A Thesis

entitled:

A Chemical Indirect Quantification Method for 5-Hydroxymethylcytosine

by

Gurdat Premnauth

Submitted to the Graduate Faculty as partial fulfillment of the requirements for

the Masters of Science Degree in Chemistry

Dr. Peter Andreana, Committee Chair

Dr. Amanda Bryant-Friedrich, Committee Member

Dr. Steve Sucheck, Committee Member

Dr. Amanda Bryant-Friedrich, Dean
College of Graduate Studies

The University of Toledo

August 2016
An Abstract of

A Chemical Indirect Quantification Method for 5-Hydroxymethylcytosine

by

Gurdat Premnauth

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the Masters of Science Degree in Chemistry

The University of Toledo

August 2016

5-Hydroxymethylcytosine (5-hmC) results from the oxidation of 5-methylcytosine (5-mC) by one of the ten-eleven translocation enzymes. This moiety is considered a major epigenetic modification. Although research on DNA methylation is well documented, there is still a lot to learn about DNA hydroxymethylation. Multiple studies have shown the variation of 5-hmC in multiple diseases including a wide range of cancers and neurodegenerative and neurodevelopmental diseases. This fact supports the hypothesis that 5-hmC plays an important role in these diseases, thus making its occurrence a major research target. Although its biological role is still under investigation, multiple quantification methods have been developed to provide a better understanding of its role in different disease states.

The method currently under development in our laboratory involves the oxidative labeling of 5-hmC. This is done through the selective oxidation of 5-hmC to 5-formylcytosine (5-foC), using efficient and mild oxidants followed by the formation of a stable conjugate with a biotin aminooxy linker, affinity enrichment, and detection.
I dedicate this thesis to my mom and to my dad that I miss so much.
**Acknowledgements**

I wanted to thank my advisors Dr. Peter Andreana and Dr. Amanda Bryant-Friedrich. They helped me through difficulties, and guided me to become a better chemist and a better researcher. They have aided in the development of my skills and the obtainment of new knowledge. I also would like to thank Dr. Yong-Yah Kim for offering me the significant opportunity to work and learn with him on Nuclear Magnetic Resonance spectroscopy. I wanted to thank my lab mates for answering my questions, and for their guidance.

Finally I would like to thank the University of Toledo and the Ecole Nationale Superieure de Chimie de Mulhouse for giving the opportunity to enroll in this exchange program and the NIH for the grant to finance my project.
Table of Contents

Abstract ........................................................................................................................................ iii

Acknowledgements .................................................................................................................. v

Table of Contents ...................................................................................................................... vi

List of Tables .............................................................................................................................. x

List of Figures ............................................................................................................................. xi

List of Abbreviations .................................................................................................................. xiii

List of Symbols .......................................................................................................................... xvi

1 Introduction .............................................................................................................................. 1

1.1 DNA Methylation ............................................................................................................... 1

1.2 DNA Hydroxymethylation ................................................................................................. 3

1.3 Hydroxymethylation in disease ......................................................................................... 6

1.3.1 Cancer ............................................................................................................................. 6

1.2 Neurodegenerative and neurodevelopmental diseases ...................................................... 9

1.3 Existing methods of quantification ...................................................................................... 12

1.3.1 Enzymatic β-glucosylation ............................................................................................ 12

1.3.2 Direct Biotinylation ....................................................................................................... 15

1.3.3 Sequencing ...................................................................................................................... 16

1.3.4 Dot blot quantification ................................................................................................. 18

1.3.5 Oxidative Labeling ........................................................................................................ 19
Results and discussion ........................................................................................................................................21
  2.1 Research overview ....................................................................................................................................21
  2.2 Parts of method development ................................................................................................................22
  2.3 Oxidation conditions ................................................................................................................................25
    2.3.1 Oxidation of the nucleobase ...............................................................................................................26
    2.3.2 Oxidation of the nucleoside ...............................................................................................................27
    2.3.3 Verification for the oligomer application .......................................................................................29
  2.4 Conjugation investigation ......................................................................................................................32
    2.1 Conjugation with the nucleobase .........................................................................................................33
    2.1 Oxidation with RNA nucleoside .........................................................................................................34
    2.1 Oxidation with DNA nucleoside .........................................................................................................36
  2.5 Development of the linker ....................................................................................................................38
    2.5.1 Development of the biotin hydroxylamine linker .........................................................................39
    2.5.2 Conjugation of the biotin hydroxylamine linker .........................................................................42

Conclusion ......................................................................................................................................................43

Future work .....................................................................................................................................................44

Experimental section ....................................................................................................................................46
  3.1 General methods .....................................................................................................................................46
  3.2 Experimental ..........................................................................................................................................47
    3.2.1 Synthesis of (E)-4-amino-2-oxo-1,2-dihydropyrimidine-5-carbaldehyde O-methyl oxime.............48
    3.2.2 Synthesis of (E)-4-amino-2-oxo-1,2-dihydropyrimidine-5-carbaldehyde O-benzyl oxime............49
3.2.3 Synthesis of (E)-4-amino-5-((2-benzylhydrazineylidene)methyl)pyrimidin-2(1H)-one

3.2.4 Synthesis of (E)-4-amino-1-((2R,3S,4R,5R)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-2-oxo-1,2-dihydropyrimidine-5-carbaldehyde O-methyl oxime

3.2.5 Synthesis of (E)-4-amino-1-((2R,3S,4R,5R)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-2-oxo-1,2-dihydropyrimidine-5-carbaldehyde O-benzyl oxime

3.2.6 Synthesis of 4-amino-5-((E)-(2-benzylhydrazineylidene)methyl)-1-((2R,3S,4R,5R)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)pyrimidin-2(1H)-one

3.2.7 Synthesis of (E)-4-amino-1-((2R,4R,5R)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-2-oxo-1,2-dihydropyrimidine-5-carbaldehyde O-methyl oxime

3.2.8 (E)-4-amino-1-((2R,4R,5R)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-1,2-dihydropyrimidine-5-carbaldehyde O-benzyl oxime

3.2.9 Synthesis of (2R,3R,5R)-5-(4-amino-5-((E)-(2-benzylhydrazineylidene)methyl)pyrimidin-1(2H)-yl)-2-(hydroxymethyl)tetrahydrofuran-3-ol

3.2.10 Synthesis of 2-((5-aminopentyl)oxy)isoindoline-1,3-dione hydrochloride
3.2.11 Synthesis of $N$-(5-((1,3-dioxoisindolin-2-yl)oxy)pentyl)-5-(2-oxohexahydro-1H-thieno[3,4-$d$]imidazol-4-yl)pentanamide ..........................55

3.2.12 Synthesis of $N$-(5-(aminoxy)pentyl)-5-(2-oxohexahydro-1H-thieno[3,4-$d$]imidazol-4-yl)pentanamide ........................................56

3.2.13 Synthesis of (Z)-$N$-(5-(((4-amino-2-oxo-1,2-dihydropyrimidin-5-yl)methylene)amino)oxy)pentyl)-5-(2-oxohexahydro-1H-thieno[3,4-$d$]imidazol-4-yl)pentanamide ........................................................................................................56

3.2.14 Synthesis of $N$-(5-(((Z)-4-amino-1-((2R,4R,5R)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-2-oxo-1,2-dihydropyrimidin-5-yl)methylene)amino)oxy)pentyl)-5-(2-oxohexahydro-1H-thieno[3,4-$d$]imidazol-4-yl)pentanamide ........................................................................................................57

References ...........................................................................................................................59

A Experimental spectra .........................................................................................................74
List of Tables

1.1 Summary of OxBS-Seq........................................................................................................11

1.2 Summary of TAB-Seq .........................................................................................................12

2.1 Improvement of the condition of oxidation of 5-hmC nucleobase.................................26

2.2 Improvement of the condition of oxidation of 5-hmC nucleoside.....................................27

2.3 Test of the oxidation conditions on other nucleosides.........................................................31

2.4 Conjugation to oximes and hydrazine of 5-formylcytosine ............................................33

2.5 Conjugation to oximes and hydrazine of 5-formylcytidine .............................................35

2.6 Conjugation to oximes and hydrazine of 2-deoxy-5-formylcytidine ...............................37
# List of Figures

1-1 Successive oxidations by TET proteins ..........................................................4  
1-2 Schematic representation of TET proteins.......................................................5  
1-3 Proposed mechanism for TET-mediated oxidation .........................................6  
1-4 IDH1/IDH2 function and mutation ................................................................9  
1-5 Enzymatic β-glucosylasation of 5-hydroxymethylcytosine..........................13  
1-6 β-glucosyl-5-hydroxymethyl-cytosine “Click Chemistry” with biotin tag ..........13  
1-7 β-glucosyl-5-hydroxymethyl-cytosine “Click chemistry” with fluorescent tag ....14  
1-8 Precipitation of β-glucosyl-5-hydroxymethyl-cytosine binding to JBP1 ..........15  
1-9 Direct Biotinylation .......................................................................................16  
1-10 Quantification using antibodies ....................................................................18  
1-11 Specificity of the anti-5-hmC and anti-CMS antisera.................................18  
1-12 Density dependence of anti-5-hmC and anti-CMS antisera .........................19  
1-13 Oxidative labeling methods .........................................................................20  
2-1 5-hmC species undergoing oxidation .............................................................23  
2-2 Linker used in project ...................................................................................24  
2-3 Labelling route on the monomer ....................................................................24  
2-4 Labelling route on the oligomer ....................................................................25  
2.5 Mechanism of the selective allylic oxidation..................................................28  
2.6 Structure of 8-oxoguanine ..............................................................................31
A-20 $^{13}$C NMR of compound 43 ................................................................. 83
A-21 LRMS of compound 43 ................................................................. 84
A-22 HRMS of compound 43 ................................................................. 84
A-23 $^1$H NMR of compound 44 ................................................................. 85
A-24 $^{13}$C NMR of compound 44 ................................................................. 85
A-25 LRMS of compound 44 ................................................................. 86
A-26 HRMS of compound 44 ................................................................. 86
A-27 $^1$H NMR of compound 52 ................................................................. 87
A-28 $^{13}$C NMR of compound 52 ................................................................. 87
A-29 LRMS of compound 52 ................................................................. 88
A-30 HRMS of compound 52 ................................................................. 88
A-31 $^1$H NMR of compound 53 ................................................................. 89
A-32 $^{13}$C NMR of compound 53 ................................................................. 89
A-33 LRMS of compound 53 ................................................................. 90
A-34 HRMS of compound 53 ................................................................. 90
A-34 LRMS of compound 23 ................................................................. 91
A-35 LRMS of compound 24 ................................................................. 91
A-36 $^1$H NMR of guanosine ................................................................. 92
A-37 $^1$H NMR of oxidized guanosine ................................................................. 92
A-38 $^1$H NMR of 2-deoxyguanosine ................................................................. 93
A-39 $^1$H NMR of oxidized 2-deoxyguanosine ................................................................. 93
A-40 $^1$H NMR of 2-deoxyadenosine ................................................................. 94
A-41 $^1$H NMR of oxidized 2-deoxyadenosine ................................................................. 94
A-42 $^1$H NMR of 2-deoxycytidine

A-43 $^1$H NMR of oxidized 2-deoxycytidine

A-44 $^1$H NMR of thymidine

A-45 $^1$H NMR of oxidized thymidine

A-46 $^1$H NMR of 2-deoxyuridine

A-43 $^1$H NMR of oxidized 2-deoxyuridine
List of Abbreviations

2-OG .......................... 2-oxoglutarate
5-caC .......................... 5-Carboxycytosine
5-foC .......................... 5-Formylcytosine
5-hmC .......................... 5-Hydroxymethylcytosine

AD .............................. Alzheimer disease
AlkB ........................... Alkane hydroxylases
ALS ............................. Amyotrophic lateral sclerosis
AdoMet ......................... Adenosyl-L-methionine
AML ............................ Acute myeloid leukemia

BER ............................ Base-excision repair

CpG ............................. Cytosine phosphate guanine
CMS ............................. Cytosine-5-methylenesulfonate
CNS ............................. Central nervous system

DCM ............................ Dichloromethane
DMF ............................ Dimethylformamide
DNA ........................... Deoxyribonucleic acid
DNMT .......................... DNA Methyl transferase
DSBH ........................... Double-stranded β-helix

FXTAS .......................... Fragile C-associated tremor/axtasia syndrome

HD .............................. Huntington disease
HPLC ........................... High-performance liquid chromatography

IHD ............................. Isocitrate dehydrogenase

JBP ............................. J-binding protein

mESC .......................... Mouse embryonic stem cell
MPN ............................ Myeloproliferative neoplasm
MLL ............................. Mixed lineage leukemia

oxBS-seq ..................... Oxidative bisulfite sequencing
PCR............................Polymerase chain reaction
PGC............................Primordial germ cell
redBS-seq..................Reduced bisulfite sequencing
ROS............................Reactive oxygen species

TAB-seq....................Tet-assisted bisulfite sequencing
TDG..........................Thymine DNA Glycosylase
TET...........................Ten eleven translocation
THF..........................Tetrahydrofuran
TLC..........................Thin Layer Chromatography

UDP-Glu......................Uridine diphosphoglucone

α-KG.........................α-Ketoglutarate
β-GT..........................β-glucosyl transferase
List of Symbols

°C ............................. Degrees Celcius

r.t. ............................. Room Temperature

h.................................. hours

M ............................. Molar
Chapter 1

Introduction of 5-hydroxymethylcytosine

1.1 DNA Methylation

Epigenetics seeks to understand the changes in gene activity or function that are not due to a change in the DNA sequence, whereas genetics is the study of the heritable changes due to a direct alteration in the DNA sequence. Genetics is consequently the study of alterations such as mutations, deletions, insertions and translocations. Epigenetics however, studies chemical modification of DNA. One of the most important epigenetic modifications is DNA methylation. [1]. It was discovered in 1944 by Avery et al, around the same time of the discovery of DNA in mammals. [2, 3] The existence of methylated cytosine was showcased for the first time by Rollin Hotchkiss in 1948 while analyzing calf thymus using paper chromatography. He submitted the hypothesis that the unknown fraction was 5-methylcytosine, because the separation from cytosine was very similar to the manner in which thymine (methyluracil) separated from uracil. [4] Studies by Holliday and Pugh in 1975 and Compere and Palmiter in 1981 suggested that DNA methylation functions in cell differentiation and gene regulation. Currently it is known that DNA methylation is a major epigenetic factor that influences gene activity. [5, 6]

DNA methylation occurs at the fifth position of a cytosine that is generally followed
by a guanine. It is catalyzed by the DNA methyltransferase (DNMT) enzymes in the presence of $S$-adenosyl-$L$-methionine, and gives 5-methylcytosine (5mC), which is also called the 5th base of DNA. [7] DNMT can be categorized in two different groups according to their function. De novo DNMTs are able to establish a new methylation pattern from unmodified DNA. They are composed of DNMT3A and DNMT3B. The other category is composed of DNMT1, which can copy the DNA methylation pattern during DNA replication from the template DNA strand. [8] The other enzyme DNMT2 is not able to catalyze the methyl transfer to DNA. However, it is known to methylate the cytosine in position 38 of transfer RNA of aspartic acid, and has a mechanism similar to the DNA methyltransferase. [9] DNA methylation happens in the entire body, but has its highest levels in the brain. In 1982, Ehrlich et al. quantified the number of 5-mC to be around 1% of the total nucleic acids in the human genome. [10] The majority of DNA methylation happens at the CpG site, which is a cytosine followed by a guanine nucleotide. These nucleotides are heavily methylated with the exception of CpG islands. [11] Methylation can also happen in non-CpG sites as seen in mouse and human embryonic stem cells; however most of it is lost in mature tissues. [12, 13] A few methylated non-CpG sites are present in adolescent and adult cortex tissues, and seem to be highly regulated. They are present in specific cell types such as pluripotent cells and neural progenitors. [14] The CpG islands are composed of approximately 1,000 base pairs with a high CpG density, and are known to be hypomethylated. [11] They retain around 70% of the gene promoters [15] and are known to have a role in gene expression and transcription. However, their exact role requires further studies in order to be fully understood. [16] Methylated CpG islands can impair the transcriptional binding
factor by recruiting repressive methyl-binding protein and effectively silence gene expression. [17]

Consequently, methylation intervenes in the regulation of gene expression and other biological process such as the inactivation of the X chromosome in women [18]. It also impacts cancer by causing a hypermethylation of the promotors of the gene suppressor of tumors that will inhibit the expression of such promotors as seen in the case of retinoblastoma. [19]

1.2 Hydroxymethylation

Often considered as an accidental deterioration of DNA, 5-hydroxymethyl-cytosine recently became a major epigenetics target due to its abundance and enzymatic selectivity. The generation of 5-hydroxymethylcytosine (5-hmC) is the result of the oxidation of 5-mC by a specific category of enzymes in the ten-eleven translocation (TET) family. TET enzymes can successfully oxidize 5-methylcytosine into 5-hydroxymethylcytosine. The oxidation can then continue to form 5-formylcytosine (5-foC) and later 5-carboxylcytosine (5-caC). [20, 21]

The levels of the different compounds decrease with the number of oxidations. The abundance of 5-hydroxymethylcytosine accounts for 0.6 % of total cytosine, 5-foC accounts for 0.02 % and of 5-caC counts for 0.003 %. [20] 5-Formylcytosine and 5-carboxylcytosine can respectively be deformylated and decarboxylated, respectively by the thymine-DNA-glycosylase enzyme that converts both back to cytosine through a base excision repair process (BER). [22] Consequently, 5-hmC is a major subject of TET family research.
Figure 1.1: Successive oxidations by TET proteins.

TET proteins are known for their fusion with mixed lineage leukemia (MLL) or myeloid/lymphoid gene in acute myeloid leukemia (AML). [23, 24] Studies in vivo of the TET genes established a better understanding of their particular expression pattern. [25-28] Indeed, the analysis of mice that are missing TET proteins resulted in better understanding the important regulatory roles that TET serve in different processes: epigenetic reprogramming, germ cell development, pluripotent stem cell differentiation, and myelopoiesis. [27-29].

TET proteins are 2-oxoglutarate (2OG) and iron-dependent dioxygenase, and require α-ketoglutarate as a cosubstrate for enzymatic activity. [30] There are three forms of TET (TET1, TET2 and TET3) that are the result of triplication from a common ancestor; all of which are capable of oxidizing 5-mC into 5-hmC. [31-33] They all contain predicted domains for the binding of iron and oxoglutarate, and they all also contain a cysteine-rich region, which is a unique domain known to be involved in DNA binding. At their termini, TET1 and TET3 contain a CXXC zinc finger domain. [31] This domain is common to other methylated proteins such as DNMT1. The other CXXC-containing proteins can also bind to unmethylated, methylated and hydroxy-methylated DNA, but they preferably bind with areas of high CpG density. [31, 34] The proteins also contain a spacer region between two parts of the
disconnected double stranded-β-helix (DSBH) with variable length bridges; however the function of this spacer is still unknown. [35]

![Schematic representation of TET proteins](image)

**Figure 1.2:** Schematic representation of TET proteins [36].

The precise mechanism of TET-mediated oxidation still needs to be optimized. The mononuclear iron (II) center transfers two electrons for the reduction of a dioxygen. α-Ketoglutarate (α-KG) plays the role of the cosubstrate that donates two other electrons. [37-41] With the available four electrons, the active site iron (II) binds and activates a dioxygen molecule to form an iron (IV)-oxo species with the α-KG conversion to succinate and causes release of a carbon dioxide. [42-46] The orientation of the iron (IV) goes in the direction of the methyl C-H group [47, 48], and activates the C-H bond through a radical rebound mechanism or a more concerted reaction. [49]
1.3 Hydroxymethylation in disease

1.3.1 Cancer

The highest levels of 5-hmC were detected in the brain within the CNS tissues [51] where it was found at levels as high as 40% of abundance in the DNA of Purkinje cells of the cerebellum. [20] In diseases such as cancer, there is a decrease in the amount of 5-hmC, which suggests that the maintenance of 5-hmC is strained in highly proliferating cells. [52]

One of the first studies that explored the levels of 5-hmC in cancer was carried out by
Ko et al., in which the focus of the study was the potential changes due to TET2 mutation in AML. The levels of 5-hmC were found to be lower in bone marrow samples in the presence of mutated TET2 compared to normal TET2. [53] More recent studies display a global loss of 5-hmC in different human solid tumors (breast, colon, gastric, liver, lung, melanoma, prostate etc.). [54, 55] The loss of 5-hmC correlates with the downregulation of TETs. Indeed, TET1 downregulation causes a 72.7% reduction of 5-hmC in CRCs, and a 75% reduction in gastric cancer. In melanoma, downregulation is principally due to TET2 and isocitrate dehydrogenase 2 (IDH2). For liver and breast cancers, all three TETs levels decreased along with 5-hmC. [55] Mass spectroscopic methods were used to quantify 5-hmC in hepatocellular carcinoma (HCC), in which 5-hmC was found to be four to five times lower in quantity than a normal tumor adjacent tissue, and an analogy was observed between the levels and the tumor stage [56]. The levels in lung squamous cancer seemed to be two to five times lower. [57]

The mechanism of the loss of 5-hmC in different disease states is still under investigation but several explanations are possible:

- The tumor cells grow faster than the healthy cells, and thus cannot compete with the levels of 5-hmC. 5-hmC does not undergo DNMT1 copy of the methylated cytosine followed by an oxidation using TET enzyme. *In vitro* studies have shown that DNMT1 is not capable of methylation of a strand opposite to a 5-hmC. [58-60] UHFR1 (the maintenance methylation cofactor of DNMT1) affinity with a 5-hmC containing CpG site is poor so it consequently does not contribute to higher methylation. [57] Other mechanisms such as BER that uses TDG following successive oxidation of 5-mC can also be the cause of a possible
loss of 5-hmC. [22] Nevertheless, the levels of 5-mC do not decrease as in the case of 5-hmC; however, the correlation is too weak in order to explain the entire loss of 5-hmC that is much greater. [57]

- The weakening of TET activity or function can also be an explanation for the decrease of 5-hmC. The levels of TET seem to be decreasing in the RNA of all targeted cancers. [55, 57, 61] Recent publications have discussed the control of TET family expression by microRNAs, which play an important role in multiple types of cancers. [62-65] Mutations of the TETs can also enter this category due to a variety of genetic aberrations in the TET2 gene: microdeletions, copy-number-neutral loss of heterozygosity, and the most common, somatic mutations. [66, 67]

- Finally, the mutations of the isocitrate dehydrogenase 1 (IDH1) and IDH2 can explain the loss of 5-hmC. The role of IDH1 and IDH2 is to catalyze the conversion of isocitrate to α-KG, the key cofactor of TET enzyme. The formation of mutations is common in a multiple range of cancers and solid tumors: gliomas, AML, Myelodysplastic Syndromes (MDS), Myeloproliferative Neoplasms (MPN), and Chronic myelomonocytic leukemia (CMM). [68-81] IDG mutation is able to transform α-KG into (R)-2-hydroxyglutarate ((R)-2-HG), which is its analogue inhibitor. [82, 83] Thus, the inhibition of TET causes the decrease in the levels of 5-hmC.
The general loss of 5-hmC in cancer shows its importance as a target for cancer therapy or as a novel biomarker. Nevertheless, there is still a significant amount to learn on the mechanism of action and the biological role of 5-hmC in cancer, especially if it is a cause or consequence of cancer.

### 1.3.2 Neurodevelopmental and neurodegenerative disease

5-hmC has been shown to be a dynamic epigenetic marker during different stages of mouse neurodevelopment from embryonic to adult brain, and has been reported to regulate transcriptional factors involved in neurodevelopment [84]. Several neurodevelopmental diseases such as Rett syndrome, autism spectrum disorder, schizophrenia, and fetal alcohol syndrome have shown variations in their levels of 5-hmC. [85] These variations imply that 5-hmC plays an important role in the
neurodevelopment and the maintenance of adult CNS functions. Several neurodegenerative diseases have been studied more closely.

Alzheimer’s disease (AD) is a disease characterized by a decline of cognitive functions and neuronal cell lost. It is also accompanied with extracellular amyloid beta plaques and intracellular neurofibrillary tangles made of hyperphosphorylated tau proteins. [86] The causes of the disease are unknown, but abnormalities in beta-amyloid precursor protein gene (APP) and the presenilin genes (PSEN1 and PSEN2) have been identified in addition to the potential alteration of epigenetic factor due to aging and environmental stress. [87-89] The quantification of 5-hmC has shown the variation of levels in the brain of AD patients. Decreases of 5-mC and 5-hmC were observed in the hippocampal region of AD brain [90]. Other studies have shown increases in levels of TET1, 5-mC, and 5-hmC in the hippocampus, frontal and temporal cortexes of AD patients. [91] Further investigations must be executed in order to understand the differences between the variations and to determine if they are a cause or a consequence of AD.

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease, which is the consequence of a selective loss of motor neurons in the brain and spinal cord [92]. The cause is unknown, but a small percentage of defective genes has been identified some of which include encoding superoxide dismutase 1 (SOD1), TAR DNA Binding Protein (TARDBP), fused in sarcoma (FUS), Ubiquilin2 (UBQLN2) and C9ORF72. Epigenetic changes can happen with environmental factors or exposure to toxins or dietary factors [93]. The levels of 5-mC and 5-hmC increase in postmortem sporadic ALS spinal cord, but not in blood samples. [94] Consequently 5-hmC cannot be used as a suitable biomarker for ALS. Regardless, the role of 5-hmC still needs to be determined.
Fragile X-associated tremor/ataxia syndrome (FXTAS) is a neurodegenerative disease for which patients become symptomatic later in life. It is caused by repeated CGG expansion mutations in the 5’ untranslated region (UTR) of the FMR1 gene [95]. There are 55-200 repeated CGG mutations that are carried by FXTAS patients. Symptoms are expressed as the development of a severe tremor, ataxia, and progressive cognitive decline, which usually occurs beginning at age 50 [96]. A study analyzed the amounts of 5-hmC in the rCGG mouse model of FXTAS, where the increase of the amount of human CGG repeats in the UTR of the FMR1 gene in Purkinje neuronal cells, which causes deficits in behavior or cell death [97]. This was accompanied with an increase of 5-hmC levels in the cerebellum-specific enhancer and in the repeated elements that correspond with the genes and transcription factors that are known to be involved in neurodevelopment. This suggests a potential role of 5-hmC in the disease mechanism of FXTAS.

Huntington’s disease (HD) is a hereditary and progressive neurodegenerative disease with symptoms including chorea, dystonia and cognitive decline. The primary targeted area of the pathology is the striatum. It is caused by a CAG repeat expansion mutation within exon 1 of the HTT gene that provokes an aberrant polyglutamine formation in the amino-terminus of the HTT protein [98]. The possible involvement of 5-hmC has been recently shown. Increases of levels of 5-mC and decreases of the levels of 5-hmC were observed in the 5’ UTR of the ADORA2A gene in the striatum of HD patients [99]. ADORA2A gene presence encodes the G-protein-coupled receptor, adenosine A2A, which is expressed in high quantities in the basal ganglia in normal individuals; whereas it is reduced in HD [100]. A decrease of 5-hmC has also been reported in the YAC128 HD mouse striatum and cortex brain [101]. 5-hmC in
HD is a potential epigenetic feature that is involved in neurogenesis, neuronal function and survival in the HD brain.

There is still a considerable sum of investigation that must be executed on the potential involvement of 5-hmC in all of these disease as a cause or as a consequence, in order to be able to have a full understanding of these diseases. 5-hmC could become a possible biomarker for those diseases and has the potential to identify new targets for therapy. The limitations today are in the poorly reproducible techniques of immunohistochemical quantification. [102] Thus it is imperative to develop new techniques such as third-generation sequencing and oxidative bisulfite (oxBs) BeadChip platforms.

1.4 Existing methods of quantification

Several methods for the quantification of 5-hmC in the DNA have been developed in the past year.

1.4.1 Enzymatic β-glucosylation

The first category uses the glycosylation process found in bacteriophage T4 that converts 5-hmC into β-glucosyl-5-hydroxymethyl- cytosine (5-g-hmC) in the presence of the catalytic β-glucosyl transferase (β-GT) enzyme. The enzyme catalyzes the transfer of β-D-glucosyl residues from uridine diphosphogluco (UDP-Glu) to the hydroxyl group of 5-hmC. This process is highly selective and has been used to develop several methods for the quantification of 5-hmC. [103]
One category of quantification method uses this principle for the attachment of 5-hmC to a modified glucose, 6-N$_3$-glucose. The azide group present on the glucose is then used in click chemistry with alkynes using a Huisgen cycloaddition.

![Chemical structure](image)

**Figure 1.5:** Enzymatic β-glucosylazation of 5-hydroxymethylcytosine.

The first click chemistry method was developed in 2010 by Chun-Xiao Song et al., and it uses the coupling of the N$_3$-5-g-hmC with a dibenzocyclooctyne-modified biotin. The affinity enrichment of the biotin with avidin–horseradish peroxidase (HRP) enables accurate quantification of 5-hmC. [104]
Another method developed by Tamar Shahal et al. in 2014 uses the click chemistry with a Dibenzocyclooctyne-Cy5 (DBCO-Cy5) followed by fluorescence detection. [105]

**Figure 1.6:** β-glucosyl-5-hydroxymethyl-cytosine “Click chemistry” using a biotin tag.

**Figure 1.7:** β-glucosyl-5-hydroxymethyl-cytosine “Click chemistry” with a fluorescent tag.
Glycosylation of 5-hmC was used by A.B. Robertson et al. to develop a new method in 2011. It uses the chemical similarity between β-glucosyl-5-hydroxymethyl-cytosine and β-glucosyl-5-hydroxymethyl-uracil, which in certain protozoa is specific to a variety of proteins. In the DNA of African trypanosomes and other kinetoplastids, modified β-glucosyl-5-hydroxymethyl-uracil, which is also known as J-base, is present. Those organisms are equipped with J-binding proteins (JBP1 and JPB2); the binding of JBP1 being totally specific to the DNA containing the modified base. The method was developed using the assumption that if JBP1 can bind to the J-base containing DNA, it should react the same way with β-glucosyl-5-hydroxymethylcytosine, since the structures are very similar.

This method displays the β-glucosyl-5-hydroxymethylcytosine placed in the presence of JBP1 proteins form Crithidia fasciculate coupled to magnetic beads after selective glycosylation of the 5-hmC. After precipitation of the 5-hmC containing DNA, the quantification was executed by Real-Time Polymerase Chain Reaction (RT-PCR) in the case of single gene analysis, and by microarray and sequencing in the case of genome wide analysis. [106]

![Figure 1.8: Precipitation of β-glucosyl-5-hydroxymethyl-cytosine bound to JBP1.](image-url)
1.4.2 Direct Biotinylation

Another method developed in 2013 by Wen-Wu Li et al. uses direct bisulfite-mediated biotinylation with N-biotinyl-L-cysteine. However, the low yield (33-55%) of the biotinylation makes this method less attractive. [107]

![Chemical structure](image)

Figure 1.9: Direct biotinylation.

1.4.3 Sequencing

5-hmC treatment with sodium bisulfite gives cytosine-5-methylsulfonate (CMS), which is resistant to deamination and is read as a cytosine (C) during DNA sequencing. 5-mC that is resistant to sodium bisulfite treatment is also read as a cytosine. Both 5-formylcytosine and 5-carboxylcytosine after bisulfite treatment and sequencing are read as thymine (T). It is impossible with normal bisulfite sequencing to determine the levels of 5-hmC. This limitation of the bisulfite sequencing method induced the generation of derivatives for sequencing such as oxidative bisulfite sequencing (oxBS-Seq), which was developed by J. Hu et al. in 2013 or TET-assisted bisulfite sequencing (TAB-seq), developed by C. He et al. in 2012. OxBS-Seq relies on the selective oxidation of 5-hmC into 5-foC. This method uses
the sensitivity of 5-foC towards the bisulfite treatment compared to 5-hmC, which does not change. 5-mC does not undergo any oxidation so is read as a C after bisulfite treatment, and the 5-hmC, which now becomes a 5-foC, is read as a T after bisulfite treatment. The comparison of BS-seq DNA sequencing and oxBS-seq DNA sequencing allows for the location of the 5-hmC as the C that became T (table 1.1). Ninety five percent of the 5-hmC present is oxidized with this procedure. This method is thus an efficient way to quantify 5-mC and 5-hmC in the DNA. [108]

Table 1.1: Summary of sequence reading changes during OxBS-Seq.

<table>
<thead>
<tr>
<th>method</th>
<th>C</th>
<th>5-mC</th>
<th>5-hmC</th>
<th>5-foC</th>
<th>5-caC</th>
</tr>
</thead>
<tbody>
<tr>
<td>BS-Seq</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>OxBS-Seq</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
</tbody>
</table>

TAB-seq is an enzymatic method that uses the specificity of the β-glucosylation of 5-hmC in the presence of UDP-Glu to form glu-5-hmC. Treatment with the TET1 enzyme converts 5-mC and 5-foC into 5-caC. After bisulfite treatment, the 5-caC and cytosine are read as T and the glu-5-hmC is the only one read as C (table 2.2). The percentage of 5-hmC detected in this method is around ninety-two percent, which can be a limitation for the accuracy of this sequencing method. [109]
Table 1.2: Summary sequence reading of TAB-Seq.

<table>
<thead>
<tr>
<th>method</th>
<th>C</th>
<th>5-mC</th>
<th>5-hmC</th>
<th>5-foC</th>
<th>5-caC</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAB-Seq</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
</tbody>
</table>

1.4.4 Dot blot quantification

![Chemical reaction and antibody interaction diagram]

**Figure 1.10:** Quantification using antibodies

This biochemical method uses the specificity of attachment of anti-5-hmC antibodies and anti-5-CMS antibodies.
Figure 1.11: Specificity of the anti-5-hmC and anti-CMS antisera. Reprinted by permission from Macmillan Publishers Ltd: Nature [53], copyright 2010.

Rabbits were immunized with 5-hmC or CMS conjugated to KLH. Test of the immunized rabbit antisera was done a the dot-blot assay for their ability to selectively detect distinct cytosine modifications on serially-diluted (0.05 pmol to 3 pmol) lambda (λ) phage DNA, lambda (λ) phage DNA methylated in vitro at CpG sites with SssI (λMe), or DNA from 5-hmC-containing T4* bacteriophage (produced by growth on the appropriate E. coli host\(^1\)) before (T4*) and after bisulfite treatment (T4*-BiS). The result shows the high specificity of the methods.
The 5-hmC antisera are limited to a minimum of sixteen 5-hmC per oligomer while the 5-CMS antisera can be very precisely used to a minimum of one 5-hmC per oligomer. This method can be very specific and precise but remains very expensive. [53]

1.4.5 Oxidative Labeling

The other category of methods developed involves oxidative labeling. The existing methods developed use ruthenium tetroxide (RuO₄) in order to oxidize 5-hmC into 5-formylcytosine (5-foC). RuO₄ was also used in the oxidative sequencing (Ox-seq) of 5-hmC developed by Alasubramanian et al. In 2013, J. Hu et al. proposed an oxidative labeling, which used a fluorescent imine tag[110], and in the same year T. Hong et al. developed an oxidative labeling method, which used a hydroxylamine-BODIPY
fluorescent tag. [111] The characterization of 5-hmC after oxidation and attachment of the tag is done by fluorescence detection.

Figure 1.13: Oxidative labeling for quantification of 5-hmC.
Chapter 2

Results and Discussion

2.1 Research overview

It is essential to have simple methods to quantify hydroxymethylation in cells and tissue in order to have a better understanding of the role and the biological relevance of 5-hmC. Several limitations were found in the existing methods of quantification. The majority of quantification methods were developed using enzymatic β-glycosylation, oxidative labeling, or antisera and are extremely expensive due to the use of enzymes, UDP-6-azido-6-deoxy-D-Glucose, antibodies, or fluorescent tags. The low conversion of direct biotinylation makes the method unusable for the quantification of the total 5-hmC in the cell. Indeed, a majority of the present 5-hmC will not be linked to the biotin, so, consequently, will not be detectable. The anti-5-hmC antisera limit of detection is too high to precisely detect the amount of 5-hmC, causing a false negative in cells with low 5-hmC levels. Sequencing method limitations are principally due to the requirement of multiple bisulfite treatments to deaminate the generated 5-fC in OxBS-Seq that could damage the DNA stained in addition to the oxidation. TAB-Seq requires a highly active TET enzyme in order to have high conversions and is also expensive due to the β-glycosylation to protect 5-hmC. [112] Another drawback while using the enzymatic β-glycosylation clicked
with a fluorescent tag could be the relatively bulky tag generated, which would be difficult to attach in the presence of vicinal 5-hmC due to sterical repulsion. These problems were answered during the development of our method. The method consists of selective oxidative labelling using an inexpensive biotin group as a tag. The possible steric interactions between the biotin group and the vicinal environment were overruled using a relatively long linker.

The development of the method was divided in four major parts.

### 2.2 Parts of the method development

The first part of the study consists of an investigation of the best possible oxidant for the allylic oxidation of the alcohol present on the nucleobase of 5-hmC. The major problem in this part will be to find allylic selective oxidants. This is due to the competition with the other alcohols present on the deoxyribose in the case of DNA, and on the ribose for RNA. Indeed, there is a competition between the primary alcohol present in the C5' position of the nucleobase and the secondary alcohols present on the 2' and 3' positions, as well as the the primary alcohol present on the 5' position. However none of these hydroxyls are in an allylic position. To be able to choose the best oxidant for the nucleobase and the nucleoside of 5-hmC, they will both be subjected to several oxidants.
The second part of the study aims to find the best system for the conjugation of the linker with the 5-foC. For the study we chose hydroxylamine and hydrazine functional groups to form an oxime and hydrazone, respectively. The option of an imine was ruled out due to the lack of stability compared to the oximes and hydrazones. Indeed, the higher nucleophilicity of the hydroxylamine and hydrazine compared to the amines makes it generate more stable products. Moreover the generation of a resonance structure, due to the presence of the oxygen, increases the stability of the oximes and hydrazones. [113] Testing of the conjugation will first be done on the oxidized nucleobase of 5-hmC followed by the oxidized nucleoside. This will allow for choosing the most stable conjugate for the next step, the development of the linker.

The following part consists of creating a long linker with biotin on one end and the hydroxylamine or hydrazine group on the other. The length of the linker is very important here. It is essential to have a linker long enough to avoid interactions between the biotin and the DNA strand, but the linker cannot be too long, as to avoid interaction with the vicinal linker in the case of nearby 5-hmC. We chose to use an 11-atom linker to fulfill those priorities. [114]
After the synthesis of the linker, it will be coupled with the oxidized nucleobase and nucleosides to test the efficiency of the labeling on the monomer.

The final part of developing the linker is to test the newly developed method on a DNA and RNA oligomer. The DNA and RNA strands will be synthesized using a DNA and RNA synthesizer, with the inclusion of 5-hmC in the sequence. The strand will then be oxidized to convert 5-hmC into 5-foC, and the biotin linker will be conjugated with the modified part of the sequence. The last step of the study is the detection of the attached biotin. Avidin and streptavidin have high affinity with biotin,
with dissociation constants on the order of $10^{-14} - 10^{-15}$ mol/L. In our technique, we chose to use streptavidin due to the affinity of avidin to undergo non-specific binding. [115] The reporter enzyme attached to streptavidin is Horseradish Peroxidase (HRP), and the substrate is 3,3',5,5'-Tetramethylbenzidine (TMB) or 3,3'-Diaminobenzidine (DAB). Another detection method after the biotin linker attachment could be LC-MS quantification.

Figure 2.4: Labelling and quantification of 5-hmC in the oligomer using the new method
2.3 Oxidation conditions

The investigation of the conversion of 5-hmC to 5-foC was done using several readily available oxidants. Common oxidants, such as cesium ammonium nitrate (Can) and potassium persulfate catalyzed with silver nitrate, were investigated, as well as other oxidants like allylic alcohol specific manganese dioxide and ruthenium dioxide [116].

2.3.1 Oxidation of the nucleobase

Table 2.1: Improvement of the condition of oxidation of 5-hmC nucleobase

<table>
<thead>
<tr>
<th>Entry</th>
<th>Oxidant</th>
<th>Nature</th>
<th>Number of equivalent</th>
<th>Solvent</th>
<th>Conditions</th>
<th>Conversion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ce(NH₄)₂(NO₃)₆</td>
<td>5</td>
<td>H₂O</td>
<td>r.t, 24 h</td>
<td>98-100%</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>K₂S₂O₈ Cat. AgNO₃</td>
<td>1.8</td>
<td>H₂O</td>
<td>r.t, 24 h</td>
<td>95%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>RuO₂.xH₂O</td>
<td>6.7</td>
<td>H₂O</td>
<td>r.t, 24 h</td>
<td>97%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>activated MnO₂</td>
<td>9.7</td>
<td>DCM/Dioxane</td>
<td>65 °C, 24 h</td>
<td>87 %</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>activated MnO₂</td>
<td>17.6</td>
<td>CDCl₃</td>
<td>r.t, 48 h</td>
<td>60 %</td>
<td></td>
</tr>
</tbody>
</table>
The CAN oxidant, potassium persulfate with catalytic silver nitrate, and ruthenium oxide (table 1, entry 1 to 3) were the best oxidants for the full conversion of the alcohol. Indeed, their conversions are respectively 98%, 95%, and 97%, which is good enough for an ideal labelling of the 5-hmC present. Manganese dioxide (table 1, entry 4 and 5) does not give a conversion effective enough to use in the development of a highly precise method. This is mainly due to the insolubility of manganese dioxide in water, which called for the use of more compatible solvents that do not give the same results as the oxidation carried out in water.

The CAN oxidant, potassium persulfate with a catalytic amount of silver nitrate, and ruthenium dioxide were chosen to investigate the oxidation on the nucleoside.

2.3.2 Oxidation of the nucleoside

Table 2.2: Improvement of the condition of oxidation of 5-hmC nucleoside

<table>
<thead>
<tr>
<th>entry</th>
<th>oxidant</th>
<th>Nature</th>
<th>Number of equivalent</th>
<th>Solvent</th>
<th>Conditions</th>
<th>Conversion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ce(NH₄)₂(NO₃)₆</td>
<td>4</td>
<td></td>
<td>H₂O</td>
<td>r.t, 15h</td>
<td>degradation</td>
</tr>
<tr>
<td>2</td>
<td>K₂S₂O₈ Cat. AgNO₃</td>
<td>1.8</td>
<td></td>
<td>H₂O</td>
<td>r.t, 15h</td>
<td>degradation</td>
</tr>
<tr>
<td>3</td>
<td>RuO₂.xH₂O</td>
<td>18.2</td>
<td></td>
<td>H₂O</td>
<td>rt, 72h</td>
<td>&gt;99%</td>
</tr>
</tbody>
</table>
Neither the CAN, nor the potassium persulfate with a catalytic amount of silver nitrate, generates the 5-formylcytidine (table 2.2, entry 1 and 2). However, the ruthenium oxide hydrate was able to oxidize the 5-hydroxymethylcytidine with a conversion above 99% (table 1, entry 8). This result is due to the selectivity of the ruthenium oxide hydrate oxidizing the allylic alcohol. The non-selectivity of the other oxidant (table 2.2, entry 1 and 2) induced the degradation of 5-hmC due to oxidations of the 2', 3', and 5' positions of the ribose. The following mechanisms of the C5 position of the nucleobase can explain the selectivity of the oxidation:

![Figure 2.5: Proposed mechanism for selective allylic oxidation.](image)

This proposed mechanism is inspired by the known selenium dioxide allylic specific oxidation. [117] The first step is a 2, 3 ene reaction between the ruthenium oxide and the 5-formylcytosine. The ruthenium dioxide attachment to the allylic position forms a highly stable compound (24) where the ruthenium metal coordinates with the alcohol in order to stabilize the intermediate. The ruthenium dioxide abstraction of the alcohol hydrogen allows the formation of the 5-foC and the release of the ruthenium (II) dihydroxide.
2.3.3 Verification of the oxidation conditions on the oligomers

After finding the oxidation conditions that work to fully convert 5-hmC into 5-foC, it is essential that these conditions do not affect the other nucleosides present in DNA and RNA to guarantee the specificity of the oxidation in the cell. The next step is to test the oxidation of DNA and RNA nucleobases, and, if necessary, change the conditions until there is no damage to any other nucleoside. Guanosine is known to be the most reactive nucleoside, so the investigation logically begins with testing guanosine.

Table 2.3: Test of the oxidation condition on other nucleosides.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Oxidized compound</th>
<th>Number of equivalent RuO$_2$·xH$_2$O</th>
<th>Conditions</th>
<th>Degradation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1" alt="化合物1" /></td>
<td>18.2</td>
<td>H$_2$O, rt, 72h</td>
<td>&gt;99% conversion</td>
</tr>
<tr>
<td>2</td>
<td><img src="image2" alt="化合物2" /></td>
<td>18.2</td>
<td>H$_2$O, r.t.72h</td>
<td>15%</td>
</tr>
<tr>
<td>3</td>
<td><img src="image3" alt="化合物3" /></td>
<td>5.0</td>
<td>H$_2$O, 40°C, 4h</td>
<td>&gt;99% conversion</td>
</tr>
<tr>
<td></td>
<td>Formula</td>
<td>Conditions</td>
<td>Product</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---------</td>
<td>------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td><img src="image1.png" alt="Image" /></td>
<td>H₂O, 40°C, 16h</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td><img src="image2.png" alt="Image" /></td>
<td>H₂O, 40°C, 16h</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td><img src="image3.png" alt="Image" /></td>
<td>H₂O, 40°C, 16h</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td><img src="image4.png" alt="Image" /></td>
<td>H₂O, 40°C, 16h</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td><img src="image5.png" alt="Image" /></td>
<td>H₂O, 40°C, 16h</td>
<td>None</td>
<td></td>
</tr>
</tbody>
</table>

The oxidation conditions found in the previous section (table 2.3, entry 1) were used on guanosine (table 2.3, entry 2, 26), and a 15% degradation of the Guanosine was observed. A common byproduct generated from the oxidation of guanosine is 8-oxoguanine, which could be the generated product here.
Figure 2.6: Structure of 8-oxoguanine

It is essential to avoid its creation by having more mild conditions. The 72 hr oxidation could be the reason for the degradation, so it is essential to shorten the duration of the reaction. The high number of equivalents of ruthenium could also be the reason for the degradation. It is necessary to reduce this to a more reasonable amount as well. To be able to reduce the quantity of oxidant and the duration of the oxidation, it is necessary to increase the temperature. The temperature needs to be monitored closely here to avoid damage to the nucleoside. Indeed, high temperature and low pH are the main cause of DNA/RNA depurination and degradation. To avoid those processes, it is recommended to avoid long exposure above 50 degrees Celsius. [118]

The ruthenium oxide conditions used were with less oxidant (5 equivalents), shorter reaction time (16 hr), and a temperature within the stability range (40 degrees celius) of DNA and RNA (table 1, entry 9). The application of the new conditions on 5-hydroxymethylcytidine gave full conversion to its oxidized form (table 2.3, entry 3). Those conditions were tested on the other nucleosides of DNA; 2-deoxyguanosine (26), 2-deoxyadenosine (27), thymidine (28), 2-deoxycytidine (29), and 2-deoxyuridine (30) (table 2.3, entry 4-8) and no degradation was observed.

The new oxidation conditions are ideal for the development of the oxidative labelling in the oligomers of DNA and RNA.
2.4 Conjugation investigation

The next step of our technique is to conjugate our 5-formylcytosine 4 to an oxime or hydrazone linker attached to a biotin motif. It is necessary to use a conjugate that is stable with full conversion to be able to validate this part of the method.

2.4.1 Conjugation with the nucleobase

In order to test the stability of the conjugation with the nucleobase, the 5-formylcytosine 22 was added to several hydroxylamines and hydrazines.

Table 2.4: Conjugation of oximes and hydrazones to 5-formylcytosines

<table>
<thead>
<tr>
<th>Entry</th>
<th>Conjugate</th>
<th>Product</th>
<th>Conditions</th>
<th>Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MeONH₂.HCl</td>
<td><img src="image1.png" alt="Image" /></td>
<td>H₂O/EtOH r.t., 15h</td>
<td>68.7%</td>
</tr>
<tr>
<td>2</td>
<td>BnONH₂.HCl</td>
<td><img src="image2.png" alt="Image" /></td>
<td>H₂O/EtOH r.t., 15h</td>
<td>62.8%</td>
</tr>
<tr>
<td>3</td>
<td>BnNHNH₂.HCl</td>
<td><img src="image3.png" alt="Image" /></td>
<td>H₂O/EtOH r.t., 38h</td>
<td>20%</td>
</tr>
</tbody>
</table>
Full conversion was obtained for the conjugation of methoxyamine hydrochloride, benzyloxyamine hydrochloride, and benzylhydrazine (table 2.4, entry1-3). A white solid was formed with the formation of compounds 32 and 33. The pure compounds were obtained by centrifuging the mixture and washing the solid with ethanol then EtOAc/Hex 1:1. Centrifuge precipitation is a sensitive process where a minimum amount of solid can be dissolved in the solvent used for the wash or where a small amount of solid can be withdrawn while removing the solvent from the vial with a pipet. This explains the yields of compound 32 and 33 to be respectively, 68.7% and 62.8%, not higher. Compound 34 was purified by recrystallization in methanol. Loss of product during recrystallization is common, and it is difficult to have high yields using this technique, whereas the main advantage is the generation of an extremely pure product. Compound 34 was obtained with a 20% yield. The conjugation works efficiently on the nucleobase, so the work can now be tested on RNA and DNA nucleosides.

2.4.2 Conjugation with RNA nucleosides

The presence of the the 2’-OH renders the ribose more acid and base labile. Acidic or alkaline conditions can lead to base elimination. The hydroxylamines and hydrazines used are all hydrochloride salts, which, once dissolved in solution, create a very acidic solution with a pH close to 1.0. For that reason, it is essential to work in a controlled environment using a phosphate buffered saline solution of pH 7.0, a commonly used buffer in biological research.
Table 2.5: Conjugation of oximes and hydrazones to 5-formylcytidine

Coupling of 5-formylcytidine (20) with methoxyamine hydrochloride, benzylamine hydrochloride, and benzylhydrazine hydrochloride was successfully done with a full conversion into compounds 35, 36, and 37 respectively (table 2.5, entry 1-3). Purification was then done using a polyacrylamide biogel p2 column. The principle

<table>
<thead>
<tr>
<th>Entry</th>
<th>Conjugate</th>
<th>Product</th>
<th>Conditions</th>
<th>Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MeONH₂.HCl</td>
<td><img src="image1.png" alt="Product 35" /></td>
<td>PBS Buffer 7.0 r.t., 24h</td>
<td>90%</td>
</tr>
<tr>
<td>2</td>
<td>BnONH₂.HCl</td>
<td><img src="image2.png" alt="Product 36" /></td>
<td>PBS Buffer 7.0 r.t., 24h</td>
<td>86%</td>
</tr>
<tr>
<td>3</td>
<td>BnNHNH₂.HCl</td>
<td><img src="image3.png" alt="Product 37" /></td>
<td>PBS Buffer 7.0 r.t., 24h</td>
<td>Not isolable</td>
</tr>
</tbody>
</table>
of this column purification, also known as size exclusion, relies on the differences in mass and size of the molecules. The small molecules are trapped in the pores, whereas the bigger molecules cannot fit and elute faster. The solvent generally used for this type of column is water, but the gel can tolerate up to 20% organic solvent. The mass difference is proportional to the quality of the separation. The purification of compound 35 (table 2.5, entry 1) gave a relatively high yield of 90 % due to the high mass difference of methoxylamine (MW=47 g/mol) and compound 35(MW=300 g/mol). The purification of compound 36 (table 2.5, entry 2) gave a yield of 86% due to a difference between benzylhydroxylamine (MW=123 g/mol) and compound 36 (MW= 376 g/mol). Compound 37 was not able to be isolated by a p2 column. The hypothesis is that the long duration of a p2 column (up to 7 hours) in aqueous environment causes the hydrazine to hydrolyze.

The conjugation investigation needs to be done on the DNA nucleoside to determine if it reacts the same way as the RNA nucleoside.

2.4.3 Conjugation with DNA nucleosides

The conjugation to the oxidized modified cytosine was done on 2-deoxy-5-formylcytidine, which is conjugated with the methoxylamine hydrochloride, benzylhydroxylamine hydrochloride, and benzylhydrazine hydrochloride as in the previous studies. The oxidations were carried out in PBS buffer at pH 7.0 for the same stability reason as the RNA nucleosides.
**Table 2.6:** Conjugation of oximes and hydrazones to 2-deoxy-5-formylcytosine.

![Chemical structures](image)

<table>
<thead>
<tr>
<th>Entry</th>
<th>Conjugate</th>
<th>Product</th>
<th>Conditions</th>
<th>Yield</th>
</tr>
</thead>
</table>
| 1     | MeONH₂·HCl         | ![Product 39](image) | PBS Buffer 7.0  
r.t., 24h | 91%    |
| 2     | BnONH₂·HCl         | ![Product 40](image) | PBS Buffer 7.0  
r.t., 24h | 68%    |
| 3     | BnNHNH₂·HCl        | ![Product 41](image) | PBS Buffer 7.0  
r.t., 24h | Not isolable |
Full conversion was obtained for the conjugation of all three species to the 2-deoxy-5-formylcytosine (table 2.6, entry 1-3). The compounds were purified on a p2 biogel column as was done previously. The lower yield for compound 40 (table 2.6, entry 2) is explained by a smaller mass difference between benzylhydroxylamine (MW=123 g/mol) and compound 40 (MW= 360 g/mol). The conjugated hydrazine 41 (table 2.6, entry 3) was still not able to be isolated. Hydrazone was referenced to be stable for a maximum of a few hours at pH 7.0, while the oxime can be stable for days. Due to the duration of the column, the hydrazine derivative degraded and could not be isolated. [113]

Consequently, oximes are the obvious choice for the development of our quantification method.

2.5 Development of the linker

The results described earlier suggest the use of an oxime linker from the perspective of stability. The biotin tag was chosen for the detection of the 5-hmC due to its economic accessibility and its sensitive detection using streptavidin coupled to horseradish peroxidase. Biotin-NHS ester (42) is a common species used for the synthesis of biotin containing molecule due the high reactivity of the ester group. The spacer we chose to use is a 5 carbon spacer with a hydroxylamine and an amine on each end (43).
Figure 2.7: Compounds for the synthesis and conjugation of the linker.

The order of coupling of the three compounds is important for the stability of the nucleoside. The spacer needs to be protected in order to control the regioselectivity of the reactions. Deprotection often requires highly acidic or basic conditions that are incompatible with the nucleosides, which is why it is essential to couple the Biotin NHS ester and the spacer first and then follow with the conjugation with the oxidized 5-hmC species. A phtalamide group is used to protect the hydroxylamine due to the ease of synthesis using an alcohol and N-hydroxyphtalamide. [119] The deprotection of this group can be done using hydrazine hydrate [120], methylamine [120], or sodium borohydride reduction followed with acetic acid treatment [121]. It was essential to choose a protecting group for the amine with different conditions of deprotection to avoid any undesired deprotection of the phtalamide. *Tert*-butyl carbamate protection (boc-protection) was chosen for the amine. A widely used and successful method of deprotection of boc groups is the use of trifluoroacetic acid, which will not affect the phtalamide protection [122].

2.5.1 Development of the biotin hydroxylamine linker

The synthesis of the linker starts with a Mitsunobu-like reaction between the N-hydroxyphthalimide (44 and 5-(boc-amino)-1-pentanol (45) to obtain compound **46** with a Boc-protected amine on one end and a phthalimide protected hydroxylamine on
the other. The next step is a Boc deprotection, which uses trifluoroacetic acid to obtain compound 47 as an amine hydrochloride salt. The amine is then coupled to biotin NHS ester (42) in order to obtain the phtalamide-protected hydroxylamine biotin linker 48. The last step is a phtalamide deprotection in order to obtain the biotin hydroxylamine linker 49.

**Figure 2.8**: Synthesis of the biotin linker

The Mitsunobu-like reaction is easy to carry out, and the compound is formed rapidly. The purification can be difficult due to the triphenylphosphine oxide and the diisopropyl azodicarboxilate (DIAD), which are difficult to remove using column
chromatography. The problem can be overcome during the deprotection step using the high affinity of the newly formed HCl salt (47) with the aqueous environment and the high affinity of DIAD with organic solvents.

In the rest of the synthesis, high polarity solvents (DMSO and DMF) are required due to the low solubility of biotin in other solvents. This is useful in the coupling step, using the high miscibility of water and DMF to obtain the phtalamide protected biotin linker (42). Indeed, compound (48) is not soluble in water so it precipitates as a pure solid upon addition of water. Phtalamide deprotection of compound 48 causes more problems due to the production of its byproduct phtalahydrazide.

![Figure 2.9: Phtalamide deprotection byproduct](image)

The solubility issues of phtalahydrazide and compound 49 make the purification difficult. Indeed, both are almost insoluble in a majority of common solvents. It was observed that the phtalahydrazide was not soluble in DMF, but after precipitation, there are still traces of up to 15% present of the byproduct in compound 49. A p2 biogel size exclusion in alkali environment to solubilize both compounds is not useful for this purification due to a small mass difference. The purification was done by transforming compound 49 into an HCl salt to make it fully soluble in water. The high solubility of compound 49 in water in contrast with the byproduct allowed the isolation of a pure product.
2.5.2 Conjugation of the linker with the nucleobase and nucleosides of 5-hmC

The newly formed linker was coupled with the different oxidized derivatives of 5-hmC.

![Conjugation diagram](image)

**Figure 2.10:** Conjugation of the biotin linker to the oxidized 5-hmC species.

The coupling was done under controlled pH (pH=5-7) with the nucleobase and the DNA nucleoside. The mass of the linker coupled to both species has been observed using mass spectrometry (ESI). Purification was attempted with a p2 biogel size exclusion column and but unfortunately delivered inconclusive results. The total solubility of the new species is still unknown in water, and partial solubility could be the reason for the failure of the purification. Monitoring conjugation by LC-MS could be a pathway to determine the extent of attachment of the linker to the oxidized species.
Chapter 3

Conclusion

A new oxidative labelling method of quantification of 5-hmC is currently under development. The first optimization study allowed to determinate the best oxidation condition for the full conversion of the allylic alcohol of the modified hydroxymethylated species into the corresponding aldehyde without affecting the other DNA and RNA nucleosides. The final conditions that will be applied to the DNA and RNA strand uses 5 equivalent of ruthenium oxide at a monitored temperature of 40 degrees Celsius. The step following the oxidation of 5-hmC into 5-foC is the conjugation with a hydroxylamine tag to obtain an oxime, the most stable conjugate according to the second optimization study comparing the stability of oximes and hydrazone. The tag chosen for the development of the method is a biotin due to its financial accessibility and its sensitive detection. The next step of the study was the synthesis of the hydroxylamine biotin, which was successfully done with the synthesis of compound 49. The newly synthesized hydroxylamine linker was successfully coupled with oxidized 5-hmC nucleobase and DNA nucleoside.

The method still need further work for validation.
Chapter 4

Future work

Validation of the linker with the oligomer

The newly developed oxidative labeling method need to be tested with DNA and RNA strand containing 5-hmC.

DNA sequence containing 5-hydroxymethylcytosine needs to be synthesized using a DNA synthesizer. It is important to test the ability for the linker to be attached to vicinal hydroxymethylated group and also to a large amount of hydroxymethylation. To carry the study the optimized oxidation conditions previously determined are used to oxidized the 5-hmC sites into 5-foC. The stability of the oligomer and the full conversion of the allylic alcohol are keys of this step. The conjugation of the oxidized DNA strand can be monitored by LC-MS to follow the full conversion of the reaction. A test using biotin enrichment and detection using strepatavidin coupled to horseradish peroxidase (HRP) should also be done to validate the biotin quantification choice.

The same work has to be done with RNA hydroxymethylated sequence. The position of the 5-hmC in the strand is important. Indeed 5-hmC needs to be added on the stem and on the loop with different concentration to be sure that the new method is totally valid to be used to quantify hydroxymethylated RNA.
Chapter 5

Experimental section

3.1 General methods

Reactions were carried out under standard laboratory conditions. All glassware were carefully washed and dried before reaction. The water used was preferentially nanopure water, which was obtained in the laboratory using a purelab ultra from ELGA.

All reagents were purchased commercially and used without further purification. Solvents were purchased from Fisher Scientific and Sigma-Aldrich and were ACS reagent grade unless otherwise indicated.

The HPLC used is a Dionex Ultimate 3000 and the solvent used was HPLC grade for the acetonitrile and filtered aqueous solution. The 1M TEAA buffer used was prepared in the laboratory using commercial triethylamine and acetic acid. The column used was a C18 reversed phase and the detector was a UV-Vis polychromatic detector.

The NMR preferentially used for the analysis was an Avance 600 constructed by Brucker, and the software used was TopSpin. The NMR solvents used were all deuterated and anhydrous stored in individual phial.
3.2 Experimental

Ceric ammonium nitrate (Ce(NH₄)₂(NO₃)₆, potassium persulfate with a catalytic amount of sodium nitrate (K₂S₂O₈ cat. AgNO₃), ruthenium oxide hydrate (RuO₂·xH₂O), and activated manganese oxide (MnO₂) (1.8 to 9.7 eq) were added to 5-hydroxymethylcytosine (1.0 eq, 18) dissolved in water. The oxidation was monitored by HPLC. 5-Formylcytosine was obtained as a white solid. (21).

Ceric ammonium nitrate, potassium persulfate with a catalytic amount of silver nitrate and ruthenium oxide hydrate (RuO₂·xH₂O) (MnO₂) (1.8 to 18.2 eq) were added to 5-hydroxymethylcytidine (1.0 eq, 20) dissolved in water. The oxidation was monitored by HPLC. 5-Formylcytidine (22) was obtained as a white solid.
3.2.1 Synthesis of (E)-4-amino-2-oxo-1,2-dihydropyrimidine-5-carbaldehyde O-methyl oxime

![Diagram of compound 32]

5-Formylcytosine (30 mg, 0.22 mmol, 21) and N-methylhydroxylamine hydrochloride (45.1 mg, 0.54 mmol) were dissolved in 2 ml of a 1:1 mixture of water and ethanol. Sodium bicarbonate (45.3 mg, 0.54 mmol) was added to the mixture, and the reaction was stirred at room temperature for 15 hours. The reaction mixture was then centrifuged, washed with 0.2 ml of ethanol and 1 ml of a 1:1 mixture of ethyl acetate and hexane, and dried under a high-pressure vacuum. Compound 32 (25 mg, 0.15 mmol) was obtained as a white solid in 69% yield.

$^1$H NMR (DMSO): $\delta$ 8.09 (1H, s), 7.86 (1H, s), 3.82 (3H, s) $^{13}$C NMR (DMSO): 157.97, 150.53, 147.51, 145.60, 98.15, 62.67 HRMS [M+1]$: \text{calc. for C}_{8}\text{H}_{9}\text{N}_{4}\text{O}_{2}$, 169.0726; found 169.0709

3.2.2 Synthesis of (E)-4-amino-2-oxo-1,2-dihydropyrimidine-5-carbaldehyde O-benzyl oxime

![Diagram of compound 33]

5-Formylcytosine (10 mg, 0.072 mmol, 21) and O-benzylhydroxylamine (20 mg, 0.162 mmol) were dissolved in 0.5 ml of a 1:1 mix of water and ethanol. The reaction was stirred at room temperature for 15 hours. The reaction mixture was then
centrifuged, washed with 0.5 ml of ethanol and 1 ml of hexane, and dried under a high-pressure vacuum. Compound 33 (22 mg, 0.090 mmol) was obtained with a 63 % yield.

$^1$H NMR (DMSO): $\delta$ 8.11 (1H, s), 7.80 (1H, s), 7.35 (5H, m), 5.08 (2H, s) $^{13}$C NMR (DMSO): 163.09, 155.03, 148.41, 147.76, 137.96, 128.79, 128.26, 128.32, 97.79, 97.79, 75.76 HRMS [M+1]$^+$: calc. for C$_{12}$H$_{13}$N$_4$O$_2$, 245.1039; found 245.1037

### 3.2.3 Synthesis of (E)-4-amino-5-((2-benzylhydrazineylidene) methyl)pyrimidin-2(1H)-one

![34](image)

5-formylcytosine (3.1 mg, 0.022 mmol, 21) and O-benzylhydrozone dihydrocholride (9.6 mg, 0.049 mmol) were dissolved in 1 ml of a 1:1 mix of water and ethanol. Sodium bicarbonate (8.2 mg, 0.098 mmol) was added, and the mixture was stirred at room temperature for 38 hours. The solvent was then removed by rotavap and dried under a high-pressure vacuum. The solid was then recrystallized in methanol. Compound 34 (1.5 mg, 0.007 mmol) was obtained in a 30 % yield.

$^1$H NMR (MeOD): $\delta$ 7.92 (1H, s), 7.67 (1H, s), 7.37 (4H, m), 5.30 (1H, s), 4.40 (2H, s) $^{13}$C NMR (MeOD): 163.09, 155.03, 148.41, 147.76, 137.96, 128.79, 128.26, 128.32, 97.79, 97.79, 75.76 HRMS [M+1]$^+$: calc. for C$_{12}$H$_{14}$N$_5$O, 244.1198; found 244.1190
3.2.3 Synthesis of (E)-4-amino-1-((2R,3S,4R,5R)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-2-oxo-1,2-dihydropyrimidine-5-carbaldehyde O-methyl oxime

5-Formylcytidine (2.71 mg, 0.010 mmol, 20) and N-methylhydroxylamine hydrochloride (2.5 mg, 0.030 mmol) were dissolved in 2 ml of PBS buffer 7.0. Triethylamine (3.1 mg, 0.030 mmol) was added to the mixture, and the reaction was stirred at room temperature for 15 hours. The triethylamine was removed by rotavap and the mixture was frozen and placed in the lyophilizer. The purification was done with on a biogel p2 size exclusion column. Compound 35 (2.07 mg, 0.007 mmol) was obtained with a 69% yield.

$^1$H NMR (D$_2$O): $\delta$ 8.12 (1H, s), 7.91 (1H, s), 5.76 (1H, d, J = 2.88 Hz), 4.17 (1H, q, J = 4.98 Hz, J= 3.0 Hz), 4.08 (1H, t, J= 5.16 Hz), 4.03 (1H, m), 3.81 (3H, s), 3.80 (2H, m). $^{13}$C NMR (D$_2$O): 162.34, 155.57, 146.57, 145.13, 100.55, 90.71, 83.50, 74.29, 68.46, 61.96, 59.97 HRMS [M+1]+: calc. for C$_{11}$H$_{16}$N$_4$O$_6$, 301.1148; found 301.1152
3.2.5 Synthesis of \((E)-4\text{-amino-1-}\text{((2R,3S,4R,5R)-3,4-dihydroxy-5-}
\text{(hydroxymethyl)tetrahydrofuran-2-yl)-2-oxo-1,2-dihydropyrimidine-5-}
\text{carbaldehyde O-benzyl oxime}

\[
\text{HO} \quad \text{OH} \quad \text{NH}_2
\]

5-Formylcytidine (3.0 mg, 0.011 mmol, 20) and benzylhydroxylamine (4.2 mg, 0.034 mmol) were dissolved in 2 ml of PBS buffer pH=7.0. The reaction was stirred at room temperature for 15 hours. The mixture was frozen and placed in the lyophilizer. The purification was done with a biogel p2 size exclusion column chromatography. Compound 36 (2.61 mg, 0.007 mmol) was obtained in 63 % yield.

\(^1\)H NMR (D\(_2\)O): \(\delta 8.11 (1\text{H, s}), 7.96 (1\text{H, s}), 7.33 (5\text{H, m}), 5.76 (1\text{H, d, } J = 2.82 \text{ Hz}), 5.09 (2\text{H, s}), 5.16 (1\text{H, m}), 4.07 (1\text{H, t, } 5.16 \text{ Hz}), 4.01 (1\text{H, m}), 3.79 (2\text{H, m}). \(\text{^13}\)C NMR (D\(_2\)O): 162.39, 155.60, 147.30, 145.34, 136.80, 128.76, 128.75, 128.54, 100.58, 90.73, 83.51, 74.24, 68.46, 59.98 HRMS [M+1]^+: calc. for C\(_{17}\)H\(_{20}\)N\(_4\)O\(_6\), 377.1461; found 377.1466
3.2.6 Synthesis of 4-amino-5-((E)-(2-benzylhydrizineylidene)methyl)-1-((2R,3S,4R,5R)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)pyrimidin-2(1H)-one

5-Formylcytidine (3.0 mg, 0.011 mmol, 20) and benzylhydrazine dihydrochloride (6.2 mg, 0.032 mmol) were dissolved in 2 ml of PBS buffer 7.0. Triethylamine (6.46 mg, 0.064 mmol) was added to the mixture and the reaction was stirred at room temperature for 15 hours. The triethylamine was removed by rotavap, and the mixture was frozen and placed in the lyophilizer. The purification was done on a biogel p2 size exclusion column chromatography. Compound 37 could not be isolated.

3.2.7 Synthesis of (E)-4-amino-1-((2R,4R,5R)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-2-oxo-1,2-dihydropyrimidine-5-carbaldehyde O-methyl oxime

2’-Deoxy-5-(formyl)cytidine (2.9 mg, 0.011 mmol, 38) and N-methylhydroxylamine hydrochloride (2.8 mg, 0.033 mmol) were dissolved in 2 ml of PBS buffer pH=7.0. Triethylamine (3.3 mg, 0.033 mmol) was added to the mixture and the reaction was
stirred at room temperature for 15 hours. The triethylamine was removed by rotavap, and the mixture was frozen and placed in the lyophilizer. The purification was done with a biogel p2 size exclusion column chromatography. Compound 39 (2.84 mg, 0.010 mmol) was obtained in 91 % yield.

$^1$H NMR (D$_2$O): $\delta$ 8.05 (1H, s), 7.96 (1H, s), 6.12 (1H, t, $J = 6.30$ Hz), 4.31 (1H, m), 3.97 (1H, m), 3.82 (3H, s), 3.71 (2H, m), 2.27 (2H, m). $^{13}$C NMR (D$_2$O): 162.41, 155.66, 146.80, 145.19, 100.56, 86.81, 61.94, 60.85, 48.78, 39.66 HRMS [M+1]$^+$: calc. for C$_{11}$H$_{16}$N$_4$O$_5$, 285.1199; found 285.1196

3.2.8 (E)-4-amino-1-((2R,4R,5R)-4-hydroxy-5-(hydroxymethyl)tetrahydrofur-2-yl)-1,2-dihydropyrimidine-5-carbaldehyde O-benzyl oxime

2'-Deoxy-5-(formyl)cytidine (3.0 mg, 0.012 mmol, 38) and benzylhydroxylamine (4.2 mg, 0.036 mmol) were dissolved in 2 ml of PBS buffer pH=7.0. The reaction was stirred at room temperature for 15 hours. The mixture was frozen and placed in the lyophilizer. The purification was done with a biogel p2 size exclusion column chromatography. Compound 40 (2.93 mg, 0.008 mmol) was obtained in 68 % yield.

$^1$H NMR (D$_2$O): $\delta$ 8.04 (1H, s), 7.98 (1H, s), 7.33 (5H, m), 6.12 (1H, t, $J = 6.30$ Hz), 5.09 (2H, s), 4.30 (1H, m), 3.96 (1H, m), 3.69 (2H, m), 2.27 (2H, m). $^{13}$C NMR (D$_2$O): 162.36, 147.40, 145.31, 136.81, 128.76, 128.75, 128.53, 100.53, 86.82, 86.63,
76.23, 70.00, 60.81, 39.66 HRMS [M+1]$^+$: calc. for C$_{17}$H$_{20}$N$_4$O$_5$, 361.1512; found 361.1521

3.2.9 Synthesis of (2R,3R,5R)-5-(4-amino-5-((E)-(2-benzylhydrazineylidene)methyl)pyrimidin-1(2H)-yl)-2-(hydroxymethyl)tetrahydrofuran-3-ol

2’-Deoxy-5-(formyl)cytidine (3.0 mg, 0.012 mmol, 38) and benzylhydrazine dihydrochloride (7.02 mg, 0.036 mmol) were dissolved in 2 ml of PBS buffer 7.0. Triethylamine (6.46 mg, 0.072 mmol) was added to the mixture, and the reaction was stirred at room temperature for 15 hours. The triethylamine was removed by rotavap, and the mixture was frozen and placed in the lyophilizer. The purification was done with a biogel p2 size exclusion column chromatography. Compound 41 could not be isolated.
3.2.10 Synthesis of \(2\)-((5-aminopentyl)oxy)isoindoline-1,3-dione hydrochloride

The synthesis of compound 47 was performed using published experimental. [119]

3.2.11 Synthesis of \(N\)-(5-((1,3-dioxoisindolin-2-yl)oxy)pentyl)-5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanamide

Compound 47 (36.5 mg, 0.15 mmol), and biotin NHS ester (50.0 mg, 0.15 mmol, 42) were dissolved in 2.5 mL of DMF. Triethylamine (21 µl, 0.15 mmol) was added to the mixture, and stirred at room temperature for 12 hours. Water was added to the mixture and a white precipitate appeared. After filtration and water wash of the precipitate, compound 48 was obtained (52.2 mg, 0.11 mmol) in 75 % yield.

\(^1\)H NMR (DMSO): δ 7.86 (4H, s), 7.77 (1H, t, 5.46 Hz), 6.43 (1H, s), 6.36 (1H, s), 4.29 (1H, m), 4.12 (3H, t, 6.48 Hz), 3.34 (2H, s), 3.10 (1H, m), 3.05 (2H, q, J=12 Hz, J=6.18 Hz), 2.81 (1H, m), 2.58 (2H, m), 2.05 (2H, t, J= 7.32 Hz), 1.68 (2H, m), 1.61 (1H, m), 1.51 (2H, m), 1.43 (4H, m), 1.30 (2H, m).\(^{13}\)C NMR (DMSO): 172.27, 163.78, 163.16, 135.23, 129.10, 123.69, 78.10, 61.49, 59.64, 55.91, 38.64, 35.69, 29.29, 28.70, 28.52, 27.87, 25.81, 23.05HRMS \([M+1]^+\): calc. for \(C_{23}H_{30}N_4O_5S\), 475.2015; found 475.2012
3.2.12 Synthesis of N-(5-(aminooxy)pentyl)-5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanamide

Compound 48 (52.2 mg, 0.11 mmol) was dissolved in 2 mL DMF. Hydrazine hydrate (20 µl, 0.64 mmol) was added, and the reaction was stirred for 20 hours at room temperature. A white precipitate appeared and the mixture was filtered out through a PTFE syringe filter. The DMF was removed by rotavap under high-pressure vacuum. The remaining solid was dissolved in a 1M HCl solution to form the corresponding HCl salt of compound 49. The aqueous phase was washed twice with ethyl acetate and the residual solid was filtered out through a cellulose acetate syringe filter. The aqueous phase was evaporated by rotavap and compound 49 was obtained as a brown solid in 90% yield (37 mg, 0.099 mmol).

\(^1\)H NMR (CH\(_3\)OD): \(\delta 4.67 \text{ (1H, m), } 4.47 \text{ (1H, m), } 4.06 \text{ (2H, t, } J= 6.42 \text{ Hz), } 3.24 \text{ (2H, t, } J= 7.02 \text{ Hz), } 3.00 \text{ (1H, m), } 2.80 \text{ (1H, m), } 2.29 \text{ (2H, m), } 1.74 \text{ (5H, m), } 1.58 \text{ (4H, m), } 1.47 \text{ (4H, m).} \)
\(^13\)C NMR (CH\(_3\)OD): 175.05, 164.36, 125.15, 74.79, 62.89, 61.40, 61.32, 55.43, 39.24, 38.95, 35.03, 31.78, 28.48, 28.34, 28.31, 27.99, 26.94, 25.49, 22.86, 22.59. HRMS [M+1]: calc. for C\(_{15}\)H\(_{28}\)N\(_4\)O\(_3\)S, 345.1960; found 345.1971
3.2.13 Synthesis of (Z)-N-(5-(((4-amino-2-oxo-1,2-dihydropyrimidin-5-yl)methylene)amino)(oxy)pentyl)-5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanamide

![Chemical structure](image)

Compound 49 (13.2 mg, 0.035 mmol) was dissolved in 2 ml water. The pH was adjusted to pH=5 using a 1 M sodium bicarbonate solution. 5-formylcytosine (4.15 mg, 0.029 mmol) was added to in the vial and the reaction was carried at room temperature for 16h. Compound 23 was formed according to the mass spectroscopy ESI.

LRMS-ESI [M+1]^+; calc. for C_{20}H_{31}N_{7}O_{4}S, 466.2; found 466.1

3.2.14 Synthesis of N-((Z)-(4-amino-1-((2R,4R,5R)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-2-oxo-1,2-dihydropyrimidin-5-yl)methylene)amino)(oxy)pentyl)-5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanamide

![Chemical structure](image)

Compound 49 (9.1 mg, 0.035 mmol) was dissolved in 2 ml water. The pH was adjusted to pH=5 using a 1 M sodium bicarbonate solution. 2-Deoxy-5-formylcytidine
(5.1 mg, 0.020 mmol) was added to in the vial and the reaction was carried at room temperature for 16h. Compound 23 was formed according to the mass spectroscopy ESI

LRMS-ESI [M+1]^+: calc. for C_{25}H_{39}N_{7}O_{4}S, 582.3; found 582.2
References


23. Lorsbach, R., et al., *TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t (10; 11)(q22; q23)*. Leukemia (08876924), 2003. **17**(3).


Appendix A

Analysis of new compounds

Figure A-1: $^1$H NMR of compound 32

Figure A-2: $^{13}$C NMR of compound 32
Figure A-3: LRMS of compound 32

Figure A-4: HRMS of compound 32
Figure A-5: $^1$H NMR of compound 33

Figure A-6: $^{13}$C NMR of compound 33
Figure A-7: HRMS of compound 33

Figure A-8: $^1$H NMR of compound 34
Figure A-9: LRMS of compound 34

Figure A-10: HRMS of compound 34
Figure A-11: $^1$H NMR of compound 35

Figure A-12: $^{13}$C NMR of compound 35
Figure A-13: LRMS of compound 35

Figure A-14: HRMS of compound 35
Figure A-15: $^1$H NMR of compound 36

Figure A-16: $^{13}$C NMR of compound 36
Figure A-17: LRMS of compound 36

Figure A-18: HRMS of compound 36
Figure A-19: $^1$H NMR of compound 37

Figure A-20: $^{13}$C NMR of compound 37
Figure A-21: LRMS of compound 37

Figure A-22: HRMS of compound 37
Figure A-23: $^1$H NMR of compound 39

Figure A-24: $^{13}$C NMR of compound 39
Figure A-25: LRMS of compound 39

Figure A-26: HRMS of compound 39
Figure A-27: $^1$H NMR of compound 48

Figure A-28: $^{13}$C NMR of compound 48
Figure A-29: LRMS of compound 48

Figure A-30: HRMS of compound 48
Figure A-31: $^1$H NMR of compound 49

Figure A-32: $^{13}$C NMR of compound 49
Figure A-33: LRMS of compound 49

Figure A-34: HRMS of compound 49
Figure A-35: LRMS of compound 23

Figure A-36: LRMS of compound 24
Oxidation of other DNA nucleobase

Figure A-37: $^1$H NMR of guanosine 26

Figure A-38: $^1$H NMR of oxidized guanosine 26 with initial oxidation condition
Figure A-39: $^{1}$H NMR of 2d-guanosine 27

Figure A-40: $^{1}$H NMR of 2d-guanosine 27 with mild oxidation condition
Figure A-41: $^1$H NMR of 2d-adenosine 28

Figure A-42: $^1$H NMR of 2d-adenosine 28 with mild oxidation condition
Figure A-43: $^1$H NMR of 2d-cytidine 29

Figure A-44: $^1$H NMR of 2d-cytidine 29 with mild oxidation condition
Figure A-45: $^1$H NMR of thymidine 30

Figure A-46: $^1$H NMR of thymidine 30 with mild oxidation condition
Figure A-47: $^1$H NMR of 2d-uridine 31

Figure A-48: $^1$H NMR of 2d-uridine 31 with mild oxidation condition