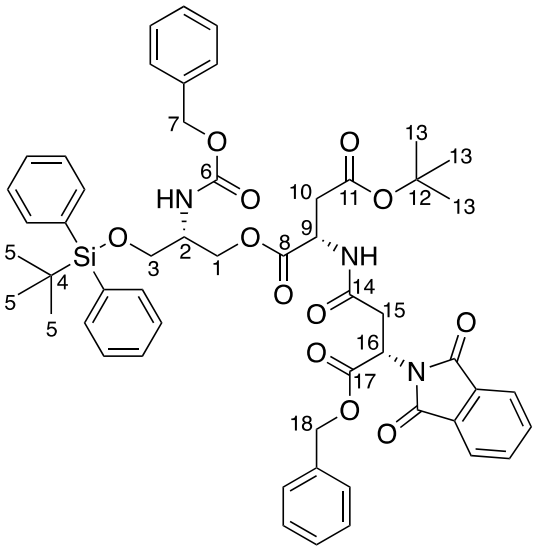
1.00 g of **20** (5.93 mmol, 1.0 eq.) was dissolved in water (5.5 mL) and 1.20 g of Na_2CO_3 (11.2 mmol, 2.5 eq.) was added. Upon addition of Na_2CO_3 gas formation was observed and mixture was cooled to 0 °C. Upon addition of 1.23 q of freshly powdered *N*(ethoxycarbonyl)phthalimide (5.63 mmol, 1.2 equiv.) colour change to white was observed. The reaction mixture was stirred at room temperature for 1.5 h and subsequently quenched with 6M HCl until no gas formation was observed and white precipitate formed. The mixture was stirred for 3 h, and subsequently was extracted with DCM (3 × 60 mL). Organic layer was dried over $MgSO_4$, and volatiles were removed under reduced pressure. The reaction yielded 602 mg of **24** as a clear oil (1.70 mmol, 68% over two steps).

R_f (DCM/MeOH = 9:1) = 0.65;

^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.93-7.17 (m, 9H, ar), 5.49-5.42 (m, 1H, H3), 5.18 (s, 2H, H5), 3.49-3.11 (m, 2H, H2);

^{13}C NMR (101 MHz, $CDCl_3$) δ 174.78 (C1), 168.28 (C4), 167.33 (C=OPht), 134.90 (ar), 134.34 (ar), 131.65 (ar), 128.53 (ar), 128.38 (ar), 128.00 (ar), 123.66 (ar), 67.89 (C5), 61.49 (C3), 33.30 (C2).

1-((*R*)-2-(((benzyloxy)carbonyl)amino)-3-((*tert*-butyldiphenylsilyl)oxy)propyl) 4-(*tert*-butyl) ((*S*)-4-(benzyloxy)-3-(1,3-dioxoisoindolin-2-yl)-4-oxobutanoyl)-*L*-aspartate (**27**)



1.05 g of HATU (2.75 mmol, 2.1 eq.) and 0.9 ml of DIPEA (677 mg, 5.24 mmol, 4.0 eq.) were added to 463 mg of **24** (1.31 mmol, 1.1 equiv.) in DMF (2 ml). Upon addition of DIPEA colour change from white to yellow was observed. The solution was covered from light and stirred at room temperature for 25 min. 755 mg of **9** (1.19 mmol, 1.0 equiv.) in DMF (2 ml) was slowly added to the first solution and the reaction mixture was stirred at room temperature for 12 h. The reaction was poured into aq. sat. NH_4Cl (5 ml). The mixture was concentrated under reduced pressure, subsequently suspended with water and extracted with DCM (3 × 30 mL). Organic layer was dried over $MgSO_4$, and volatiles were removed under reduced pressure. The crude mixture was purified through column chromatography (EtOAc/cyclohexane = 1:2, then 1:1). The reaction yielded 654 mg of **27** as a yellowish foam (0.67 mmol, 57%).

R_f (EtOAc/cyclohexane = 1:2) = 0.45;

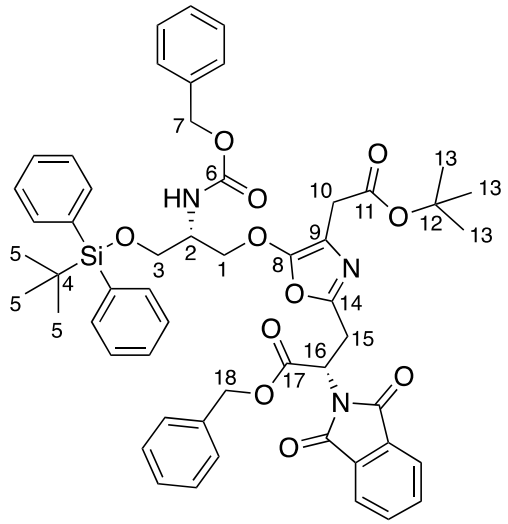
^{1}H NMR (400 MHz, CD_2Cl_2): δ 7.88-7.26 (m, 24H, ar), 6.69 (d, *J* = 7.9 Hz, 1H, NH), 5.53 (dd, *J* = 8.1, 6.4 Hz, 1H, H16), 5.45 (d, *J* = 8.9 Hz, 1H, NH), 5.22-5.01 (m, 4H, H7, H18), 4.75-4.63 (m, 1H, H9), 4.40-4.25 (m, 2H, H1), 4.08-4.01 (m, 1H, H2), 3.74 (d, *J* = 5.0 Hz, 2H, H3), 3.34-2.95 (m, 2H, H15), 2.79-2.55 (m, 2H, H10), 1.33 (s, 9H, H13), 1.08 (s, 9H, H5).

^{13}C NMR (101 MHz, CD_2Cl_2) δ 170.97 (C11), 170.40 (C8), 169.21 (C14), 167.84 (C17), 167.81 (C=OPht), 156.46 (C6), 136.07 (ar), 136.04 (ar), 134.83 (ar), 134.76 (ar), 133.36 (ar), 132.32 (ar), 132.28 (ar), 130.49 (ar), 130.43 (ar), 128.96 (ar), 128.77 (ar), 128.49 (ar), 128.48 (ar),

128.44 (ar), 128.37 (ar), 128.36 (ar), 124.76 (Pht), 123.98 (Pht), 123.95 (ar), 82.06 (C12), 68.18 (C18), 68.14 (C7), 64.76 (C1), 63.06 (C3), 51.91 (C2), 49.43 (C9), 49.10 (C16), 37.73 (C10), 36.02 (C15), 28.18 (C13), 27.13 (C5), 19.61 (C4) ;

HRMS (ESI$^-$): m/z calc. for ion $C_{54}H_{59}N_3NaO_{12}Si$ $[M+Na]^+$ = 992.3760; found m/z = 992.3794, D = 3.4 mDa.

Benzyl (*S*)-3-(5-((*R*)-2-(((benzyloxy)carbonyl)amino)-3-((*tert*-butyldiphenylsilyl)oxy)propoxy)-4-(2-(*tert*-butoxy)-2-oxoethyl)oxazol-2-yl)-2-(1,3-dioxoisoindolin-2-yl)propanoate (**33**)



Solution of 135 mg of triphenylphosphine (0.52 mmol, 5.0 equiv.) and 59 mg of hexachloroethane (0.25 mmol, 2.5 eq.) in acetonitrile (0.5 ml) was added to 100 mg of **27** (0.1 mmol, 1.0 eq.) in acetonitrile (0.8 ml). The mixture was heated to 85 °C and after 15 min 0.085 ml of triethylamine (0.6 mmol, 6 eq.) was added and colour changed from light yellow to dark orange. The reaction mixture was refluxed at 85 °C for 20 h and colour change to dark brown was observed. The solution was filtered through celite and volatiles were removed under reduced pressure. The crude mixture was purified through column chromatography (EtOAc/cyclohexane = 1:2). The reaction yielded 59 mg of **33** as a yellow oil (0.06 mmol, 62%).

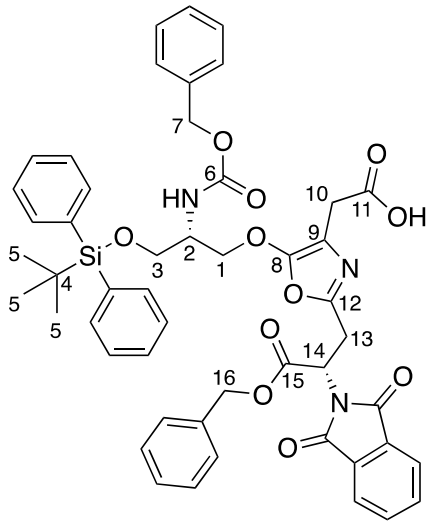
^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.99 – 7.27 (m, 24H, Ar, Pht), 5.35 – 5.28 (m, 1H, H16), 5.20 (s, 2H, H18), 5.07 (s, 2H, H7), 4.26 – 4.07 (m, 2H, H1), 4.07 – 3.99 (m, 1H, H2), 3.83 – 3.44 (m, 4H, H-3, H15), 3.13 (s, 2H, H10), 1.31 (s, 9H, H13), 1.04 (s, 9H, H5) ;

^{13}C NMR (101 MHz, $CDCl_3$) δ 171.15 (C-17), 169.47(C-11), 167.74 (C-Pht), 155.93 (C-6), 155.17 (C-8), 150.79 (C-14), 135.39 (ar), 135.35 (ar), 134.92 (Ar), 134.33 (ar), 134.19 (ar), 131.75 (ar), 131.72 (ar), 131.16 (ar), 128.57 (ar), 128.55 (ar), 128.52 (ar), 128.45 (ar), 128.19 (ar), 128.17 (ar), 124.50 (ar), 123.93 (Ar), 123.66 (Ar), 123.66 (ar), 110.38 (C-9), 81.07 (C-12), 72.32 (C-1), 68.35 (C-18), 66.31 (C-7), 61.26 (C-3), 51.65 (C-2), 49.95 (C-16), 31.82 (C-10), 29.71, 28.46 (C-15), 27.94 (C-13), 25.83 (C-5), 18.24 (C-4) ;

HRMS (ESI^+): m/z calc. for ion $C_{54}H_{58}N_3O_{11}Si$ $[M+H]^+$ = 952,3835 ; found m/z = 952.3871 D = 3.6 mDa;

Optical rotation: α_D^{23}= - 0.0089 (c = 2.14, $CHCl_3$)

2-(2-((*S*)-3-(benzyloxy)-2-(1,3-dioxoisoindolin-2-yl)-3-oxopropyl)-5-((*R*)-2-(((benzyloxy) carbonyl)amino)-3-((*tert*-butyldiphenylsilyl)oxy)propoxy)oxazol-4-yl)acetic acid (**39**)

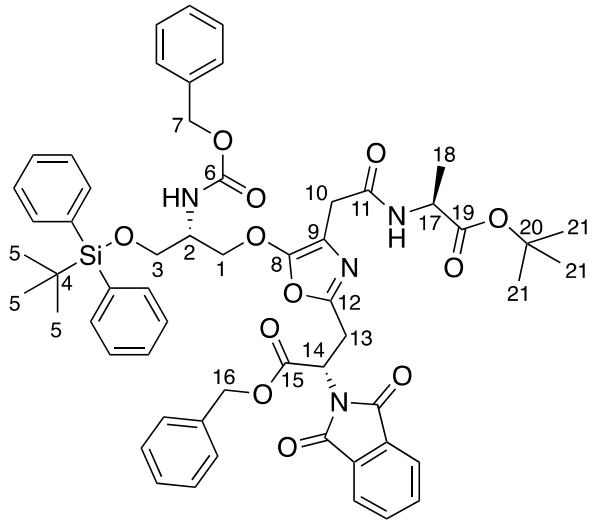


20 µl ml of 85% aq. H_3PO_4 (30 mg, 0.31 mmol, 14.8 mmol/ml, 5.0 eq) was added to 59 mg of **33** (0.06 mmol, 1.0 eq.) in DCM (0.5 ml) and the reaction mixture was stirred at room temperature for 21 h. The reaction mixture was poured in water (1 mL) and extracted with DCM (3 × 10 mL). Organic layer was dried over $MgSO_4$, and volatiles were removed under reduced pressure. The reaction yielded 38 mg of **39** as an orange oil (0.04 mmol, 69%).

R_f (EtOAc/cyclohex 2:1) = 0.17 ;

^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.84 – 7.25 (m, 24H, Ar), 5.34 – 5.27 (m, 1H, H-16), 5.18 (s, 2H, H-16), 5.07 (s, 2H, H-7), 4.31 – 4.07 (m, 2H, H-1), 4.02 (s, 1H, H-2), 3.83 – 3.42 (m, 4H, H-1, H-13), 3.22 (s, 2H, H-10), 1.04 (s, 9H, H-5) ;

Benzyl (*S*)-3-(5-((*S*)-2-(((benzyloxy)carbonyl)amino)-3-((*tert*-butyldiphenylsilyl)oxy)propoxy)-4-(2-(((*S*)-1-(*tert*-butoxy)-1-oxopropan-2-yl)amino)-2-oxoethyl)oxazol-2-yl)-2-(1,3-dioxoisoindolin-2-yl)propanoate (**40**)



33 mg of HATU (0.088 mmol, 2.1 eq) was added to 38 mg of **39** (0.042 mmol, 1.0 eq) in DMF (0.2 ml) giving a yellow solution. 0.03 ml of DIPEA (0.17 mmol, 22 mg, 4.0 eq) was added and color change to orange was observed. Reaction mixture was stirred for 20 min and 9.0 mg of L-Ala(OtBu) HCl (0.050 mmol, 1.2 eq) was added and reaction mixture was stirred at room temperature for 17 h. Subsequently aq. sat. NH_4Cl (1 ml) was added and volatiles were evaporated. The crude mixture was purified via column chromatography (EtOAc/iHex 1:1) giving 14.2 mf of **40** (0.014 mmol, 33%)

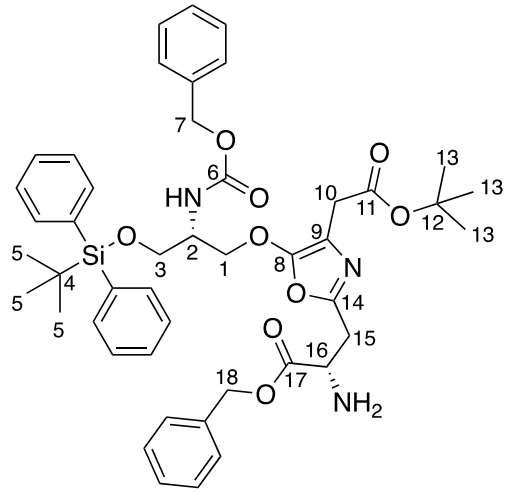
R_f (EtOAc/iHex 1:1) = 0.39;

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.88 – 7.23 (m, 24H, ar), 5.39 – 5.28 (m, 1H, H14), 5.19 (s, 2H, H16), 5.06 (s, 2H, H7), 4.40 – 4.25 (m, 1H, 17), 4.22 – 3.96 (m, 3H, H1, H2), 3.82 – 3.40 (m, 4H, H3, H13), 3.15 (s, 2H, H10), 1.40 (s, 9H, H21), 1.25 (s, 3H, H18), 1.03 (s, 9H, H5) ;

^{13}C NMR (101 MHz, $CDCl_3$) δ 172.13 (C15), 171.34 (C19), 167.44 (C11), 166.48 (C-Pht), 156.67 (C6), 156.04 (C8), 155.24 (C12), 135.69 (ar), 135.62 (ar), 134.96 (ar), 134.45 (ar), 134.32 (ar), 131.77 (ar), 130.05 (ar), 130.02 (ar), 128.70 (ar), 128.65 (ar), 128.62 (ar), 128.56 (ar), 128.36

(ar), 128.24 (ar), 128.16 (ar), 127.97 (ar), 127.94 (ar), 123.83 (ar), 123.78 (ar), 110.52 (C9), 81.84 (C20), 72.75 (C1), 68.16 (C16), 67.94 (C7), 62.24 (C3), 49.95 (C2), 49.62 (C14), 48.95 (C17), 32.06 (C10), 28.57 (C13), 28.06 (C21), 26.95 (C5), 19.37 (C4), 18.34 (C18)

Benzyl (*S*)-2-amino-3-(5-((*R*)-2-(((benzyloxy)carbonyl)amino)-3-((*tert* butyldiphenylsilyl)oxy)propoxy)-4-(2-(*tert*-butoxy)-2-oxoethyl)oxazol-2-yl)propanoate (**41**)



0.06 ml hydrazine monohydrate (61 mg, 1.22 mmol, 20 eq.) was added slowly to 0 °C solution of 58 mg of **33** (0.06 mmol, 1.0 eq.) in THF (1.7 ml). Reaction mixture was stirred at 0°C for 2 h and subsequently THF was evaporated, and the residue was suspended with water and extracted with DCM (3 x 30 ml). Organic layer was dried over $MgSO_4$ and volatiles were evaporated. The crude mixture was purified via column chromatography (EtOAc/cyclohexane 2:1->3:1) to yield 32% of compound **41** as yellow oil.

R_f (EtOAc/cyclohex 2:1) = 0.14 ;

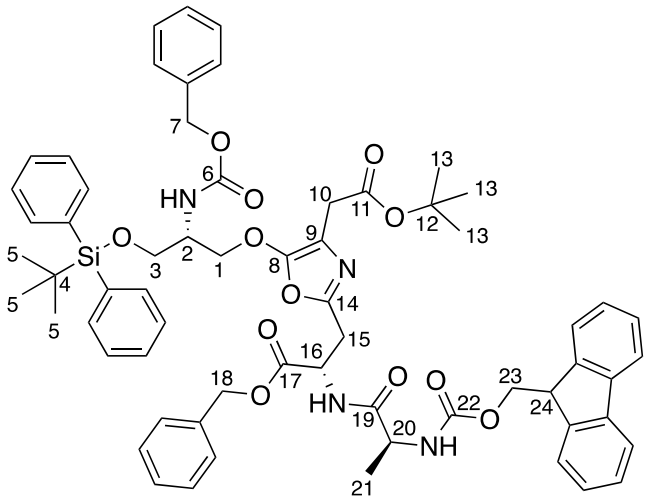
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.69 – 7.26 (m, 20H, ar), 5.46 – 5.32 (m, 1H, H16), 5.14 (s, 2H, H18), 5.07 (s, 2H, H7), 4.37-4.15 (m, 2H, H1), 4.12-4.01 (m, 1H, H2), 3.93 (ddd, *J* = 8.3, 6.4, 4.6 Hz, 1H, H16), 3.77 (ddd, *J* = 34.2, 10.7, 6.0 Hz, 2H, H3), 3.24 (s, 2H, H10), 3.14-2.86 (m, 2H, H15), 1.39 (s, 9H, H13), 1.05 (s, 9H, H5) ;

^{13}C NMR (101 MHz, $CDCl_3$) δ 174.01 (C17), 169.85 (C11), 155.28 (C6), 155.25 (C8), 150.50 (C14), 135.65 (ar), 135.52 (ar), 132.89 (ar), 132.04 (ar), 130.05 (ar), 128.74 (ar), 128.66 (ar), 128.57 (ar), 128.45 (ar), 128.28 (ar), 127.98 (ar), 127.96 (ar), 110.12 (C9), 81.41 (C12), 72.02 (C1),

67.26 (C18), 66.87 (C7), 62.15 (C3), 52.71 (C16), 51.93 (C2), 33.64 (C15), 32.08 (C10), 28.14 (C13), 26.96 (C5), 19.40 (C4) ;

HRMS (ESI$^+$): m/z calc. for ion $C_{46}H_{56}N_3O_9Si$ $[M+H]^+$ = 822.3780; found m/z = 822.3774, D = 0.6 mDa.

Benzyl (*S*)-2-((*S*)-2-((((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)propanamido)-3-(5-((*R*)-2-(((benzyloxy)carbonyl)amino)-3-((*tert*-butyldiphenylsilyl)oxy)propoxy)-4-(2-(*tert*-butoxy)-2-oxoethyl)oxazol-2-yl)propanoate (**42**)



16 mg HATU (0.041 mmol, 2.1 eq) and 14 µl DIPEA (10 mg, 0.08 mmol, 4.0 eq) were added to 13 mg of Fmoc-Ala H_2O (0.04 mmol,2.0 eq) in DMF (0.1 ml) and stirred at room temperature for 20 min. The mixture was added to 16 mg of **41** (0.02 mmol, 1.0 eq) in DMF (0.1 ml). Reaction mixture was stirred at room temperature for 22 h and subsequently poured into aq. sat. NH_4Cl (1 ml) and extracted with DCM (3 x 10 ml). Organic layer was dried over $MgSO_4$ and volatiles were evaporated under reduced pressure. The crude mixture was purified via column chromatography (EtOAc/iHex 1:1) giving 19 mg of **42** (0.017 mmol, 86 %).

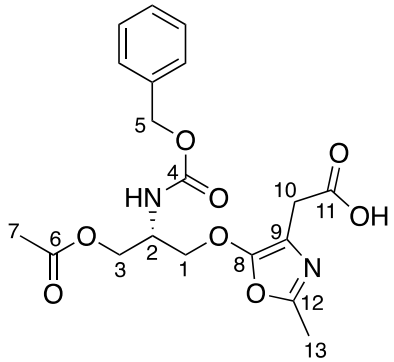
R_f (EtOAc/iHex 1:1) = 0.62

^{1}H NMR (400 MHz, $CDCl_3$) δ δ 7.80 – 7.27 (m, 24H, ar), 5.36 (ddd, *J* = 12.2, 5.7, 3.3 Hz, 1H, H16), 5.19 – 4.90 (m, 4H, H18, H7), 4.37 – 3.97 (m, 7H, H1, H2, H20, H23, H24), 3.83 – 3.59 (m, 2H, H3), 3.19 (s, 2H, H10), 2.87 – 2.56 (m, 2H, H15), 1.42 (s, 3H, H21), 1.37 (s, 10H, H13), 1.05 (s, 9H, H5).

^{13}C NMR (101 MHz, $CDCl_3$) δ 174.01 (C17), 169.85 (C11), 155.28 (C6), 155.25 (C8), 150.50 (C14), 135.65 (ar), 135.52 (ar), 132.89 (ar), 132.04 (ar), 130.05 (ar), 128.74 (ar), 128.66 (ar), 128.57 (ar), 128.45 (ar), 128.28 (ar), 127.98 (ar), 127.96 (ar), 110.12 (C9), 81.41 (C12), 72.02 (C1), 67.26 (C18), 66.87 (C7), 62.15 (C3), 52.71 (C16), 51.93 (C2), 33.64 (C15), 32.08 (C10), 28.14 (C13), 26.96 (C5), 19.40 (C4) ;

Optical rotation: α_D^{23}= - 0.0035 (c = 0.79, $CHCl_3$).

(*R*)-2-(5-(3-acetoxy-2-(((benzyloxy)carbonyl)amino)propoxy)-2-methyloxazol-4-yl)acetic acid



47 mg of **31** (0.10 mmol, 1.0 eq.) was dissolved in DCM (0.2 mL) and 34 µL H_3PO_4 was added. The reaction mixture was stirred at room temperature for 20.5 h. TLC control (EtOAc/cyclohexane = 3:1) showed full conversion of the starting material. The reaction was quenched with water (1 mL) and extracted with DCM (3 × 20 mL). Organic layer was dried over $MgSO_4$, and volatiles were removed under reduced pressure. The reaction yielded 85% of (*R*)-2-(5-(3-acetoxy-2-(((benzyloxy)carbonyl)amino)propoxy)-2-methyloxazol-4-yl)acetic acid as an orange oil (35 mg, 0.09 mmol).

R_f (EtOAc/cyclohexane = 3:1) = 0.08;

^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.37-7.29 (m, 5H, ar), 5.52 (dd, *J* = 80.0, 8.1 Hz, 1H, NH), 5.10-5.06 (m, 2H, H5), 4.30-4.13 (m, 4H, H1, H3), 4.00 (dt, *J* = 9.3, 5.0 Hz, 1H, H2), 3.40 (s, 2H, H10), 2.30 (s, 3H, H13), 2.04 (s, 3H, H7);

^{13}C NMR (101 MHz, $CDCl_3$): δ = 172.96 (C11), 171.08 (C6), 156.57 (C4), 154.89 (C8), 153.80 (C12), 132.27 (ar), 128.77 (ar), 128.68 (ar), 128.26 (ar), 109.68 (C9), 72.98 (C1), 67.21 (C5), 62.76 (C3), 51.94 (C2), 30.25 (C10), 20.82 (C7), 14.09 (C13);

HRMS (ESI^-): m/z calc. for ion $C_{19}H_{21}N_2O_8$ $[M-H]^-$ = 405.1; found m/z = 405.0 D = 0.1 Da;

Optical rotation: α_D^{23}= - 0.0040 (c = 4.02, $CHCl_3$).

10. List of abbreviations

A	adenosine
Ac	acetyl
AP	apurinic
APE1	apurinic endonuclease
5-aza-dC	5-aza-2´-deoxycytidine
6-aza-fdC	6-aza-5-formyl-2´-deoxycytidine
BER	base excision repair
Bn	benzyl
Boc	*tert*-butyloxycarbonyl
BR	biological replicate
Bz-dA	*N*-benzoyl-2´-deoxyadenosine
Bz-dC	*N*-benzoyl-2´-deoxycytidine
Bz-mdC	*N*-benzoyl-5-methyl-2´-deoxycytidine
C	cytidine
cadC	5-carboxy-2´-deoxycytidine
CASPT2	complete active space with second-order perturbation theory
Cbz	benzyloxycarbonyl
CE	collision energy
CHO-K	Chinese hamster ovary cells
CID	collision induced dissociation
CpG	dC-dG dinucleotide
dA	2´-deoxyadenosine
DAST	diethylaminosulfur trifluoride
dC	2´-deoxycytidine
DCM	dichloromethane
DCK	deoxycytidine kinase
ddH_2O	double distilled water
DIPEA	di-*iso*-propyl ethyl amine
dG	2´deoxyguanosine
DIPEA	N,N-diisopropylenethylamine

DMAP	4-(dimethylamino)-pyridine
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DMT	4,4´-dimethoxytrityl
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
Dnmt	DNA nucleotide methyltransferase
DSB	double strand break
dsDNA	double stranded DNA
dT	2´-deoxythymidine
EDTA	ethylene diamine tetra acetic acid
EMV	electron multiplier voltage
ESC	embryonic stem cell
ESI	electrospray ionization
F-cadC	5-carboxy-2´-fluoro-2´-deoxycytidine
F-cadCTP	5-carboxy-2´-fluoro-2´-deoxycytidine triphosphate
F-dC	2´-fluoro-2´-deoxycytidine
fdC	5-formyl-2´-deoxycytidine
F-dCTP	2´-fluoro-2´-deoxycytidine triphosphate
F-fdC	5-formyl-2´-fluoro-2´-deoxycytidine
F-fdCTP	5-formyl-2´-fluoro-2´-deoxycytidine triphosphate
fdU	5-formyl-2´-deoxyuridine
F-hmdC	5-hydroxy-2´-fluoro-2´-deoxycytidine
F-mdC	5-methyl-2´-fluoro-2´-deoxycytidine
Fmoc	fluorenylmethyloxycarbonyl
G	guanosine
gDNA	genomic DNA
g-cadC	genomic 5-carboxy-2´-deoxycytidine
HATU	*O*-(7-Azabenzotriazole-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate
HeLa	Henrietta Lacks
hmdC	5-hydoxymethyl-2´-deoxycytidine

HPA	3-hydroxypicolinic acid
HPLC	high-performance liquid chromatography
HRMS	high resolution mass spectrometry
hTDG	human thymine-DNA glycosylase
*i*Bu-dG	N^2-isobutyryl-2´-deoxyguanosine
IR	infrared spectroscopy
LC-MS	liquid-chromatography coupled mass spectrometry
LMU	Ludwigs-Maximilians Universität
LOD	lowest limit of detection
MALDI	matrix assisted laser desorption ionisation
mdC	5-methyl-2´-deoxycytidin
MeCN	acetonitrile
MeOH	methanol
mESCs	murine embryonic stem cells
MM	molecular mechanics
MS	mass spectrometry
N2a	neuro 2a cells
NEB	New England BioLabs
NMR	nuclear magnetic resonance
5-NO_2-F-dC	5-nitro-2´-fluoro-2´-deoxycytidine
5-NO_2-F-dU	5-nitro-2´-fluoro-2´-deoxyuridine
PDE 1	snake venom phosphodiesterase
ppm	parts per million
QQQ	triple quadrupole
Py	pyridine
R_f	retention factor
RNA	ribonucleic acid
rp LC	reverse phase liquid chromatography
RT	room temperature
SAM	S-adenosylmethionine
S-dCMP	2'-deoxycytidine 5'-monophosphorothioate
S-cadCMP	5-carboxy-2'-deoxycytidine 5'-monophosphorothioate

S_N2	nucleophilic substitution of 2nd order
TA	transient absorption
TBDPS	*tert*-butyldiphenylsilyl
TBDPSCl	*tert*-butyldiphenylsilyl chloride
TBSCl	*tert*-butyldimethylsilyl chloride
TDG	thymine-DNA glycosylase
TET	ten-eleven translocation
TFA	trifluoroacetic acid
TfOH	trifluoromethanesulfonic acid
THF	tetrahydrofuran
TOF	time of flight
UHPLC	ultra high-performance liquid chromatography
ULOD	under limit of detection
UV/Vis	ultraviolet and visible light
α-KG	α-ketoglutarate

11. References

(1) Watanabe, A.; Yamada, Y.; Yamanaka, S. *Philos. Trans. R. Soc. B Biol. Sci.* **2013**, *368* (1609).

(2) Wyatt, G. R. *Nature* **1950**, *166* (4214), 237–238.

(3) Hotchkiss, R. D. *J. Biol. Chem.* **1948**, *175* (1), 315–332.

(4) Johnson, T. B.; Coghill, R. D. *J. Am. Chem. Soc.* **1925**, *47* (11), 2838–2844.

(5) Jaenisch, R.; Bird, A. *Nat. Genet.* **2003**, *33*, 245–254.

(6) Bannister, A. J.; Kouzarides, T. *Cell Res.* **2011**, *21* (3), 381–395.

(7) Penn, N. W.; Suwalski, R.; O'Riley, C.; Bojanowski, K.; Yura, R. *Biochem. J.* **1972**, *126* (4), 781–790.

(8) Kothari, R. M.; Shankar, V. *J. Mol. Evol.* **1976**, *7* (4), 325–329.

(9) Kriaucionis, S.; Heintz, N. *Science* **2009**, *324* (5929), 929–930.

(10) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. *Science* **2009**, *324* (5929), 930–935.

(11) Loenarz, C.; Schofield, C. J. *Chem. Biol.* **2009**, *16* (6), 580–583.

(12) Globisch, D.; Münzel, M.; Müller, M.; Michalakis, S.; Wagner, M.; Koch, S.; Brückl, T.; Biel, M.; Carell, T. *PLoS One* **2010**, *5* (12), e15367.

(13) Pfaffeneder, T.; Hackner, B.; Truß, M.; Münzel, M.; Müller, M.; Deiml, C. A.; Hagemeier, C.; Carell, T. *Angew. Chemie Int. Ed.* **2011**, *50* (31), 7008–7012.

(14) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. *Science* **2011**, *333* (6047), 1300–1303.

(15) He, Y.-F.; Li, B.-Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; Sun, Y.; Li, X.; Dai, Q.; Song, C.-X.; Zhang, K.; He, C.; Xu, G.-L. *Science* **2011**, *333* (6047), 1303–1307.

(16) Iwan, K.; Rahimoff, R.; Kirchner, A.; Spada, F.; Schröder, A. S.; Kosmatchev, O.; Ferizaj, S.; Steinbacher, J.; Parsa, E.; Müller, M.; Carell, T. *Nat. Chem. Biol.* **2018**, *14* (1), 72–78.

(17) Spruijt, C. G.; Gnerlich, F.; Smits, A. H.; Pfaffeneder, T.; Jansen, P. W. T. C.; Bauer, C.; Münzel, M.; Wagner, M.; Müller, M.; Khan, F.; Eberl, H. C.; Mensinga, A.; Brinkman, A. B.; Lephikov, K.; Müller, U.; Walter, J.; Boelens, R.; Van Ingen, H.; Leonhardt, H.; Carell, T.; Vermeulen, M. *Cell* **2013**, *152* (5), 1146–1159.

(18) Iurlaro, M.; Ficz, G.; Oxley, D.; Raiber, E. A.; Bachman, M.; Booth, M. J.; Andrews, S.;

Balasubramanian, S.; Reik, W. *Genome Biol.* **2013**, *14* (10), R119.

(19) Zhou, T.; Xiong, J.; Wang, M.; Yang, N.; Wong, J.; Zhu, B.; Xu, R. M. *Mol. Cell* **2014**, *54* (5), 879–886.

(20) Münzel, M.; Globisch, D.; Brückl, T.; Wagner, M.; Welzmiller, V.; Michalakis, S.; Müller, M.; Biel, M.; Carell, T. *Angew. Chemie Int. Ed.* **2010**, *49* (31), 5375–5377.

(21) Wagner, M.; Steinbacher, J.; Kraus, T. F. J.; Michalakis, S.; Hackner, B.; Pfaffeneder, T.; Perera, A.; Müller, M.; Giese, A.; Kretzschmar, H. A.; Carell, T. *Angew. Chemie Int. Ed.* **2015**, *54* (42), 12511–12514.

(22) Kraus, T. F. J.; Globisch, D.; Wagner, M.; Eigenbrod, S.; Widmann, D.; Münzel, M.; Müller, M.; Pfaffeneder, T.; Hackner, B.; Feiden, W.; Schüller, U.; Carell, T.; Kretzschmar, H. A. *Int. J. Cancer* **2012**, *131* (7), 1577–1590.

(23) Mo, J.; Liang, Z.; Lu, M.; Wang, H. *Anal. Chem.* **2020**, *92* (11), 7430–7436.

(24) Liu, S.; Wang, J.; Su, Y.; Guerrero, C.; Zeng, Y.; Mitra, D.; Brooks, P. J.; Fisher, D. E.; Song, H.; Wang, Y. *Nucleic Acids Res.* **2013**, *41*, 6421–6429.

(25) Yu, Y.; Yuan, F.; Zhang, X. H.; Zhao, M. Z.; Zhou, Y. L.; Zhang, X. X. *Anal. Chem.* **2019**, *91* (20), 13047–13053.

(26) Carell, T.; Kurz, M. Q.; Müller, M.; Rossa, M.; Spada, F. *Angew. Chemie Int. Ed.* **2018**, *57* (16), 4296–4312.

(27) Booth, M. J.; Marsico, G.; Bachman, M.; Beraldi, D.; Balasubramanian, S. *Nat. Chem.* **2014**, *6* (5), 435–440.

(28) Leitch, H. G.; Mcewen, K. R.; Turp, A.; Encheva, V.; Carroll, T.; Grabole, N.; Mansfield, W.; Nashun, B.; Knezovich, J. G.; Smith, A.; Surani, M. A.; Hajkova, P. *Nat. Struct. Mol. Biol.* **2013**, *20* (3), 311.

(29) Tang, Y.; Zheng, S. J.; Qi, C. B.; Feng, Y. Q.; Yuan, B. F. *Anal. Chem.* **2015**, *87* (6), 3445–3452.

(30) Zhang, H. Y.; Xiong, J.; Qi, B. L.; Feng, Y. Q.; Yuan, B. F. *Chem. Commun.* **2016**, *52* (4), 737–740.

(31) Xiong, J.; Liu, X.; Cheng, Q. Y.; Xiao, S.; Xia, L. X.; Yuan, B. F.; Feng, Y. Q. *ACS Chem. Biol.* **2017**, *12* (6), 1636–1643.

(32) Jiang, H. P.; Liu, T.; Guo, N.; Yu, L.; Yuan, B. F.; Feng, Y. Q. *Anal. Chim. Acta* **2017**, *981*, 1–10.

(33) Huang, W.; Lan, M. D.; Qi, C. B.; Zheng, S. J.; Wei, S. Z.; Yuan, B. F.; Feng, Y. Q. *Chem. Sci.*

2016, *7* (8), 5495–5502.

(34) Yuan, F.; Yu, Y.; Zhou, Y. L.; Zhang, X. X. *Anal. Chem.* **2020**, *92* (1), 1605–1610.

(35) Sumino, M.; Ohkubo, A.; Taguchi, H.; Seio, K.; Sekine, M. *Bioorg. Med. Chem. Lett.* **2008**, *18* (1), 274–277.

(36) Crain, P. F. *Methods Enzymol.* **1990**, *193* (C), 782–790.

(37) Quinlivan, E. P.; Gregory, J. F. *Anal. Biochem.* **2008**, *373* (2), 383–385.

(38) Nestles, M.; Roberts, W. K. *J. Biol. Chem.* **1969**, *244* (19), 5219–5225.

(39) Desai, N. A.; Shankar, V. *FEMS Microbiol. Rev.* **2003**, *26* (5), 457–491.

(40) Yuan, F.; Bi, Y.; Zhang, J. Y.; Zhou, Y. L.; Zhang, X. X.; Song, C. X. *RSC Adv.* **2019**, *9* (50), 29010–29014.

(41) Hackett, J. A.; Sengupta, R.; Zylicz, J. J.; Murakami, K.; Lee, C.; Down, T. A.; Surani, M. A. *Science* **2013**, *339* (6118), 448–452.

(42) Mayer, W.; Niveleau, A.; Walter, J.; Fundele, R.; Haaf, T. *Nature* **2000**, *403* (6769), 501–502.

(43) Feng, Y. Y.-Q.; Xie, N.-B.; Tao, W.-B.; Ding, J.-H.; You, X.-J.; Ma, C.-J.; Zhang, X.; Yi, C.; Zhou, X.; Yuan, B.-F.; Feng, Y. Y.-Q. *CCS Chem* **2020**, *2*, 994–1008.

(44) Ngo, T. T. M.; Yoo, J.; Dai, Q.; Zhang, Q.; He, C.; Aksimentiev, A.; Ha, T. *Nat. Commun.* **2016**, *7* (1), 1–9.

(45) Fu, T.; Liu, L.; Yang, Q. L.; Wang, Y.; Xu, P.; Zhang, L.; Liu, S.; Dai, Q.; Ji, Q.; Xu, G. L.; He, C.; Luo, C.; Zhang, L. *Chem. Sci.* **2019**, *10* (31), 7407–7417.

(46) Weber, A. R.; Krawczyk, C.; Robertson, A. B.; Kusnierczyk, A.; Vågbø, C. B.; Schuermann, D.; Klungland, A.; Schar, P. *Nat. Commun.* **2016**, *7*.

(47) Schlosberg, C. E.; Wu, D. Y.; Gabel, H. W.; Edwards, J. R. *Nucleic Acids Res.* **2019**, *47* (5), e28.

(48) Ji, D.; Lin, K.; Song, J.; Wang, Y. *Mol. Biosyst.* **2014**, *10* (7), 1749–1752.

(49) Hashimoto, H.; Liu, Y.; Upadhyay, A. K.; Chang, Y.; Howerton, S. B.; Vertino, P. M.; Zhang, X.; Cheng, X. *Nucleic Acids Res.* **2012**, *40* (11), 4841–4849.

(50) Juttermann, R.; Li, E.; Jaenisch, R. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91* (25), 11797.

(51) Jones, P. A.; Taylor, S. M. *Cell* **1980**, *20* (1), 85–93.

(52) Maiti, A.; Drohat, A. C. *J. Biol. Chem.* **2011**, *266* (41), 35334–35338.

(53) Raiber, E. A.; Beraldi, D.; Ficz, G.; Burgess, H. E.; Branco, M. R.; Murat, P.; Oxley, D.; Booth, M. J.; Reik, W.; Balasubramanian, S. *Genome Biol.* **2012**, *13* (8), R69.

(54) Zhang, L.; Lu, X.; Lu, J.; Liang, H.; Dai, Q.; Xu, G. L.; Luo, C.; Jiang, H.; He, C. *Nat. Chem. Biol.* **2012**, *8* (4), 328–330.

(55) Guo, F.; Li, X.; Liang, D.; Li, T.; Zhu, P.; Guo, H.; Wu, X.; Wen, L.; Gu, T. P.; Hu, B.; Walsh, C. P.; Li, J.; Tang, F.; Xu, G. L. *Cell Stem Cell* **2014**, *15* (4), 447–459.

(56) Schön, A.; Kaminska, E.; Schelter, F.; Ponkkonen, E.; Korytiaková, E.; Schiffers, S.; Carell, T. *Angew. Chemie Int. Ed.* **2020**, *59* (14), 5591–5594.

(57) Robertson, A. B.; Klungland, A.; Rognes, T.; Leiros, I. *Cell. Mol. Life Sci.* **2009**, *66* (6), 981–993.

(58) Berti, P. J.; McCann, J. A. B. *Chem. Rev.* **2006**, *106* (2), 506–555.

(59) Izumi, T.; Wiederhold, L. R.; Roy, G.; Roy, R.; Jaiswal, A.; Bhakat, K. K.; Mitra, S.; Hazra, T. K. *Toxicology* **2003**, *193* (1–2), 43–65.

(60) Waters, T. R.; Swann, P. F. *J. Biol. Chem.* **1998**, *273* (32), 20007–20014.

(61) Morgan, M. T.; Bennett, M. T.; Drohat, A. C. *J. Biol. Chem.* **2007**, *282* (38), 27578–27586.

(62) Abu, M.; Waters, T. R. *J. Biol. Chem.* **2003**, *278* (10), 8739–8744.

(63) Raiber, E. A.; Murat, P.; Chirgadze, D. Y.; Beraldi, D.; Luisi, B. F.; Balasubramanian, S. *Nat. Struct. Mol. Biol.* **2015**, *22* (1), 44–49.

(64) Yang, N.; Galick, H.; Wallace, S. S. *DNA Repair (Amst).* **2004**, *3* (10), 1323–1334.

(65) Cannan, W. J.; Tsang, B. P.; Wallace, S. S.; Pederson, D. S. *J. Biol. Chem.* **2014**, *289* (29), 19881–19893.

(66) Rahimoff, R.; Kosmatchev, O.; Kirchner, A.; Pfaffeneder, T.; Spada, F.; Brantl, V.; Müller, M.; Carell, T. *J. Am. Chem. Soc.* **2017**, *139* (30), 10359–10364.

(67) Wu, S. C.; Zhang, Y. *Nat Rev Mol Cell Biol.* **2010**, *11* (9), 607–620.

(68) Münzel, M.; Globisch, D.; Carell, T. *Angew. Chemie Int. Ed.* **2011**, *50* (29), 6460–6468.

(69) Schiesser, S.; Hackner, B.; Pfaffeneder, T.; Müller, M.; Hagemeier, C.; Truss, M.; Carell, T. *Angew. Chemie - Int. Ed.* **2012**, *51* (26), 6516–6520.

(70) Liutkevičiutė, Z.; Kriukienė, E.; Ličytė, J.; Rudytė, M.; Urbanavičiutė, G.; Klimašauskas, S. *J. Am. Chem. Soc.* **2014**, *136* (16), 5884–5887.

(71) Schiesser, S.; Pfaffeneder, T.; Sadeghian, K.; Hackner, B.; Steigenberger, B.; Schröder, A. S.; Steinbacher, J.; Kashiwazaki, G.; Höfner, G.; Wanner, K. T.; Ochsenfeld, C.; Carell, T. *J. Am. Chem. Soc.* **2013**, *135* (39), 14593–14599.

(72) Schröder, A. S.; Kotljarova, O.; Parsa, E.; Iwan, K.; Raddaoui, N.; Carell, T. *Org. Lett.* **2016**, *18* (17), 4368–4371.

(73) Schröder, A. S.; Parsa, E.; Iwan, K.; Traube, F. R.; Wallner, M.; Serdjukow, S.; Carell, T. *Chem. Commun.* **2016**, *52* (100), 14361–14364.

(74) Traube, F. R.; Carell, T. *RNA Biol.* **2017**, *14* (9), 1099–1107.

(75) Jeltsch, A.; Jurkowska, R. Z. *Nucleic Acids Res.* **2016**, *44* (18), 8556–8575.

(76) Rahimoff, R.; Kosmatchev, O.; Kirchner, A.; Pfaffeneder, T.; Spada, F.; Brantl, V.; Müller, M.; Carell, T. *J. Am. Chem. Soc.* **2017**, *139* (30), 10359–10364.

(77) Zauri, M.; Berridge, G.; Thézénas, M. L.; Pugh, K. M.; Goldin, R.; Kessler, B. M.; Kriaucionis, S. *Nature* **2015**, *524* (7563), 114–118.

(78) Hecker, S. J.; Erion, M. D. *J. Med. Chem.* **2008**, *51* (8), 2328–2345.

(79) Gardelli, C.; Attenni, B.; Donghi, M.; Meppen, M.; Pacini, B.; Harper, S.; Di Marco, A.; Fiore, F.; Giuliano, C.; Pucci, V.; Laufer, R.; Gennari, N.; Marcucci, I.; Leone, J. F.; Olsen, D. B.; MacCoss, M.; Rowley, M.; Narjes, F. *J. Med. Chem.* **2009**, *52* (17), 5394–5407.

(80) Davey, M. S.; Malde, R.; Mykura, R. C.; Baker, A. T.; Taher, T. E.; Le Duff, C. S.; Willcox, B. E.; Mehellou, Y. *J. Med. Chem.* **2018**, *61* (5), 2111–2117.

(81) Zawada, Z.; Tatar, A.; Mocilac, P.; Buděšínský, M.; Kraus, T. *Angew. Chemie - Int. Ed.* **2018**, *57* (31), 9891–9895.

(82) Ludwig, J.; Eckstein, F. *J. Org. Chem.* **2002**, *54* (3), 631–635.

(83) Rahimoff, R.; Kosmatchev, O.; Kirchner, A.; Pfaffeneder, T.; Spada, F.; Brantl, V.; Müller, M.; Carell, T. *J. Am. Chem. Soc.* **2017**, *139* (30), 10359–10364.

(84) Zhao, M. Z.; Zhang, Y. W.; Yuan, F.; Deng, Y.; Liu, J. X.; Zhou, Y. L.; Zhang, X. X. *Talanta* **2015**, *144*, 992–997.

(85) Kamińska, E.; Korytiaková, E.; Reichl, A.; Müller, M.; Carell, T. *Angew. Chemie Int. Ed.* **2021**, *60* (43), 23207–23211.

(86) Takahashi, S.; Kobayashi, S.; Hiratani, I. *Cell. Mol. Life Sci.* **2018**, *75* (7), 1191–1203.

(87) Traube, F. R.; Schiffers, S.; Iwan, K.; Kellner, S.; Spada, F.; Müller, M.; Carell, T. *Nat. Protoc.* **2019**, *14* (1), 283–312.

(88) Smiley, J. A.; Kundracik, M.; Landfried, D. A.; Barnes, V. R.; Axhemi, A. A. *Biochim. Biophys. Acta* **2005**, *1723* (1–3), 256–264.

(89) Xu, S.; Li, W.; Zhu, J.; Wang, R.; Li, Z.; Xu, G. L.; Ding, J. *Cell Res.* **2013**, *23* (11), 1296.

(90) Plagemann, P. G. W. *J. Cell Biol.* **1972**, *52* (1), 131–146.

(91) Hüttel, R.; Büchele, F.; Jochum, P. *Chem. Ber.* **1955**, *88* (10), 1577–1585.

(92) Giziewicz, J.; Wnuk, S. F.; Robins, M. J. *J. Org. Chem.* **1999**, *64*, 2149–2151.

(93) Chehade, K. A. H.; Spielmann, H. P. *J. Org. Chem.* **2000**, *65* (16), 4949–4953.

(94) Sethi, D.; Patnaik, S.; Kumar, A.; Gandhi, R. P.; Gupta, K. C.; Kumar, P. *Bioorg. Med. Chem.* **2009**, *17* (15), 5442–5450.

(95) Canal-Martín, A.; Navo, C. D.; Sáez, E.; Molero, D.; Jiménez-Osés, G.; Pérez-Fernández, R. *Org. Biomol. Chem.* **2021**, *19* (33), 7202–7210.

(96) Heckel, A.; Seebach, D. *Chemistry (Easton).* **2002**, *8* (3), 559–572.

(97) Cunningham, A. D.; Qvit, N.; Mochly-Rosen, D. *Curr. Opin. Struct. Biol.* **2017**, *44*, 59.

(98) London, N.; Raveh, B.; Schueler-Furman, O. *Curr. Opin. Chem. Biol.* **2013**, *17* (6), 952–959.

(99) Banting F G, Best C H, Collip J B, Campbell W R, F. A. A. *Can. Med. Assoc. J.* **1922**, *12* (3), 141–146.

(100) Wang, L.; Wang, N.; Zhang, W.; Cheng, X.; Yan, Z.; Shao, G.; Wang, X.; Wang, R.; Fu, C. *Signal Transduct. Target. Ther.* **2022**, *7* (1), 1–27.

(101) Staś, M.; Bujak, M.; Broda, M. A.; Siodłak, D. *Pept. Sci.* **2016**, *106* (3), 283–294.

(102) Kaiser, D.; Videnov, G.; Maichle-Mössmer, C.; Strähle, J.; Jung, G. *J. Chem. Soc. Perkin Trans. 2* **2000**, No. 5, 1081–1085.

(103) Lucke, A. J.; Tyndall, J. D. A.; Singh, Y.; Fairlie, D. P. *J. Mol. Graph. Model.* **2003**, *21* (5), 341–355.

(104) Melby, J. O.; Nard, N. J.; Mitchell, D. A. *Curr. Opin. Chem. Biol.* **2011**, *15* (3), 369–378.

(105) Wang, W. L.; Yao, D. Y.; Gu, M.; Fan, M. Z.; Li, J. Y.; Xing, Y. C.; Nan, F. J. *Bioorg. Med. Chem. Lett.* **2005**, *15* (23), 5284–5287.

(106) Andersson, I. E.; Dzhambazov, B.; Holmdahl, R.; Linusson, A.; Kihlberg, J. *J. Med. Chem.* **2007**, *50* (23), 5627–5643.

(107) Siodłak, D.; Staś, M.; Broda, M. A.; Bujak, M.; Lis, T. *J. Phys. Chem. B* **2014**, *118* (9), 2340–2350.

(108) Singh, Y.; Stoermer, M. J.; Lucke, A. J.; Glenn, M. P.; Fairlie, D. P. *Org. Lett.* **2002**, *4* (20), 3367–3370.

(109) Liu, S.; Cai, C.; Bai, Z.; Sheng, W.; Tan, J.; Wang, H. *Org. Lett.* **2021**, *23* (8), 2933–2937.

(110) Kaur, D.; Khanna, S. *Comput. Theor. Chem.* **2011**, *963* (1), 71–75.

(111) Kim, Y. S.; Zhou, Y.; Bryant, H.; Milenic, D. E.; Baidoo, K. E.; Lewis, B. K.; Frank, J. A.; Brechbiel, M. W. *Bioorg. Med. Chem. Lett.* **2015**, *25* (10), 2056–2059.

(112) Yun, B. S.; Fujita, K. I.; Furihata, K.; Seto, H. *Tetrahedron* **2001**, *57* (48), 9683–9687.

(113) Debono, M.; Molloy, R. M.; Occolowitz, J. L.; Paschal, J. W.; Hunt, A. H.; Michel, K. H.; Martin, J. W. *J. Org. Chem.* **1992**, *57* (19), 5200–5208.

(114) Fukuhara, K.; Takada, K.; Okada, S.; Matsunaga, S. *J. Nat. Prod.* **2016**, *79* (6), 1694–1697.

(115) Lau, R. C. M.; Rinehart, K. L. *J. Am. Chem. Soc.* **1995**, *117* (29), 7606–7610.

(116) Vijaya Kumar, E. K. S.; Kenia, J.; Mukhopadhyay, T.; Nadkarni, S. R. *J. Nat. Prod.* **1999**, *62* (11), 1562–1564.

(117) Oku, N.; Adachi, K.; Matsuda, S.; Kasai, H.; Takatsuki, A.; Shizuri, Y. *Org. Lett.* **2008**, *10* (12), 2481–2484.

(118) Al-Wahaibi, L. H.; Mostafa, A.; Mostafa, Y. A.; Abou-Ghadir, O. F.; Abdelazeem, A. H.; Gouda, A. M.; Kutkat, O.; Abo Shama, N. M.; Shehata, M.; Gomaa, H. A. M.; Abdelrahman, M. H.; Mohamed, F. A. M.; Gu, X.; Ali, M. A.; Trembleau, L.; Youssif, B. G. M. *Bioorg. Chem.* **2021**, *116*.

(119) Linington, R. G.; González, J.; Ureña, L. D.; Romero, L. I.; Ortega-Barría, E.; Gerwick, W. H. *J. Nat. Prod.* **2007**, *70* (3), 397–401.

(120) Kehraus, S.; König, G. M.; Wright, A. D.; Woerheide, G. *J. Org. Chem.* **2002**, *67* (14), 4989–4992.

(121) Andersson, I. E.; Batsalova, T.; Dzhambazov, B.; Edvinsson, L.; Holmdahl, R.; Kihlberg, J.; Linusson, A. *Org. Biomol. Chem.* **2010**, *8* (13), 2931–2940.

(122) Al-Wahaibi, L. H.; Mostafa, A.; Mostafa, Y. A.; Abou-Ghadir, O. F.; Abdelazeem, A. H.; Gouda, A. M.; Kutkat, O.; Abo Shama, N. M.; Shehata, M.; Gomaa, H. A. M.; Abdelrahman, M. H.; Mohamed, F. A. M.; Gu, X.; Ali, M. A.; Trembleau, L.; Youssif, B. G. M. *Bioorg. Chem.* **2021**, *116*.

(123) Ishida, K.; Nakagawa, H.; Murakami, M. *J. Nat. Prod.* **2000**, *63* (9), 1315–1317.

(124) Lindquist, N.; Fenical, W.; Van Duyne, G. D.; Clardy, J. *J. Am. Chem. Soc.* **1991**, *113* (6), 2303–2304.

(125) Matsuo, Y.; Kanoh, K.; Yamori, T.; Kasai, H.; Katsuta, A.; Adachi, K.; Shin-ya, K.; Shizuri, Y. *J. Antibiot. (Tokyo).* **2007**, *60* (4), 251–255.

(126) Gruß, H.; Feiner, R. C.; Mseya, R.; Schröder, D. C.; Jewgiński, M.; Müller, K. M.; Latajka, R.; Marion, A.; Sewald, N. *Beilstein J. Org. Chem.* **2022**, *18*, 1–12.

(127) Hernández, J. N.; Martin, V. S. *J. Org. Chem.* **2004**, *69* (10), 3590–3592.

(128) Wasserman, H. H.; Vinick, F. J. *J. Org. Chem.* **1973**, *38* (13), 2407–2408.

(129) Thalhammer, A.; Mecinović, J.; Schofield, C. J. *Tetrahedron Lett.* **2009**, *50* (9), 1045–

1047.

(130) Phillips, A. J.; Uto, Y.; Wipf, P.; Reno, M. J.; Williams, D. R. *Org. Lett.* **2000**, *2* (8), 1165–1168.

(131) Plant, A.; Stieber, F.; Scherkenbeck, J.; Lösel, P.; Dyker, H. *Org. Lett.* **2001**, *3* (22), 3427–3430.

(132) Schäfler, L. Synthese potentieller Mannosidase und Glucosidase Inhibitoren, 1990.

(133) Yamada, H.; Tanigakiuchi, K.; Nagao, K.; Okajima, K.; Mukae, T. *Tetrahedron Lett.* **2004**, *45* (50), 9207–9209.

(134) Slobodyanyuk, E. Y.; Artamonov, O. S.; Kulik, I. B.; Rusanov, E.; Volochnyuk, D. M.; Grygorenko, O. O. *Heterocycl. Commun.* **2018**, *24* (1), 11–17.

(135) Wipf, P.; Miller, C. P. *J. Org. Chem.* **1993**, *58* (14), 3604–3606.

(136) Liang, X. Y.; Yang, J.; Qiu, Z. H.; Wang, L.; Lowary, T. L.; Shawn Fan, H. J. *Carbohydr. Res.* **2021**, *500*.

(137) Long, D. E.; Karmakar, P.; Wall, K. A.; Sucheck, S. J. *Bioorg. Med. Chem.* **2014**, *22* (19), 5279–5289.

(138) Jurczak, J.; Gryko, D.; Kobrzycka, E.; Gruza, H.; Prokopowicz, P. *Tetrahedron* **1998**, *54* (22), 6051–6064.

(139) Litvinchuk, M. B.; Bentya, A. V.; Grozav, A. M.; Yakovychuk, N. D.; Palamar, A. O.; Saliyeva, L. M.; Slyvka, N. Y.; Shishkina, S. V.; Vovk, M. V. *Monatshefte fur Chemie* **2023**, *154* (10), 1145–1159.

(140) Wang, Y.; Fan, S.; Zhong, W.; Zhou, X.; Li, S. *Int. J. Mol. Sci.* **2017**, *18* (9).

(141) Zhang, J.; Cao, J.; Zhou, J.; Zhang, J.; Zhao, Y.; Zhang, Z.; Liu, R. *J. Org. Chem.* **2023**, *88* (13), 9087–9093.

(142) Chen, Y.; Sun, X.; Wu, N.; Li, J.; Jin, S.; Zhong, Y.; Liu, Z.; Rogachev, A.; Chong, H. S. *Org. Biomol. Chem.* **2016**, *14* (3), 920–939.

(143) Ling, P. X.; Fang, S. L.; Yin, X. S.; Zhang, Q.; Chen, K.; Shi, B. F. *Chem. Commun.* **2017**, *53* (47), 6351–6354.

12. Acknowledgements

First, I am extremely grateful to *Prof. Dr. Anja Hoffmann-Röder* for the opportunity to work on this interesting project in her group as part of my PhD studies. I am thankful for *Prof. Hoffmann-Röder* for her continuous support and flexibility to find solutions to any existing challenges and always being available for scientific discussions.

I would like to thank *Prof. Dr. Oliver Trapp* to be the second reviewer of my thesis. I am also very thankful to *Prof. Dr. Roland Beckmann, Prof. Dr. Hendrik Zipse*, *Prof. Dr. Konstantin Karaghiosoff*, *Prof. Dr. Paul Knochel* for being on my defense committee.

Also, many thanks to *Dr. Markus Müller* for very useful scientific discussions, suggestions and for proofreading my thesis.

I had the pleasure of working with the members of Hoffmann-Röder group. Thanks to my lab mate *Sebastian Sander* and his *mom*, *Alexander "the dog" Herrmann, David "ghost boy" Kostadinov, Maxi Reif, Florian Mödl, Alexandra Sebeschuk and Seher Korkmaz*, for the nice work atmosphere and delightful interdisciplinary discussions. I would also like to recognize *Ursula Hülsmann* for the essential support in daily laboratory life.

Additionally, I would like to acknowledge all the students who I had the pleasure to supervise and who contributed to the success of this work: *Diana Atabaeva*, *Henrik Wilke* and *Adam O´Connell.*

Of course, many thanks go to my parents who have supported and encouraged me through the studies and after moving to Germany for PhD. Also, special thank you to *Dr. Susanne Schön* for helping with *Eero* and *Aatos*, that allowed me to proceed and finalize this work.

At last, I would like to express my deepest appreciation to my husband *Alexander* for his patience over many long, yet inspiring scientific discussions. Without your support and motivating attitude towards my work this endeavour would not have been possible. Kiitos kulta!