Synthesis of 2'-Fluorinated-2'-Deoxycytidine Derivatives to Investigate a Direct DNA Demethylation Pathway in Stem Cells

Master's Thesis
University of Jyväskylä
Department of Chemistry
Ludwig-Maximillians-Univärsität München
Department of Chemistry and Pharmacy
24.4.2018
Eveliina Ponkkonen

Abstract

This master's thesis is divided into a literature review and an experimental part. The literature review starts with an introduction of nucleic acids and epigenetics, covering both chemical- and biological aspects. These chemical- and biological processes are explained at a level that is needed to understand the aim of the experimental project and the purpose of the synthesized compounds. The main focus of this thesis is the 2'-deoxyribonucleic acid, especially 2'-deoxycytidine derivatives being in the leading role. The roles and functions of other components, like purine bases and histones, are not included in this thesis.

The latter part of the literature review, concerning the chemistry of 2'-deoxy pyrimidines and respective 2'-fluorinated derivatives, introduces their chemical properties. Based on the published data, this review will summarize different chemical approaches to introduce a variety of functional groups to the C5-postion of the pyrimidine ring to obtain C5 functionalized- 2'-deoxyuridine and 2'-deoxycytidine derivatives.

In the experimental part, the main objective was to synthesize 5-nitro-2'-fluoro-2'-deoxycytidine and 5-(butyl-4-acetoxy-benzoate)-2'-fluoro-2'-deoxycytidine to study their roles in the active demethylation process that occurs via C–C bond cleavage. The last steps of the latter mentioned compound requires further optimization and therefore remains under work. 5-Nitro-2'-fluoro-2'-deoxycytidine was further studied by feeding it to cultured mammalian cells. For the detection of its potential incorporation into DNA, UHPLC-MS/MS was used to provide quantitative data. The analysis showed no incorporation occurred into the DNA, however the nucleoside was found in the soluble pool and therefore, could have other biological implications.

This thesis gives a brief insight into the challenging field of chemical biology that is not studied during the master courses but is explained in the understandable manner for the students on master level. It will be of interest to an audience of multidisciplinary researchers in organic- and biological chemistry and it can be helpful for entering chemists to understand the chemical aspects concerning epigenetics and nucleoside chemistry as well as their connection to biology.

Tiivistelmä

Tämä opinnäytetyö jakautuu kirjallisuuskatsaukseen ja kokeelliseen osioon. Kirjallisuuskatsaus alkaa tutustumisella nukleiinihappoihin ja epigenetiikkaan kemiallisesta ja biologisesta näkökulmasta. Biologinen tausta on selitetty tasolla, mikä on tarpeellista kokeellisen osion ymmärtämiseksi, sekä ymmärtääkseen mikä on syntetisoitujen yhdisteiden tarkoitus. Tämän kirjallisuuskatsauksen pääroolissa ovat 2'-deoksyribonukleiinihapot, erityisesti 2'-deoksysytidiini johdannaiset. Muiden komponenttien, kuten puriini emästen ja histonien rooli ei sisälly tähän tutkielmaan.

Toisessa osiossa kohteena ovat 2'-deoksipyrimidiinien ja niiden 2'-fluorinoitujen johdannaisten kemia, jossa esitellään näiden yhdisteiden kemiallisia ominaisuuksia. Perustuen julkaistuihin tutkimustuloksiin, tämä kirjallisuustutkielma kokoaa erilaisia lähestymistapoja lisätä eri funktionaalisia ryhmiä C5 asemaan pyrimidiini emäksissä.

Kokeellisen osion pääkohteena oli syntetisoida 5-nitro-2'-fluoro-deoksisytidiini ja 5-(butyyli-4-asetoksi-bentsoaatti)-2'-fluoro-deoksisytidiini työkaluiksi aktiivisen demetylaation tutkimiseen, joka tapahtuu C–C sidoksen katkeamisen kautta. Jälkimmäisenä mainitun yhdisteen viimeiset vaiheet ovat vielä työnalla. 5-nitro-2'-fluoro-deoksisytidiini syötettiin nisäkkään kantasoluille, joissa sen toivottiin yhdistyvän niiden genomiin. Potentiaalinen DNA:han liittyminen analysoitiin ultra korkean erotuskyvyn nestekromatografialla, joka oli kytketty tandem massaspektrometriin. Analysiin perusteella voidaan sanoa, ettei syötetty nukleosidi liittynyt DNA:han, mutta löydettiin solulimaan liukenevasta osasta. Tämä löydös vaatii lisätutkimusta, jotta yhdisteen biologinen merkitys kantasoluissa saadaan selville.

Tämä opinnäytetyö on suunnattu poikkitieteelliselle orgaanisen- ja biologisen kemian aloista kiinnostuneille. Se helpottaa alalle tulevia uusia kemistejä ymmärtämään epigenetiikaa kemiallisesta näkökulmasta ja sen yhteydestä biologiaan.

Preface

This interdisciplinary master thesis was conducted in the Carell Group at the University of Ludwig-Maximillians-Universität München during 1st of August 2017 and 23rd of May 2018. Prof. Doc. Thomas Carell and Academic Prof. Doc. Kari Rissanen were the supervisors of the thesis.

I want to express my gratitude Prof. Thomas Carell for arranging and supervising an extremely exciting topic for the thesis. I also want to thank the Carell Group for a very welcoming atmosphere and enthusiastic attitude towards the research.

Secondly, I want to thank Prof. Doc. Petri Pihko for arranging very demanding and challenging organic chemistry courses. The exciting lectures showed interesting side of chemistry and leaded me to choose organic chemistry as my major.

I also want to thank my parents for supporting me during my studies. Special thanks to my dad for patience and providing me the mathematical and technical understanding.

I want to address the special thanks to Eva Korytiaková for mentoring me to the topic of the thesis and giving invaluable knowledge with her invincible competence. Also, thanks for taking me into the coolest lab.

Table of Contents

Acronyms, Abbreviations, Symbols and Definitions

A	adenosine
Ac	acetyl
AcOH	acetic acid
anti	alignment of two substituents on the opposite sides / faces of a compound
aq.	aqueous
Ar	aromatic group
B	base
BER	base excision repair
Bn	benzyl
Boc	*tert*-butyloxycarbon
bp	base pair
Bu	butyl
c	concentration
c.	circa
C	cytidine
calcd.	calculated
cis	prefix that describes the position of functional groups attached on the same side of the molecule
CpG	cytosine-phosphate-guanine
Cy	cyclohexyl
d	doublet
dA	2'-deoxyadenosine
DABCO	triethylamine
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
DBH	1,3-dibromo-5,5-dimethylhydantoin
DCC	*N,N'*-dicyclohexylcarbodiimide
dC	2'-deoxycytidine
DCM	dichloromethane
dd	doublet of doublets
ddd	doublet of doublet of doublets
DEAD	diethyl azodicarboxylate
dG	2'-deoxyguanosine

DIPEA	*N,N*-diisopropylethylamine, Hünig's base
DMAP	*N,N*-dimethylpyridin-4-amine
DME	1,2-dimethyoxyethane
DMF	*N,N*-dimethylformamide
DMSO	dimethylsulfoxide
DNA	2'-deoxyribonucleic acid
ds	double-stranded
dt	doublet of triplets
dT	2'-deoxythymidine
dU	2'-deoxyuridine
epi	unnatural isomer
ESI-MS	electrospray ionization mass spectrometry
ESC	embryonic stem cell
Et	ethyl
et al.	and others
EtOAc	ethyl acetate
EtOH	ethanol
EWG	electron withdrawing group
G	guanosine
GC	gas chromatography
HMDS	hexamethyldisilazane
HOMO	highest occupied molecular orbital
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
i-	iso
in vacuo	under reduced pressure
IR	infrared
IUPAC	International Union of Pure and Applied Chemistry
J	coupling constant
LUMO	lowest unoccupied molecular orbital
m/z	mass to charge ratio
M	metal, mol / L (molarity)
m	multiplet, prefix milli- (10^{-3})
Me	methyl
MeCN	acetonitrile

mESC	mouse embryonic stem cell
MNP	2-methyl-2-nitrosopropane
n-	normal, linear chain
NMR	nuclear magnetic resonance
Nu	nucleophile
[O]	oxidation
PG	protecting group
Ph	phenyl
Pr	propyl
pyr	pyridine
quin	quintet
q	quartet
R	arbitrary substituent
R	*rectus*, Latin for right, used in the nomenclature of enantiomers
RNA	ribonucleic acid
rt	room temperature
R_f	Retardation factor
s	singlet
sec-BuLi	*sec*-buthyl lithium
sex	sextet
ss	single-stranded
S	*sinister*, Latin for left, used in the nomenclature of enantiomers
sat.	saturated
syn	alignment of two substituents on the same side / face of a compound
t	triplet
t-	*tert*-, tertiary
T	thymidine
TBDMS	*tert*-butyldimethylsilyl
td	triplet of dublets
TDG	thymine DNA glycosylase
TET	ten-eleven translocation
Tf	triflyl
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TIPS	triisopropylsilyl

TLC	thin layer chromatography
TMSOTf	trimethylsilyl trifluoromethanesulfonate
Tol	toluene
trans	prefix that describes the position of functional groups attached on the opposite sides of the molecule
Ts	tosyl
TsOH	*p*-toluenesulfonic acid
U	uridine
UV	ultraviolet
α	referring to the position of the carbon in relation to a functional group, α is the position adjacent to the functionalized carbon
β	referring to the position of the carbon in relation to a functional group, β is the position one carbon further than α
Δ	difference, heating
δ	chemical shift in parts per million downfield from TMS

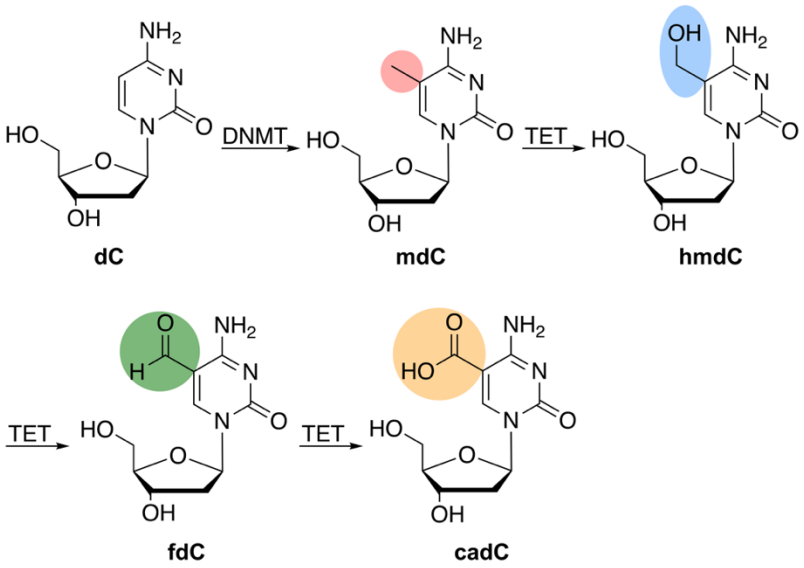
Literature Review

1. Introduction

The discovery of nucleic acids by Friedrich Miescher in 1868[1] was left under the shadow of the groundbreaking inventions of the electric light bulb and the telephone which were invented at the same time[2]. 80 years later, in 1944, Oswald Avery[3] provided the findings of the function, purpose and utility of nucleic acids. In 1953, the structure of the DNA – the double helix – was confirmed by James Watson and Francis Crick Watson[4] in collaboration with Rosalind Franklin[5,6] and Maurice Wilkins[7]. Hence, it was established that the genetic material of living organisms is constructed from four different canonical nucleobases; 2'-deoxyadenosine (**dA**), 2'-deoxycytidine (**dC**), 2'-deoxyguanosine (**dG**), and 2'-deoxythymidine (**dT**) and this genetic code is stored in the nucleus of every cell.

Specialized cells in multicellular organisms perform specific functions, for example neurons and fibroblasts. These cells differ dramatically in both function and structure but still have the same sequence of information stored in the nucleus of the cells. The differences in their functions lay in gene regulation which involves controlled activation and silencing of specific genes. Gene regulation leads to the production of the spesific proteins in one cell type but not in the others. This gene expression is operated on many levels beyond the canonical base sequence and is provided by epigenetics.[8,9]

Epigenetic modifications take place at the C5 position of the canonical dC. dC is methylated by DNMT forming a 5-methylated 2'-deoxycytidine (**mdC**) which is known to be oxidized to 5-hydroxymethyl-2'-deoxycytidine (**hmdC**) and further to 5-formyl 2'-deoxycytidine (**fdC**) and 5-carboxy-2'-deoxycytidine (**cadC**) by Ten eleven translocation (TET) family of Fe(II)/α-ketoglutarate-dependent dioxygenase-mediated process (Scheme 1).[10] These functionalized bases are considered to construct the second, more flexible information layer of nucleic acids that controls the activity of the genes.



Scheme 1. Epigenetic bases formed as a result of oxidation catalyzed by TET.

DNA methylation is well studied field whereas, DNA demethylation has not been deciphered yet. While trying to understand the active demethylation of DNA, several mechanisms involving epigenetic bases as intermediates have been proposed. To gain better insights into DNA demethylation pathways, an investigation of TET mediated demethylation processes is necessary. This makes 2'-deoxycytidine derivatives suitable target molecules and tools for investigating the DNA demethylation mechanisms.

In the first part of this literature review the basics of nucleic acid chemistry are introduced. Subsequently, the DNA repair mechanisms are explained, followed by epigenetics focusing on modified dCs. These biological chemistry aspects are reviewed concentrating on the processes in mammalian cells. For the simplicity, β-D designation of the nucleosides is omitted from their names but this configuration is implied. Also, the epigenetic modifications are mentioned as mdC, hmdC, fdC and cadC, referring always to the C5 functionality unless noted otherwise.

In the second part of the literature review, chemical properties behind 2'-deoxypyrimidine nucleosides and respective 2'-fluorinated derivatives are introduced, followed by different approaches to functionalize the C5 position of the pyrimidine ring. Presented molecules are of interest in studying epigenetics, and other compounds, like antiviral agents, are ruled out. Explanations are illustrated with figures and with mechanistic schemes.

2. Chemical Composition of DNA and RNA

2.1. Nucleotides

Nucleotides, monomers of nucleic acids, are constructed of three parts; a heterocyclic base, a sugar and of one to three phosphate groups. In Figure 1 is presented a structure of cytidine-5'-monophosphate, which is a nucleotide (base + sugar + phosphate group). This structure without a phosphate group is a nucleoside and it is constructed of a cytosine base (black) and a sugar (red), which is 2'-deoxyribose for DNA nucleosides. DNA bases are heteroaromatic compounds and they are divided into purines (adenine and guanine) and pyrimidines (cytosine and thymine). The sugar component is 2'-deoxyribose (R = H) and for RNA nucleosides, the R group is a hydroxyl group. Another difference between DNA- and RNA nucleosides is that the base thymine is found in DNA whereas in RNA it is substituted by the base uracil. The bond attaching the nucleobase to the sugar is a β-*N*-glycosidic bond (blue). The glycosidic bond is always between the C1' of the sugar and N9 of the purine or the N1 of the pyrimidine. The rotation about this bond gives rise to *anti*- and *syn*-conformations, where the *anti*-conformation is favored due to steric factors.

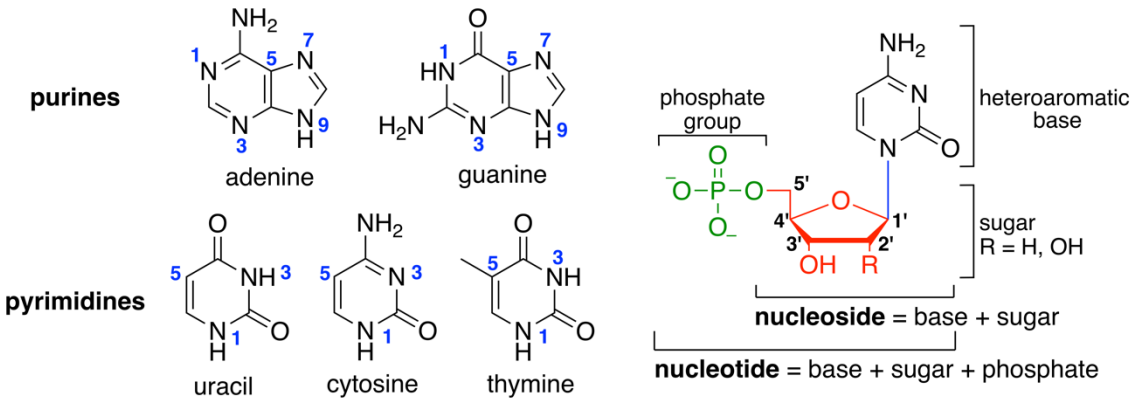
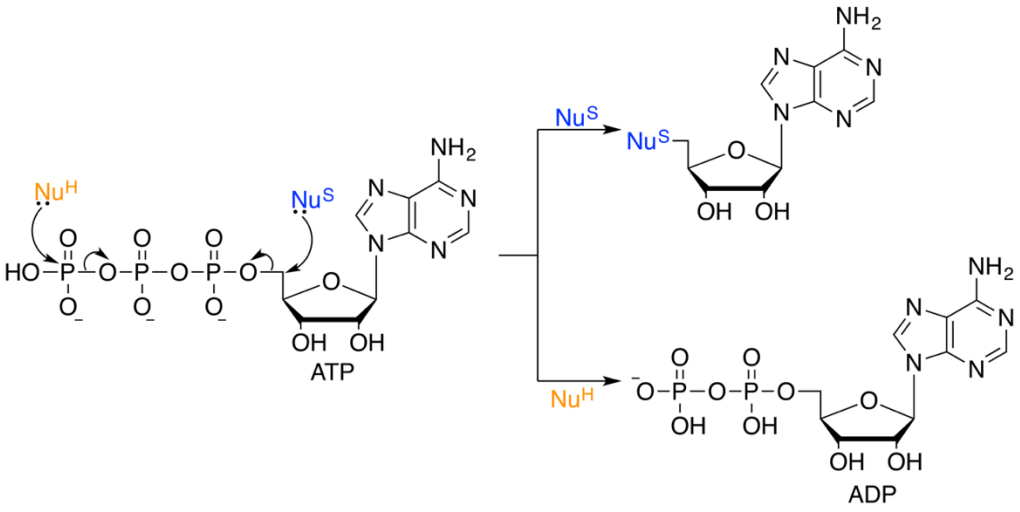


Figure 1. Structure of a nucleotide and five canonical nucleobases.

One of the most important nucleotides is the ribonucleotide adenosine triphosphate (ATP). It serves as a short-term carrier of chemical energy in cells. ATP is synthesized in a phosphorylation reaction of adenosine diphosphate (ADP). It is a highly reactive molecule because the phosphates groups are excellent leaving groups due to their very low lying LUMO.[11] The break of a

phosphoanhydride bond and formation of an energetically favorable phosphoester bond releases a large amount of energy (-ΔG). In particular, the terminal phosphate group is suitable for an attack by a hard nucleophile (Nu^H), like H_2O or a hydroxyl group from an alcohol (Scheme 2). Therefore, ATP is often hydrolyzed transferring a phosphate to other molecules, thus making them more reactive.[11] These type of reactions are involved in the synthesis of phospholipids and in the sugar catabolizing reactions. The large -ΔG is the consequence of the removal of unfavorable repulsion between adjacent negative charges from neighboring phosphate groups. The released inorganic phosphate ion (P_i) is stabilized by resonance and by favorable hydrogen bonding with water. After the first hydrolysis of ATP to ADP, the second dephosphorylation can take place and form adenosine monophosphate (AMP). Soft nucleophiles (Nu^S), like methionine, attack ATP at the C5'-position releasing the triphosphate. An example of this reaction is the *S*-adenosylmethionine (SAM) synthesis, which will be discussed in chapter 4.1.



Scheme 2. Possible reactive sites for a nucleophilic attack at ATP.

2.2. Nucleic Acids

Nucleotides form long polynucleotide chains – nucleic acids. Nucleotides are linked covalently into each other forming a sugar phosphate backbone. The link between the nucleotides consists of a phosphodiester bond between the 5'- and 3'-carbon atom of the ribose. Since nucleotides are linked together via 5'- and 3'-carbon atoms, they are always read in direction from 5' end to 3' end. The 5' end

of a nucleic acid consists of a free phosphate group and the 3’ end carries an unreacted hydroxyl group which is shown in Figure 2.

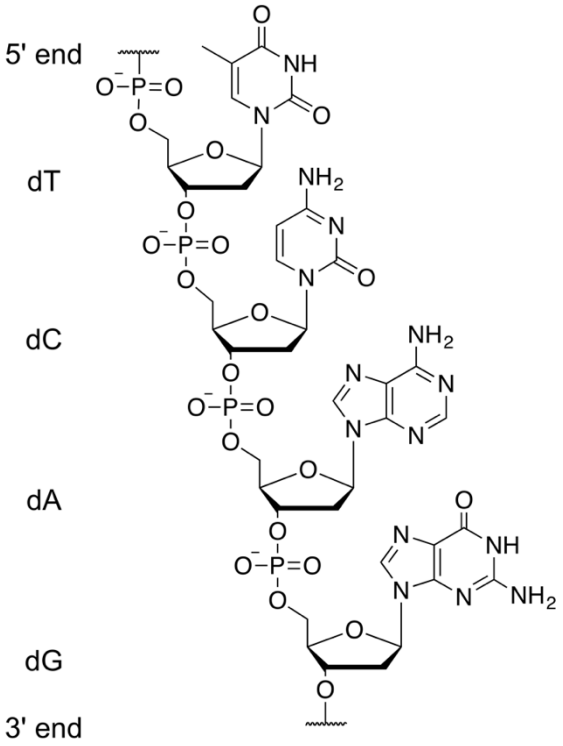


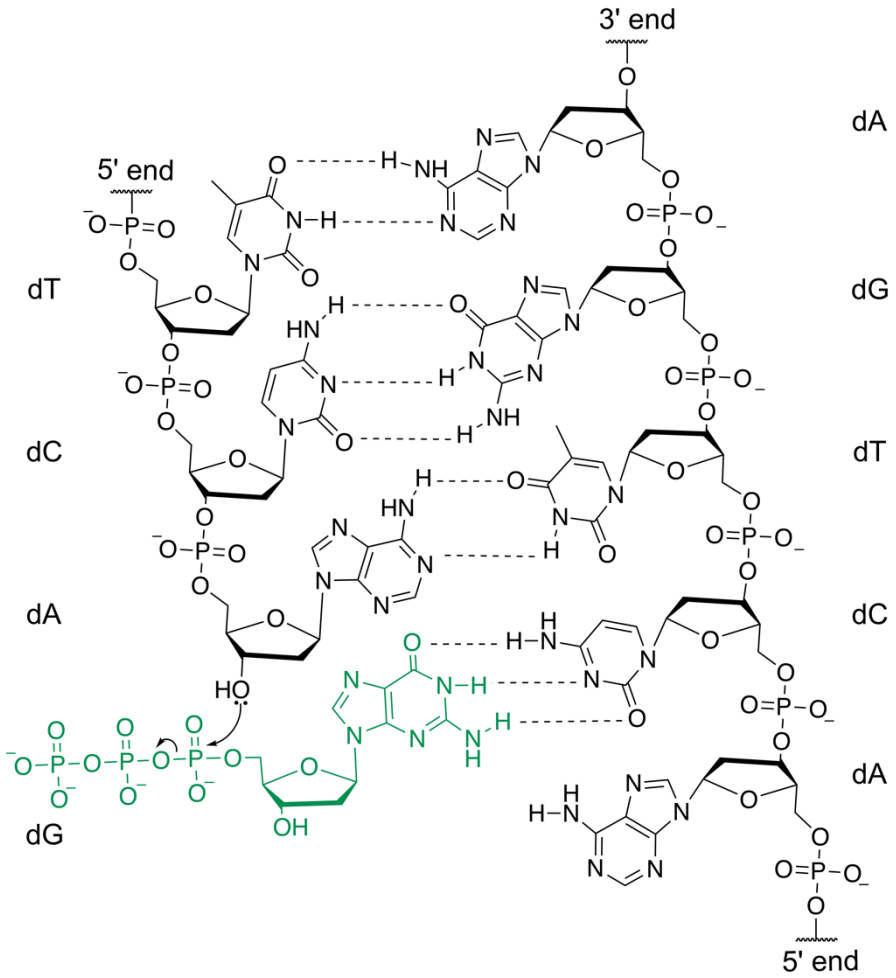
Figure 2. DNA section consisting of four nucleotides dT, dC, dA and dG, where the nucleotides are linked to each other with phosphodiester bonds.

There are two types of nucleic acids; ribonucleic acids (RNA) and deoxyribonucleic acids (DNA). RNA can be found in a nucleus and cytoplasm of a cell and DNA is situated in the nucleus and mitochondria. Nucleic acids have different functions in a cell. They encode the genetic information and are also involved in storage, transfer and expression of genetic information in all living organisms. Nucleic acids are also present in viruses.

2.3. DNA Structure

Most living cells store their hereditary information in form of double-stranded deoxyribonucleic acids (DNA). This double helix consists of two DNA single strands coiled around each other. The binding between the two single strands occurs in a strict rule defined by the complementary structures of the bases which are also called Watson-Crick base pairs. The base pairing occurs through

hydrogen bonding between these nucleotides; dT: dA and dC: dG. The covalent sugar-phosphate bonding of the DNA is strong compared to hydrogen bonding between base pairs. This allows the two DNA strands to be pulled apart while the backbone stays intact. When new DNA is synthesized, these separated strands serve as a template for new strands. Strict base pairing properties define which one of the four monomers can be added to the growing strand and therefore, newly formed strands are always complimentary (Scheme 3). This process is called DNA replication. The sequence of these four canonical bases establishes the sequence information layer in DNA, consisting of protein coding and regulatory elements.[12]



Scheme 3. Base pairing in a DNA double strand according to Watson-Crick base-pairing rules and addition of a new nucleotide, dG, during DNA replication.

2.4. Non-canonical Nucleobases

Till today day, four non-canonical, epigenetically relevant, nucleobases which play an important role in gene regulation are known. One of the most important epigenetically relevant nucleobases is the methylation product of dC – mdC. It was the first identified additional genetic element.[13–15] Furthermore, it was considered to be a stable epigenetic mark[10,16] but later it was recognized as a regulatory epigenetic element.[8] Moreover, mdCs correspond about 1% of the human genome.[17] The presence of this base in a specific promoter segment leads to silencing the transcription of the corresponding gene.[18] The other three epigenetically relevant bases, isolated from neuronal tissue and from mouse embryonic stem cells (mESC), were discovered in 2009 and 2011. In 2009, the same time two groups discovered hmdC.[10,16] It was found in granule cells, stem cells, Purkinje neurons and other tissues in vertebrate DNA.[16] Furthermore, it was found that hmdC is formed by the oxidation of mdC, catalyzed by TET.[10] A few years after the discovery of hmdC, the two other oxidation products of mdC; fdC[19,20] and cadC[20,21] were discovered and linked to epigenetic modifications (Figure 3).[19–21]

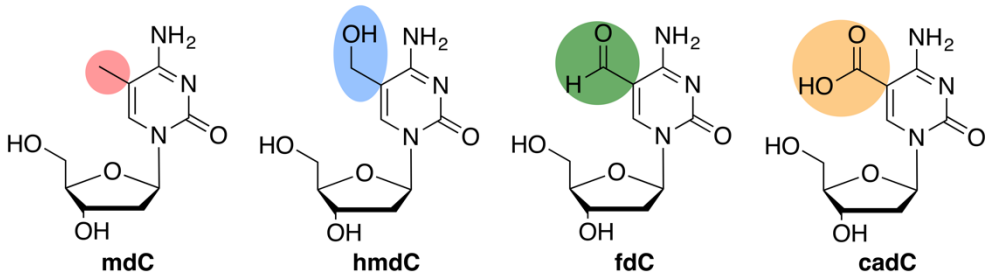


Figure 3. Epigenetically relevant nucleosides, mdC, hmdC, fdC and cadC, that take part in gene regulation.

At the time when hmdC was found, Carell and co-workers developed an MS-based method to study the distribution and absolute levels of oxidized mdC-derivatives in different tissues.[22,23] The quantitative LC-MS studies involve enzymatic digestion of the isolated DNA to nucleosides, which are then separated by liquid chromatography and analyzed with mass spectrometry. Later, high-resolution mass spectrometry was substituted with more sensitive triple quadruple mass spectrometry which detects molecule specific fragmentation

reactions. To get quantitative information of the investigated compounds, internal standards, like isotopically labelled analogues of the compounds are used.[24]

As a result of studies based on this method, it was observed that hmdC is present in many tissues[25]; highest levels were found in brains[23] (Table 1), especially in the areas associated with higher cognitive functions.[26] It was also observed that the same gene can have different hmdC levels in different tissues, which shows that the hmdC levels are determined by the tissue type instead of expression of the respective gene.[26] The levels of mdC are constant, 4-5% per dG nucleotide in different tissue types (Table 1). These observations support the view that unlike mdC, hmdC does not have a post replicative mechanism to maintain the levels in rapidly proliferating cells and that it accumulates at high levels only in post mitotic cells. Among this finding, it was confirmed that fdC is present not only in mESCs but also in human neurons.[27,28] In all investigated tissue types, fdC levels are low in comparison to mdC or hmdC. The detected cadC levels in these tissues are even lower – 1/10 of fdC (Table 1).[21] However, increased levels of cadC have been found in certain cancer tissues.[29,30]

Table 1. The absolute levels of epigenetically relevant nucleosides in different tissues and embryonic stem cells given as nucleosides per dG.

tissue	mdC (10^{-3})	hmdC (10^{-4})	fdC (10^{-7})	cadC (10^{-7})
mESC	7.3	7.3	10.9 (10^{-6})	5.2
liver	8.3	2.7	1.7	0
kidney	8.4	3.8	1.9	0
heart	8.0	4.2	1.6	0
brain	9.8	11.6	4	0

At first, fdC was discovered based on mass spectrometry data in the genome of mESCs.[19] After this finding, also small amounts of cadC were observed[20]. At the same time, a study, supported by isotope standards[22], described increased levels of cadC in stem cells lacking the enzyme thymine-DNA glycosylase (TDG).[21] Soon after this finding, another study confirmed that TDG is able to recognize and cleave cadC and fdC.[31] The function behind fdC and cadC is not fully

understood, but aforementioned studies support the possibility that they are involved in an active demethylation process. This idea is supported by the findings that mammalian TDG enzyme removes fdC and cadC. This also shows the connection between the base excision repair and epigenetics that allows regulation of transcriptional activity.[32,33] Findings of reader proteins, specific for hmdC, fdC and cadC suggest that these nucleosides have independent epigenetic functions[34,35] and that their presence is epigenetic all adjusted in regulatory elements like promoters and enhancers[36,37].

A study, using isotope- and metabolic labeling by mass spectrometry suggests fdC to be a permanent base[38], whereas according to another study, fdC is considered more a semi-permanent base with a limited lifetime.[39] In contrary to these findings, some studies claim based on single cell analysis data a high cell-to-cell heterogenity of fdC.[36] This might be caused by the fast turnover of fdC, thereby suggesting fdC to be a short-lived demethylation intermediate.

The discovery of these epigenetic bases shows how epigenetic plasticity can be achieved. The C-atom connected to the C5-position of dC has a different oxidation state in each of the four aforementioned non-canonical nucleobases. This established oxidation code can be considered to have regulatory purposes.[40,41] The role of these epigenetic bases is further discussed in chapter 4.2.

3. DNA Repair

In order to maintain the genetic stability and preventing the changes in DNA sequence, mechanisms for DNA repair are essential. The double-helix structure of DNA is suited for repair because it carries two separate copies of all the genetic information; one in each of its two strands. Thus, when one strand is damaged, the complementary strand retains an intact copy of the same information. This copy is generally used to restore the correct nucleotide sequences of the damaged strand. Possible DNA damages, like stress induced oxidation (purple), hydrolysis (blue) and uncontrolled methylation (green) are presented in Figure 4.[9] Other frequent damages like depurination[42] which releases guanine or adenine resulting in an abasic site (AP) in DNA or deamination[43] of dC to dU are severe changes as well (Scheme 4). These reactions normally occur in double

helical DNA but for convenience one strand is presented in Scheme 6. A variety of different mechanisms can remove these harmful changes in the base sequence in cells. In this thesis two pathways are introduced.

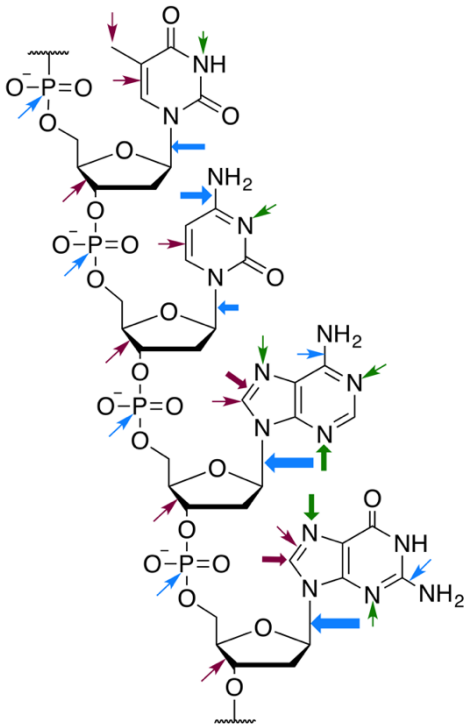
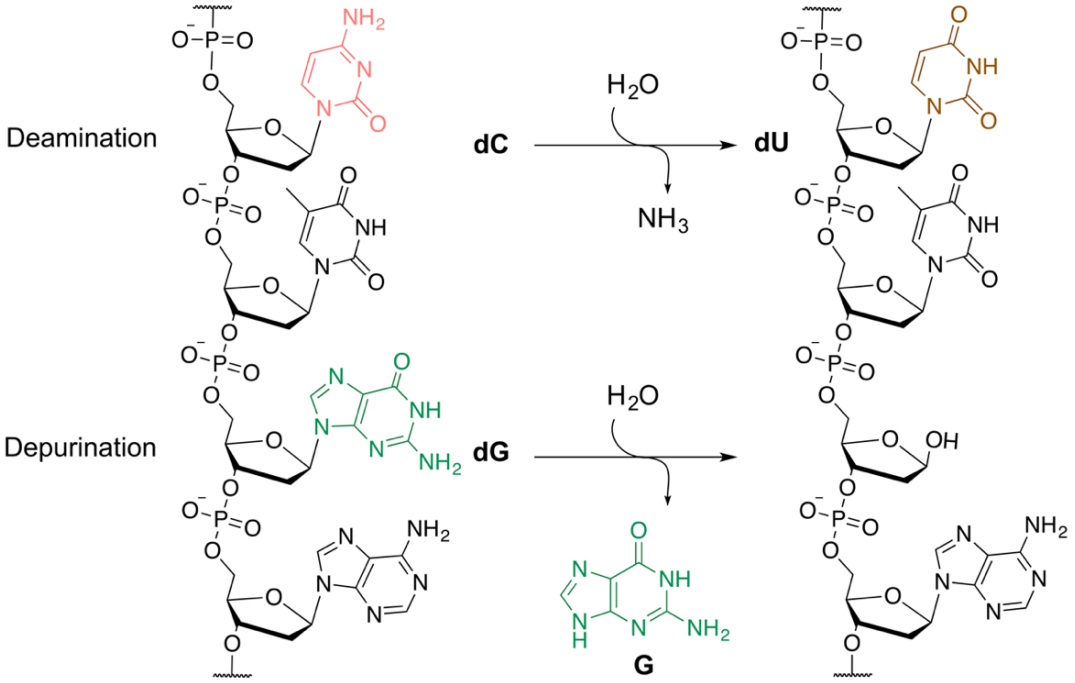


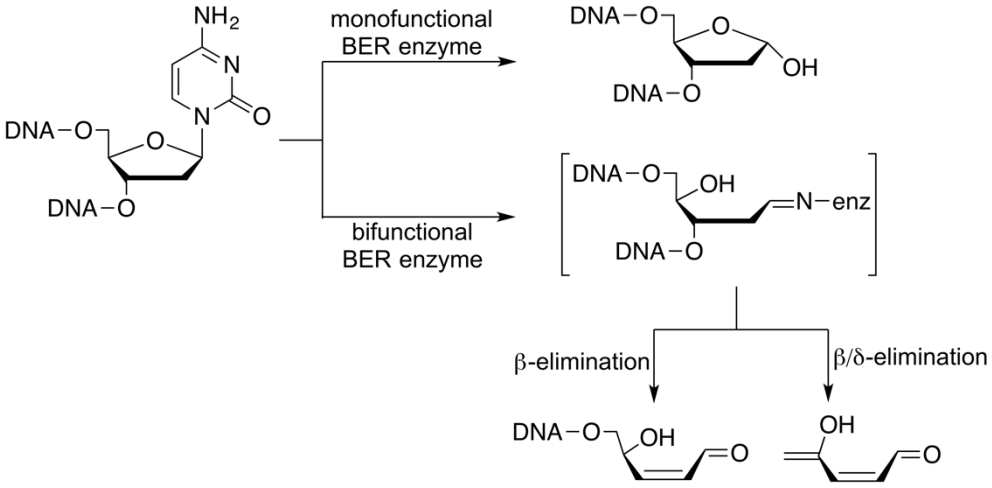
Figure 4. Sites for alterations on nucleotides presented by the relative frequency for each event, indicated by the width of the arrow. Purple color indicates oxidation, blue hydrolysis and green methylation.[44]



Scheme 4. The most frequent chemical reactions that create DNA damage in cells.

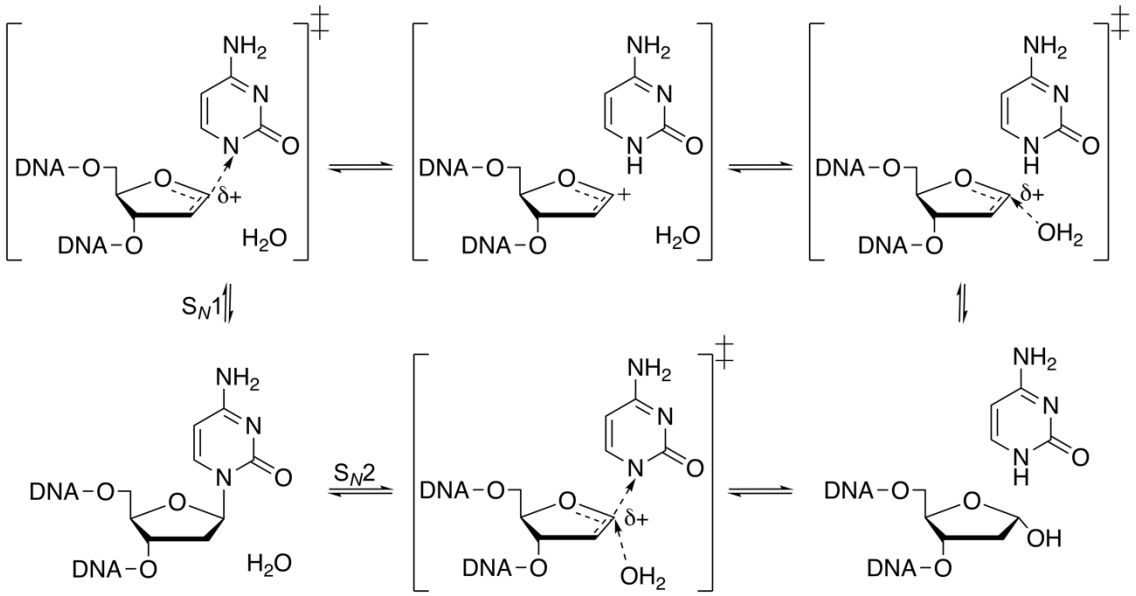
3.1. Base Excision Repair

Base excision repair (BER), corrects DNA damage caused by oxidation, deamination and alkylation. It occurs in two stages: an initial, damage specific step, carried out by a battery of enzymes, like DNA glycosylases.[45–51] Monofunctional BER enzymes are DNA glycosylases that hydrolyze the *N*-glycosidic bond creating an abasic site (Scheme 5). It is further processed by an AP endonuclease. This endonuclease cleaves the DNA backbone together with a phosphodiesterase on the 5'-side leaving a single nucleotide gap in the DNA strand. The gap is filled with a new nucleotide by DNA polymerase and DNA ligase.[48] Bifunctional BER enzymes have both glycosylase and lyase activities – they remove the damaged nucleobase by glycosyl transfer with an amine nucleophile. Formed Schiff base (imine) intermediate is removed by β-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.[49]



Scheme 5. Mono- and bifunctional BER enzymes have different mechanisms for cleaving the nucleobase.

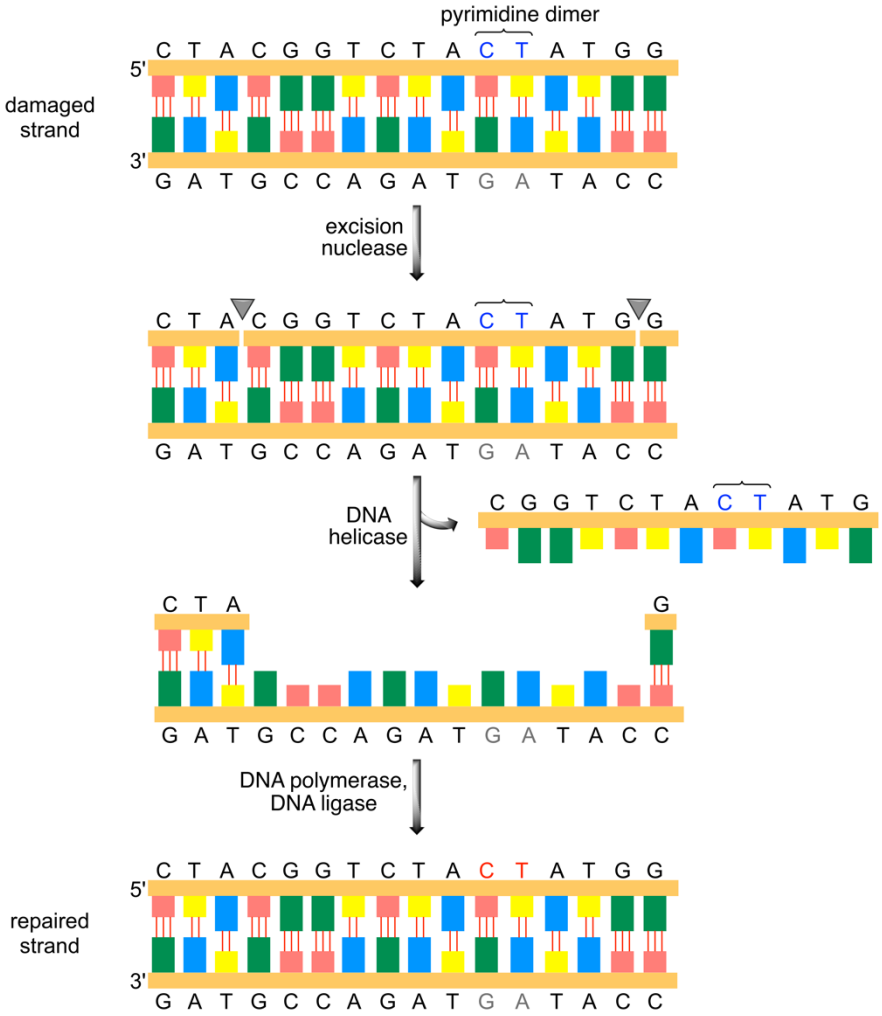
Two pathways for nucleobase hydrolysis have been suggested[48], dissociative S_N1 and associative S_N2 pathways, which are presented in Scheme 6. Calculations show these pathways to be almost isoenergetic but the S_N1 pathway to be slightly more favorable for canonical and damaged nucleosides. Furthermore, damaged nucleosides exhibit reduced glycosidic bond stability compared to undamaged ones, meaning that DNA damage does not change the deglycosylation mechanism. Protonation of nucleosides at different sites predicts the positions leading to the largest reductions in the deglycosylation barrier and they are typically used by DNA glycosylases to facilitate base excision.[47,48,51]



Scheme 6. Two possible mechanisms for nucleobase hydrolysis.

3.2. Nucleotide Excision Repair

Nucleotide excision repair is based on a process that senses distortion in the double helix of the DNA and is more general for different types of DNA damages – large damages or changes in the double helix structure. These damages can result from the reaction of DNA bases with bulky aromatic compounds, like benzopyrene or pyrimidine dimers caused by UV radiation. Once the damaged lesion is recognized, a denaturation bubble opens up around the lesion and endonuclease nicks the lesion from the phosphodiester backbone at two sites of the distortion. DNA helicase removes the oligonucleotide containing single strand. Formed large gap is repaired by DNA polymerases and the nick is sealed by DNA ligases which is depicted in Scheme 7.[9,52]



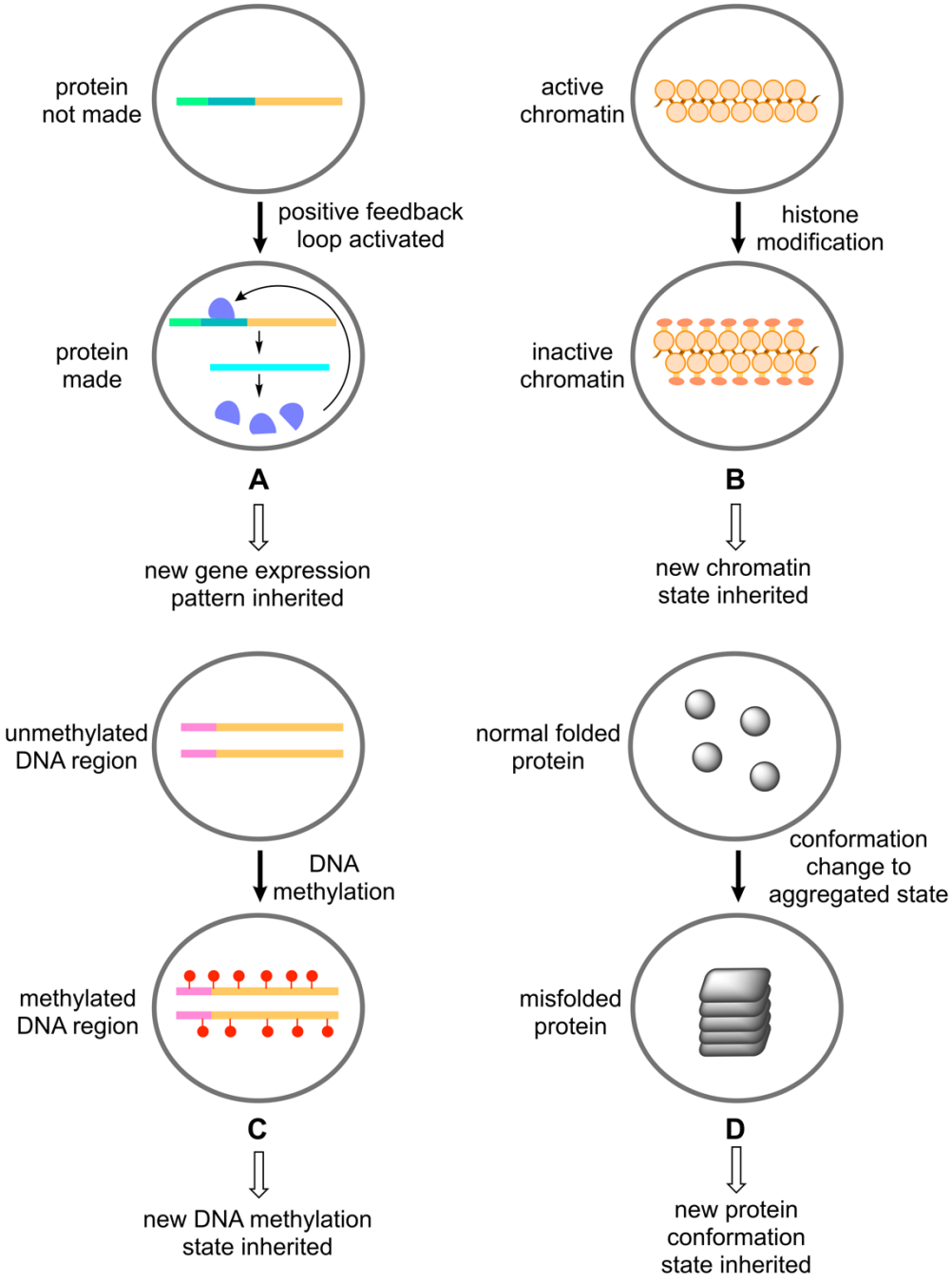
Scheme 7. Nucleotide excision repair.

4. Epigenetics

The term epigenetics refers to heritable and reversible alterations in gene regulation that aren't dependent on the DNA sequence. Epigenetic modifications alter DNA accessibility and thereby regulate the patterns of gene expression. An epigenome is the sum of chemical modifications occurring on the DNA and on histone proteins it is wrapped around. Epigenetic modifications include DNA methylation, histone methylation, acetylation, ubiquitination and phosphorylation. The presence of non-canonical bases mdC, hmdC, and fdC is firmly established in the genomes of higher eukaryotes, however cadC has been detected only in stem cells but not in other somatic cell types.[27]

Even though the same genetic content is found in all cells (except gametes), their function varies largely from each other. These cells differ in the number of active genes. This epigenetic regulation ensures expression of particular genes and transmission of stable patterns of gene expression to daughter cells. Genetic inheritance is based on the direct inheritance of the DNA sequence during the replication as discussed earlier. There are four mechanisms that can produce epigenetic forms of inheritance and they are presented in Scheme 8: positive feedback (**A**), histone modification (**B**), DNA methylation (**C**) and protein aggregation state (**D**). The molecules bound to DNA in these different mechanisms are playing the main role and are therefore less permanent than a change in DNA sequence in genetic inheritance. It is often the case that epigenetic information is erased during the formation of eggs and sperm.[9]

In this chapter, the chemistry behind dC modifications and enzymes catalyzing them are discussed. Since the studies have shown that epigenetics play an important role in many types of disease, they will be discussed later in this chapter.



Scheme 8. Different mechanisms to produce epigenetic inheritance. A) Positive feedback loop activated; protein production. B) Histone modification, where active chromatin gets inactivated. C) DNA methylation; unmethylated DNA region gets methylated. D) Protein aggregation state; normally folded protein goes through conformation change to misfolded protein.

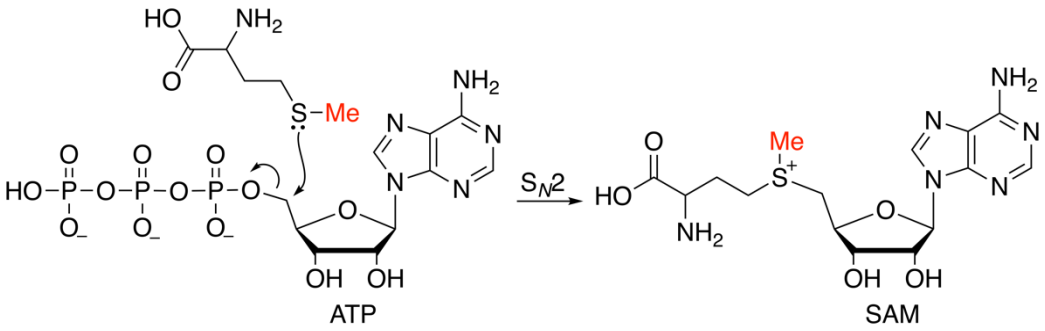
4.1. DNA Methylation

DNA methylation occurs primarily in cytosine nucleotides and especially in the context of cytosine-phosphate-guanine (CpG) dinucleotides[53]. The presence of mdC affects the transcriptional activity of genes[54], effecting transcriptional repressors, blocking the binding of transcription factor and leading to silencing of the corresponding gene.[55]

There are approximately 28 million CpGs in the human genome of which 70-80% are methylated.[56] During the course of evolution CpG sequences nucleotides tend to get eliminated which has led to significant decrease of CpGs – more than three out of four CpGs have been lost. This can be explained the way how DNA repair enzymes work; deamination of dC leads to dU, which is not a DNA base, and thus is recognized by the DNA repair enzyme – uracil DNA glycosylase. In the case of deamination of mdC, repairing does not work this way: the deamination product of mdC is dT which cannot be distinguished from the nonmutant dT nucleotides in the DNA. Even though a repair system for removing these mutant dT nucleotides exists, many deaminations are undetectable. This leads evolutionary pressure to deplete this dinucleotide.

CpG islands are dense regions of CG nucleotides. They are 300-3000 base pairs long and are often situated at transcription start sites of housekeeping and developmental regulator genes.[9] Housekeeping genes encode many proteins that are essential for cell viability and they are therefore expressed in most cells. The genes are largely resistant to DNA methylation, and therefore, constitutively hypomethylated. To maintain the hypomethylated state, the proteins in charge of methylation, DNA methyltransferases (DNMTs), need to be actively excluded.

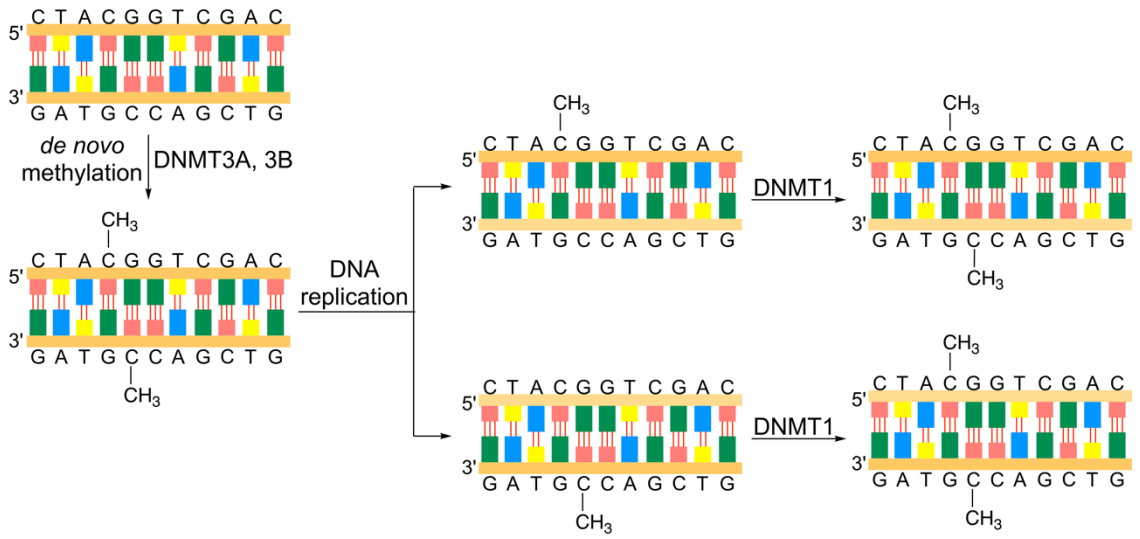
One of the nature's methylating agents is *S*-adenosyl methionine (SAM). SAM is formed in a S_N2 reaction between ATP and methionine, where sulfur in methionine acts as a soft nucleophile and attacks the primary 5' carbon bearing the triphosphate as a good leaving group as described in Scheme 9.[11]



Scheme 9. Formation of SAM from ATP and methionine.

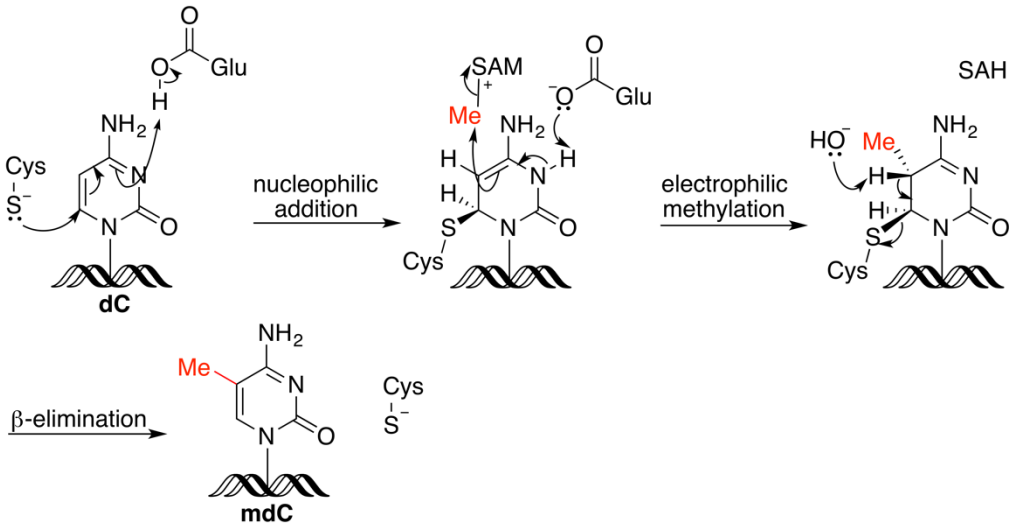
SAM is the primary methyl source for hundreds of transmethylases that methylate DNA, RNA, histones, proteins and small biological molecules. For the methylation of genomic dC, three SAM-dependent methyl transferases are needed: DNMT1[57,58], DNMT3A and DNMT3B[59,60]. DNMT3A and 3B initiate DNA methylation *de novo* in CpG dinucleotides.[61,62] DNMT1 plays an important role maintaining the methylation state in hemi methylated sites in daughter strands.[63–65]

Since mdC has the same relation to cytosine that thymine has to uracil, the modification of dC, likewise, has no effect on base-pairing. After the replication of DNA, DNMT1 acts on CpG sequences that are base paired with methylated CpG sequences.[66] During the methylation, the protein targets the cytosine within a DNA dublex. After this recognition, the mechanism is believed to be proceeded via base flipping, where the cytosine swings out of the DNA helix and rotates 180° in order to ensure a good binding to the enzyme. This guarantees that the same methylation state stays after the replication which is simplified in Scheme 10. Thus, it has been shown that DNMT3A and 3B have equal catalytic activities on all hemi modified DNA strands like mdC:dC, hmdC:dC and dC:dC sites which suggests a disability to distinguish between modifications in the hemi modified context.[67] Therefore, it has been suggested that DNMT3A and 3B complete the methylation on sites that have been passed over by DNMT1.[68,69]



Scheme 10. DNMT1 maintains the methylation state after DNA replication.

Chemistry behind the DNMT1 methylation of dC explains, how methylation happens and why it happens at C5 position. Several studies about the mechanism of methylation catalyzed by DNMT1 have been described.[70,71] It is initiated by an activated cysteine nucleophile of DNMT1 which attacks at C6-position of dC. The subsequent protonation of N3 from glutamate (Glu) activates C5 position for electrophilic attack (Scheme 11). The DNA structure is thought to influence on which face the nucleophilic attack occurs. Since the C6 position on dC base is inaccessible from the *Re* face, because of the neighboring bases, the unhindered *Si* face which is exposed in the major groove of DNA of the heterocycle is favorable for nucleophilic attack. The C6 position of dC does not have high partial positive charge but is the site of the largest coefficient in the LUMO. This makes the C6 position a soft electrophile and likely to form a bond with high energy HOMO of thiols. C5 then leads through methyl transfer from SAM (S_N2), which functions as an electrophilic methylation agent. When the nucleophilic addition and methyl transfer are expected to proceed in a stereospecific manner, the stereochemical configurations at C5 and C6 can be determined. The final mdC is obtained by elimination of cysteine; an electron pair from deprotonated C5 in dC leads to restoring the aromaticity of the ring by forming a new π-bond and cleaving the σ-bond to cysteine.[72,73] After the SAM has transferred the methyl group, S-adenosyl-homocysteine (SAH) is formed and methionine synthase refreshes the methylation agent back to SAM.



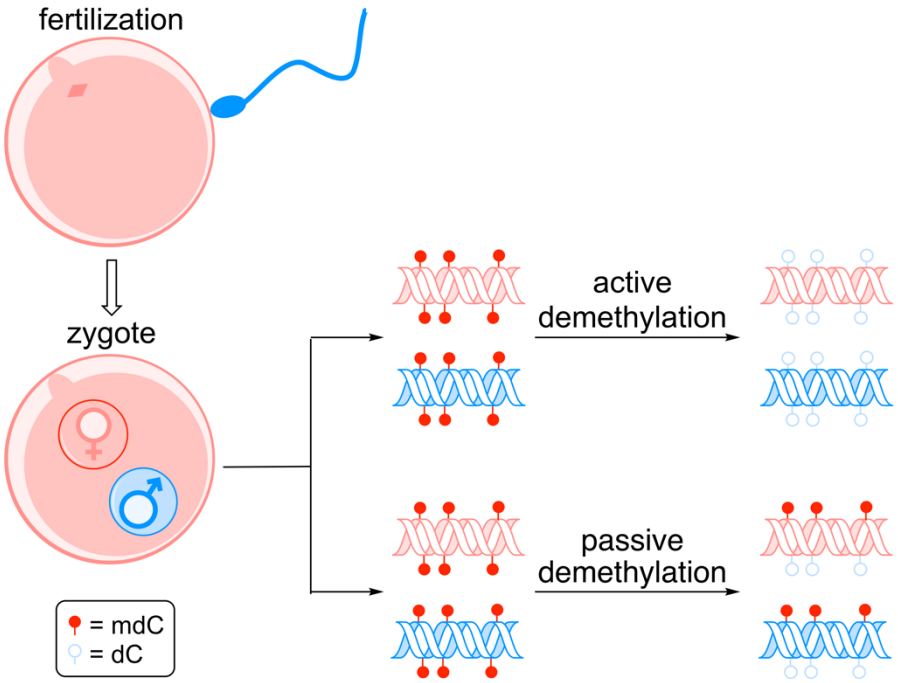
Scheme 11. Proposed mechanism for the methylation of dC by SAM.

A recent study proposes three major transition states (TS) to occur in the methylation process catalyzed by DNMT1.[71] In enzymatic reactions, formed TSs are held in Michaelis complexes.[74] A lifetime for such a TS is typically so short that no spectroscopic method is available to observe these structures. Therefore, experimental kinetic isotope effects (KIEs) have provided a chemical view into these catalytic mechanisms. TS geometries can also be reliably defined since KIEs are only influenced by covalent bond changes, not by binding interactions. This has established that enzymes act mostly on small molecules and has led to the design of enzyme inhibitors.[75,76] Aforementioned study shows that the methyl transfer (TS2) from SAM to dC in human DNMT1 methylation is the rate-limiting step, whereas studies for bacterial M.HhaI, propose that β-elimination (TS3) is the rate-limiting step of this process.[72,77]

4.2. DNA Demethylation

Enzymes catalyzing DNA methylation are already well characterized but DNA demethylation process is not understood in such details.[78,79] In order to reactivate silenced genes, removal of methyl group from mdC to restore canonical dC is required. For example, in mammalian sperm and oocytes, the epigenomes are reprogrammed to establish full developmental potential. A study showed that both maternal and paternal genomes in a zygote are passively and actively demethylated as depicted in Scheme 12.[80] Moreover, the active demethylation has been shown to occur TDG-independently, by TET3 dependent oxidation of

mdC to hmdC.[81–84] The mechanism behind active demethylation process is not completely clear and yet, three different pathways have been suggested and studied. These are introduced later in this chapter.



Scheme 12. Passive and active demethylation of zygote genome.

Passive demethylation is DNA replication dependent, which requires suppression of the process that maintain DNA methylation (DNMT-enzymes). Active demethylation in turn, is DNA replication independent, driven by enzymatic reactions, like TET enzymes that induce the oxidation of mdC[10,20,21] to hmdC[10,16], fdC[19,20] and cadC[20,21]. Studies have shown that in the genome of ESC, levels of epigenetic bases change during differentiation[10,20]. However, no clear function has yet been proved for fdC and cadC. Therefore, they are mainly considered to be intermediates in active demethylation.[36,85,86]

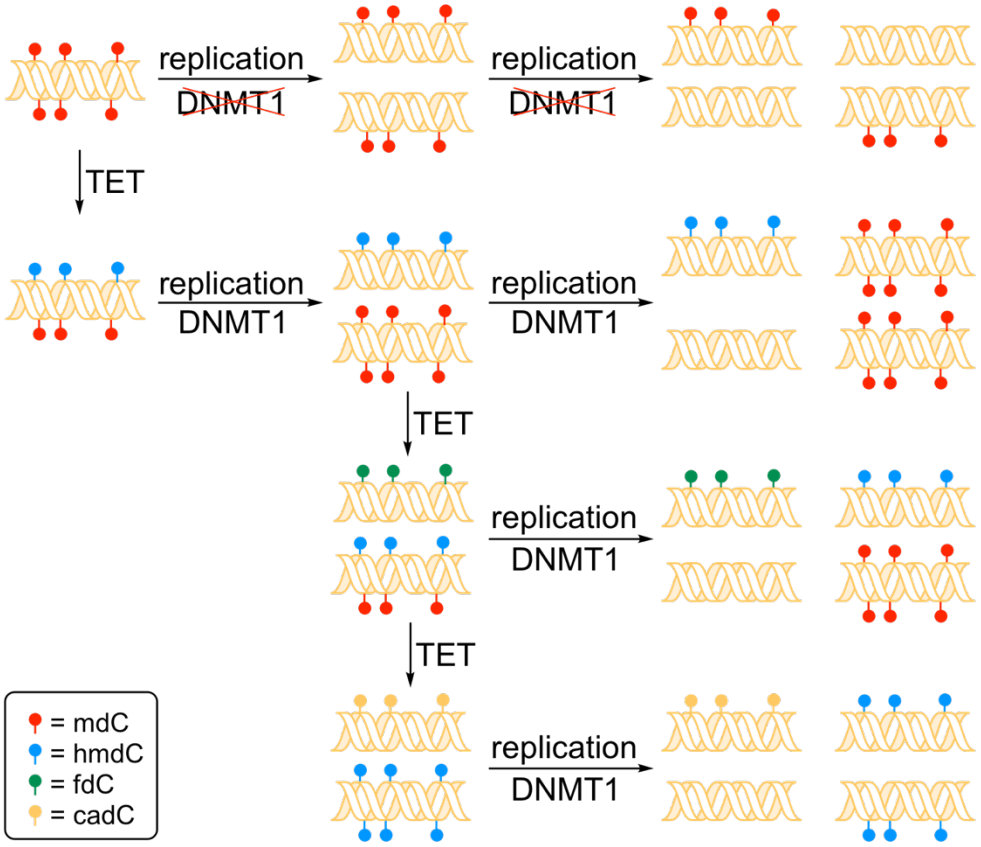
In TET proteins, the catalytic domain in the carboxylic terminus is constructed of a double-stranded β-helix domain and a cysteine-rich domain.[87] Double-stranded β-helix brings Fe(II), α-ketoglutarate and mdC close to each other. The cysteine-rich part stabilizes the whole structure as well as the TET-DNA interaction. The methyl group of mdC does not disturb the TET-DNA contact, which allows TET to adapt to the other non-canonical nucleobases.[88] Fe(II), a cofactor of the

enzyme, influences the activity of TET by generating the active site with a reactive Fe(IV)=O species. This oxidant is formed under subsequent decarboxylation of α-ketoglutarate to succinate.[89] Changes in the cellular iron concentration alter the levels of hmdC.[90] Vitamin C has been found to increase the enzymatic activity of TET, potentially as a cofactor.[91–93] Besides, it might support TET folding to make recycling of Fe(II) easier.[93] Vertebrate genomes have three TET family proteins – TET1-3, each being a different isoform. TET1 is highly expressed in the inner cell mass, blastocyst, in developing primordial germ cells and at lower levels in somatic tissues.[94–96] TET2 and TET3 are widely expressed in various stages and tissues, whereas TET3 is performing oxidation reactions in the zygote, just hours after fertilization.[97,98] Then TET3 levels drop and TET1 and TET2 level start increasing during the cleavage stages.[10,94]

4.2.1. Passive Demethylation

Absence of DNMT1 or conditions that impair DNMT1 mediated maintenance of CpG methylation are processes that happen during the passive demethylation. Under these conditions, over the course of multiple DNA replications, 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 are shown to have an inhibitory effect on maintenance methylation.[67,99–101] This suggests that genomic areas which require passive demethylation are marked with hmdC before replication starts. Through multiple rounds of DNA replication and less efficient DNMT1 activity towards hmdC:dG, fdC:dG and cadC:dG base pairs in the DNA regions, they become demethylated (Scheme 13).[67,102] Better knowledge of this process could provide a tool to target the passive demethylation to certain sites. To inreactivate silenced genes experimentally, demethylation has been achieved with chemically synthesized 5-azadC, which has been shown to inhibit the DNMT1 and so could reduce the global mdC levels.[103,104]



Scheme 13. DNMT1 inhibited pathway and epigenetically facilitated passive DNA demethylation.

4.2.2. Active Demethylation

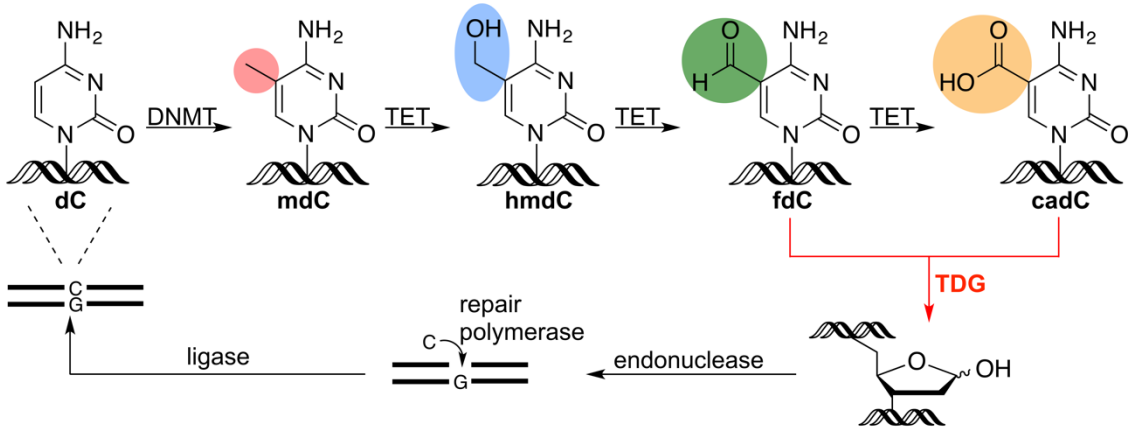
The discovery of genome wide loss of mdC in mouse zygotes was one of the first pieces of evidence of an active DNA demethylation in mammals. In this study, mdC levels on zygotic paternal genome decreased rapidly which could not be completely explained by the replication dependent dilution.[105,106] Almost ten years later it was found that human TET1 catalyzes the conversion of mdC to hmdC[10,16] but also TET2 and TET3 were shown to catalyze the oxidation of mdC to hmdC[107] and further to fdC and cadC[20,21].

The role of hmdC is linked to regulation of transcription and potentially marking the DNA of highly expressed genes.[108,109] It has been shown that in embryonic stem cells hmdC can localize at repressed and developmentally poised genes' as well as transcription starting sites[110–112] and it can be highly enriched at poised and active enhancers.[113–115] These studies suggest that hmdC is not only an intermediate product of an active demethylation but works as an epigenetic mark

and has biochemical relevance. Until today three main pathways for active demethylation have been described.

4.2.2.1. TDG-based Demethylation

The thymine-DNA glycosylase (TDG) based mechanism involves oxidation of mdC to fdC and further to cadC, cleavage of the glycosidic bond of fdC and cadC resulting in an AP site, and an activation of the base excision repair (BER) process. This leads to restoring a canonical dC (Scheme 14).[21,31] The oxidation step from hmdC to fdC and cadC is crucial for TDG in order to convert these bases into abasic sites.[21] *In vitro* studies have shown TDG to be specific for excising fdC and cadC but not hmdC.[21,31,116,117] An explanation for this is altered C-G base pairing, alteration in glycosidic bonds or varying TDG-fdC and TDG-cadC interactions.[116–120] This is defined as the active modification – active removal (AM – AR) and it is independent of DNA replication.[40] Since mdC occurs in high-density clusters, a BER-based erasure process is thought to create harmful clustered single strand breaks. Even double strand breaks can occur when both strands of a CpG site are modified with fdC or cadC. This type of a strand break caused by abasic site formation was observed in a study during the zygote development.[121] Whereas, another study showed when mixing cadC:cadC containing DNA strands *in vitro* with TDG and BER proteins, less than 1 % of double strand breaks was observed.[118] Mass spectrometry-based quantification of BER intermediates, using a specific reagent that reacts with abasic sites, revealed no increased levels of harmful BER intermediates.[122] It was found that TDG operates in a tight complex with other enzymes like APE1 and interplays with bifunctional glycosylases so that the formed AP site can be processed very quickly[123] and one strand at a time.[118]



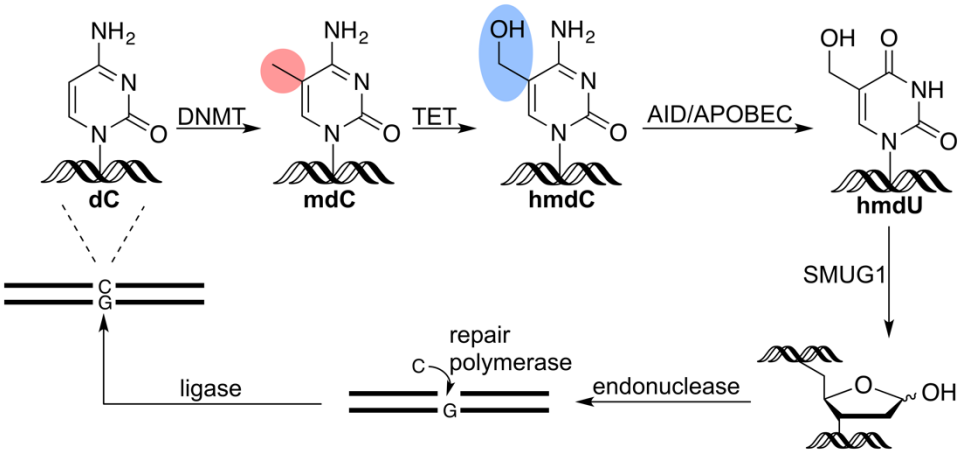
Scheme 14. TDG based demethylation.

4.2.2.2. Deamination-induced Demethylation

Instead of oxidation to fdC and cadC, deamination of mdC or hmdC followed by BER have been proposed to lead to active DNA demethylation. A family of cytosine deaminases (AID, APOBEC1,2 and 3) play a role in formation of variant antibodies in B cells, RNA editing and antiviral response.[124] It has been shown that both activation-induced deaminase (AID) and apolipoprotein B mRNA-editing enzyme complex (APOBEC1) to have a strong mdC deaminase activity *in vitro.*[125,126] This deamination would lead to T-G mismatch which then is repaired via BER (Scheme 15). It has been shown that AID and TDG require physical interaction in order to repair mismatched base pairing.[127] Interaction of TDG with AID suggests it to have another role besides enhancing proper epigenetic states: DNA demethylation in mammals through deamination and TDG mediated excision repair. Some studies have also suggested these enzymes to deaminate hmdC to hmdU *in vivo* which would lead to a hmdU:dG mismatch.[127,128] However, isotope tracing studies gave controversial results showing hmdU to be derived from oxidation of dT by TET enzymes.[129]

A study of pluripotent genes in human fibroblasts showed that AID dependent DNA demethylation is a necessary epigenetic change. When AID-dependent DNA demethylation was reduced by an AID knockdown, the initiation of nuclear reprogramming towards pluripotency was inhibited. In addition, the binding of AID was not observed in active promoters in contrast to the methylated ones, proving it to have a specific role in DNA demethylation.[130] So far, no successful deamination of hmdC have been accomplished *in vitro*[126,131,132] which leaves the role of AID in DNA demethylation unclear[133–135]. A controversial study suggests

AID to deaminate dC to dU, resulting in DNA repair in which adjacent mdC sites are replaced with canonical dC.[136,137]

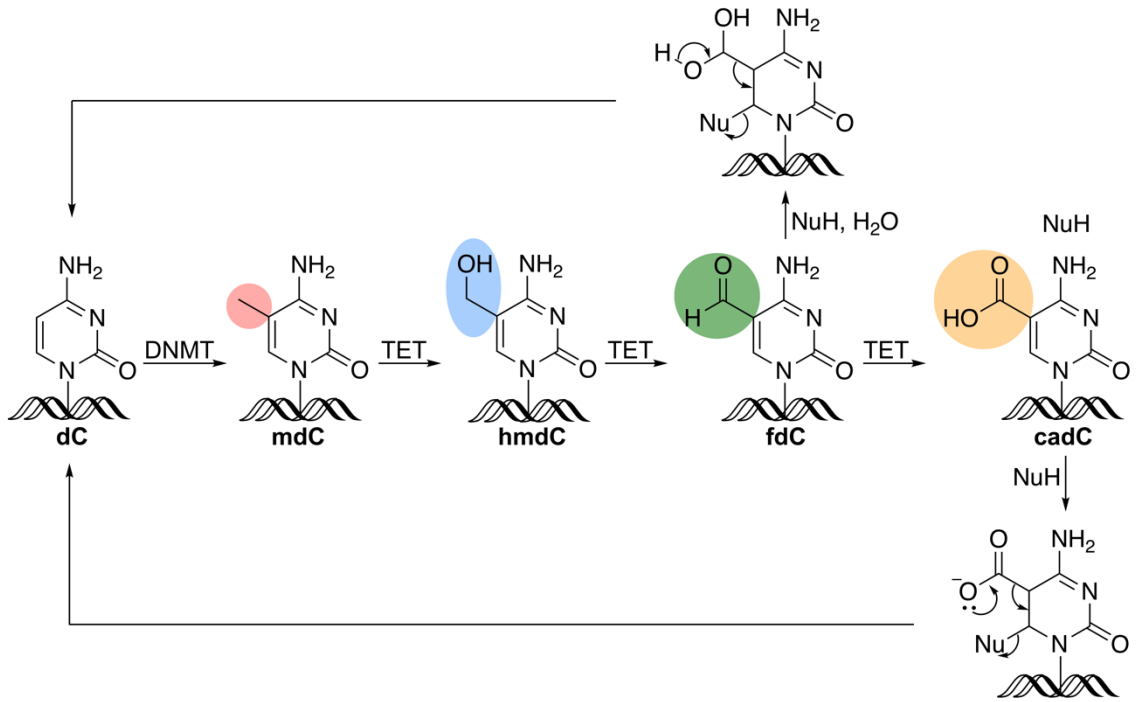


Scheme 15. Deamination induced demethylation pathway.

4.2.2.3. Active Demethylation via Direct C–C Bond Cleavage

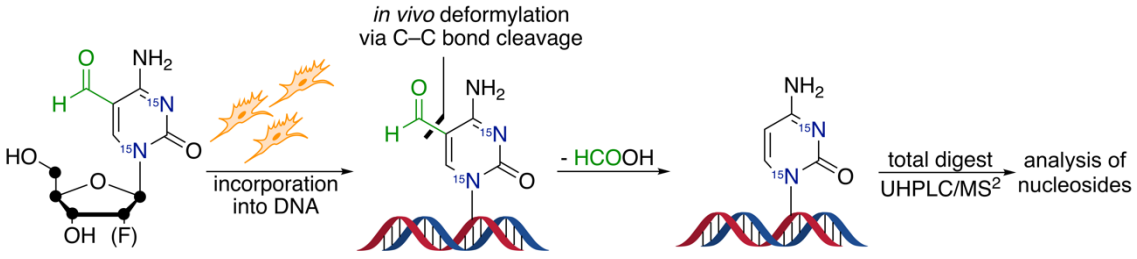
With epigenetic bases being present in neuronal tissue and in ESCs and the observations of demethylation independently from TDG in the zygote[80] it can be claimed that there are other processes for active demethylation. One possible pathway for active demethylation is a direct C–C bond cleavage via deformylation of fdC or decarboxylation of cadC resulting in dC.[25,78,138]

Decarboxylation was observed in the presence of nucleophiles in cadC containing DNA strands in stem cell extracts.[139] Later mechanistic studies proposed how such a direct C–C bond cleavage could occur.[41,140] Scheme 16 shows how deformylation and decarboxylation could occur after nucleophilic attack. C–C bond cleavage is initiated by Michael-addition type nucleophilic attack at C6 position of fdC or cadC. This leads to cleavage of the functional group followed by β-elimination of the nucleophile giving the canonical base dC.[41]



Scheme 16. Proposed pathway for active demethylation via direct C–C bond cleavage.

Iwan *et al.*[141] reported an isotope tracing study based on sensitive mass spectrometry investigating C–C bond cleavage with fdC derivatives. Isotope- and fluorine-labeled fdC derivatives were metabolically integrated in the genome of mammalian cells (Scheme 17). After isolation and enzymatical digestion of the genomic DNA, the modified nucleosides were analyzed by UHPLC-MS/MS. Analysis of the nucleoside mixtures revealed the presence of a defunctionalized isotope-labeled dC derivative. This proved that fdC is converted to dC in cultured mammalian cells while the glycosidic bond stays intact. Yet the mechanistic details of this process are unclear it is unclear.



Scheme 17. Deformylation of isotope–labelled fdC or 2'-fluoro-fdC.

Dehydroxymethylation, deformylation and decarboxylation are known and pervasive reactions in nature.[88,142,143] These reactions require enzymes such as α-ketoglutarate-, or flavin dependent oxidases. These enzymes are known to catalyze reactions where the oxidized group is attached to another heteroatom, typically a nitrogen atom (HO-CH_2-NHR). This generates an acetal structure which is fast and easily hydrolyzed through C–N bond cleavage. Enzyme catalyzed deformylation of peptides from N-terminus (OHC–NHR) occurs the same way through C-N bond cleavage.[144,142] In the case of hmdC, fdC and cadC, the functional groups are attached to a C-atom in an aromatic heterocycle and are different in terms of bond stability. The enzyme isoorotate decarboxylase is known to decarboxylate isoorotate to uracil.[143] It has also been shown that this enzyme is able to decarboxylate caC to C with a weak activity. This was the first *in vitro* evidence of caC decarboxylation catalyzed by an enzyme.[145] The closest isoorotate decarboxylase relatives found in human cells, are amino-carboxymuconate semialdehyde decarboxylase and cytosine deaminases.[24] Common deformylation reactions in nature are fatty acid degradation[146,147] and lanosterol double bond formation as an accomplishment of deformylation[148,149]. Both reactions are utilized by a nucleophilic Fe peroxyl radical which attacks the substrate. One of the Fe peroxyl radicals acts as a nucleophile and the other one as a Lewis acid activating the carbonyl group.

4.3. Diseases that are Related to Epigenetic Changes

Irregularities in the epigenome, which consists of all the epigenetic modifications, can cause severe changes in the biological integrity of a cell. mdC has been found to be involved in X-chromosome inactivation, transposon silencing and genomic imprinting.[56] Therefore, it is important to learn to understand these mechanisms in order to develop suitable treatments.

4.3.1. Cancer

Epigenetics may be the primary initiators of cancer development, like for breast cancer.[150] Epigenetic changes in the DNA methylation at certain CpG sites and in histone modifications could influence cancer progenitor cell formation, progression, and formation of metastatic cancer.[150] Inactivation of tumor suppressor genes is caused by hypermethylation of these promoter regions.

Also, a stepwise progression of lung cancer has been under investigation[151], however, the pathway of cancer progression from stem/progenitor cells to metastatic stages is poorly understood.

CpG islands are often located in promoter sites of tumor suppressor genes. Methylation of these sites silences the gene and leads to a decrease of cell-cycle inhibitors and deactivation of pro-apoptotic genes.[152] Cancer cells express DNMT1 in high levels compared to normal cells. This high regulation causes the methylation of upstream regions. In normal cells the levels of DNMT1 are cell-cycle dependent, whereas in cancer cells the regulation is always present which in turn maintains a higher methylation level. The observation of methylation levels and the nature of reversible epigenetic changes, creates a good target for epigenetic therapy.[150,153,154]

4.3.2. Cardiovascular Diseases

Histone and CpG residue modifications are in charge of many important cardiovascular functions[155,156] however, the mechanism of these processes are not completely understood. In atherosclerosis, the atheroprotective estrogen receptor genes (ESR1 and ESR2) in vascular smooth muscle cells are hypermethylated, whereas in normal cells they are expressed.[157] A high risk for coronary heart disease is linked to a methylation of cytosine in the insulin-like growth factor 2 (IGF2) gene which causes dysregulation of imprinting.[158] On the contrary hypomethylation, a loss of genomic methylation, is also linked to cardiovascular diseases like hypertension.[159] Reduced DNA methylation in Long Interspersed Nucleotide Element 1 (LINE-1) in blood cells is associated with ischemic heart and stroke.[160]

4.3.3. Metabolic Disorders

Twin-studies have shown that environmental factors can create divergent of the epigenomes. Despite sharing an identical DNA, the other twin can become more exposed to a specific disease.[161] Several studies of obesity showed correlation between the methylation of three specific genes, the vasodilator factor (affects the widening of blood vessels), and the amount of fatty tissue (adipose tissue) at birth and at the age of 9 years.[162] Leptin, an adipose derived hormone, regulates hunger and metabolism and it is epigenetically controlled. In preadiposcytes, CpG sites are heavily methylated but when maturing into an adipocyte cell

demethylation becomes predominant. Experiments on the mouse model show that the leptin expressing gene is more likely methylated in mice with diet-induced obesity compared to normal mice.[163]

Even temporary environmental changes, like a change in a diet, can lead to epigenetic irregularities. A test on the rat models showed epigenetic silencing in liver cells after exposing the test animals to a diet deficient in folic acid, L-methionine, and choline. After returning to a normal diet deregulation of hepatic DNMT1 and methyl-CpG-binding proteins also normalized. Although the changes were reversible, prolonged changes in the expression of DNMT1 and methyl-CpG-binding protein led to the development of hepatic carcinoma.[164] The understanding of the relevance of epigenetics in metabolic disorders, could allow the develop of tools to prevent and treat metabolically linked disorders.

4.3.4. Neurological Disorders

During the development of the nervous system, epigenetics retains the multipotent state by methylating CpG sites in neuronal precursor cells and during the neuronal differentiation CpG methylation is lost and demethylation gained.[165,166] Rett syndrome (RTT) is a X-linked progressive neurodevelopmental disorder, among the most common causes of serious mental retardation in females. It is known to be caused by mutation in the methyl-CpG-binding protein 2 (MeCP2) gene. When MeCP2 binds selectively in mCpG dinucleotides in the genome, it interacts with other proteins causing the condensation of the chromosome and silencing the gene. Since epigenetic gene regulation is reversible, a group of scientists discovered in experiments with mice that neurological defects could be partially or completely reversed and dysregulation of MeCP2 is not necessarily permanent.[167]

5. Synthesis of C5 Functionalized dC and dU Derivatives

Chemical transformations at the C5 position of deoxycytidine can lead to the formation of methylated and further oxidized epigenetic dC modifications; mdC, hmdC, fdC and cadC and the respective removal of them. It is known that human TDG enzymes can excise fdC and cadC[31,116], whereas other studies report these bases to also undergo potential direct deformylation or decarboxylation back to canonical dC.[41,139,140] When introducing substituents at the C5 position of the

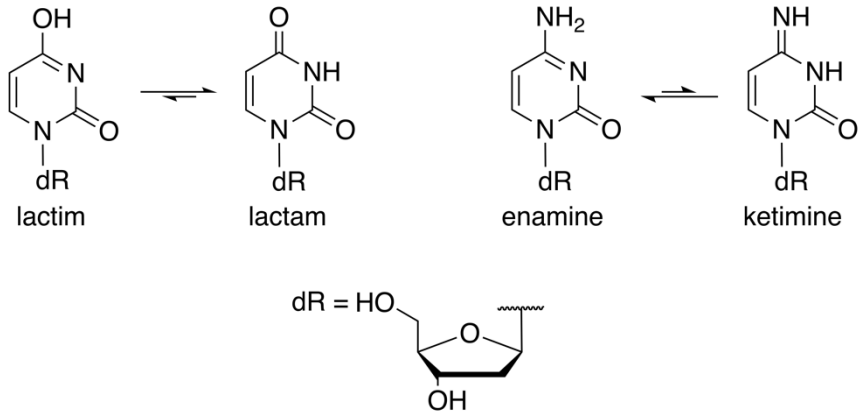
nucleobase, N1 substituent has a profound impact on its reactivity. Therefore, reactions at C5 position of a nucleobase are not always suitable for respective nucleosides or nucleotides.[168]

In this chapter, reactivity differences between of dU and dC are discussed and derivatization reactions of 2'-deoxyuridine, 2'-deoxycytidine and respective 2'-fluorinated derivatives are introduced. It is, however, advantageous to use dU as the preferred starting material since electrophilic substitutions at the C5 position of dU are easier to perform compared to dC. Additionally, several amination reactions of uridine derivatives to the corresponding cytidines are known.[169–172]

5.1. Stability of the Pyrimidine Nucleosides

The functional group at C4 in nucleobases influences the stability of the respective glycosidic bonds, for example substitution of a hydroxyl group for the amino group increases the glycosidic bond strength. When applying harsh conditions (5 % trichloroacetic acid, 100 °C, 30 min) the glycosidic bond of dC is hydrolyzed whereas dU remains mostly unreacted.[168] This can be explained with their respective tautomeric forms which affect the 3D structure and reactivity of the nucleosides, whereby hydroxyl and amino derivatives exhibit two different types of tautomerism. Lactim-lactam for hydroxyl and enamine-ketimine tautomereism for amino derivatives. The lactam form is more favorable for uridine and thymidine derivatives and the enamine form is more favorable for cytidine derivatives (Scheme 18).[173]

In the lactam form, lone pairs of the sp^2 hybridized p-orbitals of amides participate in the aromatic system, whereas the C4 carbonyl carbon does not affect the conjugation due to direct and cross conjugation of the bonds. Thus based on relative kinetic acidities, protonation at O4 forms more stable cation than the protonation at O2.[174] Furthermore, it is generally accepted that C5 and C6 substitutions at dU do not shift the tautomeric equilibrium significantly.



Scheme 18. Tautomeric forms of dU and dU.

Considering 2'-deoxythymidine the methyl group at C5 does not have a significant impact on its hydrolysis rate – dT is preferred an experimental point of view as stable as dU. Introduction of an electron withdrawing group (EWG) at C5 decreases the glycosidc bond strength i.e. 5-bromo dU derivative is hydrolyzed easier than dU. The stability of glycosidic bond of cytidine and deoxycytidine is decreased even further when the amino group at N4 is acylated. In acidic media 4-N-acyl derivatives are hydrolyzed much easier than respective unmodified cytidine nucleosides.[168] In case of DNA, electronic excited states of one of the nucleosides can undergo charge transfer process along the DNA chain, which can cause base-sugar or sugar-phosphate bond cleavages.[39]

However, the stability of the glycosidic bond is heavily dependent on the nature of substituents at 2' and 3' positions in the sugar moiety. Ribonucleosides are 100-1000 times more stable towards hydrolysis than the corresponding deoxynucleosides.[168] The glycosidic bond becomes even more stable towards hydrolysis when 2'-OH is substituted by a halogen or by an electron withdrawing group like toluenesulfo-, 2,4-dinitrobenzoyl or acetyl. Glycosidic bonds of canonical nucleosides are in general stable in neutral and basic solutions but tend to hydrolyze in the presence of mineral and organic acids. During glycosidic bond cleavage, a positively charged carbamate intermediate is formed but not preferred when the sugar unit is partially positively charged. Yet it can be stabilized by introducing an EWG protecting group at 2'-C and 5'-C positions. An observed relationship between the rate of acidic hydrolysis of the glycosidic bond in nucleosides and the pH value of the reaction mixture suggests that the cleavage of this bond is initiated by protonation of the heterocyclic base, which is in consequence as the rate determining step of this reaction.[168]

5.1.1. Distribution of the Electron Density

Understanding the electron donor and acceptor properties and in further detail the electron affinities and ionization potentials of the nucleobases is helpful for elucidation and evaluation of intermolecular forces between the bases themselves or with exogenous substances. In general, pyrimidine nucleobases are aromatic compounds and considered to be planar. Ionization of these molecules primarily occurs due to π-electrons even though lone pair electrons are present.[175] Total π-electron density distributions calculated with CNDO/2 method for cytosine, uracil and thymine are shown in Figure 6. When comparing the nucleobases, the most π-negative oxygen is found in cytosine at C2 position. Nevertheless, the C5–C6 bond is in general polarized negatively at C5 and positively at C6 for all shown pyrimidines. However, the positive charges are located at C2, C4 and on the amino group in the case of cytosine. When hydrogen atoms at C5 or C6 position of cytosine are substituted with fluorine, a change in partial electronic charges mainly occurs at these positions and in their direct surroundings, whereas a substitution of the amine or hydroxyl group at C4 does not change the charge distribution significantly.[175,176]

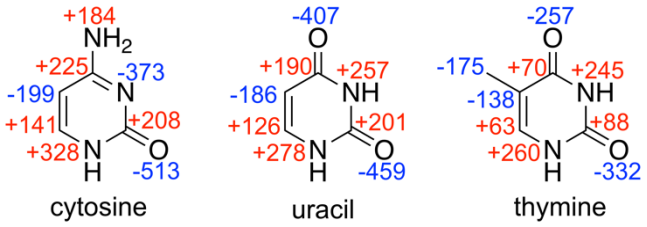


Figure 5. Total π-charges in 10^{-3} e for cytosine, uracil and thymine calculated by CNDO/2 method.

To study proton affinities and possible sites for electrophilic attacks, Bonaccorsi *et al.*[173] have proposed an index for evaluating molecular potentials interacting with an approaching reagent. The electrostatic potential V of the molecule (r_i) is formed by the nuclear charges and the electronic density distribution of the molecule. The first-order perturbation energy in a field of a point charge *q,* can be calculated from the interaction energy between the molecular potential distribution and the external point charge depending on their distance. Based on

this, isopotential maps have been constructed for studying theoretical proton affinities of the different atoms in relevant nucleobases.[176]

Thymine, has two carbonyl oxygens and the potential-energy curves indicate two spots for a possible protonation while the rest of the molecule is repulsive. The oxygen at C3 position, surrounded by two NH-groups forms nearly symmetrical potential well. Also, two minima are found in proximity to the carbonyl oxygen at C4. The minima in cytosine are observed at much lower energies than the corresponding wells in thymine which agrees with the fact that thymine and thymidine is much less basic than cytosine. Supporting these findings, it is known that protonation or alkylation of cytosine primarily occurs at N3, pKa values for cytosine, thymine and uracil are 4.4, 9.9 and 9.5 respectively.[177] Which concludes that cytosine is more basic than uracil and in consequence the same applies to the respective nucleosides[173] and explains why all the derivatives do not undergo alkylation or other electrophilic attacks when applying the same conditions as for dC derivatives.

By capturing a low-energy electron in its LUMO this excitation of a nucleobase can eventually cause damages leading to single- and double strand breaks on the DNA level. This character is also relevant to charge transfer in DNA. In these nonthermal low-energy electron-induced processes, resonance states are early key intermediates possibly resulting in the damage of DNA strands.[178] The location of π^* orbitals can be utilized in rationalizing where low-energy electron excitation preferentially could occur. Considering the modified epigenetic bases, the resonance modes are different due their substituents at C5 position and consequently inductive and mesomeric effects need to be considered when investigating the electronic excitation energies. Due to their inductive effect, electronegative substituents have the tendency to withdraw the electron density from the core π-system towards themselves, whereas electropositive groups have an electron donating tendency. Mesomeric effects occur only at conjugated π-systems and are based on overlap of a substituent's π-orbitals with the π-orbitals of the rest of the molecule. Since cytosine has only one electron withdrawing carbonyl at C2 position, resonance modes π_1^* and π_3^* need a lighter excitation energy compared to uracil, providing two carbonyl oxygen groups. (Figure 6).[178] However resonance energies are only slightly influenced comparing cytosine and dC and therefore, the resonance energies of nucleobases are

comparable to their corresponding 2'-deoxynucleosides by a applying suitable upward shift.[179]

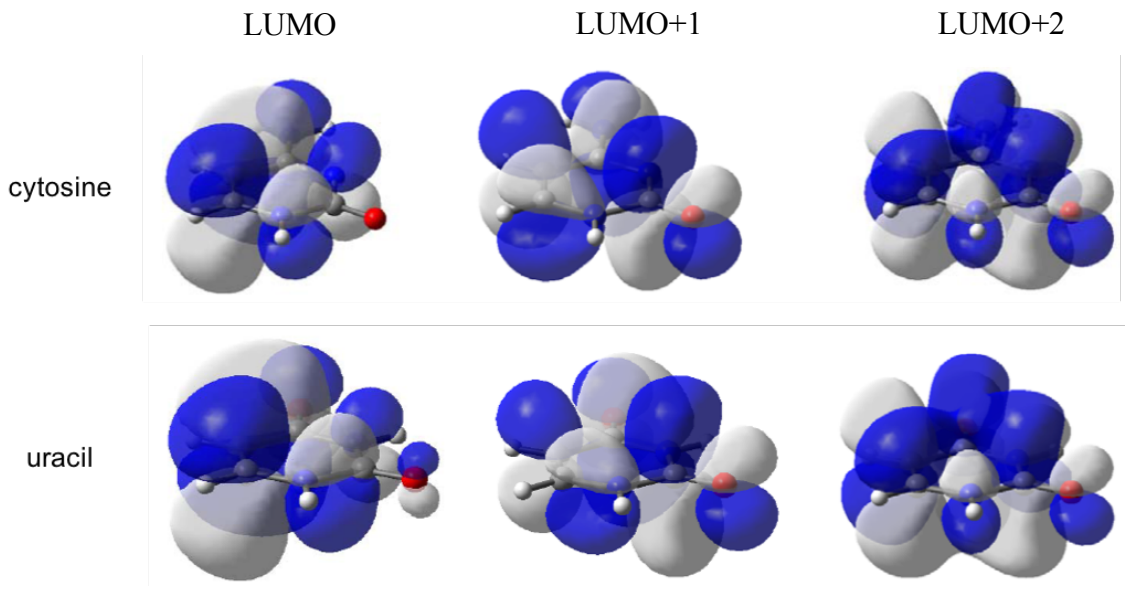


Figure 6. Energetically lowest unoccupied MOs of cytosine and uracil.[178]

When comparing the resonance modes of epigenetic bases based on CAP/SAC-CI calculations, the first difference rises in the amount of calculated resonances; cytosine, mdC and hmdC have three resonance modes whereas fdC and cadC have four. The fourth resonance mode originates from the double bond located at the C5 carbonyl/carboxyl substituent and this resonance is located at a comparable lower energy, thus allowing a lower energy electronic excitation. The π-bond of these substituent generates a mesomeric effect that attracts electron charge density and competes with the inductive effect of the present negative charge density in cytosine. The mesomeric effect can be compared to hmdC; the negative charge density over fdC is moved to the direction of the substituent whereas in hmdC the electron density remains unchanged.[180]

CNDO/2 calculations of HOMO and LUMO energies for cytosine, uracil and some other nucleobases are listed in Table 2. It shows that the HOMO of cytosine is energetically lower and its LUMO is energetically higher than the respective energies of uracil. The HOMO-LUMO gap – the energy difference between the HOMO and LUMO – is a measure for the wavelengths a molecule can absorb. Furthermore, the larger the gap, the higher kinetic stability and the lower the

chemical reactivity.[181] These observations in agreement with total π-charges indicate that cytosine should be the best π-electron donor and uracil the worst.

Table 2. HOMO and LUMO energies for C, 5-f-C, U and T calculated by CNDO/2 method.

compound	LUMO (eV)	HOMO (eV)		
cytosine	2.85	-10.79	-11.82	-13.30
5-F-cytosine	2.30	-10.98	-12.19	-13.33
uridine	2.25	-11.81	-12.86	-13.05
thymine	2.29	-11.37	-12.85	-13.10

5.2. The Purpose of 2'-fluorinated dC Derivatives

Fluorine is commonly used substituent to replace hydrogen atoms in pharmaceutically active molecules. Introducing a fluorine substituent at 2'-position of nucleosides, increases the lifetime of these nucleosides in the bloodstream. From the biochemical point of view 2'-(*R*)-configured nucleosides are well accepted by cells and fluorine often increases the lipophilicity of the molecule and therefore increases their affinities for biological targets.[182] Fluorine affects slightly the compound's reactivity towards biochemical conversions, but earlier studies confirm the effect to be small. (*R*)-2'-F-dC can be efficiently methylated by DNA methyltransferases[183] and resulting (*R*)-2'-F-mdC can be oxidized to (*R*)-2'-F-hmdC by TET enzymes with slightly reduced oxidation speed[184]. Additionally, fluorine atoms are bioisosteric to H-atoms[185] but in (*R*)-configuration of 2'-fluorinated nucleosides C3'-*endo* conformation of the sugar pucker is stabilized, shaping up the molecule RNA-like[186–188] (Figure 7). However, this conformation does not hinder base incorporation by DNA polymerases.[189] C3'-endo pucker conformation mainly leads to RNA and A-form DNA, whereas C2' endo conformation results in formation of B-form DNA. The C3'-endo structure leads to a shorter phosphate-phosphate distance thereby making the helical structure more compact which is characteristic to RNA.

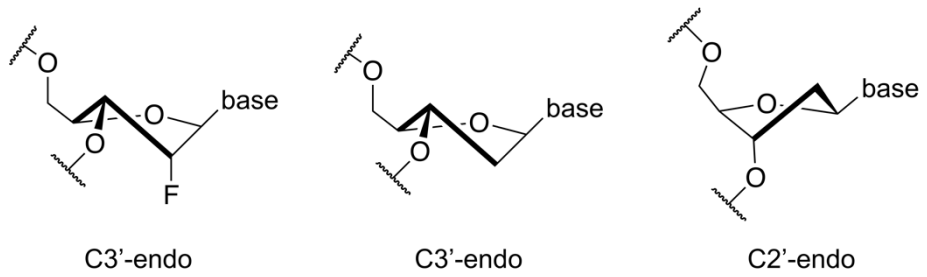


Figure 7. Sugar puckers found in natural and 2'-fluoro substituted nucleic acids.

It has been known for long that a 2'-fluorine substituent in the 2'deoxyribose ring of nucleosides efficiently inhibits the glycosidic bond cleavage by glycosylases.[183,188,190,191] From a mechanistic point of view glycosidic bond cleavage is assumed to form an intermediate oxocarbenium ion, in which substantial positive charge is accumulated, notably at O-4' and C-1' of the 2'-deoxyribose unit.[188,192] The presence of a strong electron-withdrawing fluorine at C2', is considered to destabilize this intermediate by increasing its charge density, thereby inhibiting the deglycosylation reaction.[191,193,194]

In order to detect biochemical conversions with high sensitivity, synthesized nucleosides need to be distinguished from natural ones. Substituting hydrogen with fluorine, its bioisosteric property still mimics the size of a hydrogen but makes the compound 18 atom units heavier and shifts the retention times in chromatographic experiment.[141] Also, F-atom makes the glycosidic bond more labile in the MS-fragmentation step meaning sharper signals in the UHPLC-MS/MS analysis.[141]

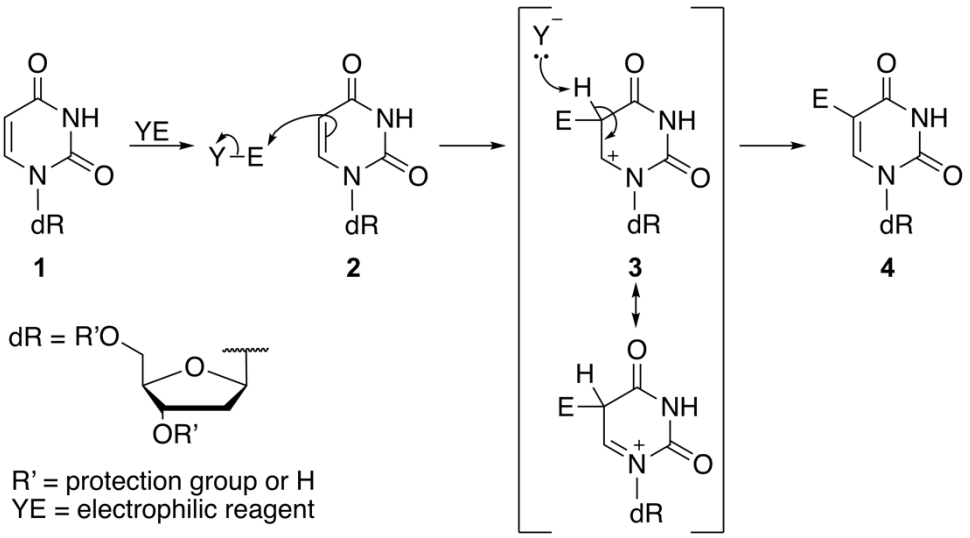
Based on these properties, stabilized 2'-fluorinated analogues to mdC, hmdC, fdC and cadC that are stable against human thymine DNA glycosylase (hTDG)[183] have been developed and are therefore considered as suitable tools for studying direct demethylation pathway.

5.3. Reaction Mechanisms to Functionalize C5 Position of dU and dC Derivatives

5.3.1. Electrophilic Aromatic Type of Substitution

Pyrimidine nucleosides are aromatic compounds, substitution at C5 position can proceed in the similar manner to that for electrophilic aromatic substitution. After the electrophilic attack, formed intermediate **3** is stabilized by electron donation from the adjacent lone pair from the p-orbital of N1 in the pyrimidine ring. Deprotonation of **3** from C5 restores the aromaticity and 5-substituted pyrimidine derivative **4** is obtained (Scheme 19).[168]

The reactivity of cytidine derivatives differs in these types of reactions. Based on calculated total π-charges the most probable site for electrophilic attack on cytosine is at N3 position[173]as discussed earlier. Therefore, under the same reaction conditions, N3 of dC the tends to get protonated/substituted reaction does not proceed in the same manner as with dU.

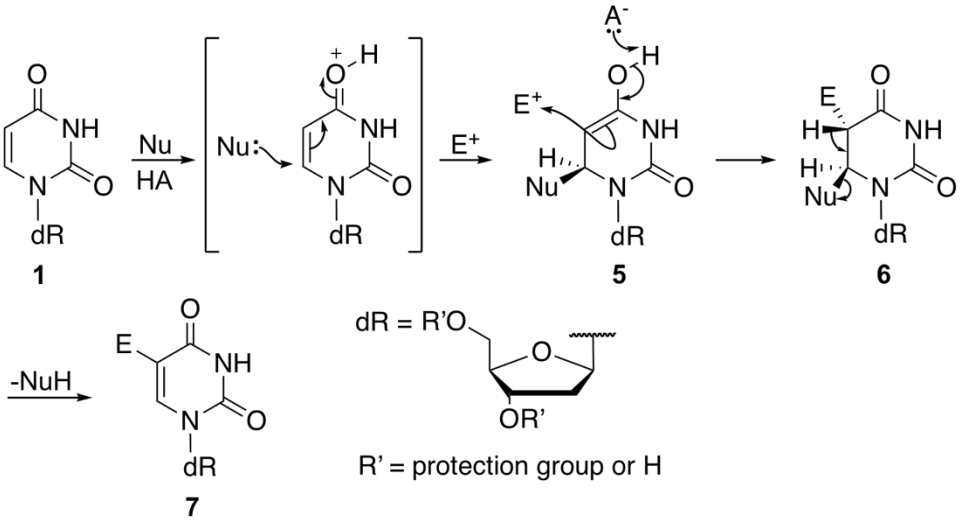


Scheme 19. General mechanism for electrophilic aromatic type substitution of a uridine derivative.

5.3.2. Michael Type of Addition

The pyrimidine nucleosides can be functionalized at the C5 position of the heterocycle via Michael type addition of a nucleophile followed by an addition of

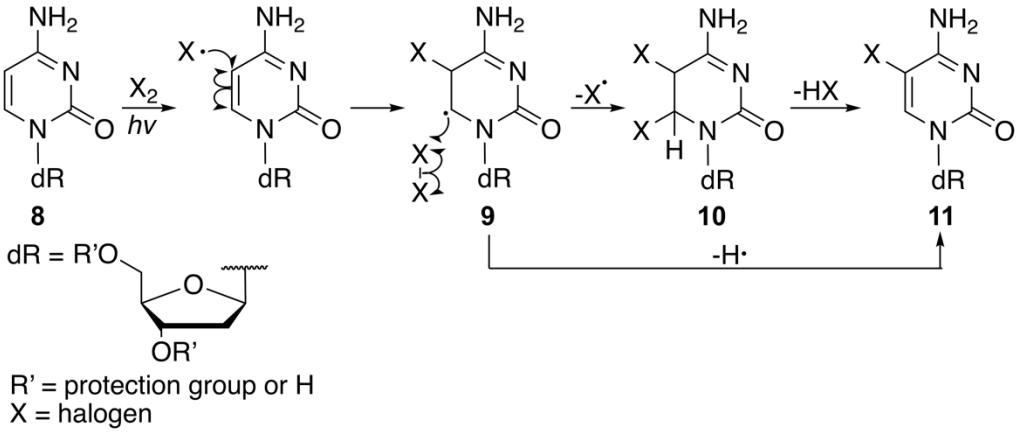
an electrophile. The reaction mechanism can be compared with the reaction mechanism of a Michael addition. After the addition of a nucleophile at C6 position, the enol-type intermediate **5** attacks the electrophile resulting in the C5/C6-functionalized nucleoside **6**. This undergoes a β-elimination leading to C5-functionalized nucleoside **7** (Scheme 20).[168] For dC, the mechanism of Michael type addition was introduced in chapter 4.1.



Scheme 20. General mechanism of Michael type addition and β-elimination.

5.3.3. Addition of Radicals

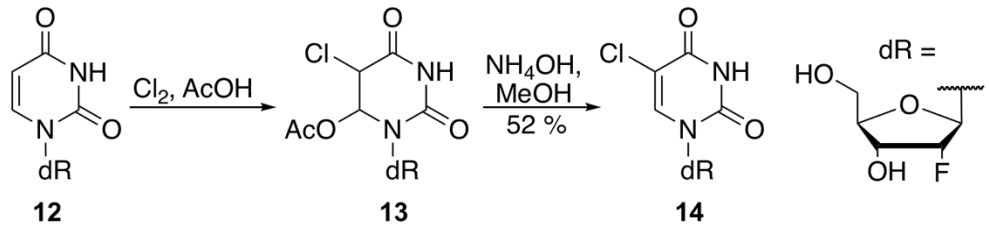
The photochemical dimerization, hydration and addition of thiols to dC and dU are well known reactions.[195,196] In Scheme 21 the mechanism is depicted for dC derivative. Addition of photochemically induced free radicals to pyrimidine nucleoside **8** gives rise to a new radical intermediate **9**. 5-substituted cytidine **11** is obtained after the loss of a hydrogen from **9**. A dihydrocytidine derivative **10** is formed when the radical **9** homolytically cleaves X–X bond and generates a new σ-bond. β-elimination needs to occur in order to regenerate the aromaticity and obtain **11**.[168,195]



Scheme 21. Radical mediated addition of dC.

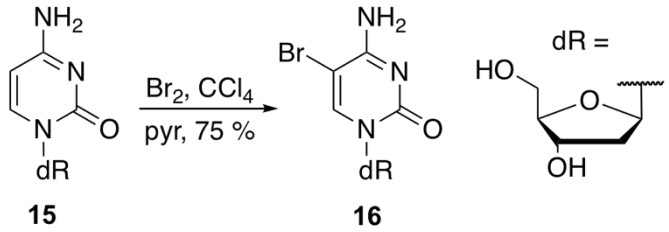
5.4. Halogenation of Pyrimidines

Electrophilic aromatic type of substitution to C5 of dU and dC derivatives have been applied using molecular halogens or N-haloamides. Chlorination of 2'-F-dU (**12**) is achieved with Cl_2 in acetic acid (AcOH). Only after 10 min, 5-chloro,6-OAc-substituted intermediate **13** is formed, which under basic conditions gives 52 % of the 5-Cl substituted nucleoside **14** (Scheme 22).[197] This method can also be applied for bromination reaction.[198,199]



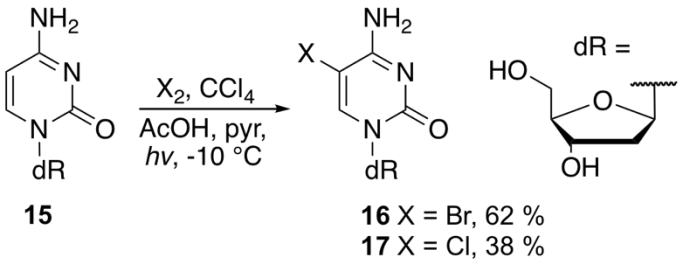
Scheme 22. Chlorination of unprotected 2'-F-dU.

The AcOH mediated halogenation is suitable for dU derivatives but dC derivatvies cannot be functionalized under these conditions. Acidic conditions protonate the N3 of the cytidine[173] and the reactivity of C5 towards electrophiles is changed. Bromination of dC **15** can be achieved with Br_2, CCl_4 in solvents like pyridine, DMF or formamide. **16** is obtained 75 % yields after 3 h (Scheme 23). Advantage of this method is that the reaction can be carried out without protecting the nucleoside.[200–202]



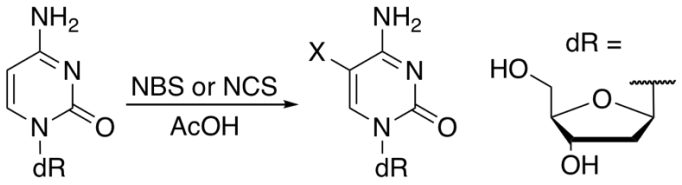
Scheme 23. Bromination of dC.

As shown before, halogenation of dC derivatives under acidic conditions is not possible, an alternative method was developed utilizing UV-radiation. The cytidine ring becomes excited and is able to undergo the halogenation reaction at the C5 position. This radiation-based method has also been applied for brominations in the same manner (Scheme 24). However, instead of an electrophilic substitution, the mechanism presumably occurs via homolytic substitution and it yields smaller.[203]



Scheme 24. Halogenation of dC using UV irradiation method.

N-halosuccinimides are good halogenating agents for unprotected dU and dC derivatives under mild conditions: *N*-chlorosuccinimide (NCS) in AcOH chlorinates dC with 65 % yields at elevated temperatures in 20 min.[203–205] The corresponding bromo-derivatives can be achieved with N-bromosuccinimide (NBS) under milder conditions (Scheme 25).

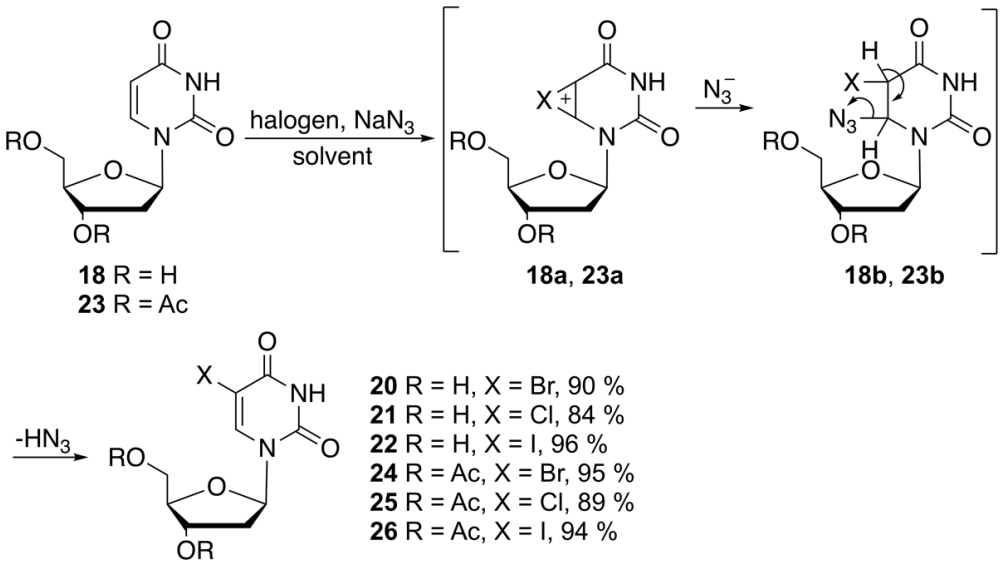


15

16 X = Br, 90 %
17 X = Cl, 95 °C, 65 %

Scheme 25. N-halosuccimide mediated halogenation of dC.

Another efficient halogenation method of pyrmidine nucleosides uses a combination of a halogen source and NaN_3. The treatment of the nucleoside **15** with the halogen source leads to formation of a halogenium ion **18a**/**23a**. The attack of the nucleophilic N_3^- opens the rigid three-membered ring and forms the 5-halogenated-6-azido-6-5,6-dihydro-intermediate (**18b**/**23b**). It is followed by a subsequent elimination of HN_3 and formation of halogenated dU **20**–**26** (Scheme 26).[206] This method is suitable for both protected and unprotected 2'-deoxyuridine derivatives. It proceeds in the similar manner as the earlier discussed halogenation of 2'-F-dU utilizing AcOH (Scheme 22). This electrophilic halogenation via the acetoxy intermediate **13** requires elevated temperatures or acid-base catalysis for restoring the aromaticity of the nucleoside. This suggests that the stability of the intermediate is dependent of the C6 substituent. Regarding to Table 3, the absence of NaN_3 gave lower yields of **22, 20** and the reaction between **18** and NCS did not occur at all. This suggests azido group to play an important role at C6 position in these C5 halogenation reactions. Therefore, utilizing NaN_3 is a good alternative halogenation method as N_3 is a better leaving group than OH, OMe or OAc. Additionally, restoring the C5–C6 double bond from intermediates **18b**/**23b**, does not require harsh conditions.[206]



Scheme 26. Halogenation of dU via the 5-halogenated-6-azido-5-6-dihydro intermediate. Reaction parameters are presented in table 3.

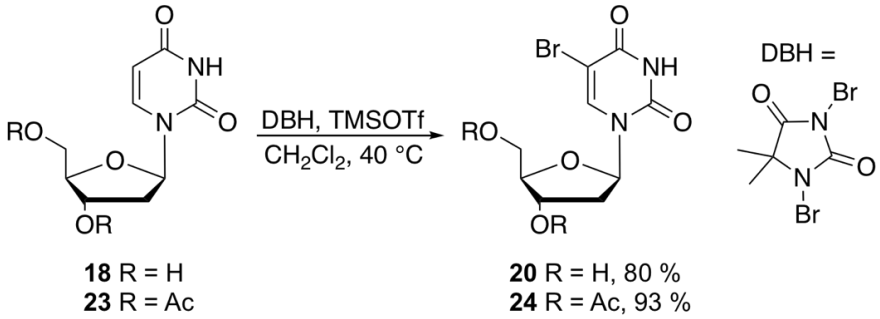
Table 3. Reaction conditions for halogenation of dU via 5-halogenated-6-azido-5-6-dihydro intermediate.

starting material	halogenation reagent	NaN_3 (eq.)	solvent	temp. (°C)	time (h)	product	yield (%)
18	ICl	4.0	CH_3CN	25	24	**22**	96
18	ICl	0	CH_3CN	25	24	**22**	20
23	ICl	4.0	CH_3CN	25	48	**26**	94
23	ICl	3.0	CH_3CN	0	12	**26**	0
18	NBS	4.0	DME	24	24	**20**	90
18	NBS	0	DME	25	20	**20**	60
23	NBS	4.0	DME	25	24	**24**	95
18	NCS	8.0	DME	45	8.0	**21**	84

starting material	halogenation reagent	NaN_3 (eq.)	solvent	temp. (°C)	time (h)	product	yield (%)
18	NCS	8.0	DME	25	48	**21**	35
18	NCS	0	DME	45	48	**21**	0
23	NCS	8.0	DME	45	3.0	**24**	89

DME = 1,2-dimethoxyethane

1,3-dibromo-5,5-dimethylhydantoin (DBH) has been reported to brominate uridine derivative **23** efficiently in aprotic solvents (Scheme 27).[207] Addition of a Lewis acid such as trimethylsilyl trifluoromethanesulfonate (TMSOTf) or *p*-toluenesulfonic acid (TsOH) improves the yield from 72 % to 93 % and shortens the reaction time to 30 min. Bromination of unprotected uridine **18** can be achieved with this method yielding 80 % of **20**. This method is also suitable for brominating cytidine ribonucleosides and therefore could be a potential brominating method for 2'-F-deoxycytidines. Also, the advantages of this method are an easy work up and no needed column chromatography purification.

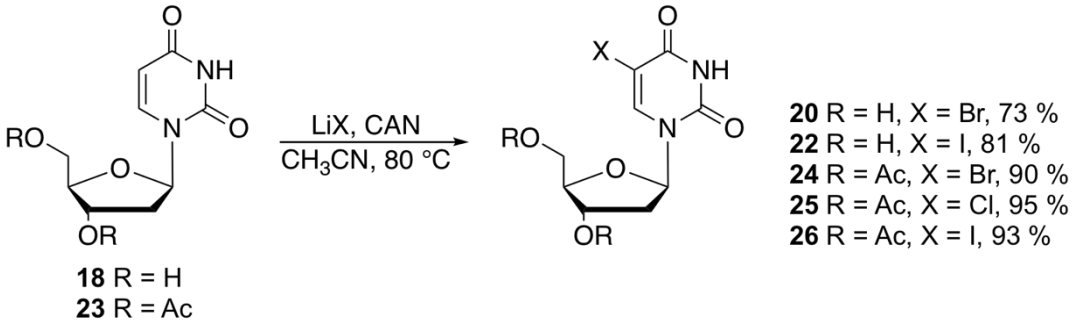


Scheme 27. Bromination of dU with DBH and Lewis acid as a catalyst.

Cerium ammonium nitrate(IV) (CAN) works as *in situ* oxidant making elemental iodine a reactive electrophilic iodo species. Acetyl protected dU, **23**, becomes iodinated at 80 °C already after 30 min giving 93 % yields (Scheme 28). Iodination of **18** to **23** using alkali-metal halogenides showed that the reaction with LiI proceeded faster than the iodination with NaI. This method can also be applied to introduce Br or Cl with an excess of CAN giving 90–95% yields. The reaction rate of the chlorination can be increased by addition of AcOH. It increases the solubility of LiCl and additionally, might increase the oxidation potential of CAN.[208]

CAN-mediated iodination and bromination with AcOH was found to be suitable also for unprotected dU (**18**). Halogenation was conducted in elevated temperatures in 30 min giving iodinated **22** and brominated **20** with 73–81% yields. Chlorination with this method was not successful for **18** as the observed compound was acetyl protected nucleoside. Chlorination attempts with conc. HCl at 70 °C did not success either leading to decomposition of the nucleoside.[209]

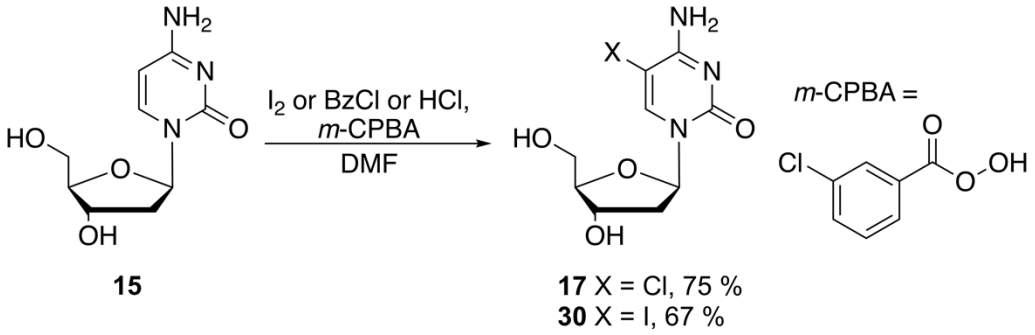
Oxidation of fluorination reagent LiF to an active species did not take place since the F_2/F^- oxidation potential (2.87 V) is higher than that of Ce(IV)/Ce(III) (1.61 V)[208]. CAN catalyzed halogenation method has also been reported to be suitable for iodination of 2'-fluorinated-2'-deoxyuridine and 2'-fluorinated-deoxycytidine at elevated temperatures giving 75 % yields.[141]



Scheme 28. CAN mediated halogenation of protected and unprotected dUs.

The oxidative iodination of unprotected **15** and its acetyl protected derivatives is an alternative method to introduce iodine or chlorine at C5 position (Scheme 29). A slight excess of elemental iodine in the presence of *m*-CPBA gives **30** with 67 % yield in 7 h[183], whereas for unprotected dU derivatives it has been reported to yield 71 % already in 30 min.[210]

Chlorination can be achieved under similar conditions using BzCl as chlorination agent. Corresponding chlorinated dC derivative is obtained with 57 % yield.[211] Chlorination with HCl, however yields **17** in 75 % after 3 h.[207] The reaction mechanism is proposed to proceed via generation of positive halogen cation which is formed from oxidation of I_2 or from BzCl–DMF complex. The corresponding halogen electrophile then undergoes a concomitant electrophilic attack at C5 affording **17** or **30**.[211]



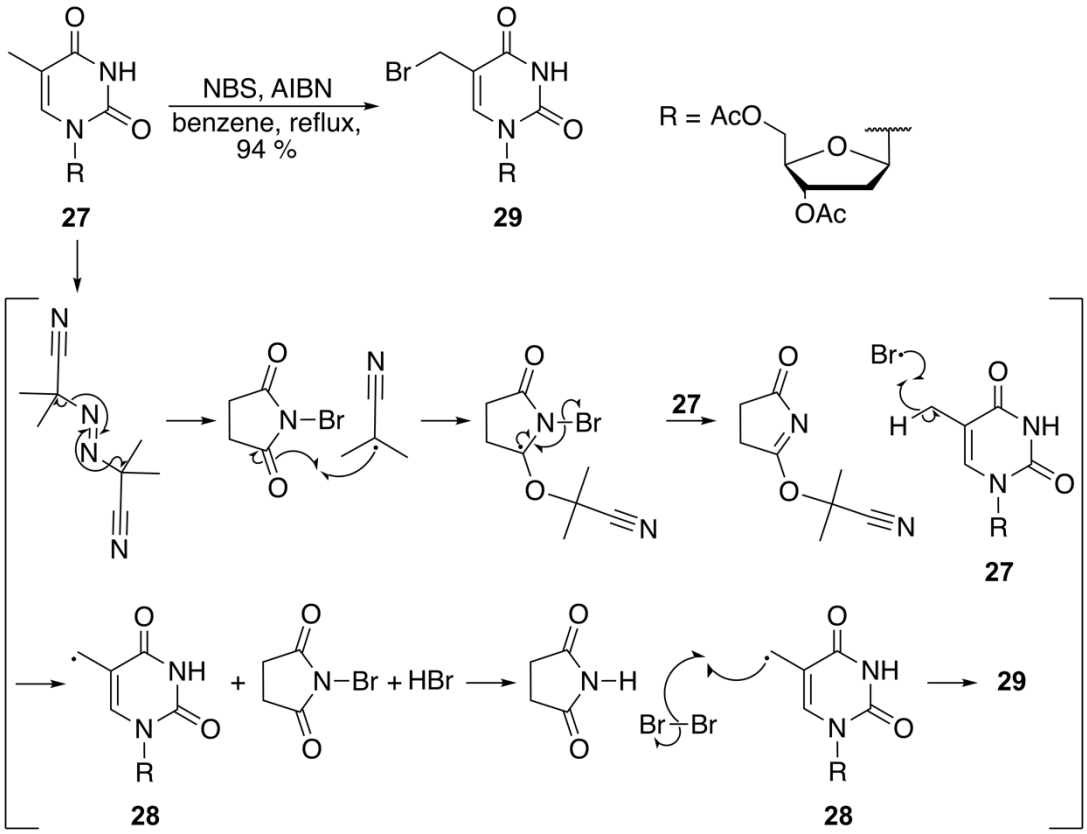
15

17 X = Cl, 75 %
30 X = I, 67 %

Scheme 29. Halogenation utilizing m-CPBA.

5.4.1. Halogenation of Deoxythymidine

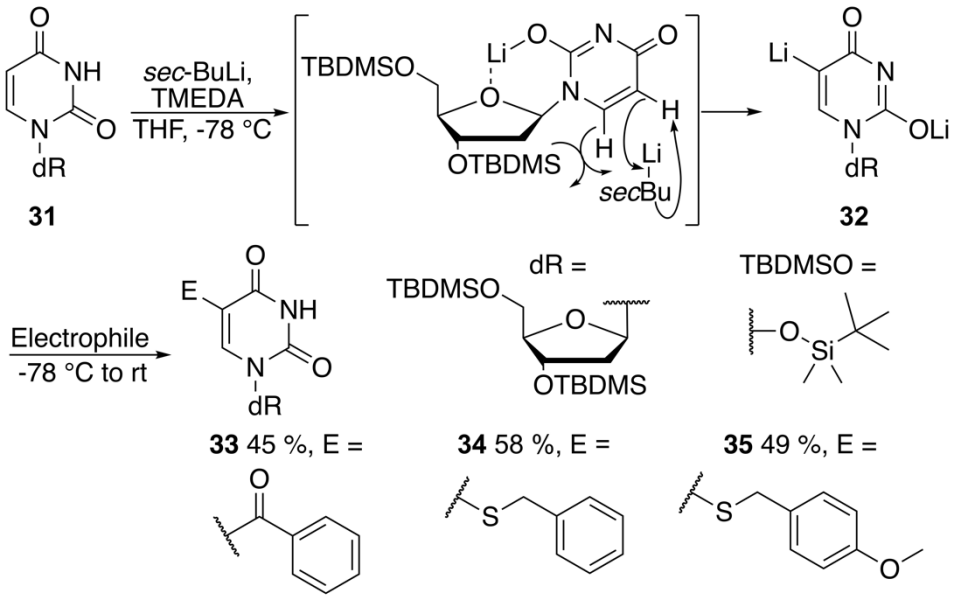
Protected dT can be brominated at the methyl group at C5 with NBS and AIBN in refluxing benzene. Radical allylic bromination is used as a selective bromination method for brominating the C5 methyl group in dT giving 90-94 % yields.[212] AIBN induces the radical formation so that the formed bromine radical can extract the hydrogen from the allylic carbon **27** forming a new primary allylic radical **28**, which is stabilized through hyperocnjugation. Formed HBr reacts with NBS forming Br_2 which reacts with radical **28** giving the brominated compound **29** (Scheme 30). However, the reaction is moisture sensitive and when water or alcohol is present, Michael type addition occurs at C6 forming a halohydrin and saturated double bond.[199,213] Additionally, the excess of bromine leads to dihalogenation and also to saturation of C5–C6 double bond.[214] Nevertheless, facile work up procedure provides **27** clean enough for the conversion into other functional groups like acetylmethyl dT by treatmeant with CH_3COOK at 40 °C in 30 min[215] and further to hmdU which is described in chapter 5.9.1.[212]



Scheme 30. Mechanism for radical bromination of dT.

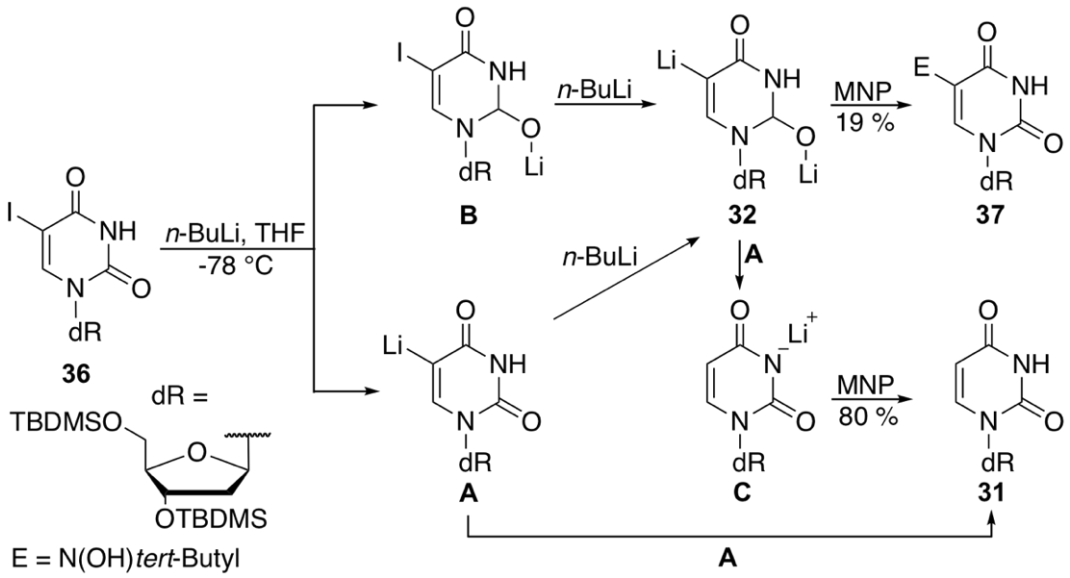
5.4.2. Functionalization via Metal-halogen Exchange

5-Lithiated dU derivative **32** is a useful intermediate for subsequent reaction with various electrophiles, like alkyl halides, carbonyl compounds and disulfides.[216,217] Armstrong R. *et al*[218] introduced a one-pot synthesis for introducing various electrophiles at C5 via lithiated intermediate **32** (Scheme 31). Regioselective lithiation of **31** using *sec*-BuLi followed by an addition of an electrophile in the presence of Tetramethylethylenediamine (TMEDA), resulted in the corresponding functionalized nucleosides **33**–**35** in moderate yields (45-58 %). TMEDA has an affinity for lithium ions and therefore enhances the reactivity of BuLi by forming a more reactive cluster. This method is reported to work more efficiently for the corresponding ribonucleosides. Therefore, an electronegative substituent at C2' position makes the reaction more efficient, hence 2'-F-dU could be a suitable compound for this reaction, yet no data for this is available.



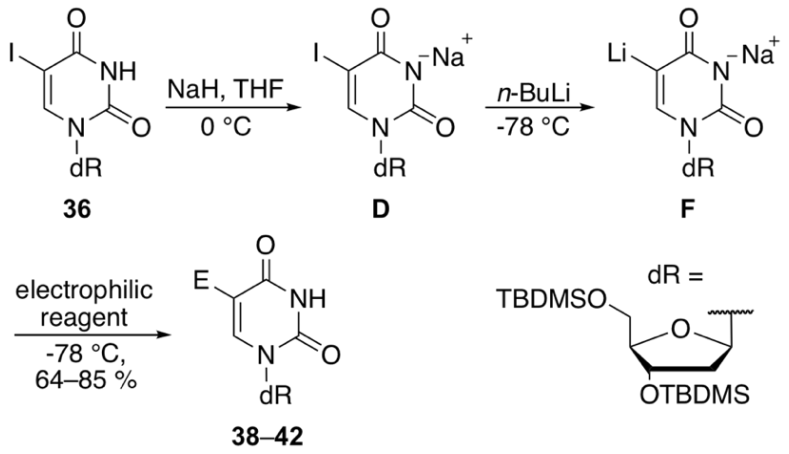
Scheme 31. One pot synthesis for introducing different electrophiles at protected dU.

A procedure to introduce new functionalities utilizing regioselective lithium-halogen exchange with *n*-BuLi has been described by Hayakawa *et al.*[217] In earlier studies, 5-lithiated dU derivative **32** was obtained with low yields.[219,220] The reason for the low yield was suspected to be caused by the unprotected imide moiety at the N3 position of the nucleoside **31** which gets deprotonated in these reaction conditions and therefore competes with the halogen metal exchange. This potential competitive imide abstraction was studied by Aso M. *et al*[221]. TBDMS-protected 5-I-dU **36** was treated with 1 eq. of n-BuLi which formed lithiated intermediates **A** and **B** (Scheme 32). Further treatment of these intermediated with *n*-BuLi gave **32**. In addition, the imide proton abstraction from **A** by **A** itself formed dehalogenated **31**. The imide proton abstraction from the lithiated intermediate **32** by intermediate **A** formed intermediate **C** which could not be converted back to reactive intermediate **32** with *n*-BuLi. The treatment of these intermediates **32** and **C** with 2-methyl-2-nitrosopropane (MNP) gave dehalogenated **31** with 80 % yield and C5 functionalized **37** with 19 % yield.



Scheme 32. Lithium halogen exchange with unprotected imide moiety.

Based on these conclusions, imide moiety of **36** was protected *in situ* by removing the imide proton with NaH which formed intermediate **D** (Scheme 33).[221] This intermediate was then treated with *n*-BuLi to obtain lithiated sodium salt **F** and followed by addition of an electrophilic reagent like MeI, PhCHO or MeSSMe giving 5-substituted dU derivatives (**38**–**42**) in improved yields (Table 4).[216–218] Noteworthy is that the reaction time including all the steps is less than two hours, it provides high regioselectivity as no C6 substitution was observed.



Scheme 33. Halogen lithium exchange with *in situ* imide protection.

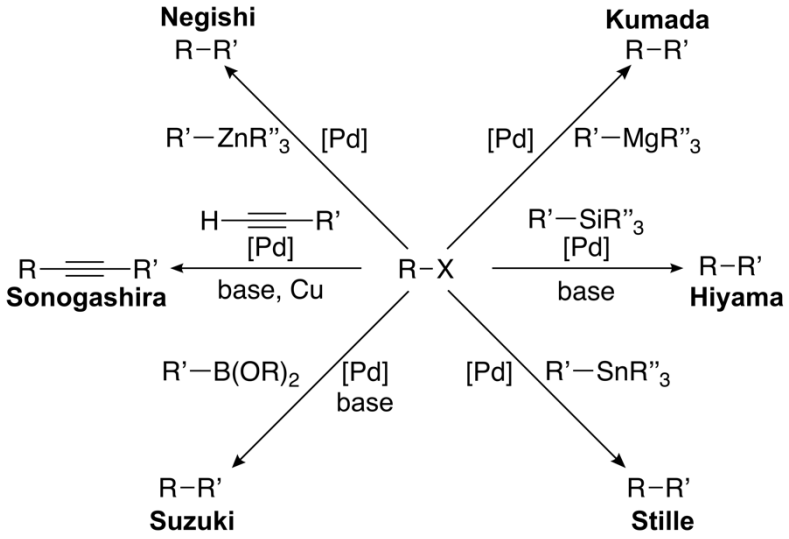
Table 4. One pot synthesis for C5 substituted dUs.

n-BuLi (eq.)	Reagent	E	Product	Yield (%)	**31** (%)
3.0	MeI	Me	38	81	7
1.2	CD_3OD	D	39	64	9
1.2	TMSCl	TMS	40	62	12
1.2	PhCHO	PhCH(OH)[a]	41	79	0
3.0	MeSSMe	SMe	42	85	0

[a] 1:1 diastereo mixture

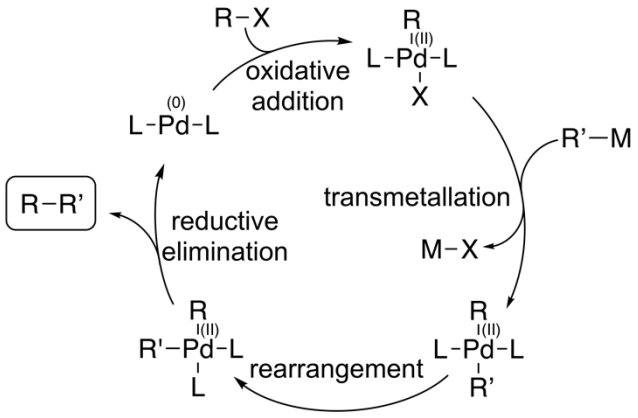
5.5. Functionalization via Palladium Catalyzed Cross-coupling Reactions

Palladium-catalyzed cross-coupling reactions have been found as a useful method to functionalize nucleosides at C5 position.[222,223] Stille (organo-stannanes), Negishi (organozinc), Hiyama (organosilicon), Suzuki (organo-boron), Sonogashira (organocopper) and Heck (organopalladium) reagents (Scheme 34) are widely used to synthesize nucleoside analogues. Pd-catalyzed coupling reactions of dU and dC derivatives can be carried under mild conditions which decrease the formation of undesired byproducts. The proper choice of ligands can increase reaction yields and decrease the reaction time or reaction temperature by lowering the energy barrier.[224] The use of an organometallic reagent R and an aryl halide R-X or pseudohalide is common for all Pd-catalyzed cross-couplings.



Scheme 34. Common Pd(0)-catalyzed cross-coupling reactions.

The general mechanism of the Pd-catalyzed cross-coupling involves three steps which are depicted in Scheme 35. The rate determining step of a catalytic cycle is believed to be the oxidative addition (first step) of the substrate R–X to the catalyst.[224] The addition of R'–M leads to transmetallation of M–X and rearrangement of the Pd(II) complex. The last step involves reductive elimination of Pd(II) to Pd(0) and a new C–C bond between the R and R'. Considering the first step, the chosen ligands have a great effect on the reaction rate and the mechanism by which the oxidative addition occurs. Pd(II) salts like $Pd(OAc)_2$, $Pd_2(dba)_3$ or $Pd(MeCN)_2Cl_2$ can be used either stoichiometric reagents or as catalysts while Pd(0) complexes like $Pd(PPh_3)_4$, $Pd(P\textit{t}\text{-}Bu_3)_2$ and $Pd_2(dba)_3$ are used only as catalysts. $Pd(OAc)_2$ can be reduced into Pd(0) species *in situ* in the presence of phosphine ligands with reducing agents such as metal hydrides, alkenes, alcohols or tertiary amines.[224] Ligands bearing a strong σ-donating ability, like trialkylphosphines, increase the electron density around the palladium, thereby accelerating the addition of a substrate. In turn, bulky ligands like phosphines with a large cone angle (Tolman angle) accelerate the elimination step.[225]

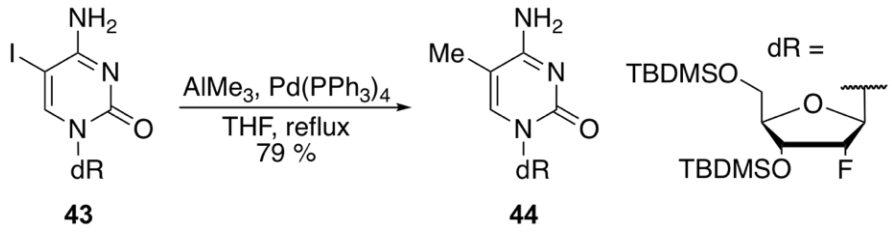


Scheme 35. General mechanism of Pd-catalyzed cross-coupling.

Stille coupling is an extremely versatile alternative to the Suzuki reaction and can be utilized in nucleoside chemistry to methylate dU and dC derivatives. It is also widely used in combination with carbonylative cross-couplings to introduce new C–C bonds at C5 position in form of a carbonyl group. Usually, this type of reaction is used with asymmetric tetraorganotin compounds e.g. allyl-, vinyl- and alkynyltin derivatives in the presence of a transitionmetal catalysts like Pd, Rh or

Ni. Generally, the most reactive group linked to the organostannane enters the coupling reaction. Therefore, it is important to consider the rates of the transmetallation step of every functional group attached to it. Symmetric tetraorganotin compounds are good cross-coupling reagents when the reductive elimination of the $R^1PdL_2R^2$ intermediate does not compete with the β-elimination. In general, a single alkyl group has the lowest transfer rate. The relative rates of transmetallation of different groups in unsymmetrical R–Sn reagents increases in order of alkyl << benzyl ~ allyl < aryl < vinyl < alkynyl.[224] The weak Sn–C bond allows an easy cleavage and subsequent formation of a new C–C bond via cross-coupling reaction. In some cases of dU functionalization via Stille coupling, a protection of N3 position of the dU derivatives is required to avoid complex mixtures.

To investigate the distribution and relative quantities of epigenetic bases in different tissues, natural and isotope-labelled nucleosides are synthesized for reference compounds to enable precise LC-MS quantification. When investigating the active DNA demethylation process that occurs via C–C bond cleavage, 2'-fluorinated dC derivatives are used as probe molecules. To introduce a methyl group at C5 position of **43**, a Pd-catalyzed cross-coupling reaction is utilized under Kumada conditions (Scheme 36). It involves the reagent trimethylaluminum and it has been reported to yield 79 % of the 5-methyl nucleoside **44**.[183,226]

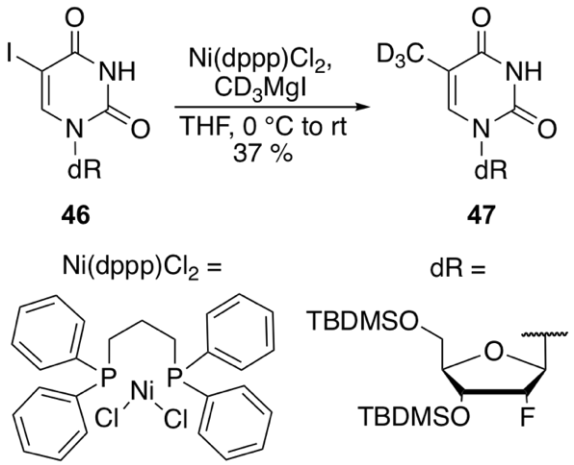


43 **44**

Scheme 36. Methylation of protected 5-iodo-2'-fluoro-2'-deoxycytidine under Kumada conditions.

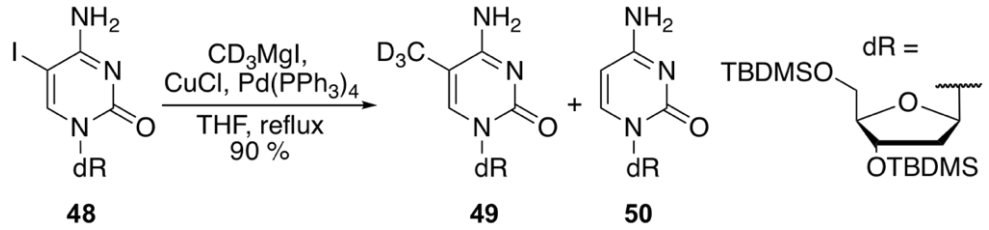
Grignard reagents like MeMgCl or CD_3MgI can also be used as methyl transferring agents to 5-halogenated dU and dC derivatives.[23] The drawback of this procedure is the chemoselectivity leading to 1:1 and 1:2 mixtures of

methylated and dehalogenated products. A recent publication introduced the CD_3 group to uridine derivative **46** using CD_3MgI as a methylating agent under Ni(II) catalysis (Scheme 37). A Ni(II) salt with bidentate C3-bridged phosphine ligand is used to catalyze this reaction yielding 37 % of isotopically lapelled dU derivative **47**.[141]



Scheme 37. Introducing an isotopically labelled methyl group to protected 5-iodo-2'-fluoro-2'-deoxyuridine.

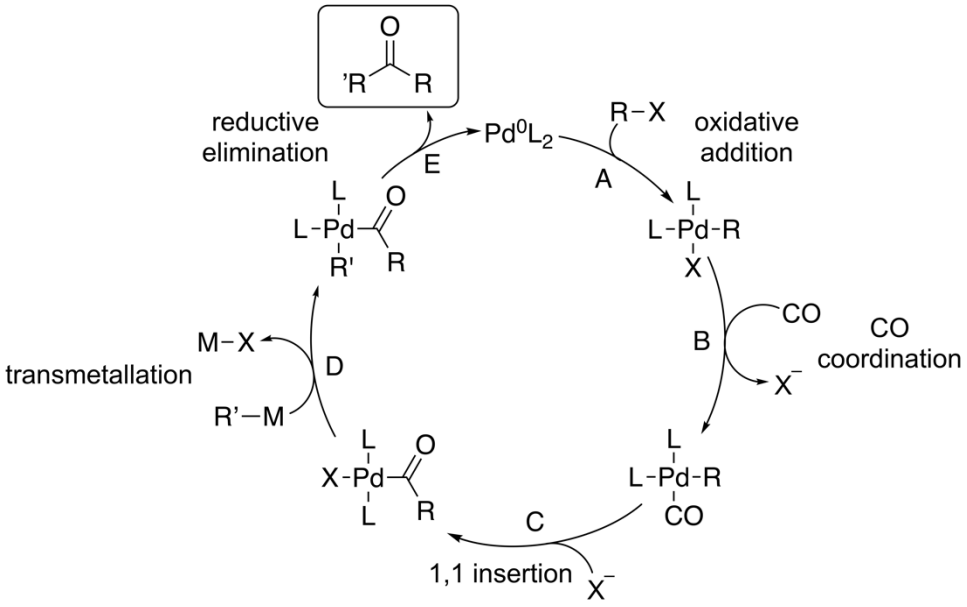
An earlier published study from the same group[183] used the same Grignard reagent to methylate a dC derivative **48** under Pd catalysis. In addition, CuCl was used in the reaction (Scheme 38). CD_3MgI undergoes transmetallation reaction with CuCl forming a organocuprate reagent. Grignard reagents prefer to react with hard electrophilic centers while organocuprates react with soft electrophiles. Magnesium is less electronegative than copper making the C–Mg bond more polarized. Due to the orbital effects, HOMO of C–Cu bond is lower in energy than the HOMO of C–Mg bond thus making it less nucleophilic. Methylation of 5-I-dC under Kumada conditions with CD_3MgX and CuCl yielded unseparable mixture of desired 5-Me-dC **49** and dehalogenated side product **50** in 90 % in an unseparable mixture.[23] In order to avoid 1,3-proton shift from the exocyclic amine to Pd-activated C5 position of the nucleoside, amine should to be protected.



48 49 50

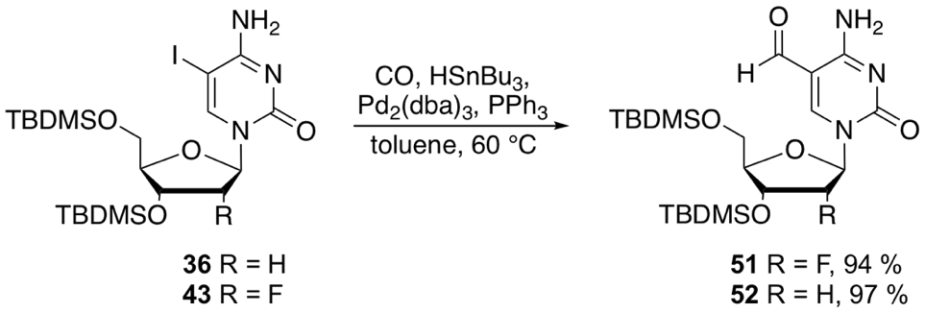
Scheme 38. Methylation of TBDMS-protected 5-iodo-2'-deoxycytidine.

Palladium catalyzed carbonylation is tolerated by many functional groups. It gives remarkable advantage over organolithium and Grignard chemistry. The actual mechanism behind the Pd-catalyzed carbonylative cross-coupling is still under debate but the potential suggestion involves five steps: oxidative addition (A), CO coordination (B), 1,1 insertion (C), transmetallation (D) and reductive elimination (E) (Scheme 39). In the first step, $Pd(0)L_2$ coordinates to the π-system of the electrophile R–X which is followed by oxidative addition that involves the cleavage of a covalent bond C–X and formation of two new bonds R–Pd–X. The oxidative addition of the aryl-halogenide to Pd changes the oxidation state of Pd from 0 to +2. This increase in the oxidation state changes the tetrahedral geometry of PdL_2 into a square planar PdL_2RX complex. In the second step CO coordinates to palladium releasing X^- and undergoes a 1,1-insertion into the Pd-organyl bond. Following transmetallation reaction between the organometallic nucleophile (M–R) and the Pd-complex, brings the organonucleophile R' to the Pd center. In order the transmetallation to occur, palladium must be the more electronegative metal since this process is driven by the electronegativity differences between the two metals involved. The formation of a new C–C bond between the electrophilic acyl group and the nucleophile R' proceeds via three-centered transition state. Regeneration of the catalytically active Pd^0L_2 occurs while the desired product is released.[227]



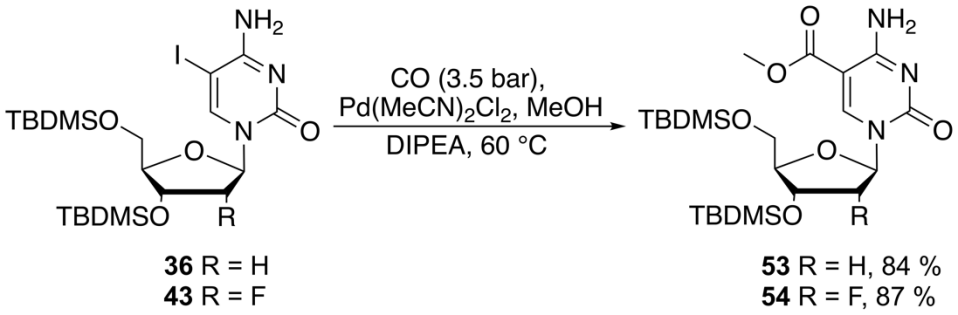
Scheme 39. General mechanism of Pd (0)-catalyzed carbonylative cross-coupling reaction.

An example of Pd-catalyzed carbonylative cross-coupling is carbonylative Stille coupling with $HSnBu_3$ that gives formylated compounds. When 5-I-dC derivatives (**36**,**43**) undergo a carbonylative Stille coupling with $HSnBu_3$, $Pd_2(dba)_3$ as a catalyst and Ph_3P as a ligand for Pd, formylated nucleosides **51 52** can be isolated in 94 and 97 % yields respectively (Scheme 40).[228] No protection of the exocyclic amino group is needed.[229,230] This reaction can also be conducted without protection of the sugar moiety, however resulting in lower yields and unpleasant purifications.[25,229,231]



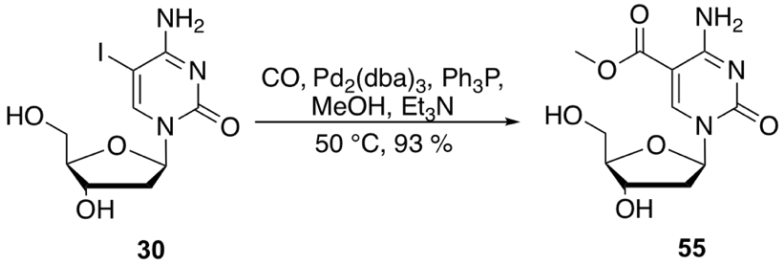
Scheme 40. Formylation of **36** and 2'-fluorinated **43** under the carbonylative Stille conditions.

Carbonylation with a Pd catalyst can also convert **36** and **43** to the corresponding esters **53** and **54** in presence of an alcohol and a base (Scheme 41).[232] When adding a ester group to the C5 position of dC derivatives, adjustment of CO pressure requires careful control in order to prevent carbonylative Buchwald coupling to the exocyclic amine.[230] Schröder *et al*[183] described a method to substitute C5 iodine with a methyl ester in 5-I-dC derivatives **36** and **43**. Carbonylative cross-coupling is conducted under 3.5 bar CO pressure and catalyzed by $Pd(MeCN)_2Cl_2$, MeOH is working as a nucleophile and deprotonated by DIPEA as a reducing agent giving **53** and **54** with 84-87 % yields.[183,230]



Scheme 41. Esterification of protected 5-I-dC derivatives.

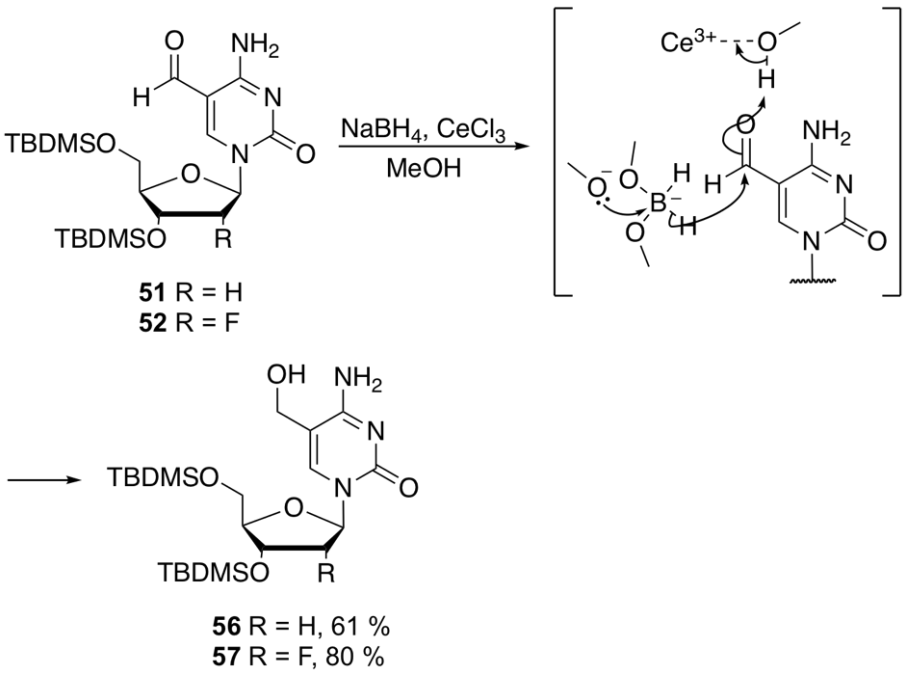
A procedure to convert unprotected 5-I-dC (**30**) into corresponding methyl ester **55** by carbonylative cross-coupling reaction has been reported to yield 93 %. The reaction is carried out under CO using $Pd_2(dba)_3$ as a catalyst and methanol as a nucleophile (Scheme 42).[228]



Scheme 42. Esterification of unprotected 5-I-dC.

5.6. Reduction of Carbonylated Nucleosides

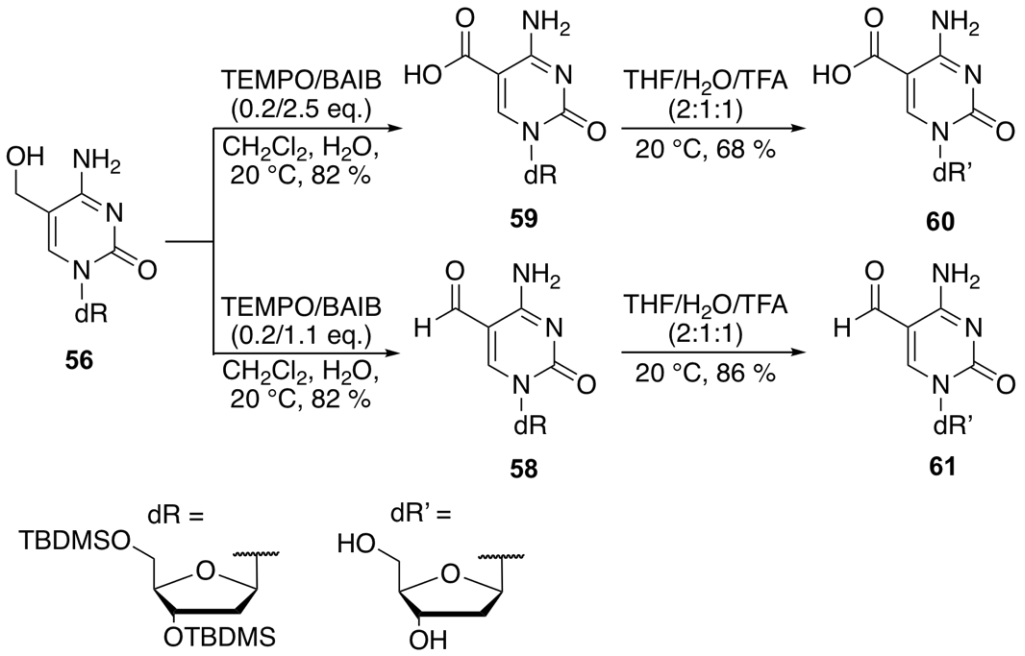
5-Formylated dC derivatives **51** and **52** can be converted into the hydroxymethyl deoxycytidines under reductive Luche conditions. Fast and efficient Luche reduction[233] with $NaBH_4$, $CeCl_3 \cdot 7\ H_2O$ and MeOH give the corresponding 1,2 reduced nucleosides **56**, **57** in good yields (Scheme 43) . $CeCl_3$ is a Lewis acid activating the methanol and a catalyst to the methanolysis of $NaBH_4$. The high selectivity of the 1,2-reduction is caused by *in situ* formed reducing agent sodium methoxyborohydrides. The replacement of the hydrides with an alkoxy groups, increases the charge density making the boron species a stronger reducing agent whose reactions are dominated by charges and electrostatic effects. The carbonyl carbon bears a substantial positive charge due to the uneven distribution of electrons at the C=O π-bond. In the unsaturated double bond between C5 and C6 positions of a nucleoside, the highest coefficient for LUMO is at C6 carbon. This is the preferred position for an attack of a soft and neutral nucleophile like $NaBH_4$ and therefore the reaction is dominated by frontier orbital effects instead of electrostatic. In the absence of the Lewis acid $CeCl_3$, the nucleophilic attack would occur at β-carbon of α,β-unsaturated carbonyl and lead to the reduction of the double bond.



Scheme 43. Reduction of a formyl group with $NaBH_4$.

5.7. Oxidation of hmdC Derivatives

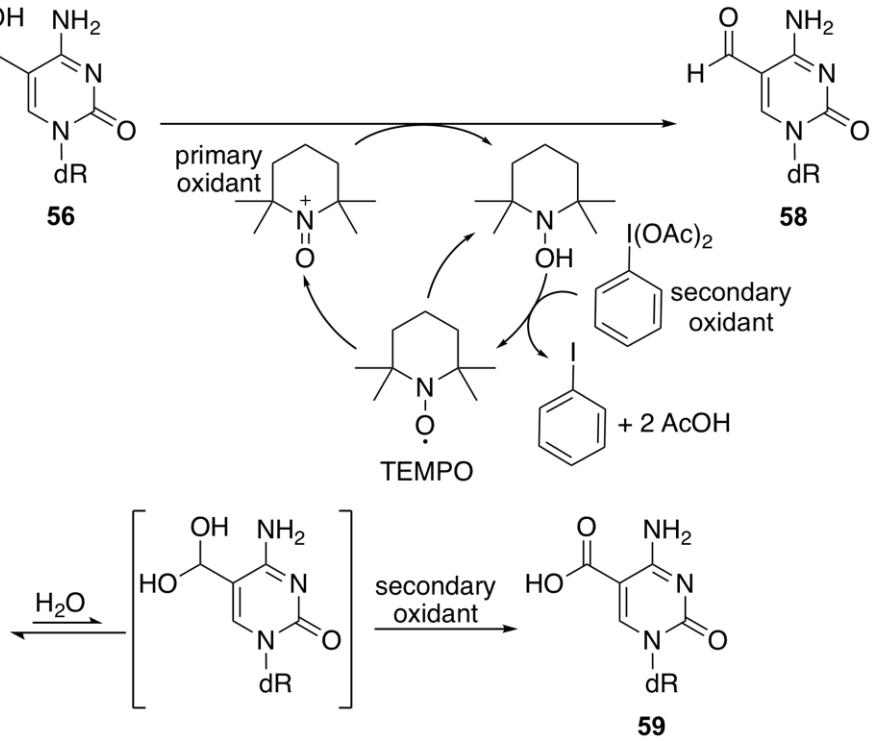
An alternative method to obtain formyl- and carboxylic functions at C5 position of dC is to use 2,2,4,4-tetramethyl-1-piperidinyloxyl / [bis(acetoxy)iodo] benzene (TEMPO/BAIB) mediated oxidation of **56** (Scheme 44).[234,235] The amount of BAIB used in the reaction correlates with the obtained compounds.[236] When stoichiometric amount of BAIB is added, formylated **58** is obtained whereas an excess of BAIB results in carboxylic function at **59**. The subsequent deprotection of **58** and **59** yields 68 % of **60** and 86 % of **61**.[215]



Scheme 44. TEMPO/BAIB oxidation of TBDMSO-protected hmdC and subsequent deprotection to fdC and cadC.

Oxoammonium salt works as a primary oxidant that converts primary alcohols to aldehydes and transforms itself to a hydroxylamine (Scheme 45). A secondary oxidant BAIB then oxidizes the hydroxylamine to TEMPO and generates AcOH. TEMPO goes through a bismutation to the hydroxylamine and oxoammonium salt. AcOH is necessary for initiation of the reaction because it is involved in formation of the oxoammonium salt. Oxoammonium salts, as primary oxidants, are in charge of oxidizing alcohols into aldehydes. The oxidation of aldehydes to carboxylic acids occurs with the oxidant BAIB which is present as an excess amount. BAIB works as a secondary oxidant when converting hydroxymethyl

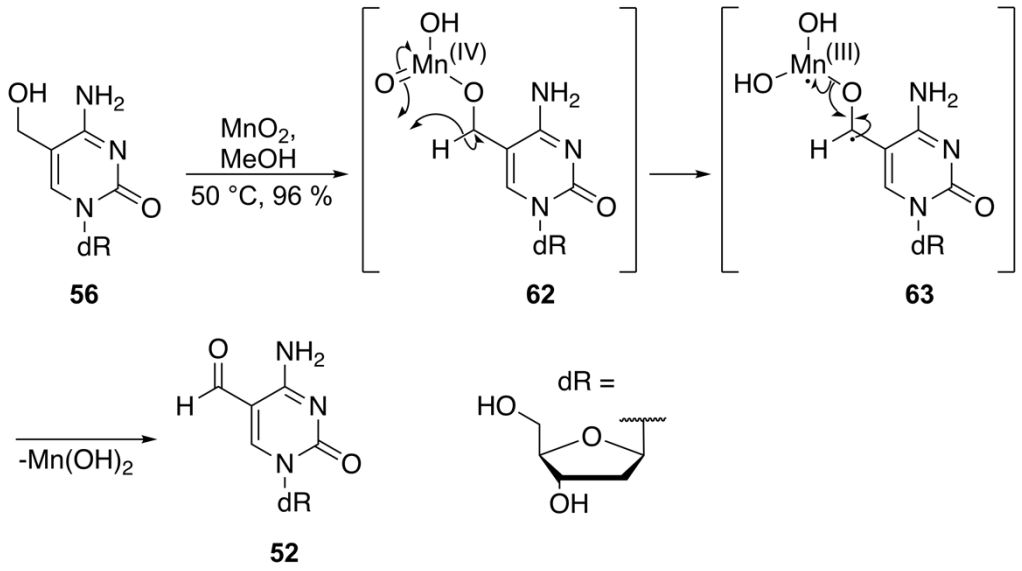
group of **56** to a formyl group **58**. When oxidizing a formyl group of **58** into a carboxyl group, BAIB works as a primary oxidant yielding **59**.[237]



Scheme 45. Mechanism for TEMPO/BAIB oxidation.

An alternative method to convert **56** to fdC **59** is the oxidation reaction utilizing activated MnO_2.[238] MnO_2 selectively oxidizes primary alcohols to aldehydes. Since allylic alcohols are more reactive, the reaction can be carried out without protection of the sugar moiety. The oxidation takes place on the surface of the MnO_2. Water competes of the surface area with an alcohol, wherefore it needs to be removed by drying the MnO_2 which results in the “activated oxidant”.

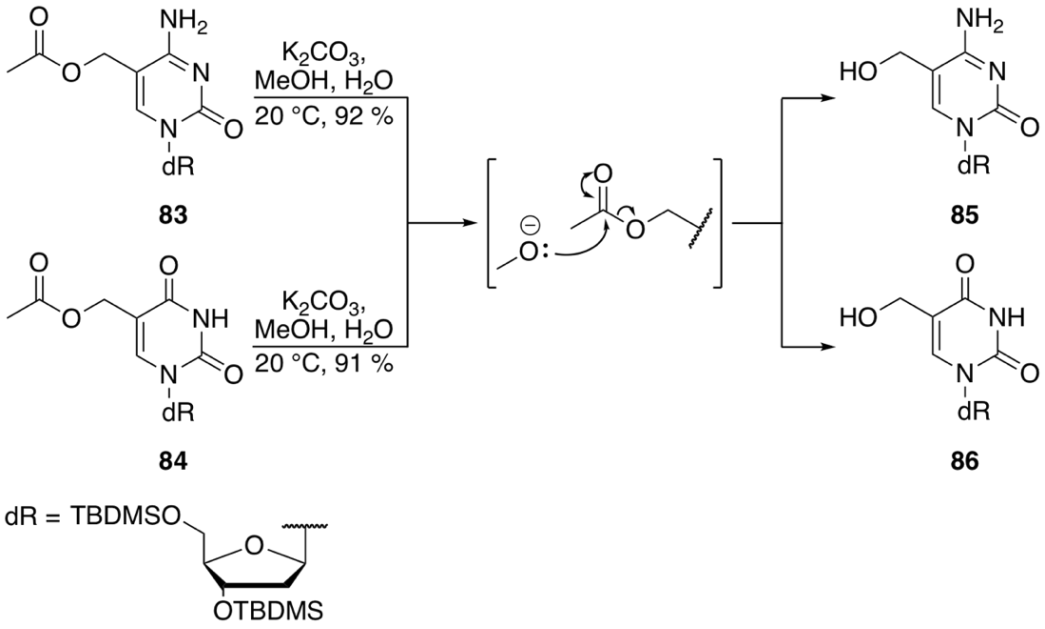
In the first step of the reaction, OH-group of the allylic alcohol binds at MnO_2 resulting in the ester functionality **62** (Scheme 46). Then Mn(IV) is reduced to Mn(III) and subsequently a hydrogen shift from the allylic carbon to the oxygen of manganese occurs. The unpaired electron at the allylic carbon **63** is resonance-stabilized. In the last step, Mn(III) is reduced to Mn(II) and a double bond between the allylic C and O is formed giving **52** with 96 % yield.



Scheme 46. Oxidation of hmdC with activated MnO_2 to fdC.

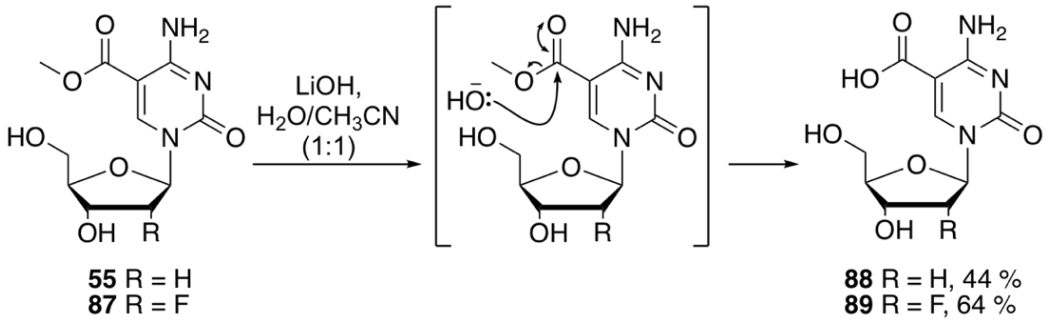
5.8. Hydrolysis of an Ester Group

Deprotection of acetyl groups from 5-acetylmethyl dC (**83**) and **5-**acetylmethyl dU (**84**) can be performed by saponification using K_2CO_3, MeOH and H_2O. K_2CO_3 acts as a base deprotonating MeOH and a nucleophilic methoxy anion is formed. This anion then attacks the carbonyl carbon eventually forming a hydroxyl function and a methyl ester (Scheme 47). This procedure deprotects **85** and **86** in good yields within two hours.[215]



Scheme 47. Deprotection of acetyl group to establish hmdC and hmdU.

In general, hydrolysis of an ester group can be performed in either basic or acidic conditions. For nucleosides, basic conditions are considered to be the preferred method.[183,250] The saponification of **55** and **87** is conducted with LiOH in a mixture of water and an organic solvent in an overnight reaction giving moderate yields of **88** and **89** (Scheme 48). The reaction proceeds in a similar manner to the earlier mentioned deprotection of **83** and **84** with K_2CO_3. A nucleophilic hydroxyl anion attacks the carbonyl carbon which is followed by subsequent removal of methoxy group forming a carboxyl group at the C5 position of the nucleoside.[183,250,256]

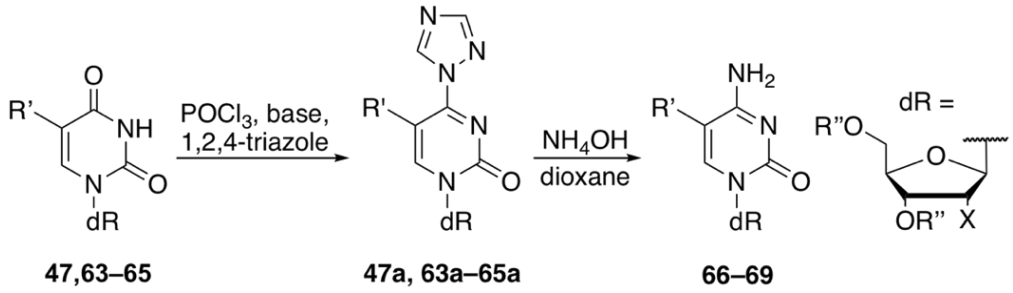


Scheme 48. Saponification of cadC esters.

5.9. Amination of 2'-deoxyuridine Derivatives

Several amination methods for converting dU derivatives to the corresponding dC derivatives have been reported. The most efficient protocols include two steps of which the first one includes the reaction of a dU nucleoside with $POCl_3$ and 1,2,4-triazole[169,241], $SOCl_2$ in DMF[242] or 2,4,6-triisopropylsulfonyl chloride (TPS-Cl)[172,243,244], which is then followed by an ammonolysis step.

The most efficient route for amination in terms of execution and purification seems to be the methdod via formed triazolide–intermediate (Scheme 49). Nucleosides **47**, **63**–**65** can be converted to triazolides **47a 63a 67a** with $POCl_3$ and 1,2,4-triazole and further to dC derivatives by ammonolysis with aqueous ammonia (NH_4OH) (Table 5).[139,141,241] Applied reaction conditions for the different nucleosides are listed in Table 5.



47,63–65 **47a, 63a–65a** **66–69**

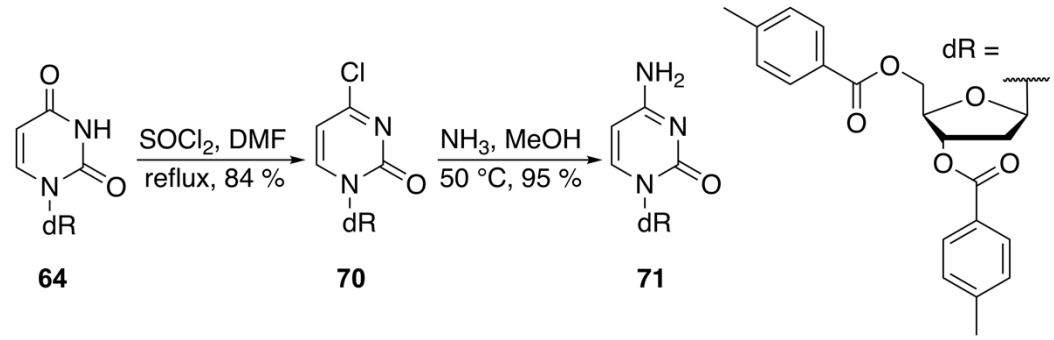
Scheme 49. Triazolide based amination of dU derivatives. Reaction conditions are shown in table 5.

Table 5. Reaction conditions for triazolide based amination of dU derivatives.

				1st step			2nd step		
sm	R'	R''	X	base	temp (°C)	time (h)	temp (°C)	time (h)	product, yield (%)
47	CD_3	TBDMS	F	pyridine	rt	2	rt	o.n	**66**, –
63	H	TBDMS	F	Et_3N	60	14	30	16	**67**, 59
64	H	4-Me-	H	Et_3N	rt	3	40 to rt	18	**68**, 85
65	I	4-Me-	H	Et_3N	30	o.n	rt	0.17	**69**, 89

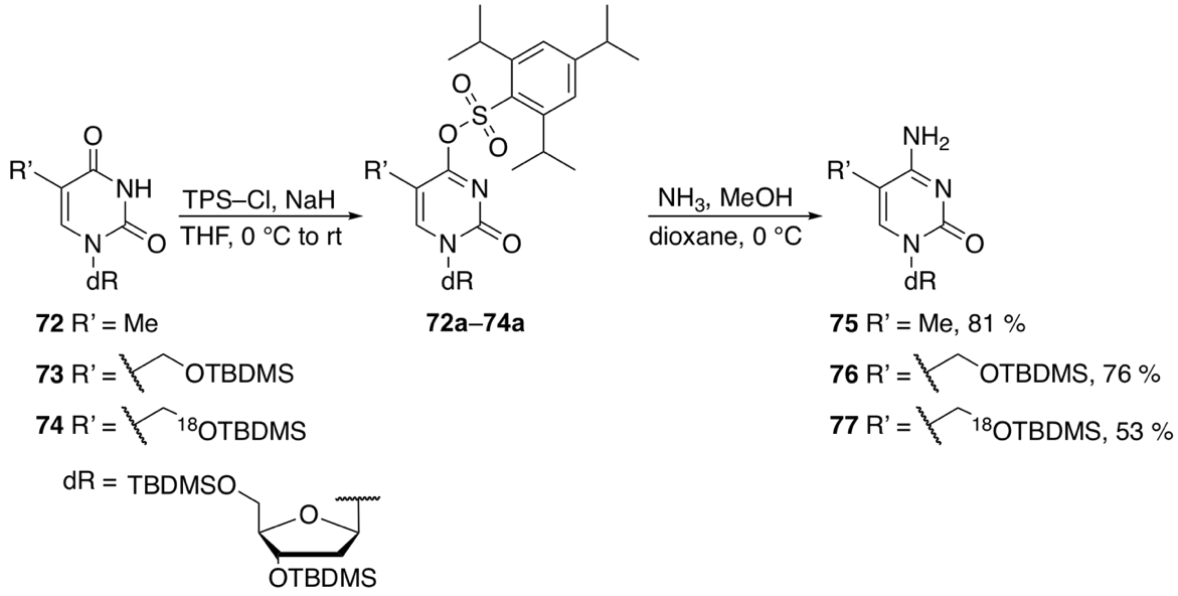
sm = starting material, rt = room temperature, o.n = over night

Another amination method that proceeds via 4-chloro derivative **70**, was first introduced by Zemlica and Sorm for the synthesis of 4-chloro-uridine derivatives.[242] This method for converting dU derivative **64** to the corresponding dC derivative **71** is described to proceed in two steps. The first step involves the activated 4-chloro derivative **70** followed by ammonolysis, yielding 95 % of the aminated **71** (Scheme 50).[241] However, the drawbacks in this method are the need for careful reaction monitoring as well as efficient quenching after the first step to avoid degradation.



Scheme 50. Amination method using $SOCl_2$ to form a 4-chloro-intermediate.

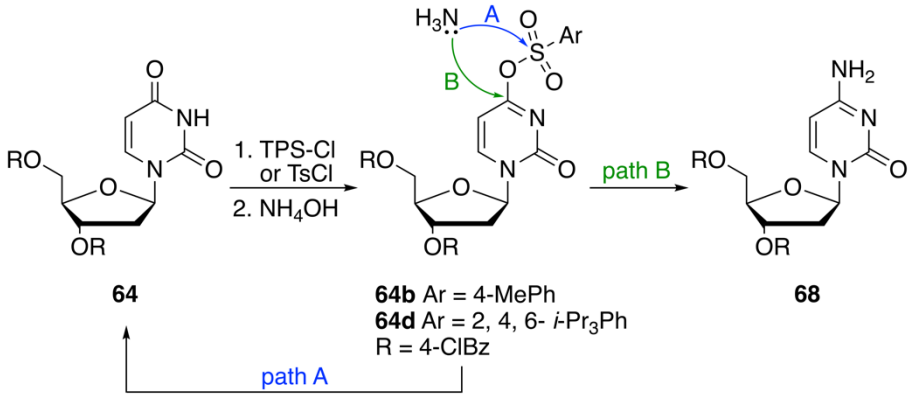
TPS–Cl mediated amination of dU derivatives **72**–**74** to corresponding dCs includes an activation of the oxygen at C4 as sulfonates (Scheme 51). These sulfonate esters **72a**–**74a** are subsequently substituted via ammonolysis which provides cytidine derivatives **75**–**77** in moderate yields. The first step of the reaction is rather slow, whereas the ammonolysis step is performed at 0 °C within three hours.[23]



Scheme 51. TPS–Cl mediated amination method.

Besides the sulfonate ester method applying TPS–Cl, does not always give satisfactory results, a method utilizing TsCl has been developed as an alternative reagent to activate C4 carbonyl group of dU derivatives. However, the ammonolysis is the critical step where the reaction of NH_3 nucleophile competes between a C4 position attack (path B) and the sulfonyl group attack (path A)

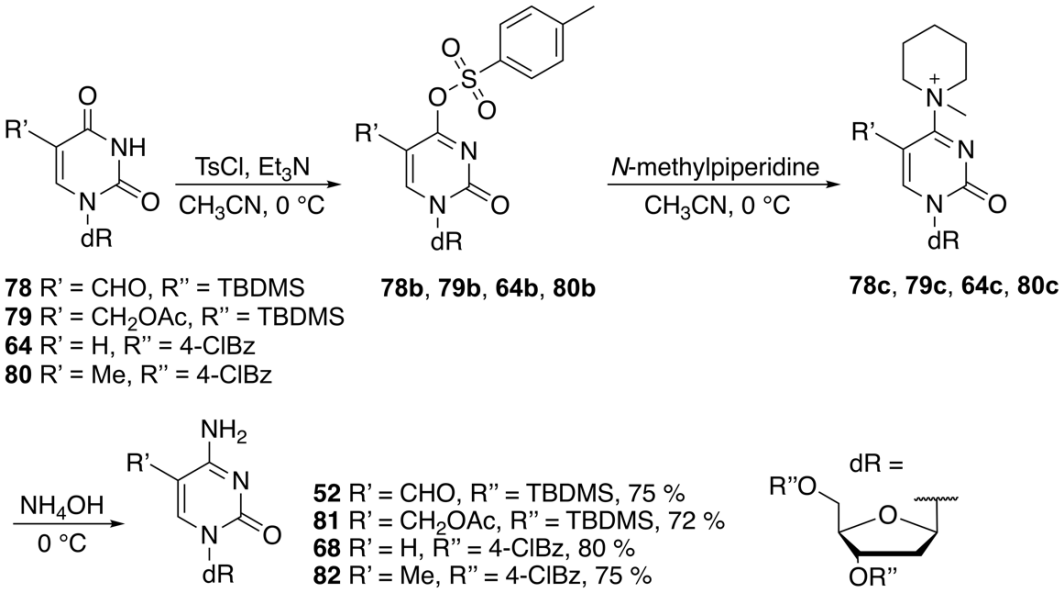
(Scheme 52). Path A reforms undesirable starting material **64.** Bulky isopropyl groups in the case of TPS-mediated method are essential for avoiding this side reaction and thereby **68** is preferentially formed. On the other hand, the triazolide or chloride based amination methods do not show this side reaction but as a drawback require long reaction times and a large excess of the respective reagents.[245]



Scheme 52. Possible sites for the nucleophilic attack of ammonia on sulfonate esters.

Moreover, the amination of different dC derivatives **64**, **78**–**80** with TsCl and Et_3N utilizing additives like 1-methylpiperidine, DMAP or DABCO increased the selectivity of the nucleophilic attack at C4 of the nucleoside rather than to the sulfonyl group. Rapid conversion of **64**, **78**–**80** to the corresponding cytidine derivatives was obtained with high selectivity with 1-methylpiperidine and subsequent ammonolysis was conducted *in situ* with NH_4OH (Scheme 53).[215,246] When the reaction was conducted with DABCO, ammonolysis with NH_4OH did not yield an aminated product. Utilizing gaseous NH_3 instead, gave high selectivity and yields, though increasing the reaction time. The reaction with DMAP proceeded slower than with the aforementioned additives, also more equivalents of DMAP were required to complete the reaction. These results can be explained with the fact that DMAP is more basic than 1-methylpiperidine and DABCO which results in a more likely protonation compared to Et_3N. This will in turn inhibit the substitution of DMAP. The slow reaction with DMAP was accelerated with the addition of K_2CO_3 but the selectivity did not improve. Based on these results, the reaction mechanism of the ammonolysis is considered to proceed via replacement of the sulfonate group of **64b**, **78b**–**80b** by the 1-

Methylpiperidine, leads to a formation of intermediates **64c**, **78c**–**80c** as quaternary ammonium salts. This quaternary salt is then substituted by NH_3 forming the desired dC derivatives **52**, **68**, **81**, **82** (Scheme 53).[215,246]



Scheme 53. Amination method using TsCl and 1-methylpiperidine.

5.10. Protection Groups

Three considerations are important when choosing an appropriate protection group (PG): the functional group to be protected, the applied conditions for the protection and conditions needed when cleaving the protecting group.[11,239] Since stability aspects of nucleosides were discussed earlier, the choice of a suitable PGs for the hydroxyl groups of the 2'-deoxyribose sugar unit is a critical point. Besides that, knowledge of the effect applying different PGs from electropositive silyl groups to EWG acyl groups can have a disctinct impact on the performed reaction. Silyl protecting groups have significantly different electronic and steric requirements than acyl and alkyl protective groups, which becomes particularly important when two neighboring alcohols are silyl protected.[219] Silyl ethers are less electron withdrawing than benzoyl (Bn) or acyl (Ac). Due to their bulkiness, they can cause a change in the sugar ring conformation and further in the stability of the glycosidic bond.[239]

Also, studies on competitive deprotonation of C5–C6 positions at uridine derivatives showed that the obtained ratio is determined by the protecting groups of the sugar moiety. Applying less bulky protection groups like 2',3'-*O*-isopropylidene acetal, the resulting derivatives exclusively undergo the thermodynamically favored C6 deprotonation.[217,247,248] On the other hand, a freely rotating tert-butyldimethylsilyloxy (TBDMS) protecting group at C2' position results in C5 deprotonation. A study of 2'-deoxyuridines showed that C5 deprotonation occurs under similar conditions but cannot be caused by the bulky silyl group. Space-filling CPK models indicate the C6 position to remain hindered regardless of the actual conformer.[218] Also, a possible exchange between C5 and C6 protons before the incorporation of an electrophile was excluded by isotope tracing study[249] in which the C6 deuterated dU was used. Applying the same deprotonation conditions followed by an alkylation, no cluttering of the isotope was observed.

In case of cytidine, its aromatic amino group can also be considered as a reactive position. Its reactivity due to substituent effects is similar to *p*-nitroaniline. The pKa value of the amino group ranges from 2.5–4. Generally, electrophilic reagents react with the amino group. A typical reaction is an amino group acylation which can also involve the acylation of hydroxyl groups of the sugar moiety. Selective *N*-acetylation reactions, however, can be conducted in a boiling alcohol solution with an equimolar amount of acyl anhydride. For selective reactions, *t*-Butyl carbamate (*t*-Boc) groups are widely used in case of heterocyclic amine protection. Such groups are inert to many nucleophiles and are not readily hydrolyzed under the basic conditions. Usually, *t*-Boc is cleaved off using strong acids but acid stability can be increased by the introduction of fluorine substituents. Selective deprotection of a single *t*-Boc group from a (*t*-Boc)$_2$-amine is possible using LiBr or Mg(ClO_4)$_2$ as suitable reagents.[250,251]

6. Conclusions

With the discovery of the modified DNA bases mdC, hmdC, fdC and cadC a new research field evolved studying the function of these bases. These modified dCs were found to be key influencers of the so called epigenetic gene regulation, a layer of information beyond the canonical genetic code.

After the relation between the TET enzymes and epigenetically relevant bases was found, several mechanisms for an active DNA demethylation process have been proposed.[78] One of the best-established pathway involves TET-mediated oxidation of mdC to hmdC, fdC and cadC followed by BER process to replace these higher oxidized derivatives.[32] However, BER-based erasure pathway involves a risk of creating clustered single and double strand breaks with potentially harmful consequences. The other potential pathway to remove fdC and cadC does not involve DNA repair. It is supported by the findings of TET3-dependent active demethylation of maternal and paternal genomes independently from TDG. Additionally, a recent study shows that mammalian cells are able to convert fdC to dC while the glycosidic bond is kept intact.[141]

Many 2'-deoxyribonucleoside derivatives, substituted at C5, are known to have biological activities and can act i.e. as antiviral or anticancer agents. Furthermore, they can be used as suitable target molecules to study DNA metabolism pathways or can provide useful information about cellular biochemistry.[141] To elucidate the biological role of the epigenetically relevant DNA bases, efficient methods to synthesize these nucleoside analogs have been developed. So far, many efficient halogenation methods provide efficient synthesic routes to C5 halogenated dU and dC derivatives which themselves are important antiviral agents or tools to study cellular biochemistry. Furthermore, they can also act as substrates for Pd-catalyzed cross-coupling reactions resulting in generation of new C–C bonds.

The first 50 years after the discovery of the double helix by Watson and Crick were devoted to studying the sequence information of the genetic code, which finally culminated in sequencing a complete human genome. Based on the recent literature, the upcoming years will be promising times to reveal the active DNA demethylation mechanism and the entity that lead the regulatory layer of genetic information – the epigenetically relevant bases – to convert back to canonical dC.

Experimental part

7. Introduction

In living organisms DNA demethylation processes of genomic mdC are required to reactivate silenced genes. Implying potential pathways from genomic mdC to canonical dC, it has been shown that TET-mediated oxidation plays a crucial role in these processes.[21,96] Currently, three main pathways (Ch.4.2.2) are intensively investigated to understand how active removal of mdC mechanistically proceeds. The TDG based BER pathway is important in many scenarios but cannot be rationalized in all of the situations, for example considering BER based demethylation of high density methylated CpG regions which would create multiple single- and double strand breaks and lead to compromised genetic integrity of the respective loci. The discovery of active demethylation to occur TDG-independently in zygotes, showed another possible pathway of active demethylation.[80] This discovery supports the earlier studies that showed cadC containing DNA strands to be decarboxylated in the presence of thiols and a base. Furthermore, it was shown that the mESC extract can convert cadC to dC in artificial DNA strands.[139] Besides, thiols mimic the active site residues of potential DNA modifying enzymes which could lead to an alternative active DNA demethylation mechanism via deformylation of fdC or decarboxylation of cadC.

In fungi, demethylation of thymine to uracil is a multistep process including several oxidation steps similar to mammalian cells. In turn, the decarboxylation of carboxuracil to uracil within the thymine salvage pathway is known to be catalyzed by isoorotate decarboxylase (IDCase). However, in mammals the oxidation of mdC to hmdC[10,16], fdC[19,20] and cadC[20,21] is catalyzed by TET enzymes but the entity behind deformylation and decarboxylation is unknown. It is suggested that fungi IDCase and a potential mammalian DNA decarboxylase may share similarities in their sequence, structure and catalytic mechanism. In fact, it has already been studied that IDCases can catalyze decarboxylation of carboxycytosine to cytosine. Nevertheless, the observed decarboxylation activity was weak but still it is the first *in vitro* evidence for decarboxylation of a modified DNA base catalyzed by a distinct enzyme.

Further, DNA C5-methyltransferases (C5-MTases) have also been shown to have an ability to promote decarboxylation via C–C bond cleavage of cadC to dC

in vitro.[140] Moreover, these enzymes have also been shown to dehydroxymethylate hmdC to dC but not to have an impact on fdC. This might indicate a different processing mechanism for cadC and fdC. Additionally, an earlier *in vitro* study of deformylation and decarboxylation in the presence of thiols and acid confirmed decarboxylation to proceed 11 times more efficiently than deformylation.[41] However, a recent study presented evidence of deformylation to occur via direct C–C bond cleavage *in vivo*.[141]

7.1. Decarboxylation via a Vinyl Anion Type Intermediate

The closest human relatives to fungi IDCase are aminocarboxymuconate semialdehyde decarboxylase (ACMSD) and cytosine deaminases.[24] They belong to the amidohydrolase superfamily, ACMSD more specifically to carboxy-lyases that cleave C–C bonds. Cytosine deamines act as hydrolases on C–N bonds, potentially also others than peptide bonds. In general, The IDCase shares a similar active site structure with these human enzymes. This active site consists of a divalent metal ion coordinated by four residues that are strictly conserved. The main character of the catalytic mechanism for these amidohydrolase superfamily members is a nucleophilic attack of a metal-bound hydrate on a carbon of the substrate in order to form a tetrahedral intermediate. However, an analysis of the catalytic mechanism of IDCase shows that in the substrate binding complex with carboxyuracil, there is no room for a water molecule to bind to Zn^{2+} and act as a nucleophile in the catalysis. Therefore, another potential nucleophile to conduct this reaction could be a Zn^{2+}-coordinated side chain carboxyl of aspartic acid (**Asp**) (Figure 9). It is oriented towards the carboxyl group of carboxyuracil providing a suitable distance (3.0 Å), angle (103°) and, moreover, a negative pKa of -2.6 which is lower than any other residue in proximity (Scheme 54).[145]

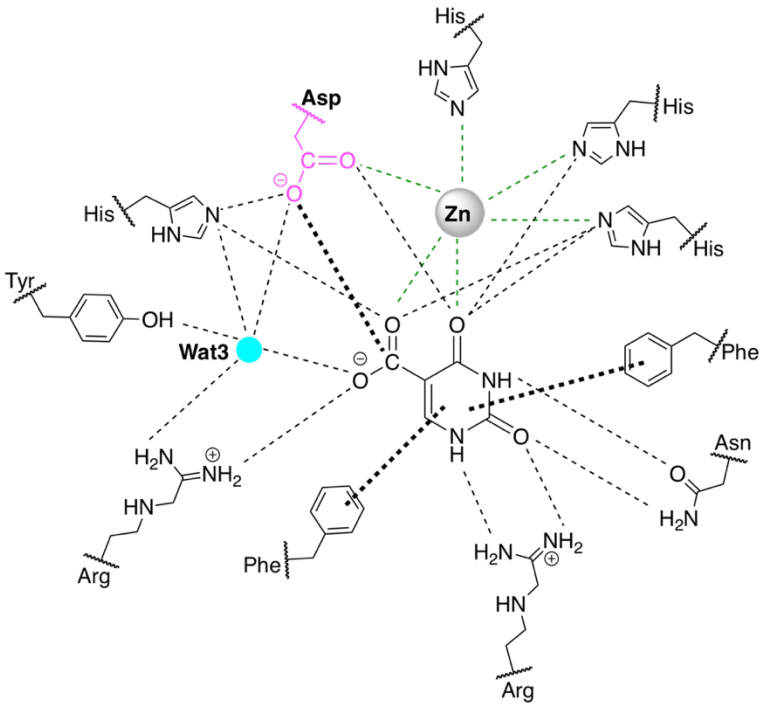
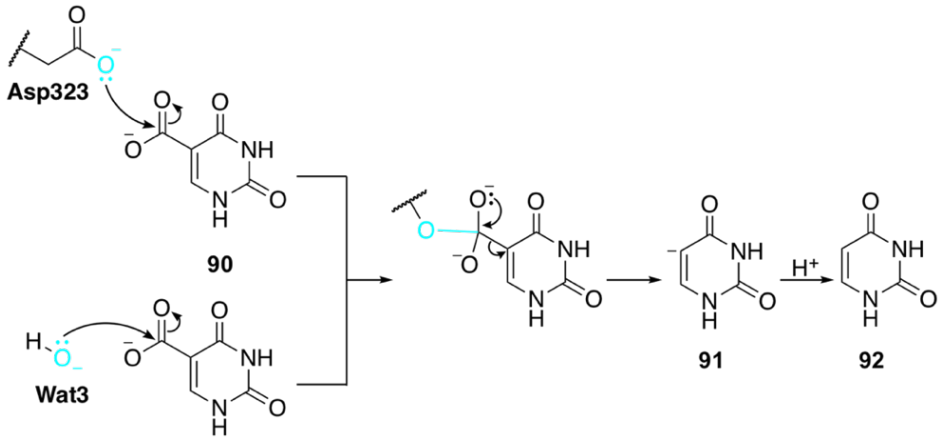


Figure 8. The IDCase substrate binding complex in presence of carboxyuracil.

Another potential nucleophilic species could be formed if Asp323 acts as a base deprotonating **Wat3.** A thereby formed hydroxide ion acts as the nucleophile (Scheme 53).[145] Subsequently, one of these nucleophiles initiates the attack at the carboxylic carbon in uracil followed by elimination of CO_2 and a formation of a vinyl anion intermediate **91**. Followed by protonation of **91** gives decarboxylated uracil **92**. However, quantum chemical calculations show this pathway not to be suitable for carboxycytosine; the energy for the vinyl anion intermediate formed in case of cytosine is +47 kcal mol^{-1}, whereas in case of uracil it is less unfavorable +34 kcal mol^{-1}.[41]



Scheme 54. Simplified mechanism of IDCase catalyzed decarboxylation via a vinyl anion intermediate. The proposed mechanisms include two possible nucleophilic species Asp323 or Asp323 activated Wat3.

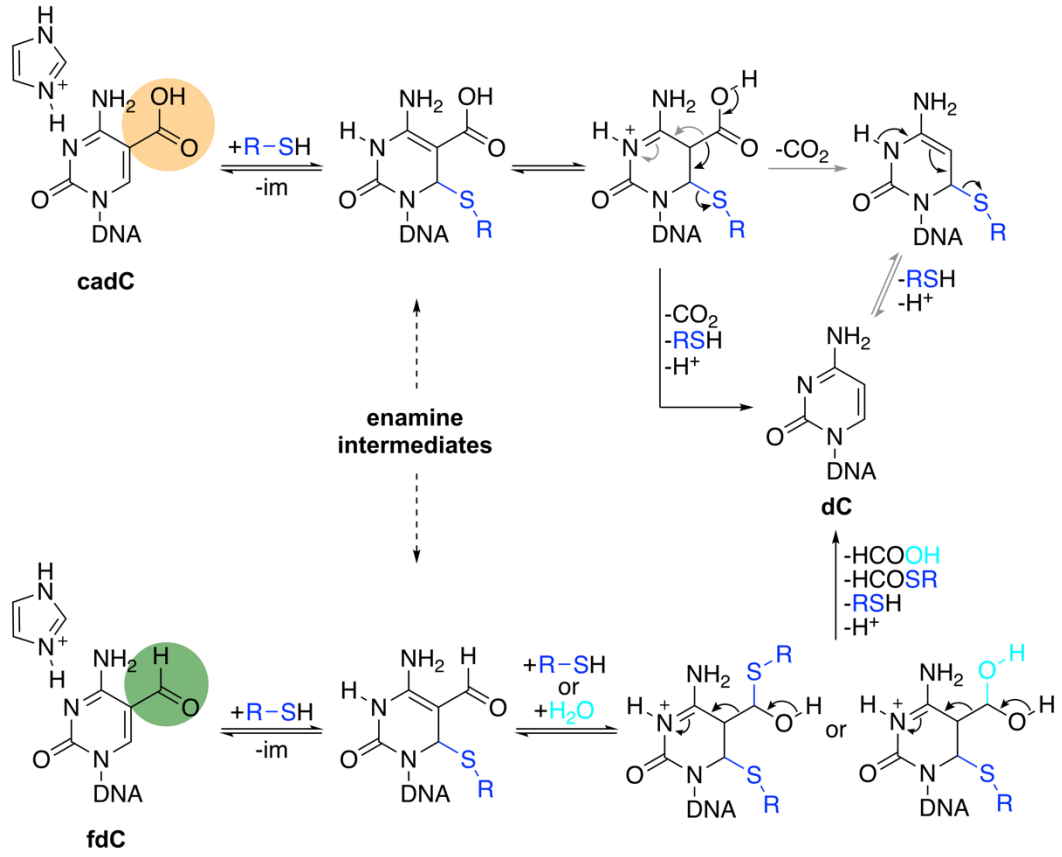
7.2. Decarboxylation via a Covalent Enamine Intermediate

The first evidence that stem cell extracts have the ability to decarboxylate cadC containing DNA was prosecuted by Sciesser *et al.*[139] Experiments were performed with synthetic oligonucleotides containing a double ^{15}N-labelled cadC. The reactions that would allow stem cells to decarboxylate cadC were then investigated. Incubating $^{15}N_2$-cadC containing DNA strand with amino acid cysteine, decarboxylation was detected. Based on chemical model study, it was suggested that decarboxylation is accelerated by saturation of C5–C6 double bond of cadC. This saturation was proposed to occur after the nucleophilic attack at C6 by a thiol. This would then further decarboxylate and re-aromatize which points the direction to an alternative active DNA demethylation via decarboxylation.[139]

Later the same research group described a study of C–C bond cleavage reactivity investigating hmdC, fdC and cadC.[41] Oligonucleotides containing epigenetically relevant nucleosides were investigated in a thiol-mediated reaction which mimics enzymatic catalysis; a high concentration of β-mercapto ethanol (β-ME) was used. In more detail, it was shown that a lower temperature and a higher pH resulted in decreasing dC yields which indicates towards a reaction catalyzed by proton activation. This is supported by the fact that decarboxylation was observed most efficiently at pH 5.0 and at 50 °C. Further, it was shown that decarboxylation

of cadC is more efficient (28 %) than deformylation (2.5 %) of fdC, whereas dehydroxymethylation of hmdC was only observed 0.5 %.

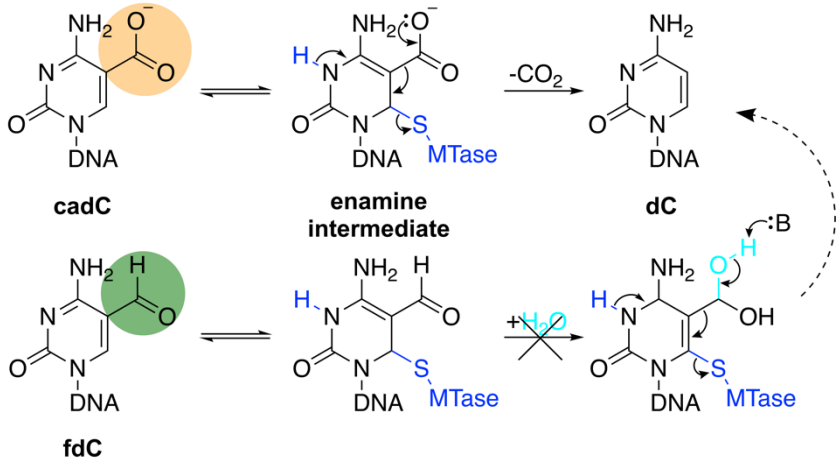
The proposed deformylation mechanism of **fdC** requires an additional nucleophile (water or thiol) to attack the formyl group, before the formic acid can be eliminated (Scheme 55). In general, the first step involves protonation of N3 of **cadC** or **fdC** and subsequent addition of a thiol at C6. This is followed by the release of formic acid (HCOOH) or carbon dioxide (CO_2) and finally by the elimination of the thiol bound at C6.[41]



Scheme 55. Dexarboxylation of cadC and deformylation of fdC via covalent enamine intermediates. The gray pathway indicates CO_2 and thiol elimination to occur in two steps. R-SH = β-mercaptoethanol; im = imidazole.

Liutkevičiūtė *et al.*[140] described potential enzymatic C5-MTase decarboxylation pathway. This reaction requires the presence of a catalytic cysteine residue in the active site which attacks at C6 of cadC, thereby allowing the reaction to proceed via the proposed covalent enamine intermediate (Scheme 56). It is

assumed that for the covalent intermediate to separate via C–C bond cleavage, the exocyclic C5 functionality needs to be deprotonated. Since the study claims that fdC is not likely to deprotonate, it might explain why these enzymes do not deformylate fdC.



Scheme 56. Proposed mechanism for MTase mediated decarboxylation via enamine intermediates.

7.3. 5-Nitrouracil as a IDCase Inhibitor

5-Nitrouracil, a potential substrate analogue for carboxyuracil, was shown to work as an inhibitor against IDCase. The crystal structure of IDCase revealed the interactions required for the formation of a substrate binding complex. Considering apo IDCase, Zn^{2+} is coordinated by six ligands in an octahedral geometry including two water molecules. When the Asp ligand coordinates to uracil, Zn^{2+} maintains its coordination with four residues but Wat is replaced by O4 of uracil resulting in a distorted square pyramidal geometry. When bound to 5-nitrouracil, both positions of water molecules are replaced due to a second coordination to the other oxygen from nitro group. In case of the uracil binding complex, its pyrimidine ring is sandwiched between the side chains of two phenylalanines (Phe). One Phe builds up an edge-to-face aromatic interaction and the other Phe creates a parallel ring stacking π-π interaction. Two hydrogen bonds are formed between both the N1 and O2 of the pyrimidine and the respective side chain of asparagine (Asn).The C4 carbonyl carbon is coordinated

to Zn^{2+} and is further stabilized by three hydrogen bonds with two histidine (His) and one aspartic acid (Asp).[145]

The uracil moiety of the 5-nitrouracil-IDCase complex forms identical interactions with the surrounding residues compared to the carboxyuracil binding complex. Further interactions are formed because of the 5-nitro group. One coordinative bond is formed between the oxygen of the nitro group and Zn^{2+} as well as an additional hydrogen bond with the side chain of His. The second oxygen of the nitro group forms two hydrogen bonds, one with the arginine (Arg) and a second one with a tyrosine (Tyr) activated water molecule (Figure 9).[145] These extra interactions indicate that carboxyuracil and 5-nitrouracil bind tighter to the complex than uracil which in consequence is an advantage for releasing decarboxylated uracil and reloading new carboxyuracil.

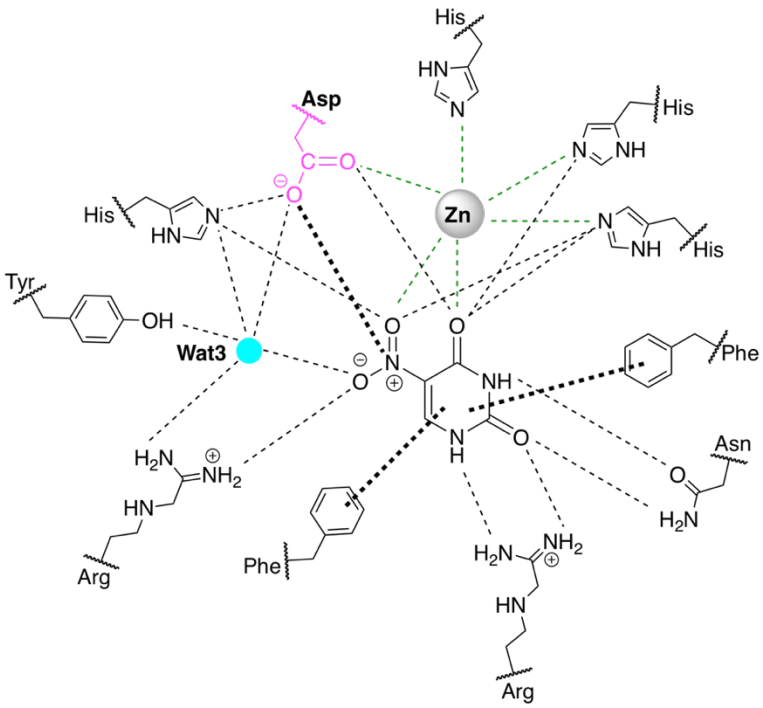


Figure 9. IDCase–nitrouracil complex.

The mechanism of this decarboxylase is suggested to proceed via the vinyl anion type intermediate discussed previously. The inhibitory effect of the nitro group is based on its lower electrophilicity compared to the carbonyl carbon of the carboxyl group in carboxyuracil. Therefore, the nucleophilic attack at N of the nitro group is prevented and a nitrogroup cleavage cannot be performed as a

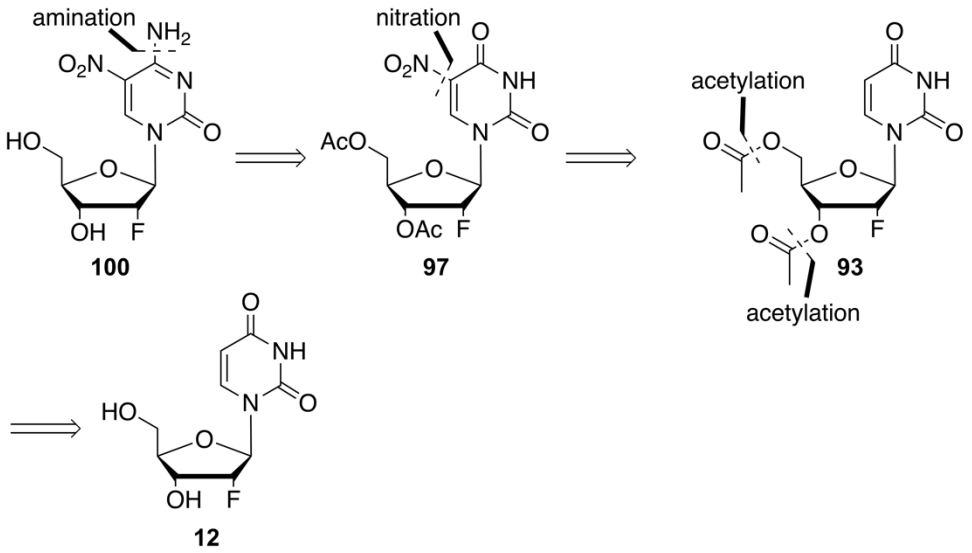
comparable C–C bond cleavage would occur in case of carboxyuracil. Since the nitro group was shown to inhibit decarboxylation of carboxyuracil, it was assumed that the nucleophilic attack cannot occur at other positions like at C6 position. Otherwise it would expose the bond between the C5 position and the functional group to electrophile and thereby lead to C–C bond cleavage. This would designate both carboxyuracil and 5-nitrouracil as substrates.[145] However, when the cadC is considered as a substrate in decarboxylation reaction in mammalian cells, it is bound to DNA which means that also the potential decarboxylase should contain DNA binding domain and be larger in size. Nevertheless, these findings shed a light on the existence of potential DNA decarboxylase and provide useful hints for the search.[145]

8. Aim of the Work

Since the entity and circumstances behind an active demethylation process that occurs via a direct C–C bond cleavage are unclear, it is a great deal of importance and a major research goal to identify the missing details. One way to study the mechanism behind this pathway is to find an inhibitor that suppresses deformylation and/or decarboxylation reactions. Such compound could give valuable information about the mechanism behind the deformylation and decarboxylation reactions as well as about the corresponding enzyme.

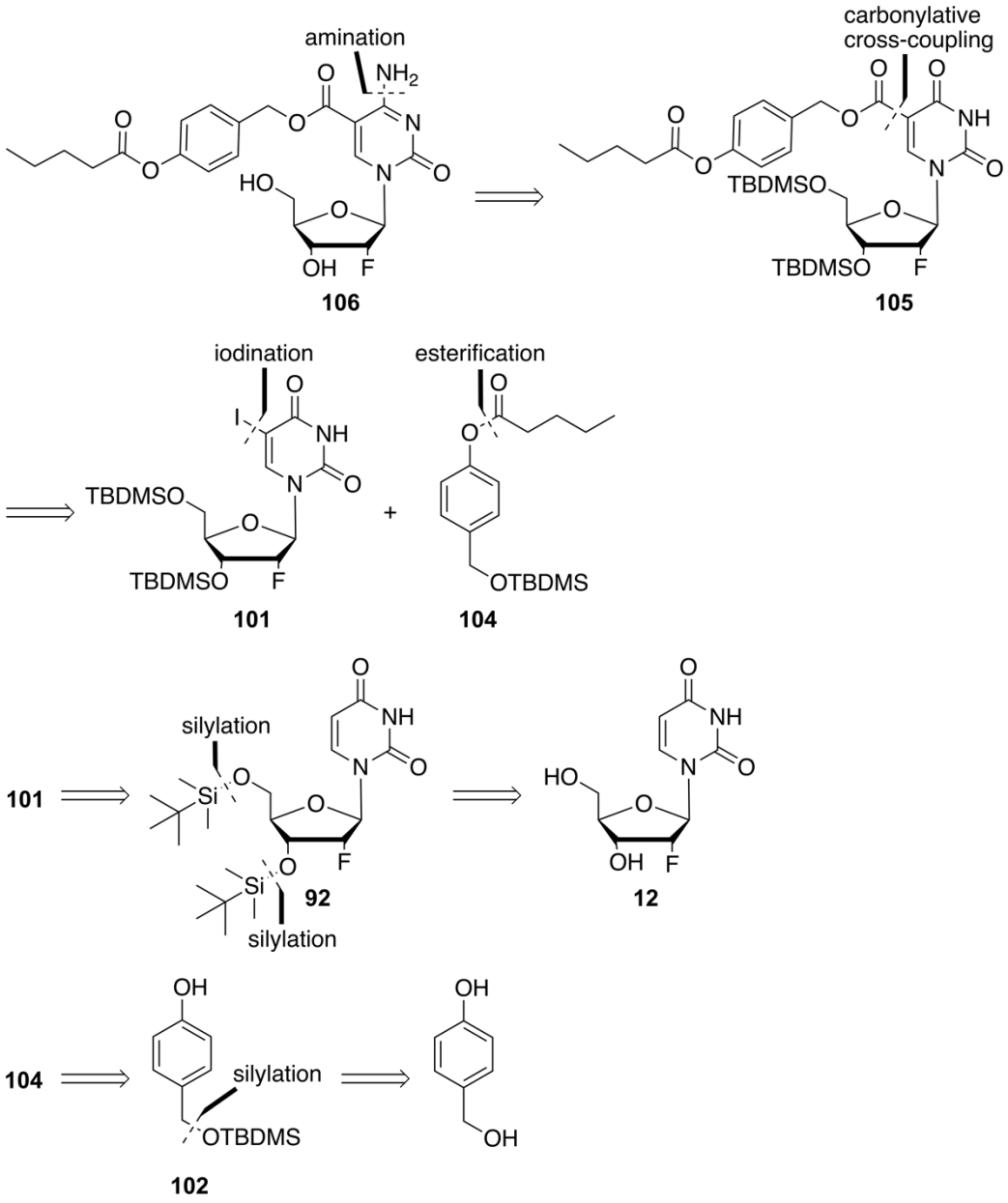
Because of its electronic properties of the nitro group has been recognized as a biologically important functionality. NO_2 has been shown to have inhibitory effects on viral agents, for example, it has been attributed to the inhibition of thymidylate synthetase. Especially noteworthy is, that it has been shown to inhibit the enzyme IDCase that catalyzes decaboxylation of carboxyuracil to uracil in fungi. Therefore, it is imaginable that this functionality could be a potential inhibitor in deformylation or decarboxylation process in mammal cells. Furthermore, the target molecule of this work, 5-nitro-2'-(*R*)-fluoro-2'-deoxycytidine, is an analogue for cadC and therefore expected to fit well in cell feeding experiments.

In the retrosynthetic point of view (Scheme 57), the target molecule **100** can be obtained from the nitrated uridine derivative **97** via amination. **97** can be synthesized from the unfunctionalized dU **93** via an electrophilic nitration which requires an acetylated sugar moiety that can be obtained from **12**.



Scheme 57. Synthesis route used to synthesize 5-nitro-2'-(*R*)-fluoro-2'-deoxycytidine presented retrosynthetically.

To study decarboxylation of cadC as a potential pathway for active demethylation, a 2'-F-cadC derivative, 5-(butyl-4-acetoxy-benzoate)-2'-fluoro-2'-deoxycytidine (**106**) should be synthesized additionally. This compound has two ester groups which are suggested to get cleaved gradually before incorporation into DNA. This compound would provide a tool to study the C–C bond cleavage process together with **100**. In the retrosynthetic approach to this target molecule the final compound **106** can be obtained by amination of the dU derivative **105** (Scheme 58). **105** can be obtained in a carbonylative cross-coupling that introduces ester function at C5 of the uridine derivative **101**. The coupling partner for the C–C cross coupling reaction, **104**, could be obtained by esterification of the alcohol **102** and an acyl chloride. Introducing the ester functionality to **104** can be achieved after selectively silylating the primary alcohol of **102** leaving the secondary alcohol available for the reaction.



Scheme 58. Retrosynthetic analysis of the target **106**.

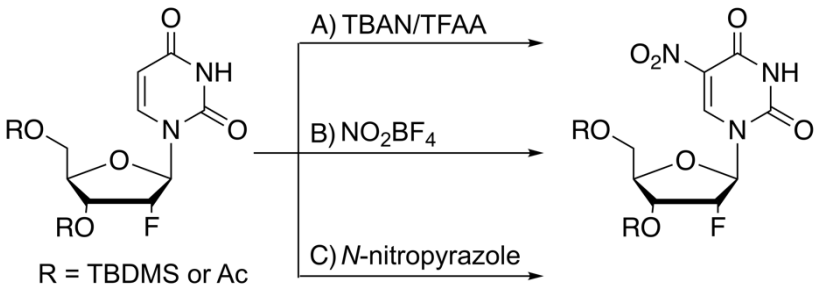
The final goal is to use these fluorine-labelled compounds to investigate the active DNA demethylation that occurs via direct C–C bond cleavage and furthermore the potential inhibitory properties of the synthesized 5-nitro-2'-(*R*)-fluoro-2'-deoxycytidine. The medium of cultured mammalian stem cells would be supplemented with these synthesized compounds in order to metabolically integrate them into their genome. The DNA would be isolated, digested into individual nucleosides and obtained nucleoside mixture would be analyzed by UHPLC-MS/MS.

9. Results and Discussion

The purpose of the target compounds **100** and **106** is to facilitate as tools to study active DNA demethylation that occurs via direct deformylation of fdC or decarboxylation of cadC. Traces of canonical 2'-deoxycytidine must be avoided, therefore, the synthesis was started from 2'-fluoro-deoxyuridine. Fluorinated derivatives were used since they are known to block the activity of glycosylases and thereby to stabilize the nucleotide during the BER process. Fluorinated compounds are also considered as useful to detect the compounds in the mass spectrometric analysis; the F-atom makes the compound 18 units heavier and changes the retention times slightly. Thus, fluorine causes the glycosidic bond to become more labile in the MS-fragmentation step and therefore providing sharper signals in UHPLC-MS/MS analysis. In general, 2'-(*R*)-configured compounds are well tolerated by cells and do not affect to the biochemical reactions that occur within the cells.[141]

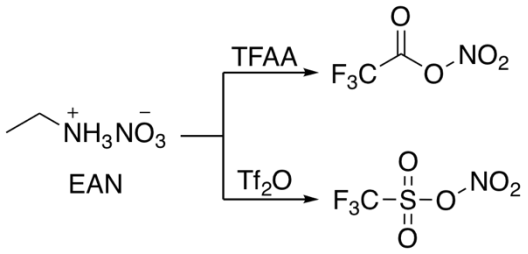
9.1. Nitration of 2'-deoxyuridine

Nitration of the C5 position was attempted with several nitrating agents trying to utilize the electrophilic substitution pathway (Scheme 59). Based on known nitrations targeted from literature, nitration agents NO_2BF_4[252] (pathway B), ethylammonium nitrate (EAN) with TFAA and EAN with Tf_2O[253] (pathway A) and *N*-nitropyrazole[254] (pathway C) been used for nitrating different nucleobases, nucleotides and other heteroaromatic compounds. Nitration experiments and reaction conditions are listed in Table 6.



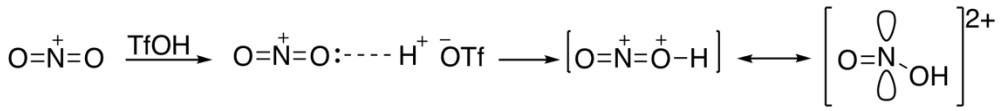
Scheme 59. Attempted nitrations of 2'-fluorinated-dU derivatives.

At first nitration of TBDMS-protected 2'-F-dU was attempted with *in situ* generated nitrating agents trifluoroacetyl nitrate (CF_3COONO_2) and triflyl nitrate ($TfONO_2$.). These are generated from EAN/TFAA and EAN/Tf_2O systems, via salt metathesis using the ionic liquid (IL EAN) as the solvent (Scheme 60). Triftyl nitrate has been described to be extremely efficient for strongly deactivated systems.[253] TBDMS-protected dU **92** was not considered to be highly deactivated; the EAN/TFAA system was used first (entry 1, Table 7). After three hours, TLC control showed full conversion of the starting material, however, the obtained NMR spectra did not show typical nucleoside signals which suggested decomposition of the nucleoside. Nitration with EAN/Tf_2O (entry 2, Table 7) system also led to nucleoside decomposition.



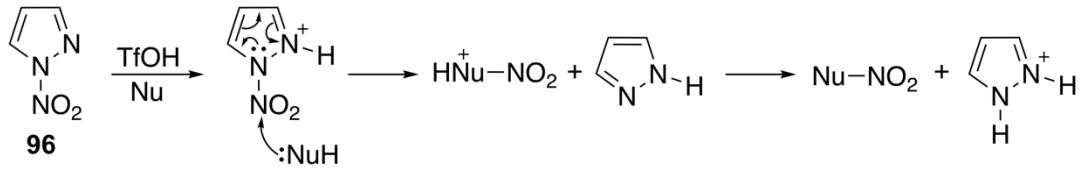
Scheme 60. Generation of nitrating agent via salt metathesis of EAN/TFAA and EAN/Tf_2O.

Further, the commercially available nitration agent NO_2BF_4 in sulfolane was used for the dU derivative. The reaction with **92** was tried first with 2.5 eq. of NO_2BF_4 (entry 3, Table 7). TLC control showed conversion but after the flash chromatographic purification, the obtained product was decomposed. With 1 eq. of NO_2BF_4 no reaction occurred (TLC control) (entry 4, Table 7). To change the reactivity of the nitrating agent, an activating acid is required; the linear nitronium ion, in its ground state, does not have a suitable low lying LUMO to overlap with the HOMO of C5 in **92**. Regardless of formal positive charge at nitrogen, there is no accessible p-orbital. The activation of nitronium ion can be achieved by protonation of the oxygen lone pair by a protic acid like TfOH or by complexation with a Lewis superacid like SbF_5. Protonation would weaken the N–O π-bond character which causes a partial electron deficiency at the p-orbital of N and thereby leads to a non-linear nitronium-ion. This lowers the activation barrier[255] for binding to π donor (Scheme 61).



Scheme 61. Activation of a nitronium ion with TfOH.

N-nitropyrazol has been reported as an efficient transfer nitrating agent for aromatic compounds.[256] It requires the presence of an acid which protonates the nitrogen of the *N*-nitropyrazol and thereby activates the adjacent nitramine group for a nucleophilic displacement (Scheme 62). When using an acid like TfOH, it is possible that *N*-nitropyrazole gets diprotonated making it an effective nitrating agent.



Scheme 62. Acid activation of *N*-nitropyrazole and subsequent nucleophilic displacement.

However, the attempt to nitrate **92** utilizing this method did not succeed (entry 5, Table 7). Protection groups can have an effect on the reactivity of the nucleoside. Electropositive TBDMS groups were thought to make the nucleobase less nucleophilic and the glycosidic bond more unstable, and therefore TBDMS protecting groups were changed to acetyl groups (Ac). Ac groups are more electron withdrawing, thus expected to make the glycosidic bond more stable under the acidic conditions and C5 positon more suitable for electrophilic attack (Figure 10). The acetylation of **12** was achieved with acetic anhydride and by DMAP activation after two hours yielding 93 % of the protected nucleoside **93** (Scheme 63).

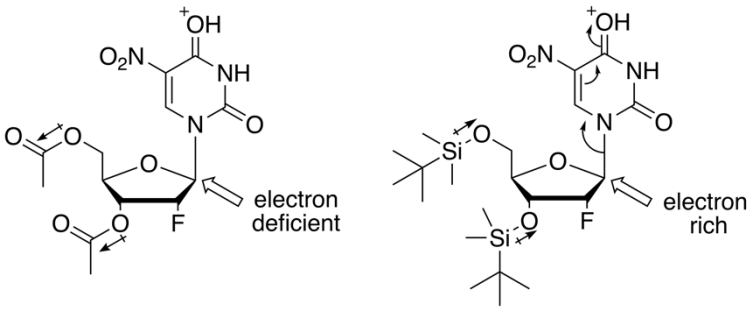
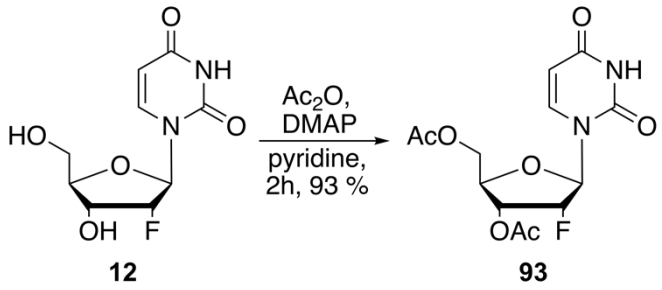


Figure 10. Effect of the protection group on the stability of the glycosidic bond under acidic conditions.

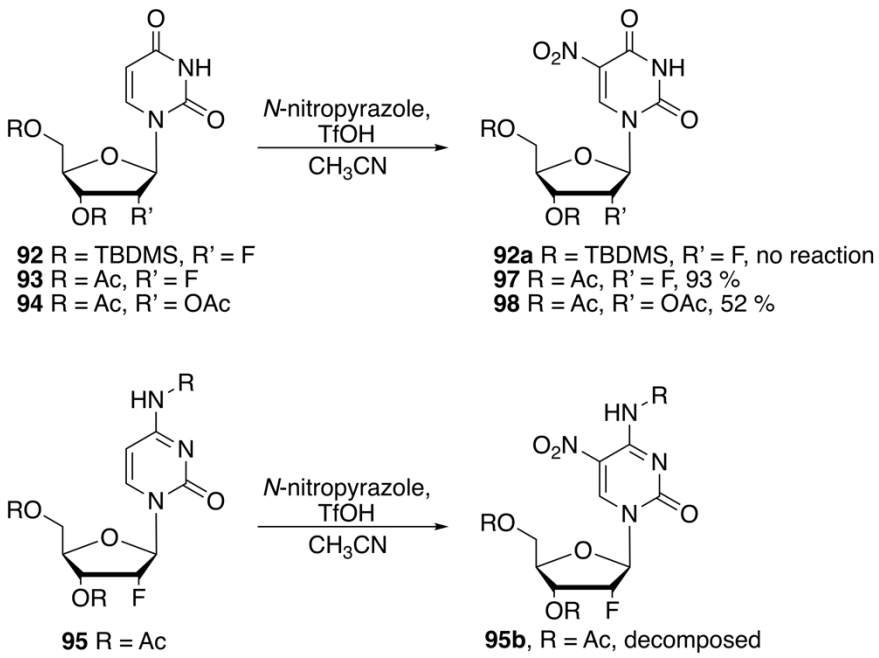


Scheme 63. Acetylation of 2'-fluoro-deoxyuridine.

Entries 5–13 in Table 7 show different reaction conditions for the nitration attempts of **93** with *N*-nitropyrazole. With 1.5 eq. of *N*-nitropyrazole and 1.5 eq. of TfOH no reaction was observed after four days (Scheme 64). The increase in the reaction temperature to 40 °C for 8 h led to conversion of the starting material (TLC). However, the nitrated nucleoside **97** was isolated with a low yield of 30 %. Presence of **97** was confirmed by ^{1}H NMR which showed disappearance of the signal of C5 proton at 5.69 ppm and the C6 signal showed a singled which was shifted downfield from 7.72 ppm to 9.14 ppm. This downfield shift can be explained by decreased shielding of the C6 proton caused by the adjacent NO_2 group. The reaction conditions were further optimized the best reaction conditions, 2.3 eq. TfOH and 2.3 eq. *N*-nitropyrazole at room temperature, yielded the desired product in 93 % yield (entry 11, Table 7).

Additionally, the 2'-fluorinated nucleoside **93** as well as the acetylated uridine derivative **94** were nitrated under the same reaction conditions (entry 12, Table 7) (Scheme 64). After 28 h, **93** was fully converted to the corresponding nitrated nucleoside (entry 10, Table 7), whereas **94** did not show full conversion after the

same reaction time (TLC). Obtained yields for **97** and **98** were 86 % and 52 %, respectively. The different reaction rates can be explained with the different electronic properties of a F-substituent in comparison to a OAc. Same reaction conditions were then applied for the protected dC derivative **95** (Scheme 64), however no nitrated compound was observed. Further analysis suggested that the nucleoside was decomposed from the glycosidic bond. This might be explained by the basic nature of the cytosine ring; highly negative π-charge on N3, is the preferred site for protonation under acidic conditions[173] which changes the electronic properties of the dC derivative thereby interfering the reaction. Also, the acetylated exocyclic amine changes the electron density in the heterocycle.



Scheme 64. Nitration with *N*-nitropyrazole. Reaction conditions and yields listed in table 7.

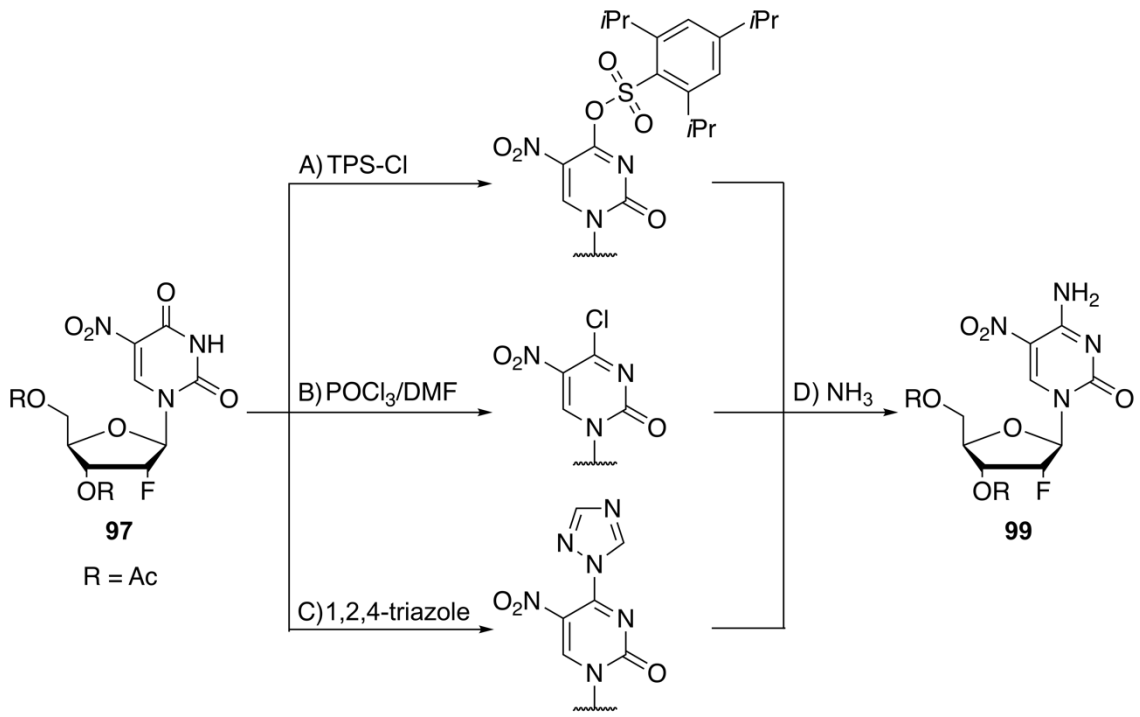
Table 6. Nitration reactions utilizing different nitrating agents.

entry	sm	nitrating agent (eq.)	acid (eq.)	time (h)	temp. (°C)	product, yield (%)
1	**92**	EAN	TFAA	3	0 to 35	decomposed
2	**92**	EAN	Tf_2O	3	0 to 35	decomposed
3	**92**	NO_2BF_4, 2.5	-	16	rt	decomposed
4	**92**	NO_2BF_4, 1.0	-	20	rt	no reaction
5	**92**	*N*-nitropyrazole,1.5	TfOH, 1.5	4 d	rt	**92a,** no reaction
6	**93**	*N*-nitropyrazole,1.5	TfOH, 1.5	4 d	rt	**97**, no reaction
7	**93**	*N*-nitropyrazole,1.5	TfOH, 1.5	3 d	rt to 40	**97**, 30
8	**93**	*N*-nitropyrazole,1.5	TfOH, 2.0	28	rt	**97**, 67
9	**93**	*N*-nitropyrazole,1.5	TfOH, 2.0	28	rt to 40	**97**, 86
10	**93**	*N*-nitropyrazole,1.5	TfOH, 2.2	28	rt	**97**, 89
11	**93**	*N*-nitropyrazole,1.5	TfOH, 2.3	28	rt	**97**, 93
12	**94**	*N*-nitropyrazole,1.5	TfOH, 2.0	28	rt to 40	**98**, 52
13	**95**	*N*-nitropyrazole,1.5	TfOH, 2.0	28	rt	**95b**, decomposed

sm = starting material, rt =room temperature

9.2. Amination of 5-nitro-2'-fluoro-2'-deoxyuridine

The next step in the synthetic route towards the target molecule was the amination of the 5-nitrated dU **97**. Amination strategies using TPS-Cl (A), 1,2,4-triazole (C) and $POCl_3$/DMF (B) followed by amination step (D) (Scheme 65) did not result in the formation of the aminated product **99** (entries 1–6, Table 7). The extremely electron withdrawing nitro group at the C5 position, might be the main reason for the challenging amination step. Earlier reports show that aminations of other dU derivatives, like 3',5'-OTBDMS-5-CD_3-2'-F-dU **47** using 1,2,4-triazole[141] or 3',5'-TBDMS-5-Me-2'-f-dU **72** with TPS-Cl[23] provide intermediates that can be isolated. These methods used with **93** followed by an aqueous work up, yielded the starting material, which suggests an unusual high sensitivity towards water (entries 1-3, 5-8, Table 7).

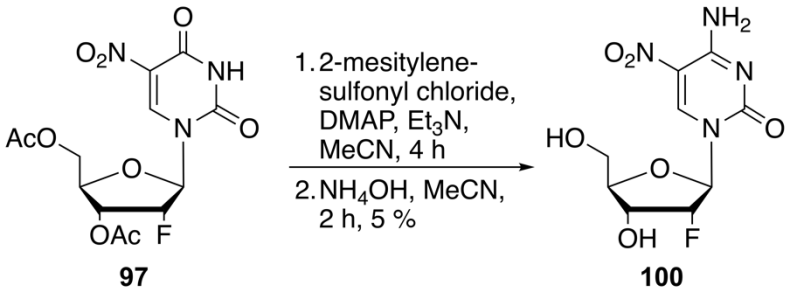


Scheme 65. Attempted aminations of **96**.

The isopropyl groups in TPS-Cl might be too bulky for this reaction, therefore an alternative less bulky reagent 2-mesitylenesulfonyl chloride could show better reactivity. Amination attempts (entries 7–10, Table 7) with 2-mesitylenesulfonyl chloride showed the desired aminated product (regarding to LC-MS), however, despite several purification attempts, no clean product was obtained. The acetyl groups in the sugar moiety are base labile and are getting cleaved during the ammonolysis step, making a part of the nuceosides deacetylated. Therefore, the amount of NH_4OH was increased in order to obtain the fully deprotected product **100**. The amination of **96** to the desired target compound **100** was achieved by tosylating the oxygen at C4 with 2-mesitylenesulfonyl chloride followed by an ammonolysis with NH_4OH. Reaction was performed as a one pot synthesis without purifications of the tosylated compound (Scheme 66). The isolation of the final compound **100** was confirmed with ^1H NMR spectra, which shows signals of the two NH groups of the ketimine form of the nucleoside. These results are supported by HR-MS-ESI analysis.

Low yields obtained from this amination pathway, utilizing 2-mesitylenesulfonyl chloride with DMAP and Et_3N, might evolve from the fact that DMAP as a stronger base than Et_3N, captures HCl that is formed during the reaction, and thereby inhibits the substitution. To increase the yield, 1-methylpiperidine could be used

instead of DMAP and TsCl could be used as an activator for C4 position of **96**. Also, protection groups could be changed to 4-ClBz or tolyol groups which would, yet, establish a stable glycosidic bond under the nitration conditions and furthermore would not be so labile under the amination conditions which deprotect **93**, thus yielding inseparable mixture and thereby causing low yields.



Scheme 66. Amination of 5-nitro-3',5'-di-*O*-acetyl-2'-(*R*)-fluoro-2'-deoxyuridine (**97**) to the final compound **100**.

Table 7. Conducted aminations for 5-nitro-3',5'-di-*O*-acetyl-2'-(*R*)-fluoro-2'-deoxyuridine.

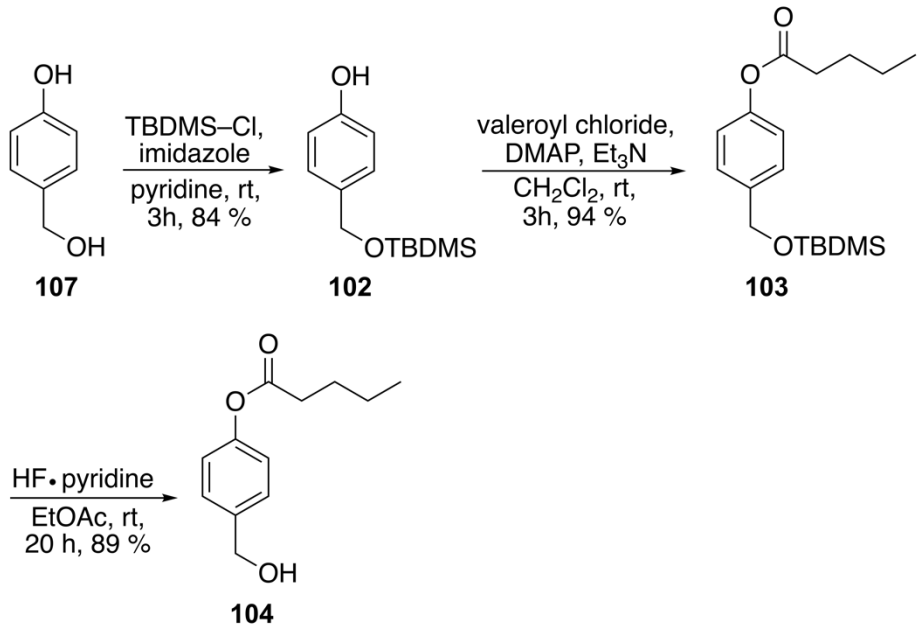
entry	reagent	base	eq.	t (h)	T (°C)	ammono-lysis	t (h)	T (°C)	isolated product
1	1,2,4-triazole, $POCl_3$	Et_3N	8.9	20	0 to rt	NH_4OH/ MeCN, 1:1	2	40	starting mat.
2	1,2,4-triazole, $POCl_3$	Et_3N	8.9	20	0 to rt	NH_3 in MeOH	20	rt	starting mat.
3	1,2,4-triazole, $POCl_3$	Et_3N	9.3	20	0 to rt	NH_3 in MeOH	20	rt	starting mat.
4	$POCl_3$, DMF	Et_3N	2.0	48	0 to rt	NH_3 in MeOH	65	rt to 60	decom-posed
5	TPS-Cl	NaH	5.0	40	0 to rt	NH_3 in MeOH	3	rt	starting mat.
6	TPS-Cl	NaH	5.0	20	0 to 40	NH_3 in MeOH	3h	40	starting mat.
7	2-MSCl	Et_3N,	2.5	3	rt	NH_4OH/ MeCN,	3	rt	starting mat.

entry	reagent	base	eq.	t (h)	T (°C)	ammono-lysis	t (h)	T (°C)	isolated product
8	2-MSCl	Et_3N, DMAP	2.1, 2.1	3	0	NH_4OH/ MeCN, 1:1	3	0 to rt	**99***
9	2-MSCl	Et_3N, DMAP	2.1, 2.1	22	0 to rt	NH_4OH/ MeCN, 1:1	3	0 to rt	**99***
10	2-MSCl	Et_3N, DMAP	10, 2.1	24	0 to 50	NH_4OH/ MeCN, 1:1	2	rt	**99***
11	2-MSCl	Et_3N, DMAP	2.1, 2.1	3	0 to rt	NH_4OH/ MeCN, 2:1	3	0 to rt	**100**, 5%
12	2-MSCl	Et_3N, DMAP	2.1, 2.1	3	rt	NH_4OH/ MeCN, 2.5:1	1.5	rt	**100**, 5%
13	2-MSCl	Et_3N, DMAP	2.1, 2.1	3	rt	NH_4OH/ MeCN,	3	rt	**100**, 7%

2-MSCl = 2-mesitylenesulfonyl chloride, * = inseparable mixture of product and side products

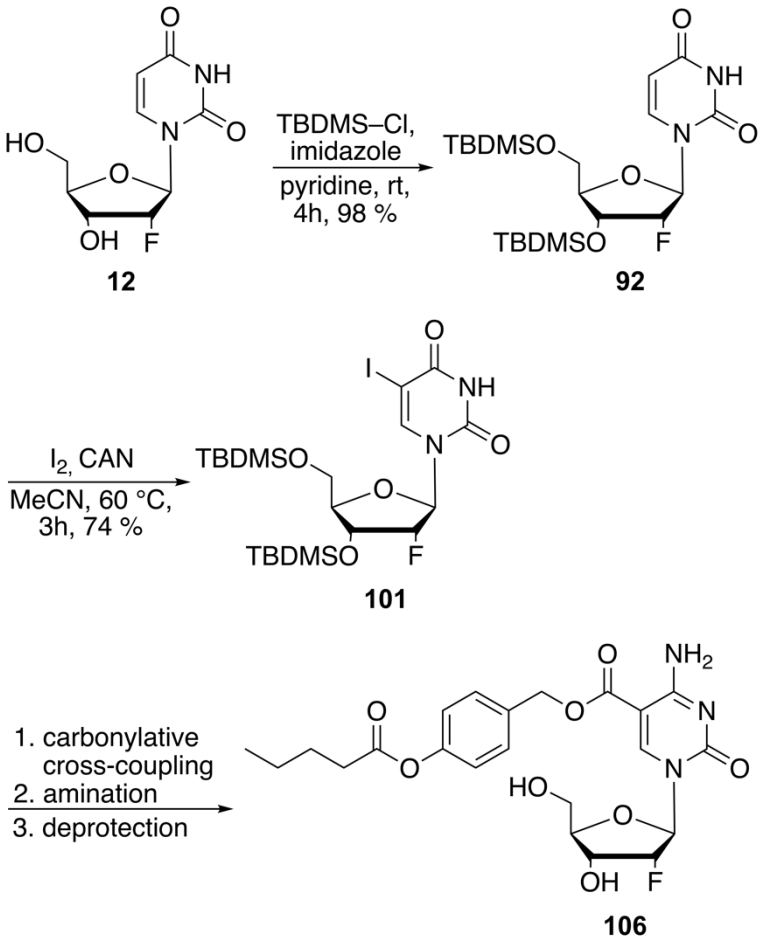
9.3. Synthesis of 4-(hydroxymethyl)phenyl pentanoate and 3',5'-bis-*O*-(*tert*-butyl(dimethyl)silyl)-5-iodo-2'-(*R*)-fluoro-2'-deoxyuridine

For the synthesis of the 5-ester substituted 2'-FdC derivative (5-(butyl-4-acetoxy-benzoate)-2'-fluoro-2'-deoxycytidine) **106**, the corresponding alcohol 4-(hydroxy-methyl)phenyl pentanoate **104** needs to be synthesized for the carbonylative cross coupling. The primary alcohol of 4-hydroxy-benzyl alcohol **107** was selectively silylated yielding **102** in 84 % (Scheme 67). The secondary hydroxyl group was converted into an ester group with an acyl chloride yielding 94 % of **103** which was then deprotected giving **104** in 89 % yield.



Scheme 67. Synthesis of 4-(hydroxymethyl)phenyl pentoate.

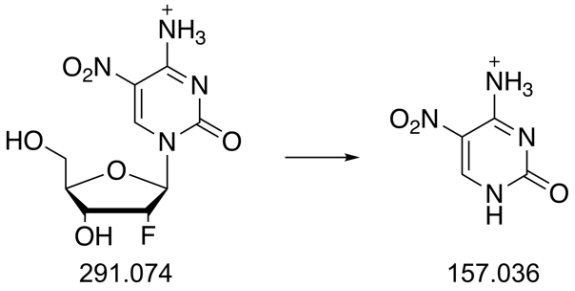
The other precursor required for the synthesis of the nucleoside derivative is thes 5-I-2'-F-dU derivative **101**. For the synthesis of **101**, the free nucleoside **12** was protected with *tert*-butyldimethylsilyl (TBDMS) groups on the sugar moiety (Scheme 68). The TBDMS-protected nucleoside **92** was then regioselectively iodinated at C5 position by an oxidative iodination yielding 74 % of **101**. 5-Iodinated nucleoside 101 is a useful precursor for Pd-catalyzed cross coupling reactions and can be converted to the desired compound **106** in two steps.



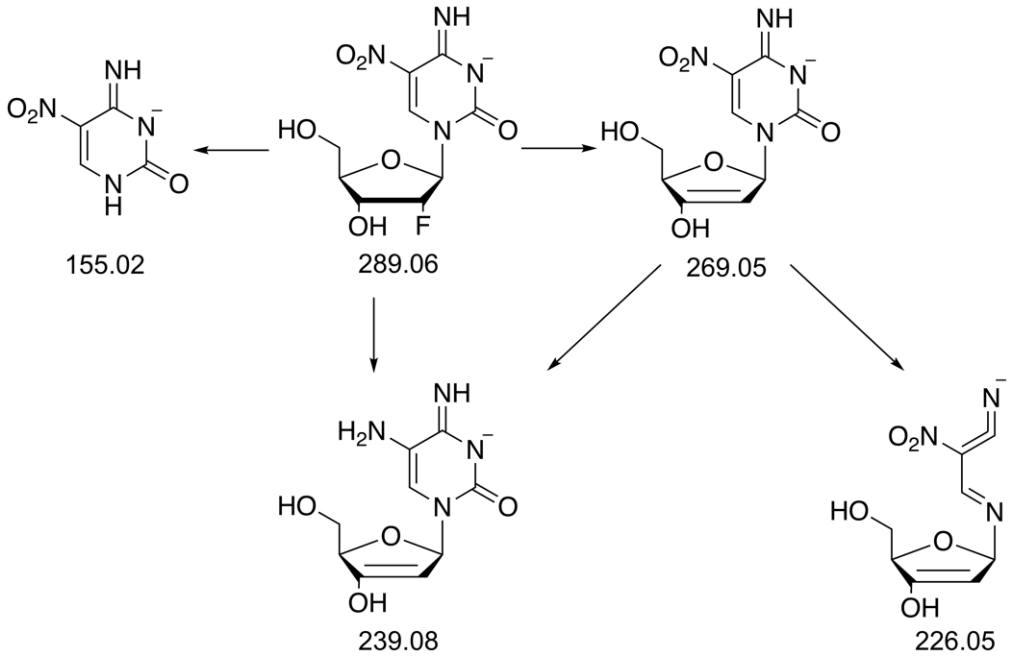
Scheme 68. Iodination of 2'-(*R*)-fluoro-2'-deoxyuridine.

9.4. Developing a Method for UHPLC–MS/MS Analysis

The development of a measuring method for an MS/MS analysis was performed by mass spectrometry specialist Katharina Iwan. Possible fragmentations of 5-nitro-2'-fluoro-2'-deoxycytidine (**100**) were suggested for the positive- and negative fragments which are presented in Scheme 69 and in Scheme 70.



Scheme 69. Possible positive fragmentation pathway.



Scheme 70. Possible fragmentations for the negative mode.

Based on the analysis of the MS patterns of **100,** the quantifier was determined based on the most dominant fragmentation pathway. It was observed for the cleavage of the glycosidic bond which led to the clearly detectable fingerprint mass transition of *m/z* = 289.1 to *m/z* = 154.9 (Figure 11) which was later used for tracking **100**. 5-NO_2-2'-F-dC was fed to mESCs and after 3 d, the genomic DNA was isolated and digested to individual nucleosides using a standard protocol.[141] Obtained nucleoside mixture was analyzed using UHPLC coupled to a triple quadrupole mass spectrometer. However, the mass transition of **100** was not detected and therefore no incorporation into the DNA was observed. One possible explanation for this can be that the phosphokinases in the cells, do not phosphorylate the nucleoside and therefore it cannot become incorporated into the DNA.

Next, the soluble nucleoside/nucleotide pool was analyzed in order to detect **100**. The quantitative data provided by UHPLC-MS/MS showed its incorporation and the signal for **100** at the expected mass transition (*m/z* = 289.1 to *m/z* = 154.9) at a retention time of 5.85 min. The presence of **100** in the soluble pool might indicate this nucleoside to have other biological implications. The Quantitative data of **100** in the soluble nucleoside pool of mESC is presented in Figure 12.

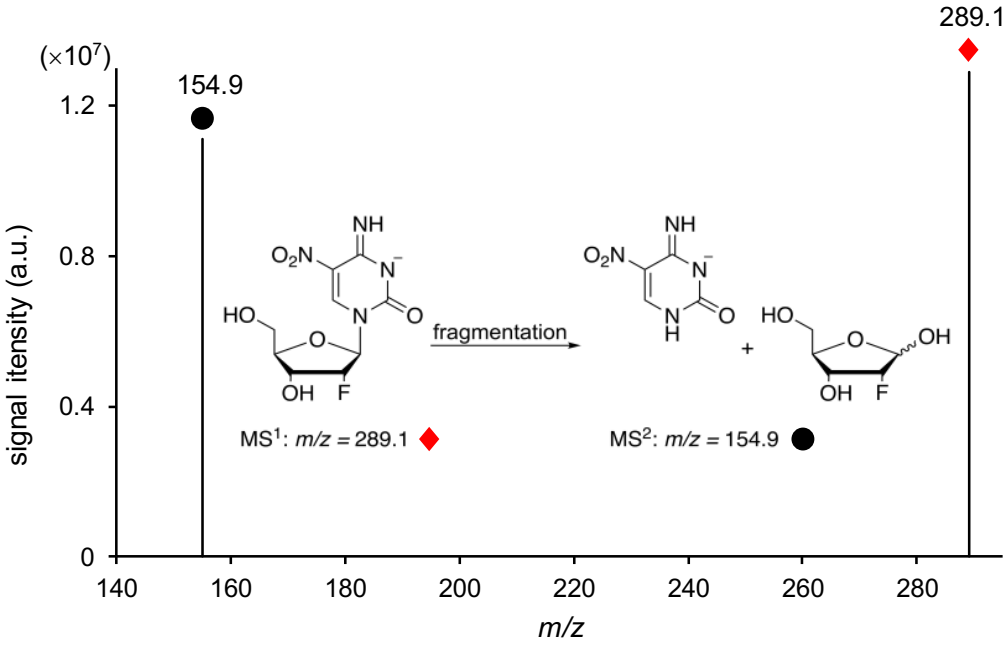


Figure 11. MS pattern of **100** shows that cleavage of the glycosidic bond is the dominant fragmentation pathway.

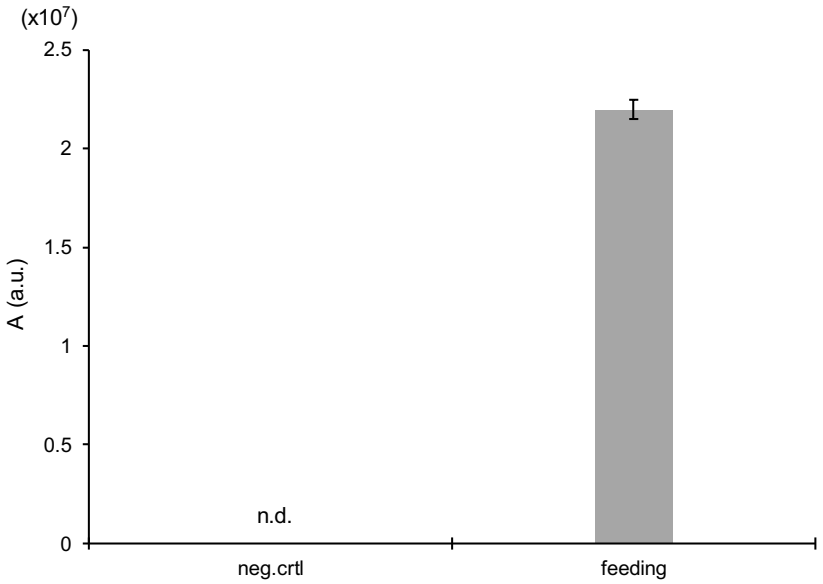


Figure 12. Quantitative data of **100** in the soluble pool after feeding with 500 µM. Mean value and s.d. of technical triplicated from a single culture is shown. a.u. refers to arbitrary units.

10. Conclusions

The main objective of this thesis was the synthesis of 5-nitro-2'-fluoro-2'-deoxycytidine to investigate the active DNA demethylation process in detail. Furthermore, to study the deformylation and decarboxylation reactions, synthesis of 5-(butyl-4-acetoxy-benzoate)-2'-fluoro-2'-deoxycytidine was planned. This compound provides a tool to investigate the potential decarboxylation reaction when incorporated into the DNA of mammalian stem cells. The synthesis for this compound was performed excluding the last two steps.

Utilizing *N*-nitropyrazole as a nitrating agent for 2'-fluoro-2'-deoxyuridine gave the nitrated compound in high yields. However, converting this 5-nitro-2'-deoxyuridine derivative to the corresponding cytidine derivative proved to be challenging and therefore requires further optimization to obtain higher yields. A potential alternative pathway to synthesize 5-nitro-2'-fluoro-2'-deoxycytidine could be a halometallation reaction between C5 iodinated 2'-deoxycytidine derivative and a organolithium reagent. C5-lithiated cytidine derivate would make the C5 position much more nucleophilic and an attack of an electrophile to occur more easily. This would afford a pathway that avoids using strong acids and circumvents the cleavage of the glycosidic bond of the cytidine derivative.

Due to the nature of the nitro group as an inhibitor in many biological processes and as an analogue to the carboxyl group, it was considered an interesting functionality to study active DNA demethylation *in vivo*. 5-Nitro-2'-(*R*)-fluoro-2'-deoxycytidine was synthesized as a potential tool for gaining information about the details of the active DNA demethylation process that occurs via direct C–C bond cleavage. 5-nitro-2'-(*R*)-fluoro-2'-deoxycytidine was added to the cultured medium of mammalian stem cells in order to incorporate it into their genome. After isolation and digestion of the genomic DNA, the obtained nucleoside mixture was analyzed using UHPLC-MS/MS provided quantitative data. Based on the analysis, no incorporation of the target compound into the DNA of the mammalian stem cells was found. However, the target molecule was observed in the soluble pool.

11. Experimental Procedures

11.1. General Information and Methods

Unless otherwise noted all the reactions were performed using flame dried glassware under an atmosphere of argon. Non-aqueous reagents were transferred under argon with a syringe or cannula. Chemicals were purchased from Sigma-Aldrich, Acros and TCI and used without further purification. Solutions were concentrated in vacuo on a Heidolph rotary evaporator. Technical grade solvents were distilled before use. Chromatographic purification of products was accomplished using flash column chromatography on Merck Geduran Si 60 (40 – 63 µm) silica gel (normal phase). Thin layer chromatography (TLC) was performed on Merck 60 (silica gel F254) plates. Visualization of the developed TLC plates was achieved through UV absorption or through staining with Hanessian's stain*, $KMnO_4$**.

NMR spectra were recorded in deuterated solvents on Varian VXR400S, Varian Inova 400, Bruker AMX 600, Bruker Ascend 400 and Bruker Avance III HD and chemical shifts (δ, ppm) were relative to used solvent's residual solvent peak***.[257] Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br. = broad.

High-resolution ESI spectra were obtained on the mass spectrometers Thermo Finnigan LTQ FT-ICR. Theoretic monoisotopic masses *m/z* values were calculated with Chemcalc software[258].

HPLC purifications were performed on a Waters Breeze system (2487 dual array detector; 1525 binary HPLC pump) using a Nucleosil VP 250/10 C18 column from Macherey Nagel, HPLC-grade MeCN was purchased from VWR. Water was

---

* 2.5 g ammonium molybdate, 1 g cerium ammonium sulfate, conc. H_2SO_4, 90 ml H_2O

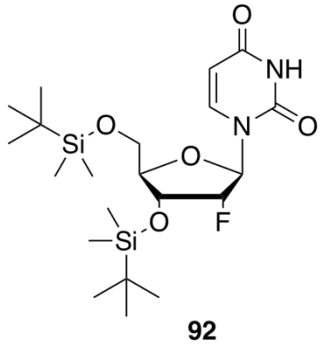
** 1 g $KMnO_4$, 6.7 g K_2CO_3, 1.7 ml 1 M NaOH, 100 ml H_2O

*** ^1H calibration with the respect of ($CDCl_3$), $(CD_3)_2SO$ or D_2O at 7.26 ppm, 2.50 ppm and 4.79 ppm

purified by a Milli-Q Plus system from *Merck Millipore.* ddH_2O refers to double distilled water obtained by MILLI-Q® Plus water purification system from *Millipore* using a QPAK® 2 cartidge.

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. Prior to every measurement series, external calibration curves were measured to quantify the levels of the F-nucleoside. 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). Chromatography was performed by a Poroshell 120 SB-C8 column (Agilent, 2.7 μm, 2.1 mm Å~ 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 effluent up to 1.5 min and after 12 min was diverted to waste by a Valco valve. The autosampler was cooled to 4 °C. The injection volume was 39 μL.[141]

11.2. 3',5'-bis-*O*-(*tert*-butyl(dimethyl)silyl)-2'-(*R*)-fluoro-2'-deoxyuridine[141]



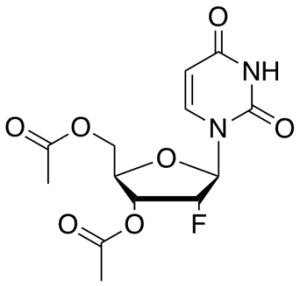
92

2'-deoxy-2'-(*R*)-fluoro-uridine (580 mg, 2.36 mmol, 1.0 eq), TBSCl (1.1 g, 7.07 mmol, 3.0 eq) and imidazole (720 mg, 10,60 mmol, 4.0 eq) were dissolved in pyridine (10 ml) and stirred under argon for 18 h. Pyridine was evaporated under reduced pressure and the residue was poured into sat. aq. $NaHCO_3$ (50 ml) and

extracted with CH_2Cl_2 (3 x 50 ml). Combined organic layers were washed with sat. aq. NaCl (100 ml) and dried over Na_2SO_4. Residual pyridine was coevaporated with toluene (3 x 25 ml). Volatiles were removed under reduced pressure to yield **92** (1.04 g, 2.19 mmol, 98%) as white foam.

R_f (MeOH/DCM 1:4)=0.77; ^{1}H NMR ($CDCl_3$, 7.26 ppm) δ 8.67 (br, 1H, N-H), 7.92 (d, 1H, *J*=8.2 Hz, H-6), 6.06 (dd, 1H, J^1=15.4 Hz, J^2=1.7 Hz, H-1'), 5.69 (dd, J^1=8.2 Hz, J^2=2.6 Hz, 1H, H-5), 4.77 (ddd, J^1=52.3 Hz, J^2=4.3 Hz, J^3=2.0 Hz, 1H, H-2'), 4.29 (ddd, J^1=18.9 Hz, J^2=7.2 Hz, J^3=4.2 Hz, 1H, H-3'), 4.08 (dd, J^1=7.2 Hz, J^2=1.6 Hz, 1H, H-4'), 4.05 (dd, J^1=11.7 Hz, J^2=2.1 Hz, 1H, H-5'), 3.78 (dd, J^1=11.7 Hz, J^2=1.8 Hz, 1H, H-5'), 0.94–0.88 (m, 18H, $Si(CH_3)_2(\underline{t}\text{-Bu})$), 0.13–0.09 (m, 12H, $Si(C\underline{H_3})_2(^t\text{Bu})$); ^{13}C ($CDCl_3$, 77.16 ppm) δ 152.99 (C-4), 149.98 (C-2), 139.86 (C-2), 102.46 (C-5), 93.20 (d, C-2'), 87.97 (C-1'), 87.74 (C-4'), 83.92 (C-1'), 68.72 (C-3'), 60.86 (C-5'), 26.05 ($Si(CH_3)_2\underline{C}(CH_3)_3$), 18.33 ($Si(CH_3)_2C(\underline{CH_3})_3$), 4.89 ($Si(\underline{CH_3})_2C(CH_3)_3$); HRMS ($ESI^+$): *m/z* calc. for ion $C_{21}H_{40}FN_2O_5Si_2^+$ $[M+H]^+$=475.2459; found *m/z*=475.2452, Δ=0.7 mDa; HRMS (ESI^-): *m/z* calc. for ion $C_{21}H_{38}FN_2O_5Si_2^-$ $[M-H]^-$=473.2303; found *m/z*=473.2325, Δ=2.2 mDa.

11.3. 3',5'-di-*O*-acetyl-2'-(*R*)-fluoro-2'-deoxyuridine

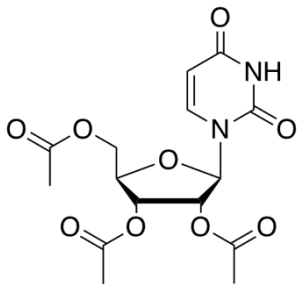


93

To a solution of 2'-deoxy-2'-(*R*)-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) was added Ac_2O (4.4 g, 4.10 ml, 43.0 mmol, 5.3 eq.) 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, crude mixture was purified with flash chromatography (8% MeOH/DCM). Product **93** was obtained as white foam (2.5 g, 7.60 mmol, 93%)

R_f (EtOAc/*i*Hex 1:1)=0.35; ^{1}H NMR ($DMSO_{d6}$): δ 11.49 (s, 1H, N-H), 7.72 (d, 1H, J=8.08 Hz, H-6), 5.87 (dd, 1H, J^1=22.47 Hz, J^2=1.96 Hz, H-1'), 5.68 (dd, 1H, J^1=8.03, J^2=2.20, H-5), 5.53 (ddd, 1H, J^1=54.58 Hz, J^2=5.28, J^3=2.00 Hz, H-2'), 5.26 (ddd, 1H, J^1=13.28 Hz, J^2=7.96 Hz, J^3=2.60 Hz, H-3'), 4.34 (dd, 1H, J^1=11.95, J^2=2.86, H-4'), 4.29–4.25 (m, 1H, H-5'), 4.16 (dd, 1H, J^1=12.08 Hz, J^2=5.71 Hz, H-5'), 2.11 (s, 3H,CO-C$\underline{H_3}$), 2.04 (s, 3H, CO-C$\underline{H_3}$); ^{13}C ($CDCl_3$, 77.16 ppm) δ 170.33 (O$\underline{C}$OCH_3), 169.91 (O$\underline{C}$OCH_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 (OCO$\underline{CH_3}$), 20.54 (OCO$\underline{CH_3}$); HRMS (ESI$^+$): *m/z* calc. for ion $C_{13}H_{16}FN_2O_7$ $[M+H]^+$=331.0942; found *m/z*=331.0936, Δ=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, Δ=0.9 mDa.

11.4. 2',3',5'-tri-*O*-acetyl-uridine



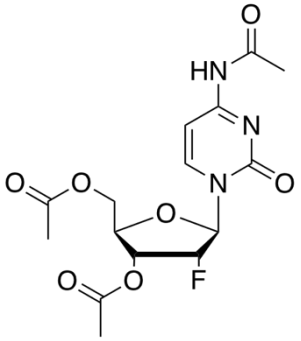
94

2'-deoxy-2'-(*R*)-fluoro-uridine (500 mg, 2.00 mmol, 1.0 eq.) and DMAP (110 mg, 0.90 mmol, 0.45 eq.) were dissolved in pyridine (10 ml) followed by addition of Ac_2O (1.45 g, 1.30 ml, 14.2 mmol, 7.1 eq.). Reaction mixture was stirred untill full conversion at room temperature for 18 h. Pyridine was evaporated under reduced pressure, the residue was diluted with H_2O (20 ml) and extracted with EtOAc (3x20 ml). Combined organic layers were washed with sat. aq. $NaHCO_3$, dried over Na_2SO_4 and volatiles were evaporated under reduced pressure. Crude mixture was purified with flash chromatography (increasing gradient EtOAc/*i*-Hex 60% to 100% EtOAc). Product was obtained as white foam (660 mg, 1.79 mmol, 89%)

R_f (EtOAc/iHex 1:1)=0.22; ^{1}H NMR ($DMSO_{d6}$): δ 11.48 (s, 1H, NH), 7.71 (d, 1H, *J*=7.9 Hz, H-6), 5.89 (d, 1H, *J*=5.1 Hz, H-1'), 5.72 (d, 1H, *J*=8.1 Hz, H-5), 5.34 (dt, 1H, J^1=50.0 Hz, J^2=5.5 Hz, H-2'), 4.31 (dd, 1H, J^1=11.4 Hz, J^2=2.9 Hz, H-4'),

4.27–4.18 (m, 1H, H-3'), 2.08 (s, 3H, C5'-COCO-C$\underline{H_3}$), 2.05 (d, 6H, *J*=2.9, C2'-OCO-C$\underline{H_3}$ +C3'-OCO-C$\underline{H_3}$); HRMS (ESI$^+$): *m/z* calc. for ion $C_{15}H_{19}N_2O_9$ $[M+H]^+$=371.1090; found *m/z*=371.1087, ∆=0.3 mDa; HRMS (ESI$^-$): *m/z* calc. for ion $C_{15}H_{17}N_2O_9$ $[M-H^-]$=369.0934; found *m/z*=369.0944, ∆=1.0 mDa.

11.5. 3',5'-di-*O*-acetyl-2'-(*R*)-fluoro-2'-deoxycytidine

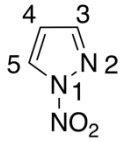


95

2'-deoxy-2'-(*R*)-fluoro-cytidine (500 mg, 2.00 mmol, 1.0 eq.) and DMAP (110 mg, 0.90 mmol, 0.45 eq.) were dissolved in pyridine (10 ml) followed by addition of Ac_2O (1.02 g, 1.00 ml, 10.0 mmol, 5.0 eq.). Reaction mixture was stirred untill full conversion at room temperature for 3h. Pyridine was evaporated, the residue was diluted with H_2O (20 ml) and extracted with EtOAc (3 x 20 ml). Combined organic layers was washed with sat. aq. $NaHCO_3$ (60 ml) and dried over Na_2SO4. Crude mixture was purified with flash chromatography (9 % MeOH/DCM). Product **95** was obtained as white foam (630 mg, 1.92 mmol, 85 %).

R_f (MeOH/DCM 1:9)=0.62; ^{1}H NMR ($DMSO_{d6}$): δ 11.02 (s, 1H, NH), 8.75 (d, 1H, *J*=4.2, H-6), 5.91 (d, 1H, *J*=22.8 Hz, H-5), 5.51 (dd, 1H, J^1=52.3 Hz, J^2=5.5 Hz, H-1'), 5.30 (ddd, 1H, J^1=19.3 Hz, J^2=7.9 Hz, J^3=5.1 Hz, H-2'), 4.39–4.32 (m, 2H, H-4' + H-5'), 4.20 (dd, 1H, J^1=11.5 Hz, J^2=5.2 Hz, H-5'), 2.10 (d, 6H, *J*=1.1 Hz, C3'-OCO-C$\underline{H_3}$ + C5'-OCO-C$\underline{H_3}$), 2.04 (s, 3H, HN-CO-C$\underline{H_3}$)

11.6. *N*-nitropyrazole[259]

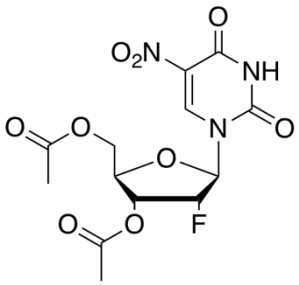


96

Acetyl nitrate: Fuming HNO_3 (0.90 ml, 22.0 mmol) was added dropwise to a stirred solution of Ac_2O (2.10 ml, 22.0 mmol) at 0 °C. Mixture was let to warm up to room temperature. After 30 min, freshly prepared acetyl nitrate (3.00 ml) was added to a solution of pyrazole (500 mg, 7.5 mmol, 1.0 eq.) in AcOH (1.40 ml) and stirred for 1 h. Reaction mixture was poured into ice cold H_2O (20 ml) and filtered through sintteri with H_2O (30 ml) with the help of 450 bar yielding **96** as white crystals (400 mg, 3.57 mmol, 48 %).

R_f (EtOAc/iHex 2:1)=0.77; ^{1}H NMR ($DMSO_{d6}$): δ 8.79 (dd, 1H, J^1=3.00 Hz, J^2=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); EI-GC: *m/z* calc. for $C_3H_3N_2O$=83.0245; found *m/z*=83.0240, Δ=0.5 mDa.

11.7. 5-nitro-3',5'-di-*O*-acetyl-2'-(*R*)-fluoro-2'-deoxyuridine[254]

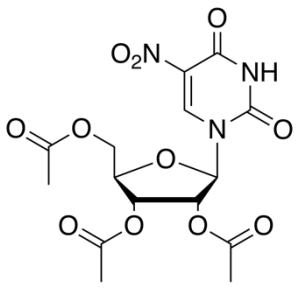


97

TfOH (1.0 g, 610 µl, 6.90 mmol, 2.3 eq.) was added to the solution of **93** (1.0 g, 3.00 mmol, 1.0 eq.) and **96** (510 mg, 4.50 mmol, 1.5 eq.) in MeCN (30 ml) and stirred at room temperature for 15 h. After evaporation of all 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, crude mixture was purified with flash chromatograph (EtOAc/iHex 1:1 increasing the gradient to 2:1) yielding **97** (1.05 g, 2.80 mmol, 93%) as white foam.

R_f (EtOAc/iHex 2:1)=0.59; ^{1}H NMR ($DMSO_{d6}$): δ 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, J^1=51.9 Hz, J^2=5.1 Hz, H-2'), 5.24 (ddd, 1H, J^1=19.6 Hz, J^2=8.4 Hz, J^3=5.1 Hz, H-3'), 4.42–4.35 (m, 2H, H-5' + H-4'),4.25 (dd, 1H, J^1=12.4 Hz, J^2=4.4 Hz, H-5'), 2.11 (s, 3H, C-OCO-C$\underline{H_3}$), 2.04 (s, 3H, C-OCO-C$\underline{H_3}$); ^{13}C ($CDCl_3$, 77.16 ppm) δ 177.95 (O$\underline{C}$OCH_3), 170.09 (O$\underline{C}$OCH_3), 154.73 (C-4), 148.15 (C-2), 144.58 (C-6), 126.10 (C-5), 91.66 (d, C-2'), 79.37 (C-1'), 68.71 (C-4'), 61.09 (C-3'), 60.60 (C-5'), 20.69 (OCO$\underline{CH_3}$), 20.49 (OCO$\underline{CH_3}$); HRMS (ESI$^+$): *m/z* value calc. for ion $C_{13}H_{18}FN_4O_9$ $[M+NH_4]^+$=393.3024; found *m/z*=393.1052, Δ=0.197 Da. HRMS (ESI$^-$): *m/z* calc. for $C_{13}H_{13}FN_3O_9$ $[M-H]^-$=374.0636; found *m/z*=374.0645, Δ=0.9 mDa

11.8. 5-nitro-2',3',5'-tri-*O*-acetyl-uridine[254]



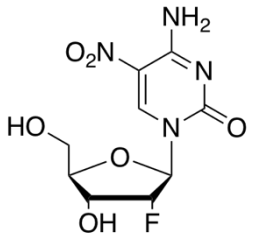
98

TfOH (900 mg, 530 µl, 6.00 mmol, 2.5 eq.) was added to the solution of **94** (1.0g, 3.00 mmol, 1.0 eq.) and **96** (510 mg, 4.50 mmol, 1.5 eq.) in MeCN (30 ml) and stirred at room temperature for 15 h. After evaporation of 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, crude mixture was purified with flash chromatograph (EtOAc/iHex 1:1, increasing the gradient to 2:1) yielding **98** (125 mg, 0.301 mmol, 52 %) as white foam.

R_f (EtOAc/iHex 2:1)=0.41; ^{1}H NMR ($DMSO_{d6}$): δ 12.22 (s, 1H, NH), 9.08 (s, 1H, H-6), 5.95 (d, 1H, *J*=5.4 Hz, H-1'), 5.56 (ddd, 1H, J^1=51.9 Hz, J^2=3.5 Hz, H-2'), 5.38 (t, 1H, *J*=6.3 Hz, H-3'), 4.38–4.32 (m, 2H, H-4' + H-5'), 4.28 (dd, 1H, J^1=12.7 Hz, J^2=5.0 Hz, H-5'), 2.09 (s, 3H, C-OCO-C$\underline{H_3}$), 2.05 (d, 6H, *J*=2.2 Hz, C-OCO-C$\underline{H_3}$+C-OCO-C$\underline{H_3}$); HRMS (ESI$^+$): *m/z* calc. for ion $C_{15}H_{21}N_4O_{11}$

$[M+NH_4]^+$=433.1207; found *m/z*=433.1201, ∆=0,6 mDa; HRMS (ESI^-): *m/z* value calc. for ion $C_{15}H_{16}N_3O_{11}$ $[M-H]^-$=414.0785; found *m/z*=414.0798; ∆=1.3 mDa.

11.9. 5-nitro-2'-(*R*)-fluoro-2'-deoxycytidine

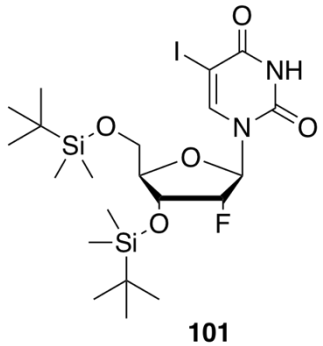


100

To the solution of **97** (100 mg, 0,27 mmol, 1.0 eq.) and DMAP (69 mg, 0.57 mmol, 2.1 eq.) in MeCN (5 ml), Et_3N (57 mg, 80.0 µl, 2.1 eq.) and mesitylene sulfonyl chlorine (88 mg, 0.41 mmol, 1.5 eq.) were added. Reaction mixture was stirred at room temperature for 2 h. A 3:1 mixture of 30-32% NH_4OH/MeCN (4 ml) was added to the reaction mixture and stirring was continued for 2.5 h. Reaction mixture was poured into H_2O (20 ml) and extracted with DCM (3x20 ml). Volatiles were evaporated, and aqueous layer was purified with HPLC (0 to 5 % MeCN/H_2O in 60 min) to give **100** as a white powder.

R_f (MeOH/DCM 1:9)=0.2; ^1H NMR (D_2O): δ 9.69 (s, 1H, H-6), 6.05 (d, 1H, *J*=16.6 Hz, H-1'), 5.14 (dd, 1H, J^1=52.0 Hz, J^2=3.9 Hz, H-2'), 4.33 (ddd, 1H, J^1=25.1 Hz, J^2=9.2 Hz, J^3=4.2 Hz, H-3'), 4.21 (d, 1H, *J*=9.4 Hz, H-4'), 3.99 (ddd, 1H, J^1=15.5 Hz, J^2=13.4 Hz, J^3=2.4 Hz, H-5'); ^{13}C NMR ($DMSO_{d6}$): δ 158.1 (C-4), 153.9 (C-2), 147.3 (C-6), 120.6 (C-4'), 93.6 (d, C-2'), 89.5 (d, C-1'), 82.5 (C-4'), 66.6 (d, C-3'), 58.2 (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, ∆=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, ∆=0.8mDa

11.10. 3',5'-bis-*O*-(*tert*-butyl(dimethyl)silyl)-5-iodo-2'-(*R*)-fluoro-2'-deoxyuridine[141]

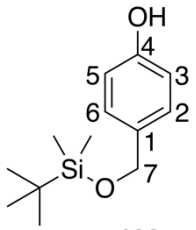


101

CAN (2.8 g, 5.15 mmol, 2.2 eq.) and I_2 (1.3 g, 5.15 mmol, 2.2 eq.) were added to a solution of **92** (1.0 g, 2.34 mmol, 1.0 eq.) in MeCN (100 ml). Brownish reaction mixture was stirred at 60 °C for 1.5 h. Reaction was quenched with sat. aq. $NaHCO_3$ (50 ml) and sat. aq. $Na_2S_2O_3$ (50 ml). After MeCN was evaporated in vacuo, the residue was extracted with EtOAc (3x100 ml). Combined organic layers were washed with sat. NaCl (300 ml) and dried over $MgSO_4$. Crude mixture was purified with flash chromatography (30% EtOAc/iHex) yielding **101** as white foam (1.0 g, 1.71 mmol, 74 %).

R_f (EtOAc/iHex 1:2)=0.3; ^{1}H NMR ($CDCl_3$, 7.26 ppm): δ 8.28 (s, 1H, NH), 7.98 (s, 1H, H-6), 6.09 (dd, J^1=14.3 Hz, J^2=3.8 Hz, 1H, H-1'), 4.82 (dt, J^1= 52.5 Hz, J^2= 4.2 Hz, 1H, H-2'), 4.28 (dt, J^1=12.9 Hz, J^2=4.9 Hz, 1H, H-3'), 4.08–4.06 (m, 1H, H-4'), 3.98 (dd, J^1=11.9 Hz, J^2=1.8 Hz, 1H, H-5'), 3.76 (dd, 1H, J^1=11.9 Hz, J^2= 2.1 Hz, H-5'), 0.97–0.89 (m, 18H, $Si(CH_3)_2(\underline{t\text{-}Bu})$), 0.18–0.09 (m, 12H, $Si(C\underline{H_3})_2(^tBu)$); ^{13}C NMR ($CDCl_3$): δ 159.9 (C-4), 149.8 (C-2), 144.1 (C-6), 93.1 (d,C-2'), 87.6 (dC-1'), 85.3 (C4'), 69.8 (C-3'), 69.3 (C-5'), 61.9 (C-5), 26.4 ($Si(C(CH_3)_2C(\underline{CH_3})_3$), 25.8 ($Si(C(CH_3)_2C(\underline{CH_3})_3$), 18.8 ($Si(CH_3)_2\underline{C}(CH_3)_3$), 18.3 ($Si(C(CH_3)_2\underline{C}(CH_3)_3$), -4.5 ($Si(\underline{CH_3})_2C(CH_3)_3$), -5.0 ($Si(\underline{CH_3})_2C(CH_3)_3$); HRMS ($ESI^+$): calc. for $C_{21}H_{39}FIN_2O_5Si_2^+$ $[M+H]^+$=601.1421, found *m/z*=601.1419, Δ=0.2 mDa; HRMS (ESI^-): calculated for $C_{21}H_{37}FIN_2O_5Si_2^-$ $[M-H]^-$=599.1270, found *m/z*=599.1287, Δ=1.7 mDa. Spectroscopic data corresponds to literature values.[141]

11.11. 4-((*tert*-butyldimethylsilyl)oxy)methyl)phenol

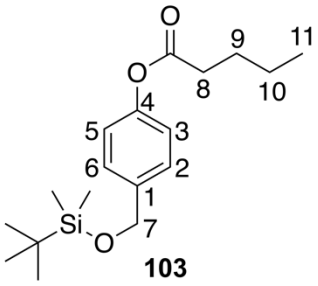


102

4-Hydroxybenzylalcohol (2.0 g, 16.1 mmol, 1.0 eq.), TBSCl (2.7 g, 17.7 mmol, 1.1 eq) and imidazole (2.4 g, 35.4 mmol, 2.2 eq.) were dissolved in DMF (16 ml) and stirred for 20 h. DMF was evaporated and the residue was taken up in sat. aq. $NaHCO_3$ (100 ml). The mixture was extracted with DCM (3x250 ml), combined organic layers were washed with sat. aq. NaCl (250 ml) and dried over Na_2SO_4. Purification of the crude mixture with flash chromatography (11% EtOAc/*i*Hex) yielded **102** (2.8 g, 11.9 mmol, 84 %) as a colorless liquid.

R_f (EtOAc/iHex 1:8)=0.37; ^{1}H NMR ($CDCl_3$, 7.26 ppm): δ 7.19 (m, 2H, H-2+H-6), 6.78 (m, 2H, H-3+H-5), 4.81 (s, 1H, OH), 4.66 (s, 2H, C$\underline{H_2}$-O-Si-$(CH_3)_2$(*t*-Bu)), 0.94–0.91 (m, 9H, Si-$(CH_3)_2$($\underline{t}$-Bu)), 0.10–0.07 (m, 6H, Si-($\underline{CH_3}$)$_2$(*t*-Bu)); GC–EI: *m/z* value calc. for $C_{12}H_{19}O_2Si$ $[M\text{-}CH_3]^-$=223.1154; found *m/z*=223.1147, Δ=0.7 mDa.

11.12. 4-((*tert*-butyldimethylsilyl)oxy)methyl)phenyl pentanoate

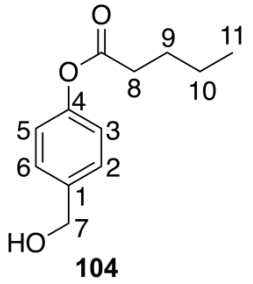


103

Pentanoyl chloride (2.2 g, 1.8 ml, 12.6 mmol, 2.0 eq) was added to a solution of **102** (1.5 g, 6.30 mmol,1.0 eq) in DCM (19 ml). After 4 min Et_3N (color change to white) and DMAP (color change to yellow) were added and the solution was stirred for 3 h. Reaction mixture was taken up with H_2O (50 ml), extracted with CH_2Cl_2 (3 x 60 ml) and dried over Na_2SO_4. Crude mixture was purified with flash chromatograph (8% EtOAc/*i*Hex increasing gradient to 11% EtOAc/*i*Hex) yielding **103** (1.9 g, 5.92 mmol, 94%) as milky liquid.

R_f (EtOAc/iHex 1:6)=0.44 ; ^{1}H NMR ($CDCl_3$, 7.26 ppm): δ 7.33–7.29 (m, 2H, H-2, H-6), 7.05–7.00 (m, 2H, H-3, H-5), 4.72 (s, 2H, H-7), 2.55 (t, 2H, *J*=7.6 Hz, H-8), 1.74 (quin, 2H, *J*=7.6 Hz, H-9), 1.44 (sex, 2H, *J*=7.6 Hz, H-10), 0.97 (t, 3H, *J*=7.3 Hz,H-11), 0.95–0.92 (m, 9H, CH_2-O-Si-(CH_3)$_2$(*t*-Bu)), 0.11–0.08 (m, 6H, CH_2-O-Si-(CH_3)$_2$(*t*-Bu)).

11.13. 4-(hydroxymethyl)phenyl pentanooate



104

In a 50-ml falcon tube containing **103** (450 mg, 1.40 mmol, 1.0 eq.) in EtOAc (14 ml) HF·pyridine (210 mg, 180 µl, 10.5 mmol, 7.5 eq. 70 V-%) was added. After 20 h of stirring, reaction mixture was cooled to 0 °C and quenched with methoxy trimethylsilane (2.34 g, 3.10 ml, 22.4 mmol, 16.0 eq.) and stirred additional hour. Volatiles were evaporated, and crude mixture was purified with flash chromatography (12.5 % EtOAc/iHex, increasing the gradient to 15 %) yielding **104** (250 mg, 1.20 mmol, 86%) as colorless liquid.

R_f (EtOAc/iHex 1:6)=0.6; ^{1}H NMR ($CDCl_3$, 7.26 ppm): δ 7.39–7.34 (m, 2H, H-2, H-6), 7.08–7.03 (m, 2H, H-2, H-5), 4.67 (s, 2H, CH_2-OH), 2.56 (t, 2H, *J*=7.6 Hz, H-8), 1.74 (quin, 2H, *J*=7.6, H-9), 1.45 (sex, 2H, *J*=7.6 Hz, H-10), 0.97 (t, 3H, *J*=7.4, H-11); GC–EI: *m/z* value calc. for $C_7H_8O_2$=124,0524; found *m/z*=124.0519, Δ=0,5 mDa.

12. References

(1) Dahm, R. *Dev. Biol.* **2005**, *278* (2), 274–288.

(2) Adair, G. *Thomas Alva Edison : Inventing the Electric Age: EBSCOhost*; Oxford University Press: New York, 1996.

(3) Avery, O. T.; MacLeod, C. M.; McCarty, M. *Mol. Med.* **1995**, *1* (4), 344–365.

(4) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171* (4361), 964–967.

(5) Franklin, R. E.; Gosling, R. G. *Nature* **1953**, *171* (4356), 740–741.

(6) Franklin, R. E.; Gosling, R. G. *Nature* **1953**, *172* (4369), 156–157.

(7) Wilkins, M. H. F.; Stokes, A. R.; Wilson, H. R. *Nature* **1953**, *171* (4356), 738–740.

(8) Jaenisch, R.; Bird, A. *Nat. Genet.* **2003**, *33* (3s), 245–254.

(9) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell*, 5th ed.; Anderson, M., Granum, S., Eds.; Garland Science, Taylor & Francis Group: New York, 2008.

(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*, 930–935.

(11) Jonathan Clayden, N. G. ja S. W. *Organic Chemistry*, 2nd ed.; Oxford University Press, 2011.

(12) Diehl, A. G.; Boyle, A. P. *Trends Genet.* **2016**, *32* (4), 238–249.

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

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

(15) WYATT, G. R. *Nature* **1950**, *166* (4214), 237–238.

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

(17) Ehrlich, M.; Gama-Sosa, M. A.; Huang, L. H.; Midgett, R. M.; Kuo, K. C.; McCune, R. A.; Gehrke, C. *Nucleic Acids Res.* **1982**, *10* (8), 2709–2721.

(18) Law, J. A.; Jacobsen, S. E. *Nat. Rev. Genet.* **2010**, *11* (3), 204–220.

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

(20) 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.

(21) 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.

(22) Brückl, T.; Globisch, D.; Wagner, M.; Müller, M.; Carell, T. *Angew. Chem. Int. Ed.* **2009**, *48* (42), 7932–7934.

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

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

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

(26) 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. Chem. Int. Ed.* **2015**, *54* (42), 12511–12514.

(27) Condliffe, D.; Wong, A.; Troakes, C.; Proitsi, P.; Patel, Y.; Chouliaras, L.; Fernandes, C.; Cooper, J.; Lovestone, S.; Schalkwyk, L.; Mill, J.; Lunnon, K. *Neurobiol. Aging* **2014**, *35* (8), 1850–1854.

(28) Gackowski, D.; Zarakowska, E.; Starczak, M.; Modrzejewska, M.; Olinski, R. *PLoS One* **2015**, *10* (12), e0144859.

(29) Eleftheriou, M.; Pascual, A. J.; Wheldon, L. M.; Perry, C.; Abakir, A.; Arora, A.; Johnson, A. D.; Auer, D. T.; Ellis, I. O.; Madhusudan, S.; Ruzov, A. *Clin. Epigenetics* **2015**, *7* (1), 88.

(30) Ramsawhook, A.; Lewis, L.; Coyle, B.; Ruzov, A. *Clin. Epigenetics* **2017**, *9* (1), 18.

(31) Maiti, A.; Drohat, A. C. *J. Biol. Chem.* **2011**, *286* (41), 35334–35338.

(32) Wu, X.; Zhang, Y. *Nat. Rev. Genet.* **2017**, *18* (9), 517–534.

(33) Kitsera, N.; Allgayer, J.; Parsa, E.; Geier, N.; Rossa, M.; Carell, T.; Khobta, A. *Nucleic Acids Res.* **2017**, *45* (19), 11033–11042.

(34) 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.

(35) 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.

(36) Zhu, C.; Gao, Y.; Guo, H.; Xia, B.; Song, J.; Wu, X.; Zeng, H.; Kee, K.; Tang, F.; Yi, C. *Cell Stem Cell* **2017**, *20* (5), 720–731.e5.

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

(38) Bachman, M.; Uribe-Lewis, S.; Yang, X.; Burgess, H. E.; Iurlaro, M.; Reik, W.; Murrell, A.; Balasubramanian, S. *Nat. Chem. Biol.* **2015**, *11* (8), 555–557.

(39) Su, M.; Kirchner, A.; Stazzoni, S.; Müller, M.; Wagner, M.; Schröder, A.; Carell, T. *Angew. Chemie Int. Ed.* **2016**, *55* (39), 11797–11800.

(40) Kohli, R. M.; Zhang, Y. *Nature* **2013**, *502* (7472), 472–479.

(41) 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.

(42) An, R.; Jia, Y.; Wan, B.; Zhang, Y.; Dong, P.; Li, J.; Liang, X. *PLoS One* **2014**, *9* (12), e115950.

(43) Duncan, B. K.; Miller, J. H. *Nature* **1980**, *287* (5782), 560–561.

(44) Pullman, A.; Berthod, H. *Int. J. Quantum Chem.* **2009**, *12* (S4), 327–336.

(45) Robertson, A. B.; Klungland, A.; Rognes, T.; Leiros, I. *Cell. Mol. Life Sci.* **2009**, *66* (6), 981–993.

(46) Jacobs, A. L.; Schär, P. *Chromosoma* **2012**, *121* (1), 1–20.

(47) Beard, W. A.; Wilson, S. H. *Chem. Rev.* **2006**, *106* (2), 361–382.

(48) Berti, P. J.; McCann, J. A. B. *Chem. Rev.* **2006**, *106* (2), 506–555.

(49) 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.

(50) Huffman, J. L.; Sundheim, O.; Tainer, J. A. *Mutat. Res. Mol. Mech. Mutagen.* **2005**, *577* (1–2), 55–76.

(51) Lenz, S. A. P.; Kellie, J. L.; Wetmore, S. D. *J. Phys. Chem. B* **2015**, *119* (51), 15601–15612.

(52) Scharer, O. D. *Cold Spring Harb. Perspect. Biol.* **2013**, *5* (10), a012609–a012609.

(53) Bird, A. P. *Nature* **1986**, *321* (6067), 209–213.

(54) Schübeler, D. *Nature* **2015**, *517* (7534), 321–326.

(55) Miranda, T. B.; Jones, P. A. *J. Cell. Physiol.* **2007**, *213* (2), 384–390.

(56) Smith, Z. D.; Meissner, A. *Nat. Rev. Genet.* **2013**, *14* (3), 204–220.

(57) Hermann, A.; Goyal, R.; Jeltsch, A. *J. Biol. Chem.* **2004**, *279* (46), 48350–48359.

(58) Song, J.; Rechkoblit, O.; Bestor, T. H.; Patel, D. J. *Science* **2011**, *331* (6020), 1036–1040.

(59) Okano, M.; Bell, D. W.; Haber, D. A.; Li, E. *Cell* **1999**, *99* (3), 247–257.

(60) Jurkowska, R. Z.; Jurkowski, T. P.; Jeltsch, A. *ChemBioChem* **2011**, *12* (2), 206–222.

(61) Ooi, S. K. T.; O'Donnell, A. H.; Bestor, T. H. *J. Cell Sci.* **2009**, *122* (16), 2787–2791.

(62) Pawlak, M.; Jaenisch, R. *Genes Dev.* **2011**, *25* (10), 1035–1040.

(63) Chuang, L. S.; Ian, H. I.; Koh, T. W.; Ng, H. H.; Xu, G.; Li, B. F. *Science* **1997**, *277* (5334), 1996–2000.

(64) Schermelleh, L.; Haemmer, A.; Spada, F.; Rösing, N.; Meilinger, D.; Rothbauer, U.; Cardoso, M. C.; Leonhardt, H. *Nucleic Acids Res.* **2007**, *35* (13), 4301–4312.

(65) Spada, F.; Haemmer, A.; Kuch, D.; Rothbauer, U.; Schermelleh, L.; Kremmer, E.; Carell, T.; Längst, G.; Leonhardt, H. *J. Cell Biol.* **2007**, *176* (5), 565–571.

(66) Song, J.; Teplova, M.; Ishibe-Murakami, S.; Patel, D. J. *Science* **2012**, *335* (6069), 709–712.

(67) 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.

(68) Jones, P. A.; Liang, G. *Nat. Rev. Genet.* **2009**, *10* (11), 805–811.

(69) Jones, P. A.; Liang, G. *Nat. Rev. Genet.* **2009**, *10* (11), 805–811.

(70) Cedar, H.; Bergman, Y. *Annu. Rev. Biochem.* **2012**, *81* (1), 97–117.

(71) Du, Q.; Wang, Z.; Schramm, V. L. *Proc. Natl. Acad. Sci.* **2016**, *113* (11), 2916–2921.

(72) Wu, J. C.; Santi, D. V. *J. Biol. Chem.* **1987**, *262* (10), 4778–4786.

(73) Linscott, J. A.; Kapilashrami, K.; Wang, Z.; Senevirathne, C.; Bothwell, I. R.; Blum, G.; Luo, M. *Proc. Natl. Acad. Sci.* **2016**, *113* (52), E8369–E8378.

(74) PAULING, L. *Chem. Eng. News* **1946**, *24* (10), 1375–1377.

(75) Cleland, W. W. *Methods Enzymol.* **1995**, *249*, 341–373.

(76) Schramm, V. L. *Annu. Rev. Biochem.* **2011**, *80* (1), 703–732.

(77) Yang, J.; Lior-Hoffmann, L.; Wang, S.; Zhang, Y.; Broyde, S. *Biochemistry* **2013**, *52* (16), 2828–2838.

(78) Wu, S. C.; Zhang, Y. *Nat. Rev. Mol. Cell Biol.* **2010**, *11* (9), 607–620.

(79) Zheng, G.; Fu, Y.; He, C. *Chem. Rev.* **2014**, *114* (8), 4602–4620.

(80) 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.

(81) Gu, T.-P.; Guo, F.; Yang, H.; Wu, H.-P.; Xu, G.-F.; Liu, W.; Xie, Z.-G.; Shi, L.; He, X.; Jin, S.; Iqbal, K.; Shi, Y. G.; Deng, Z.; Szabó, P. E.; Pfeifer, G. P.; Li, J.; Xu, G.-L. *Nature* **2011**, *477* (7366), 606–610.

(82) Iqbal, K.; Jin, S.-G.; Pfeifer, G. P.; Szabó, P. E. *Proc. Natl. Acad. Sci.* **2011**, *108* (9), 3642–3647.

(83) Wossidlo, M.; Arand, J.; Sebastiano, V.; Lepikhov, K.; Boiani, M.; Reinhardt, R.; Schöler, H.; Walter, J. *EMBO J.* **2010**, *29* (11), 1877–1888.

(84) Rougier, N.; Bourc'his, D.; Gomes, D. M.; Niveleau, A.; Plachot, M.; Pàldi, A.; Viegas-Péquignot, E. *Genes Dev.* **1998**, *12* (14), 2108–2113.

(85) Hill, P. W. S.; Amouroux, R.; Hajkova, P. *Genomics* **2014**, *104* (5), 324–333.

(86) Wu, X.; Inoue, A.; Suzuki, T.; Zhang, Y. *Genes Dev.* **2017**, *31* (5), 511–523.

(87) Pastor, W. A.; Aravind, L.; Rao, A. *Nat. Rev. Mol. Cell Biol.* **2013**, *14* (6), 341–356.

(88) Hu, L.; Li, Z.; Cheng, J.; Rao, Q.; Gong, W.; Liu, M.; Shi, Y. G.; Zhu, J.; Wang, P.; Xu, Y. *Cell* **2013**, *155* (7), 1545–1555.

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

(90) Zhao, B.; Yang, Y.; Wang, X.; Chong, Z.; Yin, R.; Song, S.-H.; Zhao, C.; Li, C.; Huang, H.; Sun, B.-F.; Wu, D.; Jin, K.-X.; Song, M.; Zhu, B.-Z.; Jiang, G.; Rendtlew Danielsen, J. M.; Xu, G.-L.; Yang, Y.-G.; Wang, H. *Nucleic Acids Res.* **2014**, *42* (3), 1593–1605.

(91) Blaschke, K.; Ebata, K. T.; Karimi, M. M.; Zepeda-Martínez, J. A.; Goyal, P.; Mahapatra, S.; Tam, A.; Laird, D. J.; Hirst, M.; Rao, A.; Lorincz, M. C.; Ramalho-Santos, M. *Nature* **2013**, *500* (7461), 222–226.

(92) Minor, E. A.; Court, B. L.; Young, J. I.; Wang, G. *J. Biol. Chem.* **2013**, *288* (19), 13669–13674.

(93) Yin, R.; Mao, S.-Q.; Zhao, B.; Chong, Z.; Yang, Y.; Zhao, C.; Zhang, D.; Huang, H.; Gao, J.; Li, Z.; Jiao, Y.; Li, C.; Liu, S.; Wu, D.; Gu, W.; Yang, Y.-G.; Xu, G.-L.; Wang, H. *J. Am. Chem. Soc.* **2013**, *135* (28), 10396–10403.

(94) Zhang, W.; Xia, W.; Wang, Q.; Towers, A. J.; Chen, J.; Gao, R.; Zhang, Y.; Yen, C.; Lee, A. Y.; Li, Y.; Zhou, C.; Liu, K.; Zhang, J.; Gu, T.-P.; Chen, X.; Chang, Z.; Leung, D.; Gao, S.; Jiang, Y.; Xie, W. *Mol. Cell* **2016**, *64* (6), 1062–1073.

(95) Hajkova, P.; Jeffries, S. J.; Lee, C.; Miller, N.; Jackson, S. P.; Surani, M. A. *Science* **2010**, *329* (5987), 78–82.

(96) Ito, S.; D'Alessio, A. C.; Taranova, O. V.; Hong, K.; Sowers, L. C.; Zhang, Y. *Nature* **2010**, *466* (7310), 1129–1133.

(97) Wossidlo, M.; Nakamura, T.; Lepikhov, K.; Marques, C. J.; Zakhartchenko, V.; Boiani, M.; Arand, J.; Nakano, T.; Reik, W.; Walter, J. *Nat. Commun.* **2011**, *2*, 241.

(98) Gu, T.-P.; Guo, F.; Yang, H.; Wu, H.-P.; Xu, G.-F.; Liu, W.; Xie, Z.-G.; Shi, L.; He, X.; Jin, S.; Iqbal, K.; Shi, Y. G.; Deng, Z.; Szabó, P. E.; Pfeifer, G. P.; Li, J.; Xu, G.-L. *Nature* **2011**, *477* (7366), 606–610.

(99) Valinluck, V.; Tsai, H.-H.; Rogstad, D. K.; Burdzy, A.; Bird, A.; Sowers, L. C. *Nucleic Acids Res.* **2004**, *32* (14), 4100–4108.

(100) Otani, J.; Kimura, H.; Sharif, J.; Endo, T. A.; Mishima, Y.; Kawakami, T.; Koseki, H.; Shirakawa, M.; Suetake, I.; Tajima, S. *PLoS One* **2013**, *8* (12), e82961.

(101) Giehr, P.; Kyriakopoulos, C.; Ficz, G.; Wolf, V.; Walter, J. *PLOS Comput. Biol.* **2016**, *12* (5), e1004905.

(102) Ji, D.; Lin, K.; Song, J.; Wang, Y. *Mol. Biosyst.* **2014**, *10* (7), 1749.

(103) Jüttermann, R.; Li, E.; Jaenisch, R. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91* (25), 11797–11801.

(104) Jones, P. A.; Taylor, S. M. *Cell* **1980**, *20* (1), 85–93.

(105) Wuts, P. G. M. *Greene's Protective Groups in Organic Synthesis*, 5th ed.; Wiley: New Jersey, 2014.

(106) Mayer, W.; Niveleau, A.; Walter, J.; Fundele, R.; Haaf, T. *Nature* **2000**, *403* (6769), 501–502.

(107) Ito, S.; D'Alessio, A. C.; Taranova, O. V.; Hong, K.; Sowers, L. C.; Zhang, Y. *Nature* **2010**, *466* (7310), 1129–1133.

(108) Ficz, G.; Branco, M. R.; Seisenberger, S.; Santos, F.; Krueger, F.; Hore, T. A.; Marques, C. J.; Andrews, S.; Reik, W. *Nature* **2011**, *473* (7347), 398–402.

(109) Tsagaratou, A.; Aijo, T.; Lio, C.-W. J.; Yue, X.; Huang, Y.; Jacobsen, S. E.; Lahdesmaki, H.; Rao, A. *Proc. Natl. Acad. Sci.* **2014**, *111* (32), E3306–E3315.

(110) Pastor, W. A.; Pape, U. J.; Huang, Y.; Henderson, H. R.; Lister, R.; Ko, M.; McLoughlin, E. M.; Brudno, Y.; Mahapatra, S.; Kapranov, P.; Tahiliani, M.; Daley, G. Q.; Liu, X. S.; Ecker, J. R.; Milos, P. M.; Agarwal, S.; Rao, A. *Nature* **2011**, *473* (7347), 394–397.

(111) Wu, H.; D'Alessio, A. C.; Ito, S.; Wang, Z.; Cui, K.; Zhao, K.; Sun, Y. E.; Zhang, Y. *Genes Dev.* **2011**, *25* (7), 679–684.

(112) Szulwach, K. E.; Li, X.; Li, Y.; Song, C.-X.; Han, J. W.; Kim, S.; Namburi, S.; Hermetz, K.; Kim, J. J.; Rudd, M. K.; Yoon, Y.-S.; Ren, B.; He, C.; Jin, P. *PLoS Genet.* **2011**, *7* (6), e1002154.

(113) Yu, M.; Hon, G. C.; Szulwach, K. E.; Song, C.-X.; Zhang, L.; Kim, A.; Li, X.; Dai, Q.; Shen, Y.; Park, B.; Min, J.-H.; Jin, P.; Ren, B.; He, C. *Cell* **2012**, *149* (6), 1368–1380.

(114) Stroud, H.; Feng, S.; Morey Kinney, S.; Pradhan, S.; Jacobsen, S. E. *Genome Biol.* **2011**, *12* (6), R54.

(115) Hon, G. C.; Song, C.-X.; Du, T.; Jin, F.; Selvaraj, S.; Lee, A. Y.; Yen, C.; Ye, Z.; Mao, S.-Q.; Wang, B.-A.; Kuan, S.; Edsall, L. E.; Zhao, B. S.; Xu, G.-L.; He, C.; Ren, B. *Mol. Cell* **2014**, *56* (2), 286–297.

(116) Hashimoto, H.; Hong, S.; Bhagwat, A. S.; Zhang, X.; Cheng, X. *Nucleic Acids Res.* **2012**, *40* (20), 10203–10214.

(117) 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.

(118) Weber, A. R.; Krawczyk, C.; Robertson, A. B.; Kuśnierczyk, A.; Vågbø, C. B.; Schuermann, D.; Klungland, A.; Schär, P. *Nat. Commun.* **2016**, *7*, 10806.

(119) Malik, S. S.; Coey, C. T.; Varney, K. M.; Pozharski, E.; Drohat, A. C. *Nucleic Acids Res.* **2015**, *43* (19), 9541–9552.

(120) Maiti, A.; Michelson, A. Z.; Armwood, C. J.; Lee, J. K.; Drohat, A. C. *J. Am. Chem. Soc.* **2013**, *135* (42), 15813–15822.

(121) Wossidlo, M.; Arand, J.; Sebastiano, V.; Lepikhov, K.; Boiani, M.; Reinhardt, R.; Schöler, H.; Walter, J. *EMBO J.* **2010**, *29* (11), 1877–1888.

(122) 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.

(123) Schomacher, L.; Han, D.; Musheev, M. U.; Arab, K.; Kienhöfer, S.; von Seggern, A.; Niehrs, C. *Nat. Struct. Mol. Biol.* **2016**, *23* (2), 116–124.

(124) Gehring, M.; Reik, W.; Henikoff, S. *Trends Genet.* **2009**, *25* (2), 82–90.

(125) Morgan, H. D.; Dean, W.; Coker, H. A.; Reik, W.; Petersen-Mahrt, S. K. *J. Biol. Chem.* **2004**, *279* (50), 52353–52360.

(126) Schutsky, E. K.; Nabel, C. S.; Davis, A. K. F.; DeNizio, J. E.; Kohli, R. M. *Nucleic Acids Res.* **2017**, *45* (13), 7655–7665.

(127) Cortellino, S.; Xu, J.; Sannai, M.; Moore, R.; Caretti, E.; Cigliano, A.; Le Coz, M.; Devarajan, K.; Wessels, A.; Soprano, D.; Abramowitz, L. K.; Bartolomei, M. S.; Rambow, F.; Bassi, M. R.; Bruno, T.; Fanciulli, M.; Renner, C.; Klein-Szanto, A. J.; Matsumoto, Y.; Kobi, D.; Davidson, I.; Alberti, C.; Larue, L.; Bellacosa, A. *Cell* **2011**, *146* (1), 67–79.

(128) Guo, J. U.; Su, Y.; Zhong, C.; Ming, G.; Song, H. *Cell* **2011**, *145* (3), 423–434.

(129) Pfaffeneder, T.; Spada, F.; Wagner, M.; Brandmayr, C.; Laube, S. K.; Eisen, D.; Truss, M.; Steinbacher, J.; Hackner, B.; Kotljarova, O.; Schuermann, D.; Michalakis, S.; Kosmatchev, O.; Schiesser, S.; Steigenberger, B.; Raddaoui, N.; Kashiwazaki, G.; Müller, U.; Spruijt, C. G.; Vermeulen, M.; Leonhardt, H.; Schär, P.; Müller, M.; Carell, T. *Nat. Chem. Biol.* **2014**, *10* (7), 574–581.

(130) Bhutani, N.; Brady, J. J.; Damian, M.; Sacco, A.; Corbel, S. Y.; Blau, H. M. *Nature* **2010**, *463* (7284), 1042–1047.

(131) Nabel, C. S.; Manning, S. A.; Kohli, R. M. *ACS Chem. Biol.* **2012**, *7* (1), 20–30.

(132) Rangam, G.; Schmitz, K.-M.; Cobb, A. J. A.; Petersen-Mahrt, S. K. *PLoS One* **2012**, *7* (8), e43279.

(133) Kumar, R.; DiMenna, L.; Schrode, N.; Liu, T.-C.; Franck, P.; Muñoz-Descalzo, S.; Hadjantonakis, A.-K.; Zarrin, A. A.; Chaudhuri, J.; Elemento, O.; Evans, T. *Nature* **2013**, *500* (7460), 89–92.

(134) Bhutani, N.; Brady, J. J.; Damian, M.; Sacco, A.; Corbel, S. Y.; Blau, H. M. *Nature* **2010**, *463* (7284), 1042–1047.

(135) Foshay, K. M.; Looney, T. J.; Chari, S.; Mao, F. F.; Lee, J. H.; Zhang, L.; Fernandes, C. J.; Baker, S. W.; Clift, K. L.; Gaetz, J.; Di, C.-G.; Xiang, A. P.; Lahn, B. T. *Mol. Cell* **2012**, *46* (2), 159–170.

(136) Franchini, D.-M.; Chan, C.-F.; Morgan, H.; Incorvaia, E.; Rangam, G.; Dean, W.; Santos, F.; Reik, W.; Petersen-Mahrt, S. K. *PLoS One* **2014**, *9* (7), e97754.

(137) Santos, F.; Peat, J.; Burgess, H.; Rada, C.; Reik, W.; Dean, W. *Epigenetics Chromatin* **2013**, *6* (1), 39.

(138) Münzel, M.; Globisch, D.; Carell, T. *Angew. Chem. Int. Ed.* **2011**, *50* (29), 6460–6468.

(139) Schiesser, S.; Hackner, B.; Pfaffeneder, T.; Müller, M.; Hagemeier, C.; Truss, M.; Carell, T. *Angew. Chem. Int. Ed.* **2012**, *51* (26), 6516–6520.

(140) Liutkevičiūtė, Z.; Kriukienė, E.; Ličytė, J.; Rudytė, M.; Urbanavičiūtė, G.; Klimašauskas, S. *J. Am. Chem. Soc.* **2014**, *136* (16), 5884–5887.

(141) 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.* **2017**.

(142) Markolovic, S.; Leissing, T. M.; Chowdhury, R.; Wilkins, S. E.; Lu, X.; Schofield, C. J. *Curr. Opin. Struct. Biol.* **2016**, *41*, 62–72.

(143) Smiley, J. A.; Kundracik, M.; Landfried, D. A.; Barnes, V. R.; Axhemi, A. A. *Biochim. Biophys. Acta* **2005**, *1723* (1–3), 256–264.

(144) Kooistra, S. M.; Helin, K. *Nat. Rev. Mol. Cell Biol.* **2012**, *13* (5), 297–311.

(145) Xu, S.; Li, W.; Zhu, J.; Wang, R.; Li, Z.; Xu, G.-L.; Ding, J. *Cell Res.* **2013**, *23* (11), 1296–1309.

(146) Aukema, K. G.; Makris, T. M.; Stoian, S. A.; Richman, J. E.; Münck, E.; Lipscomb, J. D.; Wackett, L. P. *ACS Catal.* **2013**, *3* (10), 2228–2238.

(147) Jia, C.; Li, M.; Li, J.; Zhang, J.; Zhang, H.; Cao, P.; Pan, X.; Lu, X.; Chang, W. *Protein Cell* **2015**, *6* (1), 55–67.

(148) Hargrove, T. Y.; Wawrzak, Z.; Liu, J.; Nes, W. D.; Waterman, M. R.; Lepesheva, G. I. *J. Biol. Chem.* **2011**, *286* (30), 26838–26848.

(149) Lepesheva, G. I.; Hargrove, T. Y.; Kleshchenko, Y.; Nes, W. D.; Villalta, F.; Waterman, M. R. *Lipids* **2008**, *43* (12), 1117–1125.

(150) Byler, S.; Goldgar, S.; Heerboth, S.; Leary, M.; Housman, G.; Moulton, K.; Sarkar, S. *Anticancer Res.* **2014**, *34* (3), 1071–1077.

(151) Vogelstein, B.; Kinzler, K. W. *Trends Genet.* **1993**, *9* (4), 138–141.

(152) Monier, R. *Reprod. Nutr. Dev.* **1990**, *30* (3), 445–454.

(153) Sarkar, S.; Horn, G.; Moulton, K.; Oza, A.; Byler, S.; Kokolus, S.; Longacre, M. *Int. J. Mol. Sci.* **2013**, *14* (10), 21087–21113.

(154) Sarkar, S.; Goldgar, S.; Byler, S.; Rosenthal, S.; Heerboth, S. *Epigenomics* **2013**, *5* (1), 87–94.

(155) Ordovás, J. M.; Smith, C. E. *Nat. Rev. Cardiol.* **2010**, *7* (9), 510–519.

(156) Lund, G.; Zaina, S. *Curr. Atheroscler. Rep.* **2011**, *13* (3), 208–214.

(157) Post, W. S.; Goldschmidt-Clermont, P. J.; Wilhide, C. C.; Heldman, A. W.; Sussman, M. S.; Ouyang, P.; Milliken, E. E.; Issa, J. P. *Cardiovasc. Res.* **1999**, *43* (4), 985–991.

(158) Sharma, P.; Kumar, J.; Garg, G.; Kumar, A.; Patowary, A.; Karthikeyan, G.; Ramakrishnan, L.; Brahmachari, V.; Sengupta, S. *DNA Cell Biol.* **2008**, *27* (7), 357–365.

(159) Zhang, D.; Yu, Z.; Cruz, P.; Kong, Q.; Li, S.; Kone, B. C. *Kidney Int.* **2009**, *75* (3), 260–267.

(160) Baccarelli, A.; Wright, R.; Bollati, V.; Litonjua, A.; Zanobetti, A.; Tarantini, L.; Sparrow, D.; Vokonas, P.; Schwartz, J. *Epidemiology* **2010**, *21* (6), 819–828.

(161) Fraga, M. F.; Ballestar, E.; Paz, M. F.; Ropero, S.; Setien, F.; Ballestar, M. L.; Heine-Suner, D.; Cigudosa, J. C.; Urioste, M.; Benitez, J.; Boix-Chornet, M.; Sanchez-Aguilera, A.; Ling, C.; Carlsson, E.; Poulsen, P.; Vaag, A.; Stephan, Z.; Spector, T. D.; Wu, Y.-Z.; Plass, C.; Esteller, M. *Proc. Natl. Acad. Sci.* **2005**, *102* (30), 10604–10609.

(162) Godfrey, K. M.; Sheppard, A.; Gluckman, P. D.; Lillycrop, K. A.; Burdge, G. C.; McLean, C.; Rodford, J.; Slater-Jefferies, J. L.; Garratt, E.; Crozier, S. R.; Emerald, B. S.; Gale, C. R.; Inskip, H. M.; Cooper, C.; Hanson, M. A. *Diabetes* **2011**, *60* (5), 1528–1534.

(163) Milagro, F. I.; Campión, J.; García-Díaz, D. F.; Goyenechea, E.; Paternain, L.; Martínez, J. A. *J. Physiol. Biochem.* **2009**, *65* (1), 1–9.

(164) Ghoshal, K.; Li, X.; Datta, J.; Bai, S.; Pogribny, I.; Pogribny, M.; Huang, Y.; Young, D.; Jacob, S. T. *J. Nutr.* **2006**, *136* (6), 1522–1527.

(165) Hirabayashi, Y.; Gotoh, Y. *Nat. Rev. Neurosci.* **2010**, *11* (6), 377–388.

(166) Meissner, A.; Mikkelsen, T. S.; Gu, H.; Wernig, M.; Hanna, J.; Sivachenko, A.; Zhang, X.; Bernstein, B. E.; Nusbaum, C.; Jaffe, D. B.; Gnirke, A.; Jaenisch, R.; Lander, E. S. *Nature* **2008**, *454* (7205), 766–770.

(167) Zoghbi, H. Y.; Amir, R. E.; Van den Veyver, I. B.; Wan, M.; Tran, C. Q.; Francke, U. *Nat. Genet.* **1999**, *23* (2), 185–188.

(168) Bradshaw, T. K.; Hutchinson, D. W. *Chem. Soc. Rev.* **1977**, *6* (1), 43.

(169) Matsuda, A.; Obi, K.; Miyasaka, T. *Chem. Pharm. Bull. (Tokyo).* **1985**, *33* (6), 2575–2578.

(170) Sung, W. L. *J. Org. Chem.* **1982**, *47* (19), 3623–3628.

(171) Komatsu, H.; Morizane, K.; Toshiyuki Kohno, A.; Tanikawa, H. *Org. Process Res. Dev.* **2004**, *8* (4), 564–567.

(172) Reese, C. B.; Ubasawa, A. *Tetrahedron Lett.* **1980**, *21* (23), 2265–2268.

(173) Kwiatkowski, J. S.; Pullman, B. In *Advances in Heterocyclic Chemistry*; R., K. A., Boulton, A. J., Eds.; 1975; Vol. 18, pp 199–335.

(174) Dale Poulter, C.; Anderson, R. B. *Tetrahedron Lett.* **1972**, *13* (36), 3823–3826.

(175) Pullman, A.; Rossi, M. *Biochim. Biophys. Acta* **1964**, *88*, 211–212.

(176) Pullman, A. In *Stereo- and Theoretical Chemistry*; Springer-Verlag: Berlin/Heidelberg, 1972; pp 45–103.

(177) Dawson, R. M. C.; Elliott, D. C.; Jones K. M. *Data for Biochemical Research*, 3rd ed.; Wood, E., Ed.; Wiley-Blackwell: Oxford, 1987.

(178) Kanazawa, Y.; Ehara, M.; Sommerfeld, T. *J. Phys. Chem. A* **2016**, *120*, 1545–1553.

(179) Winstead, C.; Mckoy, V.; D ', S.; Sanchez, A. *J. Chem. Phys. J. Chem. Phys. J. Chem. Phys. J. Chem. Phys. J. Chem. Phys. J. Chem. Phys. J. Chem. Phys.* **2007**, *127* (125), 85105–244302.

(180) Nunes, F. B.; Bettega, M. H. F.; Sanchez, S. d'Almeida. *J. Chem. Phys.* **2017**, *146* (24), 244314.

(181) Aihara, J. *J. Phys. Chem.* **1999**, *103* (37), 7487–7495.

(182) Gillis, E. P.; Eastman, K. J.; Hill, M. D.; Donnelly, D. J.; Meanwell, N. A. *J. Med. Chem.* **2015**, *58* (21), 8315–8359.

(183) Schröder, A. S.; Kotljarova, O.; Parsa, E.; Iwan, K.; Raddaoui, N.; Carell, T. *Org. Lett.* **2016**, *18* (17), 4368–4371.

(184) Schröder, A. S.; Parsa, E.; Iwan, K.; Traube, F. R.; Wallner, M.; Serdjukow, S.; Carell, T. *Chem. Commun.* **2016**, *52* (100), 14361–14364.

(185) R. Scott Rowland; Taylor, R. *J. Phys. Chem.* **1996**, *100* (18), 7384–7391.

(186) Berger, I.; Tereshko, V.; Ikeda, H.; Marquez, V. E.; Egli, M. *Nucleic Acids Res.* **1998**, *26* (10), 2473–2480.

(187) Williams, A. A.; Darwanto, A.; Theruvathu, J. A.; Burdzy, A.; Neidigh, J. W.; Sowers, L. C. *Biochemistry* **2009**, *48* (50), 11994–12004.

(188) Sinnott, M. L. *Chem. Rev.* **1990**, *90* (7), 1171–1202.

(189) Lee, S.; Bowman, B. R.; Ueno, Y.; Wang, S.; Verdine, G. L. *J. Am. Chem. Soc.* **2008**, *130* (35), 11570–11571.

(190) Dai, Q.; Lu, X.; Zhang, L.; He, C. *Tetrahedron* **2012**, *68* (26), 5145–5151.

(191) York, J. L. *J. Org. Chem.* **1981**, *46* (10), 2171–2173.

(192) Schrammt, V. L.; Horensteins, B. A.; Kline, P. C. *J. Biol. Chem.* **1994**, *269* (28), 18259–18262.

(193) Marquez, V. E.; Tseng, C. K. H.; Mitsuya, H.; Aoki, S.; Kelley, J. A.; Ford, H.; Roth, J. S.; Broder, S.; Johns, D. G.; Driscoll, J. S. *J. Med. Chem.* **1990**, *33* (3), 978–985.

(194) Schaerer, O. D.; Verdine, G. L. *J. Am. Chem. Soc.* **1995**, *117* (43), 10781–10782.

(195) Burr, J. G. In *Advances in the Photochemistry of Nucleic Acid Derivatives*; W.A. Noyes, G. S. Hammond, J. N. Pitts, Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1968; pp 193–299.

(196) Yang, N. C.; Okazaki, R.; Liu, F.-T. *J. Chem. Soc. Chem. Commun.* **1974**, *0* (12), 462.

(197) Mercer, J. R.; Knaus, E. E.; Wiebe, L. I. *J. Med. Chem.* **1987**, *30* (4), 670–675.

(198) Levene, P. A.; La Forge, F. B. *Berichte der Dtsch. Chem. Gesellschaft* **1912**, *45* (1), 608–620.

(199) Beltz, R. E.; Visser, D. W. *J. Am. Chem. Soc.* **1955**, *77* (3), 736–738.

(200) Gislason, K.; Sigurdsson, S. T. *European J. Org. Chem.* **2010**, *2010* (24), 4713–4718.

(201) Goldman, D.; Kalman, T. I. *Nucleosides and Nucleotides* **1983**, *2* (2), 175–187.

(202) Matsuda, A.; Inoue, H.; Ueda, T. *Chem. Pharm. Bull. (Tokyo)*. **1978**, *26* (8), 2340–2345.

(203) Frisch, D. M.; Visser, D. W. *J. Am. Chem. Soc.* **1959**, *81* (7), 1756–1758.

(204) Grover, R. K.; Pond, S. J. K.; Cui, Q.; Subramaniam, P.; Case, D. A.; Millar, D. P.; Wentworth, P. *Angew. Chem. Int. Ed.* **2007**, *46* (16), 2839–2843.

(205) de Kort, M.; de Visser, P. C.; Kurzeck, J.; Meeuwenoord, N. J.; van der Marel, G. A.; Rüger, W.; van Boom, J. H. *European J. Org. Chem.* **2001**, *2001* (11), 2075–2082.

(206) Kumar, R.; Wiebe, L. I.; Knaus Edward E. *Can. J. Chem.* **1994**, *72*, 2005–2010.

(207) Rayala, R.; Wnuk, S. F. *Tetrahedron Lett.* **2012**, *53* (26), 3333–3336.

(208) Cotton, A. F.; Geoffrey, W. *Advanced Inorganic Chemistry*, 3rd ed.; John Wiley & Sons: New York, 1972.

(209) Asakura, J.; Robins, M. J. *J. Org. Chem.* **1990**, *55* (16), 4928–4933.

(210) Hwang, C. H.; Park, J. S.; Won, J. H.; Kim, J. N.; Ryu, E. K. *Arch. Pharm. Res.* **1992**, *15* (1), 69–72.

(211) Ryu, E. K.; Kim, J. N. *Nucleosides and Nucleotides* **1989**, *8* (1), 43–48.

(212) No, Z.; Shin, D.-S.; Song, B. J.; Ahn, M.; Ha, D.-C. *Synth. Commun.* **2000**, *30* (21), 3873–3882.

(213) Fukuhara, T. K.; Visser D. W. *J. Biol. Chem.* **1951**, *190*, 95.

(214) Wang, Z. In *Comprehensive Organic Name Reactions and Reagents*; Wang, Z., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2010.

(215) Sun, Q.; Sun, J.; Gong, S.-S.; Wang, C.-J.; Pu, S.-Z.; Feng, F.-D. *RSC Adv.* **2014**, *4* (68), 36036–36039.

(216) Ahmadian, M.; Bergstrom, D. E. *Nucleosides and Nucleotides* **1998**, *17* (7), 1183–1190.

(217) Hayakawa, H.; Tanaka, H.; Obi, K.; Itoh, M.; Miyasaka, T. *Tetrahedron Lett.* **1987**, *28* (1), 87–90.

(218) Armstrong, R. W.; Gupta, S.; Whelihan, F. *Tetrahedron Lett.* **1989**, *30* (16), 2057–2060.

(219) Alauddin, M. M.; Conti, P. S. *Tetrahedron* **1994**, *50* (6), 1699–1706.

(220) Coe, P. L.; Harnden, M. R.; Jones, A. S.; Noble, S. A.; Walker, R. T. *J. Med. Chem.* **1982**, *25* (11), 1329–1334.

(221) Aso, M.; Kaneko, T.; Nakamura, M.; Koga, N.; Suemune, H. *Chem. Commun.* **2003**, *0* (9), 1094–1095.

(222) Robins, M. J.; Barr, P. J. *Tetrahedron Lett.* **1981**, *22* (5), 421–424.

(223) De Clercq, E.; Descamps, J.; Balzarini, J.; Giziewicz, J.; Barr, P. J.; Robins, M. J. *J. Med. Chem.* **1983**, *26* (5), 661–666.

(224) Luigi A. Agrofoglio; Isabelle Gillaizeau, A.; Saito, Y. *Chem. Rev.* **2003**, *103*, 1875–1916.

(225) Tolman, C. A. *Chem. Rev.* **1977**, *77* (3), 313–348.

(226) Hirota, K.; Kitade, Y.; Kanbe, Y.; Isobe, Y.; Maki, Y. *Synthesis (Stuttg).* **1993**, *1993* (2), 213–215.

(227) Sävmarker, J. Palladium-Catalyzed Carbonylation and Arylation Reactions, Uppsala Universitet, 2012.

(228) Dai, Q.; He, C. *Org. Lett.* **2011**, *13* (13), 3446–3449.

(229) Münzel, M.; Globisch, D.; Trindler, C.; Carell, T. *Org. Lett.* **2010**, *12* (24), 5671–5673.

(230) Münzel, M.; Lischke, U.; Stathis, D.; Pfaffeneder, T.; Gnerlich, F. A.; Deiml, C. A.; Koch, S. C.; Karaghiosoff, K.; Carell, T. *Chem. - A Eur. J.* **2011**, *17* (49), 13782–13788.

(231) Chentsova, A.; Kapourani, E.; Giannis, A. *Beilstein J. Org. Chem.* **2014**, *10*, 7–11.

(232) Nomura, Y.; Haginoya, N.; Ueno, Y.; Matsuda, A. *Bioorg. Med. Chem. Lett.* **1996**, *6* (23), 2811–2816.

(233) Luche, J. L. *J. Am. Chem. Soc.* **1978**, *100* (7), 2226–2227.

(234) Leendert J. van den Bos; Jeroen D. C. Codée; John C. van der Toorn; Thomas J. Boltje; Jacques H. van Boom; Herman S. Overkleeft, A.; Marel, G. A. van der. *Org. Lett.* **2004**, *6* (13), 2165–2168.

(235) Epp, J. B.; Widlanski, T. S. *J. Org. Chem.* **1999**, *64* (1), 293–295.

(236) Mico, A. De; Margarita, R.; Parlanti, L.; Andrea Vescovi, A.; Giovanni Piancatelli. *J. Org. Chem.* **1997**, *62*, 6974–6977.

(237) Tojo, G.; Fernández, M. *Oxidation of primary alcohols to carboxylic acids : a guide to current common practice*; Springer Science & Business Media, 2007.

(238) Guo, P.; Yan, S.; Hu, J.; Xing, X.; Wang, C.; Xu, X.; Qiu, X.; Ma, W.; Lu, C.; Weng, X.; Zhou, X. *Org. Lett.* **2013**, *15* (13), 3266–3269.

(239) Wuts, P. G. M. *Greene's protective groups in organic synthesis*, 5th ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2014.

(240) Steigenberger, B.; Schiesser, S.; Hackner, B.; Brandmayr, C.; Laube, S. K.; Steinbacher, J.; Pfaffeneder, T.; Carell, T. *Org. Lett.* **2013**, *15* (2), 366–369.

(241) Hodge, R. P.; Brush, C. K.; Harris, C. M.; Harris, T. M. *J. Org. Chem.* **1991**, *56* (4), 1553–1564.

(242) Žemlička, J.; Šorm, F. *Collect. Czechoslov. Chem. Commun.* **1965**, *30* (6), 2052–2067.

(243) Awano, H.; Shuto, S.; Miyashita, T.; Ashida, N.; Machida, H.; Kira, T.; Shigeta, S.; Matsuda, A. *Arch. Pharm. (Weinheim).* **1996**, *329* (2), 66–72.

(244) Sekine, M. *J. Org. Chem.* **1989**, *54* (10), 2321–2326.

(245) Shi, J.; Du, J.; Ma, T.; Pankiewicz, K. W.; Patterson, S. E.; Tharnish, P. M.; McBrayer, T. R.; Stuyver, L. J.; Otto, M. J.; Chu, C. K. *Bioorg. Med. Chem.* **2005**, *13* (5), 1641–1652.

(246) Komatsu, H.; Morizane, K.; Toshiyuki Kohno, A.; Tanikawa, H. *Org. Proc. Res. Dev.* **2004**, *8* (4), 564–567.

(247) Hayakawa, H.; Tanaka, H.; Maruyama, Y.; Miyasaka, T. *Chem. Lett.* **1985**, *14* (9), 1401–1404.

(248) Tanaka, H.; Hayakawa, H.; Obi, K.; Tadashi, M. *Tetrahedron* **1986**, *42* (15), 4187–4195.

(249) Chonan, T.; Wakasugi, D.; Yamamoto, D.; Yashiro, M.; Oi, T.; Tanaka, H.; Ohoka-Sugita, A.; Io, F.; Koretsune, H.; Hiratate, A. *Bioorg. Med. Chem.* **2011**, *19* (5), 1580–1593.

(250) Bengt Erik Haug, and; Daniel H. Rich. *Org. Lett.* **2004**, *6* (25), 4783–4786.

(251) Xu, Z.; Hu, W.; Liu, Q.; Zhang, L.; Jia, Y. *J. Org. Chem.* **2010**, *75* (22), 7626–7635.

(252) Huang, G.-F.; Torrence, P. F. *J. Org. Chem.* **1977**, *42* (24), 3821–3824.

(253) Aridoss, G.; Laali, K. K. *J. Org. Chem.* **2011**, *76* (19), 8088–8094.

(254) Jerzy Giziewicz; Stanislaw F. Wnuk; Robins, M. J. *J. Org. Chem.* **1999**, *64* (6), 2149–2151.

(255) Olah, G. A.; Laali, K. K.; Sandford, G. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (15), 6670–6672.

(256) Olah, G. A.; Narang, S. C.; Fung, A. P. *J. Org. Chem.* **1981**, *46* (13), 2706–2709.

(257) Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. *J. Org. Chem.* **1997**, *62* (21), 7512–7515.

(258) Patiny, L.; Borel, A. *J. Chem. Inf. Model.* **2013**, *53* (5), 1223–1228.

(259) Hüttel, R.; Büchele, F. *Chem. Ber.* **1955**, *88* (10), 1586–1590.

13. Appendices

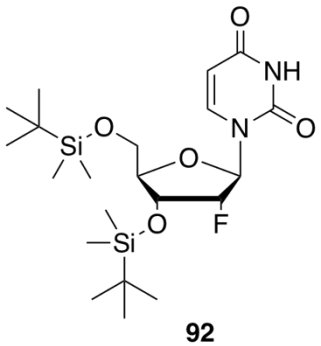
Appendices 1–13: ^{1}H spectra of synthesized compounds **92**–**104**

Appendix 13: ^{13}C NMR spectra of **100**

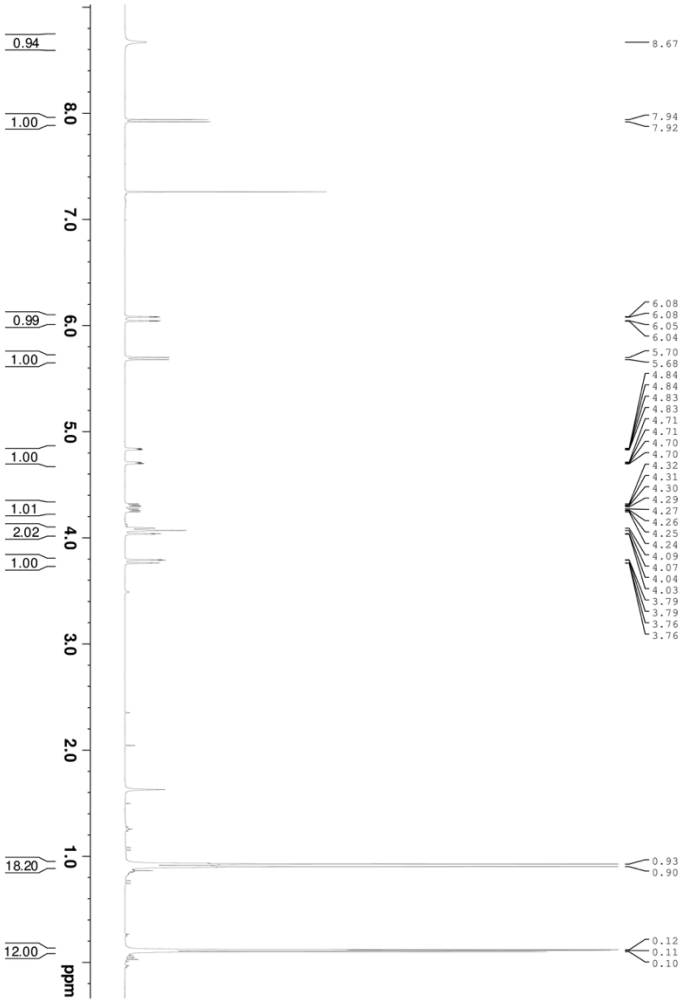
Appendices 14–15: negative and positive product ion scans of **100**

Appendix 16: defining the extinction coefficient

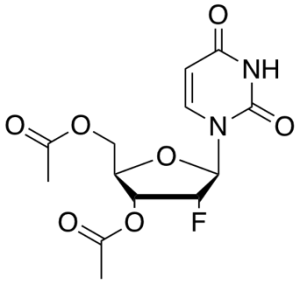
Appendix 1



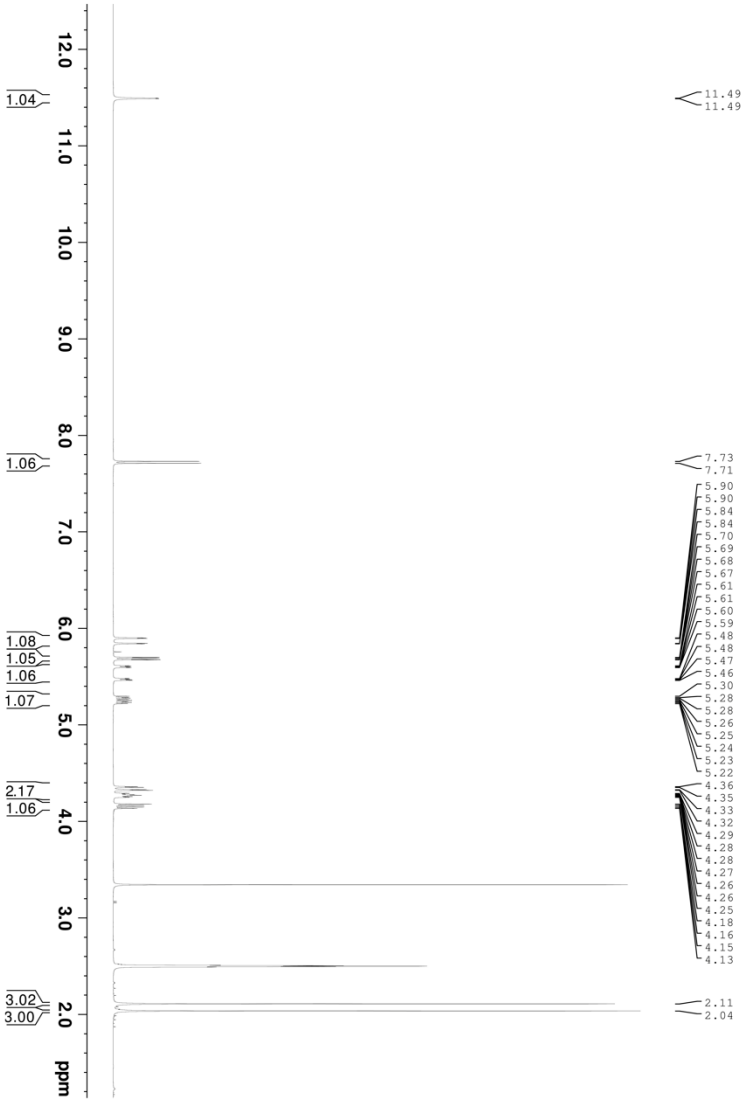
92



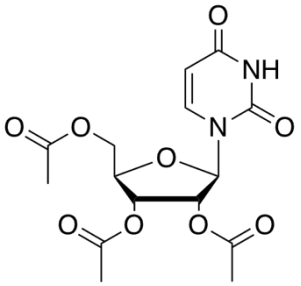
Appendix 2



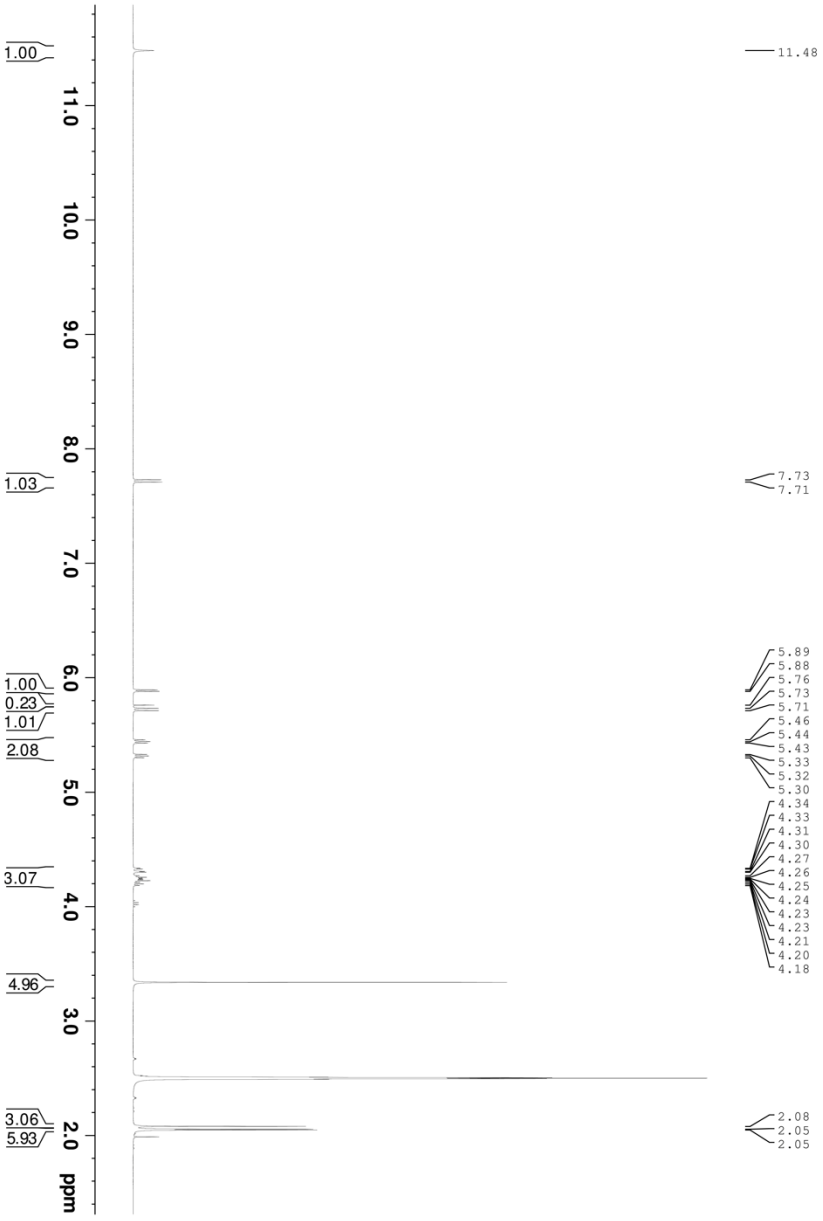
93



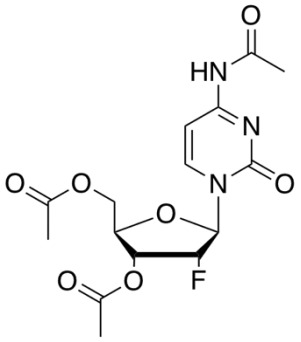
Appendix 3



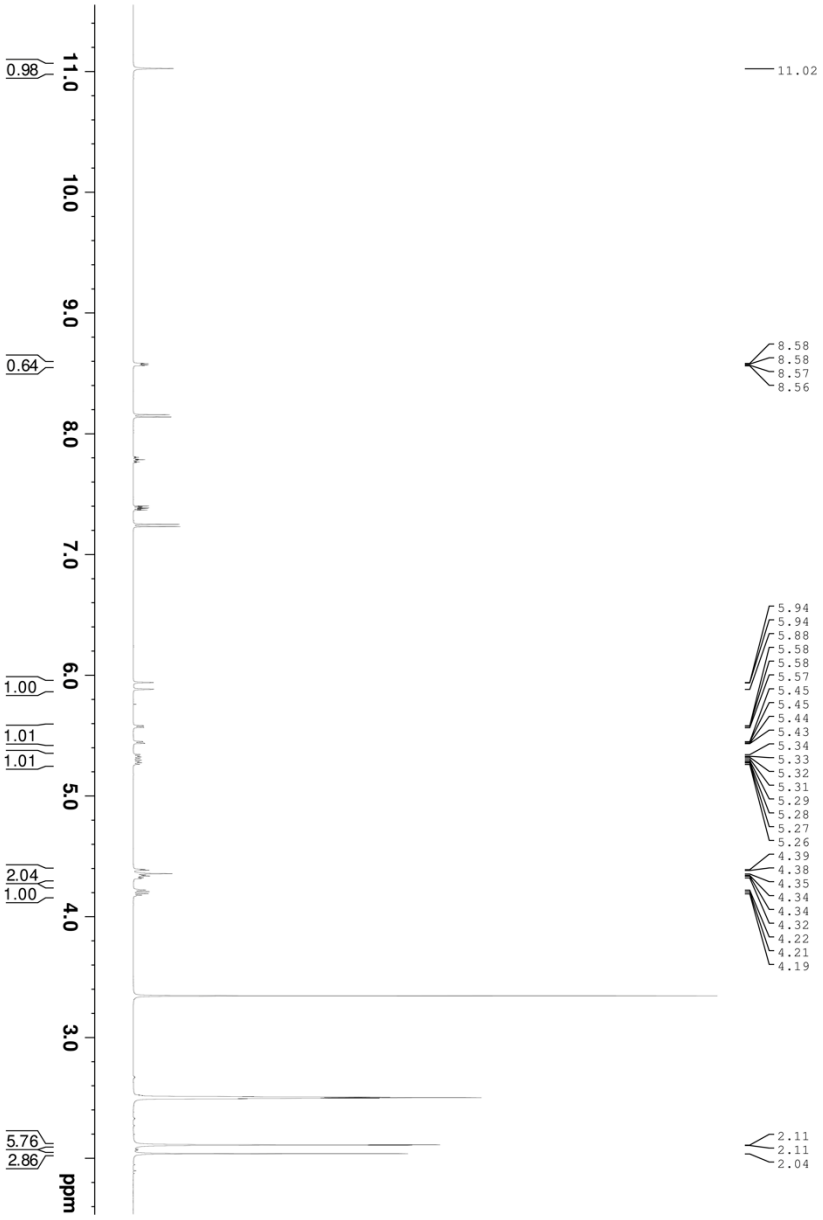
94



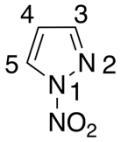
Appendix 4



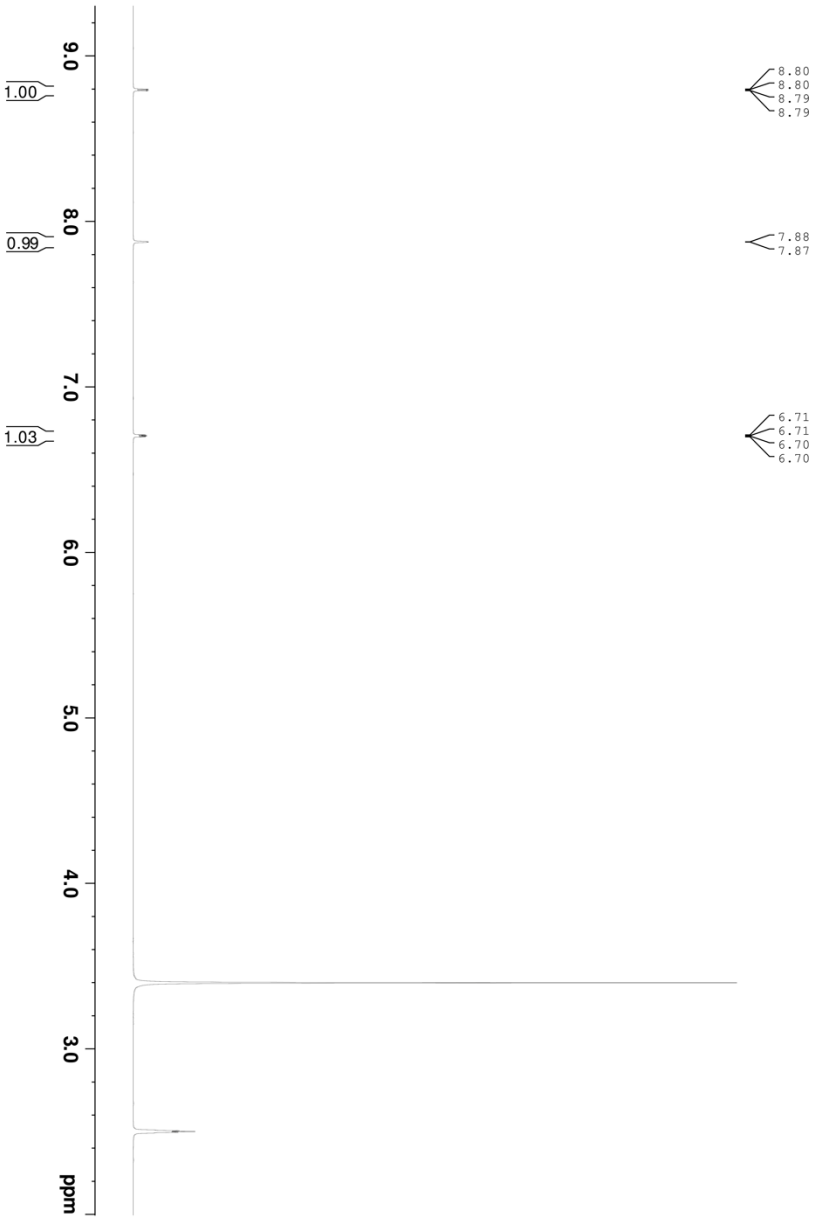
95



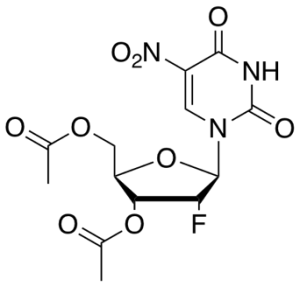
Appendix 5



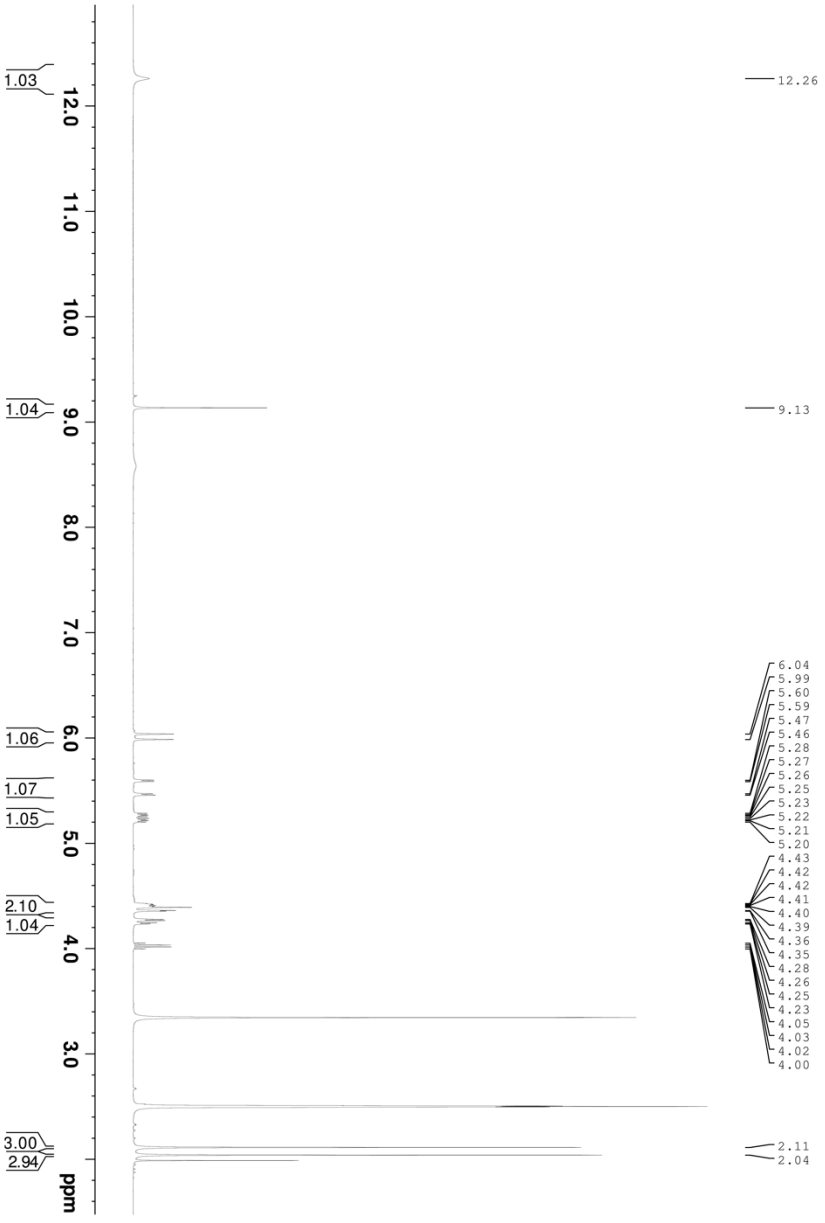
96



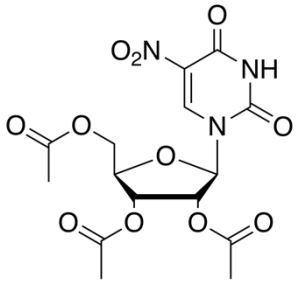
Appendix 6



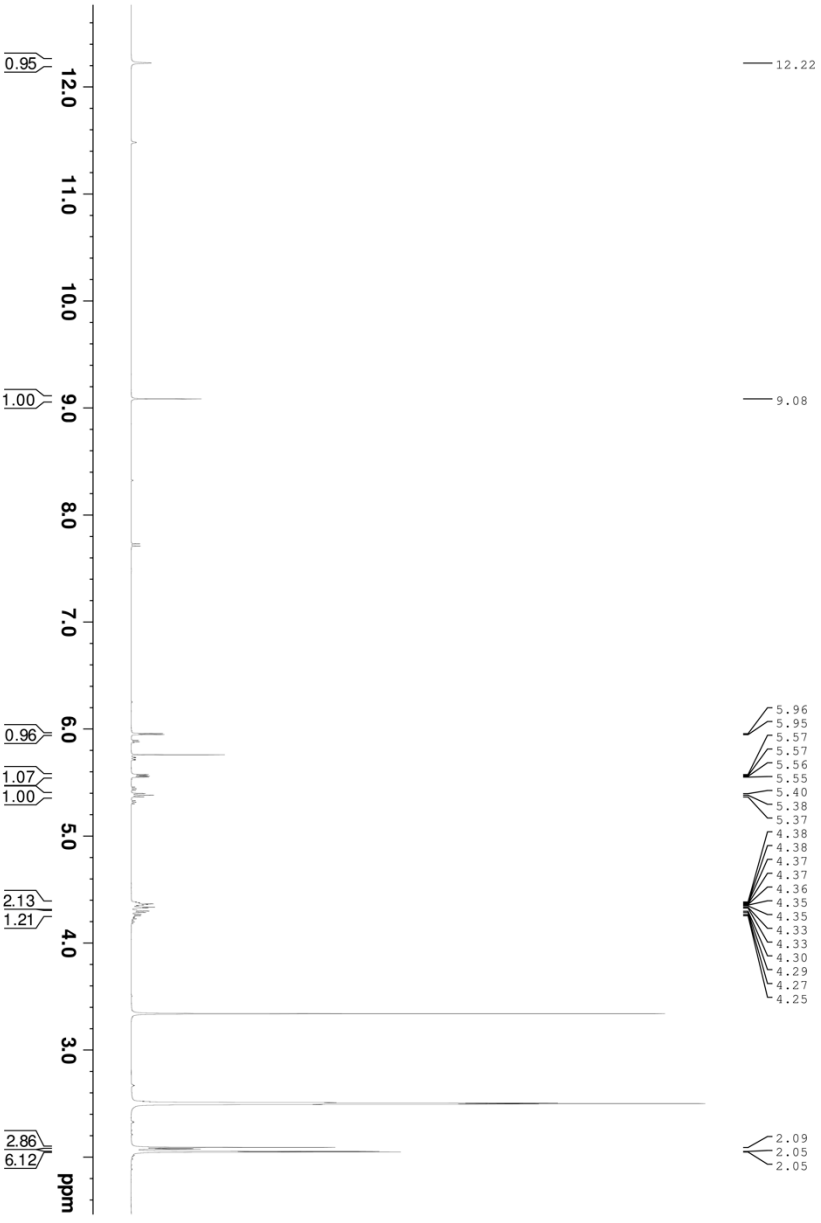
97



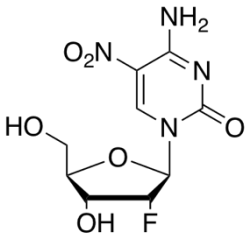
Appendix 7



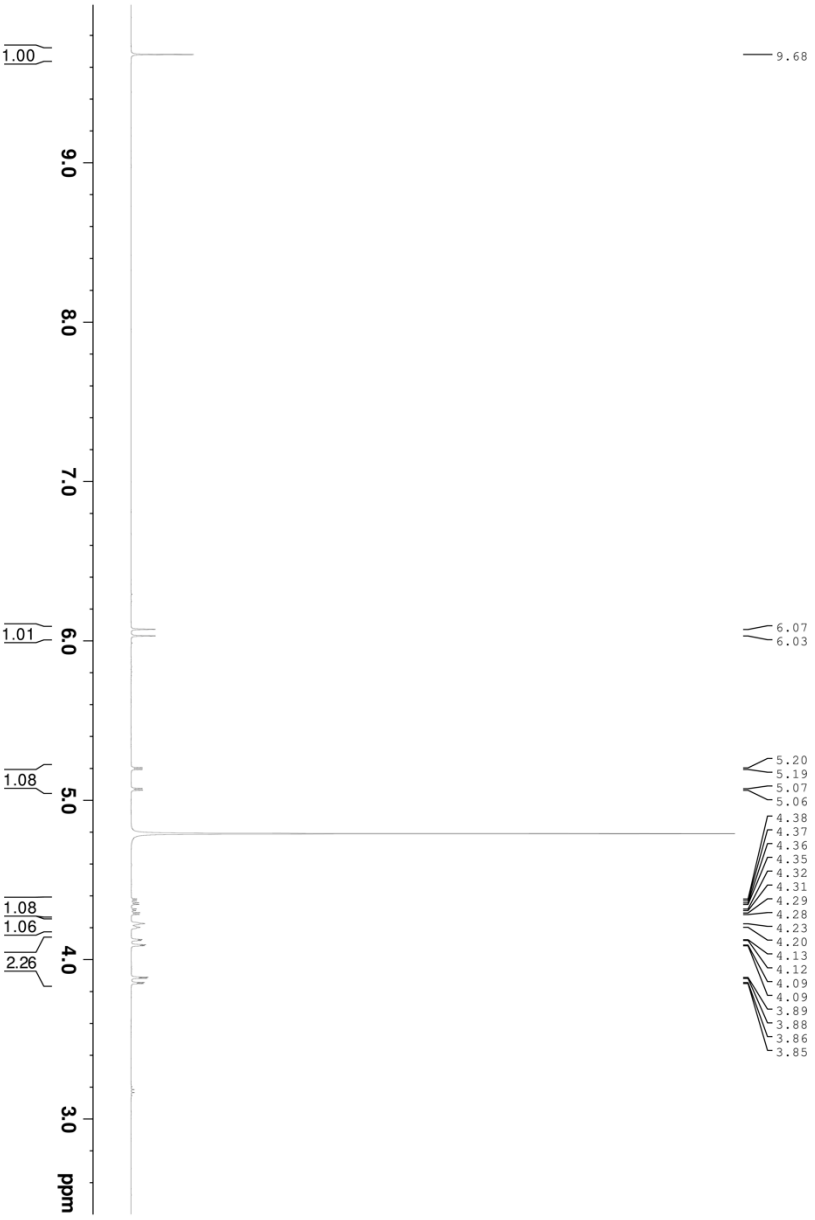
98



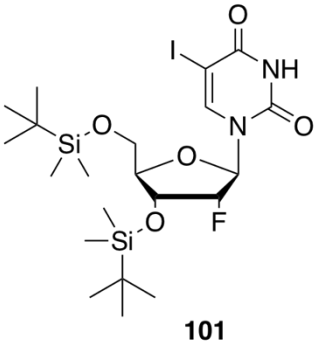
Appendix 8



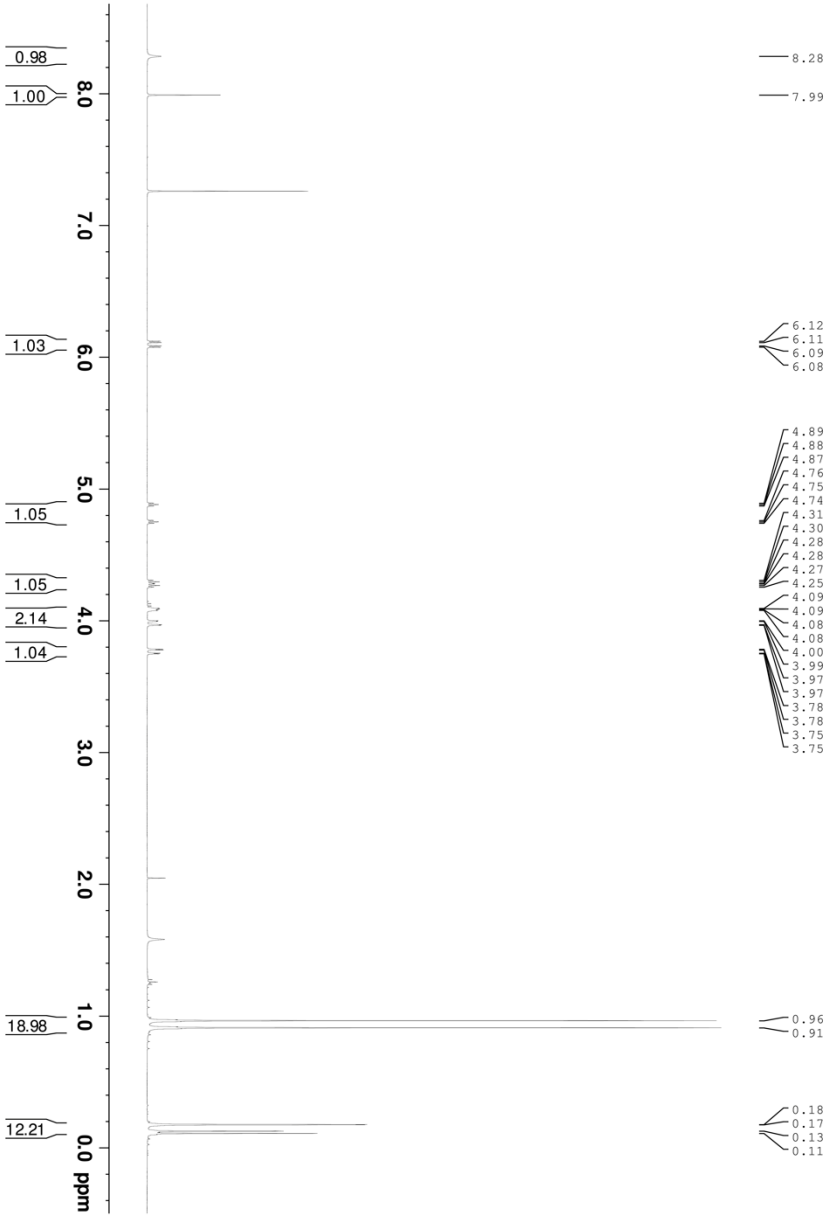
100



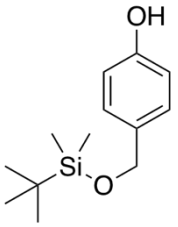
Appendix 9



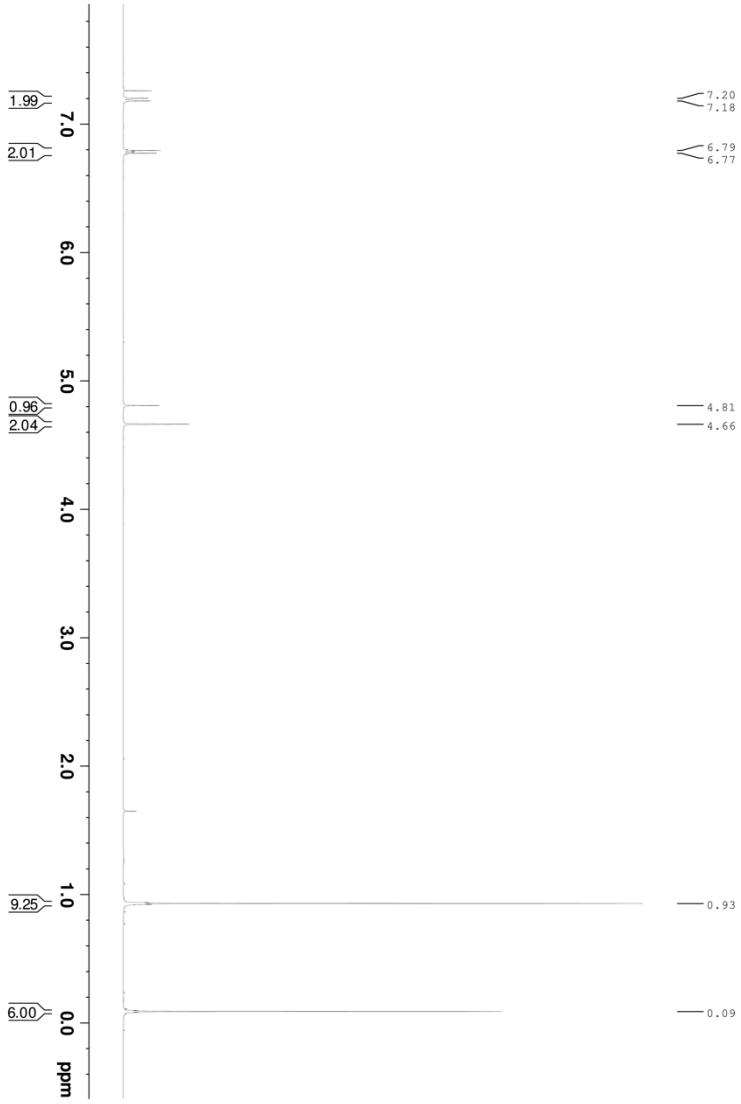
101



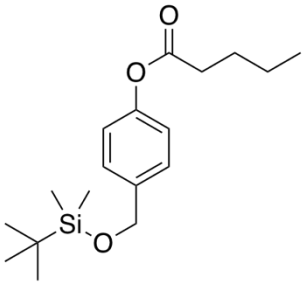
Appendix 10



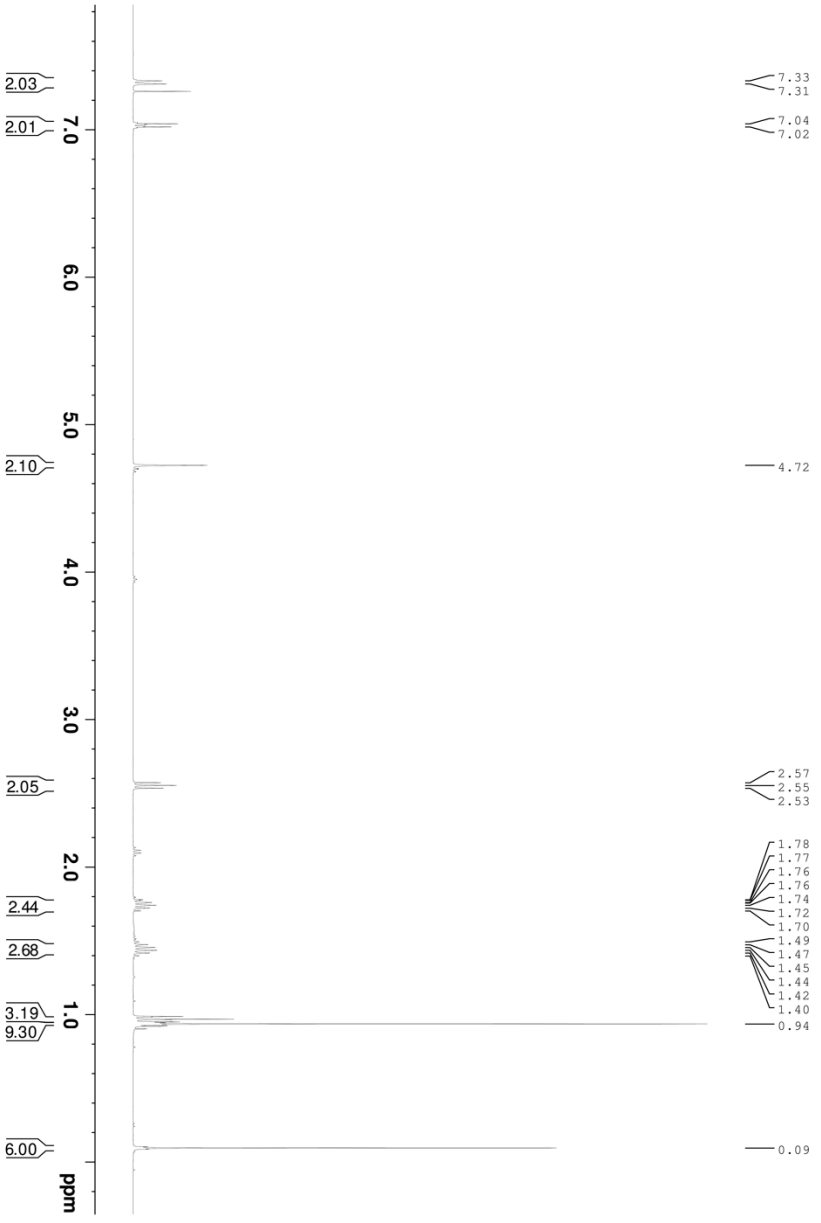
102



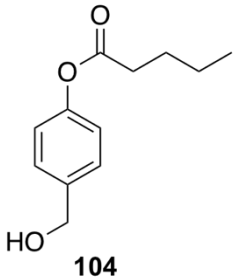
Appendix 11



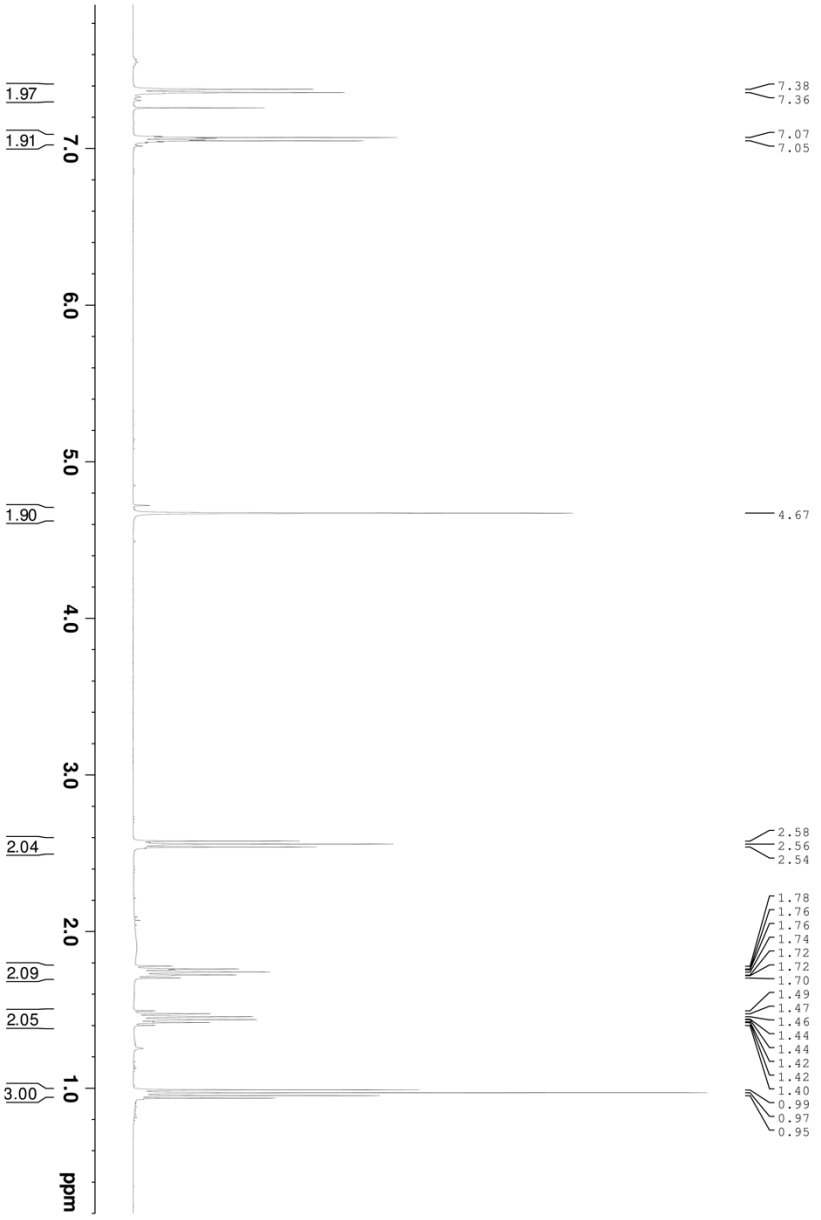
103



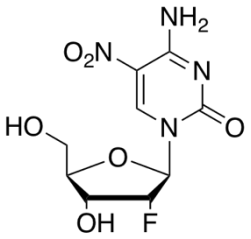
Appendix 12



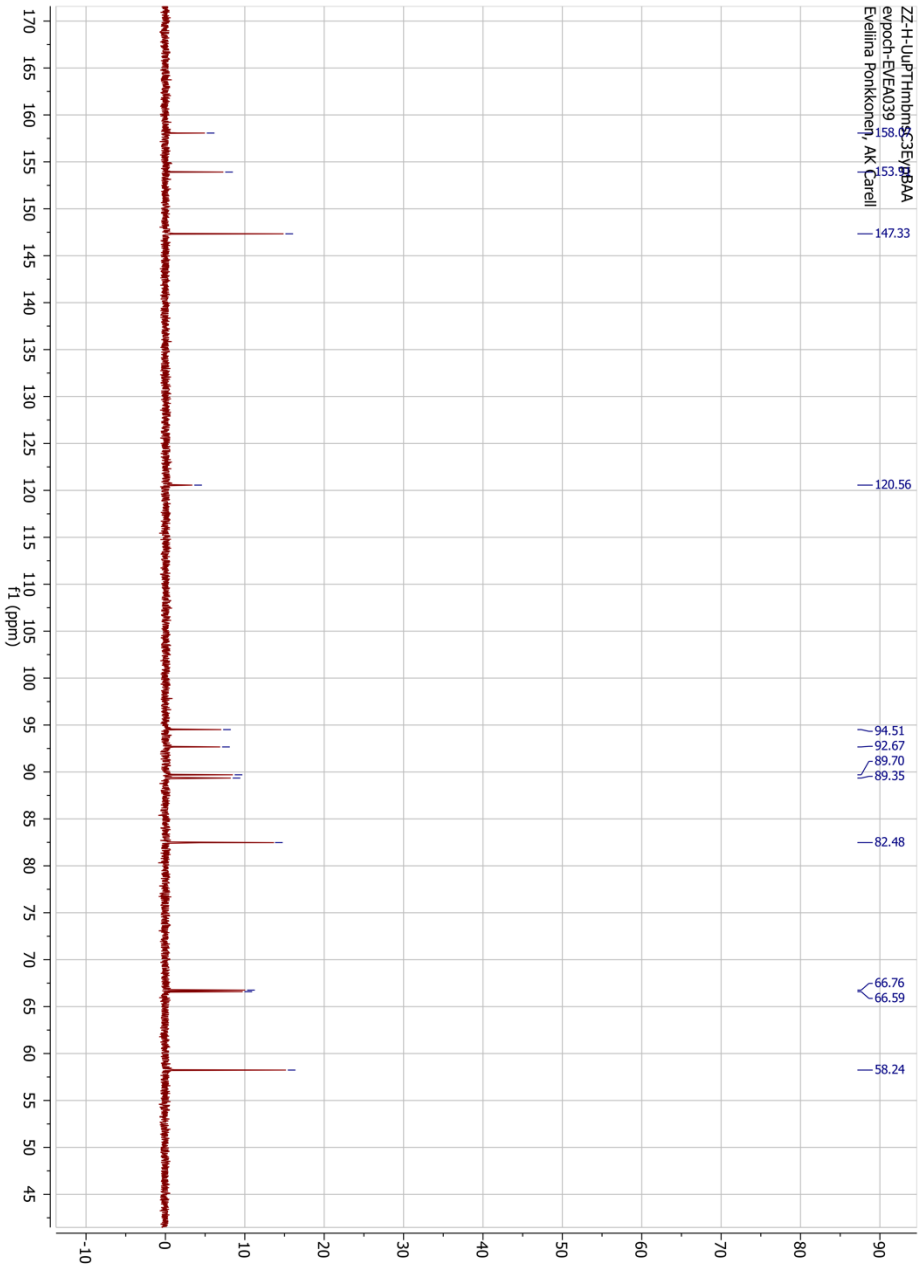
104



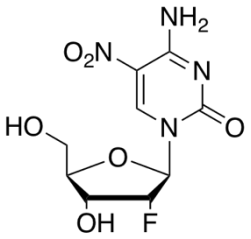
Appendix 13



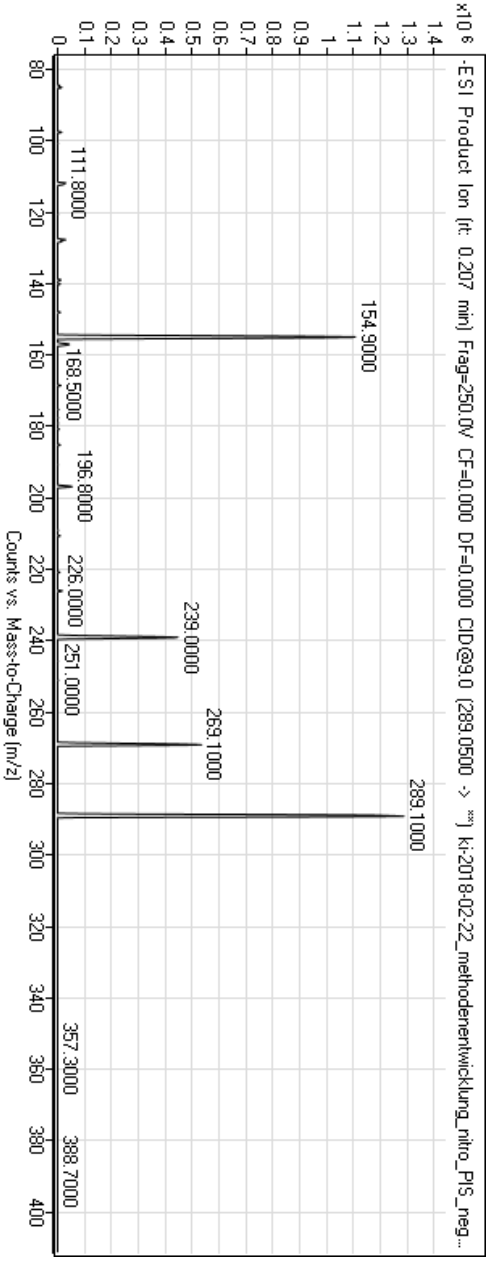
100



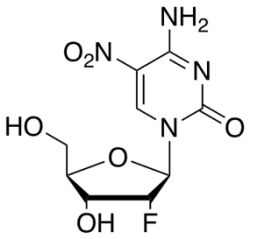
Appendix 14



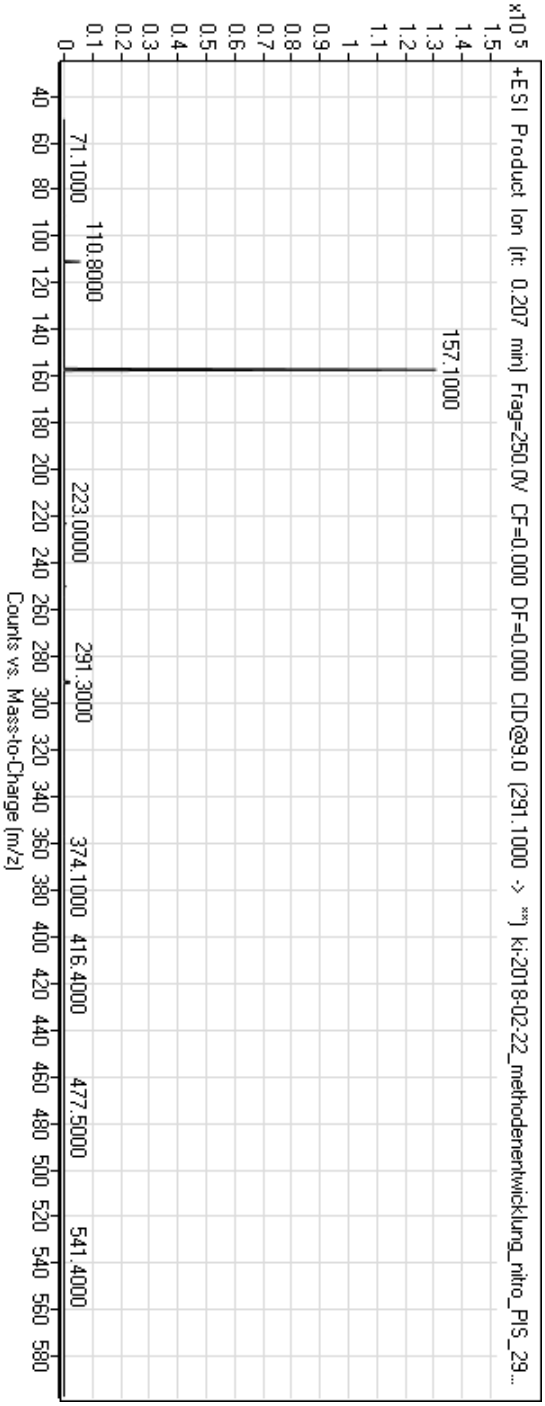
100



Appendix 15



100



Appendix 16

For 5-nitro-2'-(*R*)-fluoro-2'-deoxycytidine, absorptions at three different wave lengths, λ_{max} = 227 nm, 260 nm and λ_{min} 280 nm, were measured from 10 different concentrations. For the calculation of extinction coefficient, Beer-Lambert Law was utilized. It relates the weakening of light to the properties of the material through which the light is travelling.

The absorbance is directly proportional to the concentration, c (mol/l) of the solution of the sample and to the length of the light path, l (cm) which is equal to the width of the cuvette. The common formula for the Beer-Lambert Law can be written as:

$$A = \log_{10}\left(\frac{I_0}{I}\right) = \varepsilon l c, \qquad (1)$$

where the constant ε is molar extinction coefficient, a measure of the probability of the electronic transition. Mostly the absorbance ranges from 0 to 1 but it can also go higher. 0 absorbance at a certain wavelength refers to no light of the particular wavelength is absorbed. When intensities of the sample and reference are the same, $\frac{I_0}{I}$ is 1 and $\log_{10}$ 1 is zero. For calculating the extinction coefficient formula can be written as:

$$\varepsilon = \log_{10}\frac{\left(\frac{I_0}{I}\right)}{lc}. \qquad (2)$$

When defining ε for a compound the measured absorptions are plotted in a function of concentration. The concentration is an important factor since the proportion of the light absorbed depends on how many molecules it interacts with. Plotted values formed a straight line and for which an equation of form y = kx + b was calculated with Microsoft Excel 2017, 16.9 (180116). An example of measured absorptions as function of concentration at 280 nm is presented in table 9. The slope was determined as arithmetic mean of two calculated slopes and for wave lengths 227 nm, 260 nm and 280 nm calculated values were 7718.05 $mmol^{-1} \cdot l \cdot cm^{-1}$, 3085.95 $mmol^{-1} \cdot l \cdot cm^{-1}$, 1369.9 $mmol^{-1} \cdot l \cdot cm^{-1}$ respectively.

Table 8. Calculated equation as function of concentration at 280 nm.