

# Characterisation of DNA damage arising from nitrosated amines and aldehydes.

Thesis submitted for the degree of Doctor of Philosophy



Alec Guilhem Johns

Supervisor: Dr. D. M. Williams

The University of Sheffield- Department of Chemistry

June 2019

# Table of Contents

Acknowledgements.....	6
Abstract.....	7
Abbreviations.....	8
1. Introduction.....	12
1.1. Structure of DNA.....	12
1.2. DNA Replication.....	17
1.3. DNA Transcription and Translation.....	19
1.4. Chemical synthesis of DNA.....	20
1.5. DNA Damage and Mutations.....	23
1.6. Alkylation of Nucleobases.....	27
1.6.1. Alkylation at N-sites.....	29
1.6.2. <i>O</i> <sup>6</sup> -Alkylguanines.....	29
1.7. Colorectal Cancer and <i>N</i> -Nitroso Compounds.....	31
1.8. Synthesis of ODNs containing <i>O</i> <sup>6</sup> -Alkylguanine Adducts.....	33
1.9. DNA Repair Mechanisms.....	34
1.10. MGMT.....	36
1.11. MGMT Radioactive Assay.....	37
1.12. Alkyltransferase-Like (ATL) Proteins.....	39
1.13. Characterisation of DNA Containing Alkylated Base Adducts.....	40
1.14. Detection of DNA base adducts using Mass Spectrometry.....	42
1.15. Stable Isotope Dilution (SID).....	44
1.16. Previous Studies Using SID.....	45
1.17. Interstrand Crosslinks.....	46
1.18. Formaldehyde.....	47
1.19. Fanconi Anemia and ICL Repair.....	50
1.20. Project Aims.....	52
2. Recognition and Repair of <i>O</i> <sup>6</sup> -Alkylguanines Derived From Nitrosated Amino Acids	58
2.1. Introduction.....	58
2.2. Synthesis of <i>O</i> <sup>6</sup> -CEG Phosphoramidite.....	62
2.3. ODN deprotection and purification.....	67
2.4. Nucleoside composition analysis of ODNs containing <i>O</i> <sup>6</sup> -CMG, <i>O</i> <sup>6</sup> -CEG, <i>O</i> <sup>6</sup> -CMG carboxamide and <i>O</i> <sup>6</sup> -CEG carboxamide.....	71

2.4.1.	Nucleoside composition analysis of $O^6$ -CEG .....	73
2.4.2.	Nucleoside composition analysis of $O^6$ -CEG carboxamide .....	75
2.4.3.	Nucleoside composition analysis of $O^6$ -CMG .....	77
2.4.4.	Nucleoside composition analysis of $O^6$ -CMG carboxamide .....	78
2.4.5.	Analysis of nucleoside composition spectra .....	80
2.5.	MGMT assays.....	81
3.	Use of $^{15}\text{N}_5$ -labelled $O^6$ -Alkyl-2'-Deoxyguanosine For Quantification of Alkylation Damage Using MS.....	86
3.1.	Synthesis of $O^6$ -alkylguanine adducts .....	86
3.1.1.	Introduction.....	86
3.1.2.	Synthesis of labelled standards using 3', 5' silyl protecting groups.....	88
3.1.3.	Synthesis of labelled standards using 3', 5' acetyl protecting groups .....	90
3.1.4.	Synthesis of $O^6$ -methyl-2'-deoxyguanosine .....	93
3.1.5.	Synthesis of $O^6$ -EtG.....	96
3.1.6.	Synthesis of $O^6$ -CMG .....	96
3.1.7.	Synthesis of $O^6$ -CEG .....	97
3.2.	Synthesis of $^{15}\text{N}_5$ -labelled $O^6$ -alkylguanine adducts .....	98
3.2.1.	Introduction.....	98
3.2.2.	Synthesis of $^{15}\text{N}_5$ -labelled 6-mesitylenesulfonyl-3',5'-bis-O-acetyl-2'-deoxyguanosine.....	98
3.2.3.	Synthesis of $^{15}\text{N}_5$ -labelled $O^6$ -MeG .....	98
3.2.4.	Synthesis of $^{15}\text{N}_5$ -labelled $O^6$ -EtG.....	99
3.2.5.	Synthesis of $^{15}\text{N}_5$ -labelled $O^6$ -CMG .....	100
3.2.6.	Synthesis of $^{15}\text{N}_5$ -labelled $O^6$ -CEG .....	101
3.3.	Validation of SID method.....	103
3.3.1.	Introduction.....	103
3.3.2.	Calibration of SID method.....	104
3.3.3.	Limit of Quantification .....	108
4.	Synthesis of Formaldehyde-Induced Interstrand Crosslinks Between Adenine Nucleobases in Double-Stranded DNA .....	112
4.1.	Introduction .....	112
4.2.	Synthesis of bis-( $N^6$ -2'-deoxyadenosyl) methane using methylene diamine .....	112
4.3.	Synthetic route to formaldehyde-induced interstrand crosslinks using $N^6$ -[(dibutylamino)methylene]-2'-deoxyadenosine .....	115
4.4.	Synthesis of d(GpA <sup>dbf</sup> ).....	127
4.5.	Direct synthesis of formaldehyde-induced interstrand crosslink .....	133

4.5.1. Initial synthesis attempt .....	133
4.5.2. Secondary synthesis route.....	138
4.5.3. Nucleoside composition analysis .....	143
4.6. Formaldehyde-induced Interstrand Crosslink Within Drew-Dickerson Dodecamer 145	
4.7. <sup>1</sup> H NMR stability study of <i>bis</i> -( <i>N</i> <sup>6</sup> -2'-deoxyadenosyl) methane and <i>N</i> <sup>6</sup> - hydroxymethyl-2'-deoxyadenosine .....	148
4.7.1. Introduction.....	148
4.7.2. Synthesis of <i>N</i> <sup>6</sup> -hydroxymethyl-2'-deoxyadenosine and <i>bis</i> -( <i>N</i> <sup>6</sup> -2'- deoxyadenosyl) methane.....	149
4.7.3. <sup>1</sup> H NMR analysis of <i>N</i> <sup>6</sup> -hydroxymethyl-2'-deoxyadenosine .....	150
4.7.4. <sup>1</sup> H NMR analysis of <i>bis</i> -( <i>N</i> <sup>6</sup> -2'-deoxyadeosyl) methane.....	153
5. Conclusions and Future Work .....	160
5.1. Conclusions .....	160
5.2. Future Work.....	164
5.2.1. Detection of <i>O</i> <sup>6</sup> -CEG in human DNA .....	164
5.2.2. Development of SID methods for <sup>15</sup> N <sub>5</sub> -labelled SID internal standards .....	165
5.2.3. Determination of the crystal structure of ODN containing fICL.....	165
5.2.4. Development of antibodies for <i>bis</i> -( <i>N</i> <sup>6</sup> -2'-deoxyadenosyl) methane .....	166
6. Experimental.....	168
3', 5'- <i>Bis</i> - <i>O</i> - <i>tert</i> butyldimethylsilyl-2'-deoxyguanosine. ....	170
2-Amino-9-[2-deoxy-3,5- <i>bis</i> - <i>O</i> -( <i>tert</i> butyldimethylsilyl)-β-D-erythro-pentofuranosyl]-6- mesitylenesulfonyl-purine (2) .....	171
3',5'- <i>Bis</i> - <i>O</i> - <i>tert</i> -butyldimethylsilyl- <i>O</i> <sup>6</sup> -(methoxycarbonylethyl)-2'-deoxyguanosine (3) .....	172
<i>N</i> <sup>2</sup> -[(Dimethylamino)methylidene]-3',5'- <i>bis</i> - <i>O</i> - <i>tert</i> -butyldimethylsilyl- <i>O</i> <sup>6</sup> - (methoxycarbonylethyl)-2'-deoxyguanosine (4).....	173
<i>N</i> <sup>2</sup> -Formyl - <i>O</i> <sup>6</sup> -(methoxycarbonylethyl)-2'-deoxyguanosine (5).....	174
5'- <i>O</i> -(4,4'-Dimethoxytriphenylmethyl)- <i>N</i> <sup>2</sup> -formyl- <i>O</i> <sup>6</sup> -(methoxycarbonylethyl)-2'- deoxyguanosine (6).....	175
5'- <i>O</i> -(4,4'-Dimethoxytriphenylmethyl)- <i>N</i> <sup>2</sup> -formyl- <i>O</i> <sup>6</sup> -(methoxycarbonylethyl)-2'- deoxyguanosine-3'- <i>O</i> -(2-cyanoethyl- <i>N,N</i> -diisopropylamino)-phosphoramidite (7) .....	176
3',5'- <i>Bis</i> - <i>O</i> -acetyl-2'-deoxyguanosine (9) .....	177
6-Mesitylenesulfonyl-3', 5'- <i>bis</i> - <i>O</i> -acetyl-2'-deoxyguanosine (10) .....	178
<i>O</i> <sup>6</sup> -Methyl-2'-deoxyguanosine (13).....	179
<i>O</i> <sup>6</sup> -Ethyl-2'-deoxyguanosine (14).....	180
<i>O</i> <sup>6</sup> -Carboxymethyl-2'-deoxyguanosine (15) .....	181

<i>O</i> <sup>6</sup> -Carboxyethyl-2'-deoxyguanosine (16).....	182
<sup>15</sup> N <sub>5</sub> -Labelled-3',5'- <i>bis</i> -O-acetyl-2'-deoxyguanosine (17) .....	183
<sup>15</sup> N <sub>5</sub> -Labelled-6-mesitylenesulfonyl-3', 5'- <i>bis</i> -O-acetyl-2'-deoxyguanosine (18) .....	184
<sup>15</sup> N <sub>5</sub> -Labelled- <i>O</i> <sup>6</sup> -methyl-2'-deoxyguanosine (19).....	185
<sup>15</sup> N <sub>5</sub> -Labelled- <i>O</i> <sup>6</sup> -ethyl-2'-deoxyguanosine (20).....	186
<sup>15</sup> N <sub>5</sub> -Labelled- <i>O</i> <sup>6</sup> -carboxymethyl-2'-deoxyguanosine (21).....	187
<sup>15</sup> N <sub>5</sub> -Labelled- <i>O</i> <sup>6</sup> -carboxyethyl-2'-deoxyguanosine (22).....	188
<i>N</i> <sup>6</sup> -[(Dimethylamino)methylene]-2'-deoxyadenosine (23).....	189
<i>N</i> <sup>6</sup> -[(Dimethylamino)methylene]-3',5'- <i>bis</i> -O-( <i>t</i> -butyldimethylsilyl)-2'-deoxyadenosine (25).....	190
<i>N,N</i> -Di- <i>n</i> -butylformamide dimethyl acetal (26) .....	191
<i>N</i> <sup>6</sup> -[(Dibutylamino)methylene]-2'-deoxyadenosine (27).....	192
9-[2'-Deoxy-5'dimethoxytrityl-β-D-erythro-pentofuranoyl]-6- <i>N</i> ', <i>N</i> '- <i>n</i> -dibutylformamidinepurine (28).....	193
9-[2'-Deoxy-5'dimethoxytrityl-β-D-erythro-pentofuranoyl]-6- <i>N</i> ', <i>N</i> '- <i>n</i> -dibutylformamidinepurine (29).....	194
9-[2'-Deoxy-β-D-erythro-pentofuranosyl]-6-(hydroxymethyl)purine (31).....	195
<i>Bis</i> -( <i>N</i> <sup>6</sup> -2'-deoxyadenosyl) methane (32) .....	196
ODN Synthesis Chemistry .....	197
RP-HPLC methods for purification of ODNs.....	198
General procedure for synthesis of ODNs containing formaldehyde-induced ICL .....	199
General procedure for enzymatic digestion of ODNs.....	201
Stable-Isotope Dilution method validation .....	202
7. References .....	206

## **Acknowledgements**

I would first like to thank David Williams for all his support and guidance throughout the project, I have enjoyed my time completing the PhD and I hope I have contributed to science in some small way. I would also like to thank group members past and present, especially Sally and Esther, for making the lab an enjoyable environment and that I shouldn't be too stressed about things.

Thank you to EPSRC for funding the project as well as the support from University of Sheffield staff, with a special mention to Mark Dickman, Andrea Hounslow, Rob Hanson, Simon Thorpe and Sandra van Meurs for their technical expertise in the analysis of my various samples. Also to Geoff Margison for his MGMT analysis.

To all my numerous Sheffield friends, both in and out of the department, for keeping me sane throughout the four years. Without karaoke, pub quizzes, pizza nights and other activities, the past four years would have been a far greater struggle and I am grateful to have been able to rely on you all to brighten my day. Special mention to Joe and Rosie who have been there from the start when I first came to Sheffield all those many moons ago.

I'd like to thank my family for all their love and support. To my mum and dad for trusting my decision to complete a PhD and for housing and feeding me for the last six months. To my brother, Marcus, for his knowledge on getting through a PhD and his reassurances that I was on the right track, plus the pestering to get the writing done. To my sister, Stef, for her chats about the future post-PhD that went very far to getting me the work I have lined up from this September. And to my grandad, who was always happy to listen and offer advice whilst ensuring I was wine and dined.

Finally, I'd like to thank my wonderful Chloë for always being there no matter my mood and for always making myself feel better about things. I can't wait for our move to London and our future there together.

## **Abstract**

Reaction of DNA with alkylating agents can lead to the formation of mutagenic  $O^6$ -alkylguanine adducts that can lead to colorectal cancer (CRC). In humans,  $O^6$ -methylguanine-DNA-methyltransferase (MGMT) repairs  $O^6$ -alkylguanines in an irreversible de-alkylation reaction. Nitrosated glycine (the formation of which may be linked to red meat intake) leads to the formation of  $O^6$ -(carboxymethyl)guanine. This thesis describes the chemical synthesis and characterisation of oligodeoxyribonucleotides (ODN) containing the alanine-derived adduct  $O^6$ -(carboxyethyl)guanine,  $O^6$ -(carboxymethyl)guanine and their corresponding carboxamides that can potentially form during ODN deprotection. Both carboxamides and the standard  $O^6$ -methylguanine are repaired notably more efficiently by MGMT than the carboxylate-containing adducts. This is likely due to the charged side chain affecting either DNA binding or alkyl transfer, suggesting that both  $O^6$ -CEG and  $O^6$ -CMG are likely to be more persistent *in vivo* and potentially more toxic. An LC-MS method is developed for quantification and structural identification of these adducts in DNA using stable-isotope dilution methods. Four  $^{15}\text{N}_5$ -labelled internal standards have been synthesised:  $O^6$ -methylguanine ( $O^6$ -MeG),  $O^6$ -ethylguanine ( $O^6$ -EtG),  $O^6$ -CMG and  $O^6$ -CEG. Using  $^{15}\text{N}_5$ -labelled  $O^6$ -MeG as the internal standard, the initial limit of quantification in a synthetic ODN was calculated as 100 fmol.

DNA can also be damaged by formaldehyde that forms interstrand crosslinks (ICLs) between adenine bases on opposite strands of a duplex that block DNA replication. However, the synthesis of such DNA for biological studies is challenging. Synthesis of a formaldehyde-induced ICL (fICL) between two adenine bases has been achieved by reacting two annealed ODNs directly with aqueous formaldehyde. NMR studies have shown that the crosslinked adenine dinucleoside, *bis*-( $N^6$ -2'-deoxyadenosyl) methane is relatively stable (several weeks) whilst its precursor  $N^6$ -hydroxymethyl-2'-deoxyadenosine has a relatively short half-life of 36.5 hours. Attempts to develop chemistry to generate site-specific fICLs in ODN duplexes using ODNs containing formamidine protected adenine have been investigated.

## **Abbreviations**

A- Adenine

Ac- Acetyl

ALDH5- Aldehyde dehydrogenase family 5

*O*<sup>6</sup>-AlkylG- *O*<sup>6</sup>-Alkylguanine

AGT- *O*<sup>6</sup>-Alkylguanine-DNA-transferase

Aq- Aqueous

ATL- Alkyltransferase-like proteins

BER- Base excision repair

C- Cytosine

CID- Collision induced dissociation

CNL- Constant neutral loss

Conc.- Concentrated

CPG- Controlled pore glass

CPS- counts per second

CRC- Colorectal cancer

dA- 2'-deoxyadenosine

DABCO- 1,4-Diazabicyclo[2.2.2]octane

Dbf- *N,N*-Dibutylformamide

dC- 2'-deoxycytidine

DCI- Deuterium chloride

DCM- Dichloromethane

dG- 2'-deoxyguanosine

DIPEA- Diisopropylamine

DMAP- 4-Dimethylaminopyridine

DMF- *N,N*-Dimethylformamide

Dmf- *N,N*-Dimethylformamide

DMSO- Dimethyl sulfoxide

DMT- 4,4'-dimethoxytrityl

DNA- 2'-Deoxyribonucleic acid



D<sub>2</sub>O- Deuterium oxide  
FA- Fanconi anemia  
FANC- Fanconi anemia complementation group  
E. Coli- Escherichia coli  
ELISA- enzyme-linked immunosorbent assay  
ESI- Electrospray ionisation  
EtOAc- Ethyl acetate  
fICL- Formaldehyde-induced interstrand crosslink  
G- Guanine  
h- Hour  
HPLC- High-pressure liquid chromatography  
ICL- Interstrand crosslink  
K<sub>D</sub>- Dissociation constant  
LC-MS- Liquid chromatography- mass spectrometry  
LOD- Limit of detection  
LOQ- Limit of quantification  
MALDI- Matrix Assisted Laser Desorption/Ionization  
MeOH- Methanol  
MeCN- Acetonitrile  
MGMT- O<sup>6</sup>-Methylguanine DNA methyltransferase  
mRNA- Messenger ribonucleic acid  
M/z- Mass-to-charge ratio  
NaOH- Sodium hydroxide  
NER- Nucleotide excision repair  
NMR- Nuclear magnetic resonance  
O<sup>6</sup>-CEG- O<sup>6</sup>-(Carboxyethyl)guanine  
O<sup>6</sup>-CMG- O<sup>6</sup>-(Carboxymethyl)guanine  
O<sup>6</sup>-EtG- O<sup>6</sup>-Ethylguanine  
O<sup>6</sup>-MeG- O<sup>6</sup>-Methylguanine  
ODNs- Oligodeoxyribonucleotides

pac- Phenoxyacetyl

PAGE- Polyacrylamide gel electrophoresis

$R_f$ - Retention factor

RNA- Ribonucleic acid

RP-HPLC- Reversed phase high-pressure liquid chromatography

SID- Stable isotope dilution

SIL- Stable isotopically labelled

SIM- Secondary ionisation mass spectrometry

SRM- Selected reaction monitoring

T- Thymine

TBAF- Tetrabutylammonium fluoride

TBDMS- Tert-Butyldimethylsilyl

TEA- Triethylamine

TEAA- Triethylammonium acetate

TEAB- Triethylammonium bicarbonate

THF- Tetrahydrofuran

TLC- Thin layer chromatography

tRNA- Transfer ribonucleic acid

U- Uridine

UHPLC- Ultra high-pressure liquid chromatography

## **Chapter 1- *Introduction***

# 1. Introduction

## 1.1. Structure of DNA

Deoxyribonucleic acid (DNA) is the hereditary material vital for living organisms because it stores all genetic information. In eukaryotes, every cell holds DNA within its nucleus. Nucleotides are the building blocks joined together to form DNA and these are made up of three components: a phosphate diester, 2-deoxyribose and a heterocyclic base. These bases are guanine (G), adenine (A), cytosine (C) and thymine (T). Figure 1 shows a single-strand of DNA. As seen, the nucleotide units are connected through the 3'- and 5'- positions on adjacent deoxyribonucleosides. The base is bonded to the sugar in the N9 position for G and A whilst T and C bond through the N1 position.

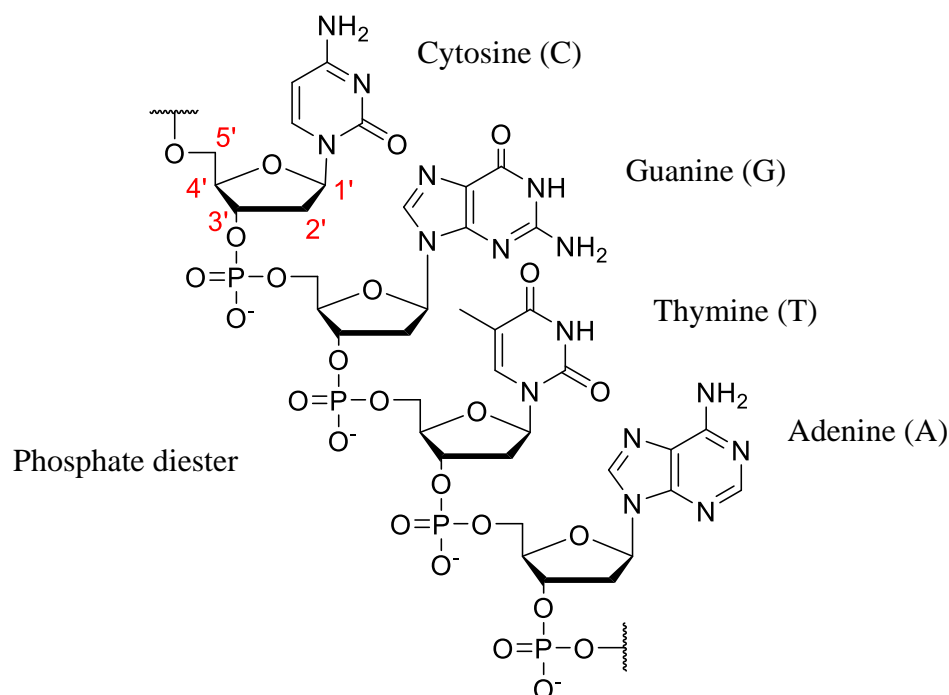


Figure 1. Primary DNA structure including heterocyclic bases and phosphate diester backbone.

Figure 2 shows the four bases of DNA. These comprise of two purines, G and A, and two pyrimidines, C and T. Complementary base-pairing leads to the double-stranded DNA structure held together by hydrogen bonds between the bases.

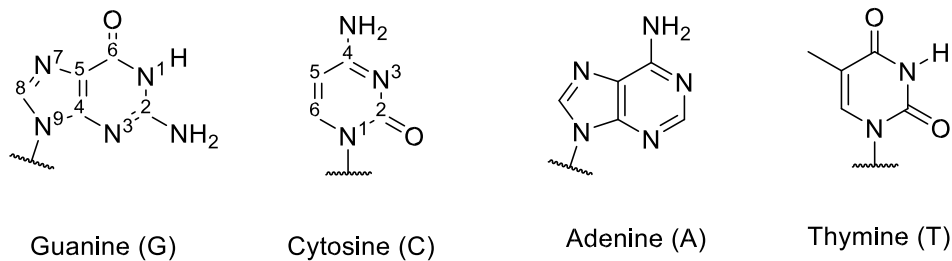


Figure 2. The four bases of DNA.

Figure 3 shows the complementary base pairings between the respective bases in double helical DNA. This was discovered by James Watson and Francis Crick and is therefore referred to as Watson-Crick base pairing<sup>1</sup>. Due to the specific tautomeric structures of each base, preferences occur. The purine-pyrimidine preference gives DNA its regular shape and selective hydrogen bonding between A-T and G-C results in the complementary base pairings. The bases are stacked inside the helix, away from the hydrophilic conditions of the cell and are held together through hydrogen bonds.  $\pi$ - $\pi$  stacking interactions of the base pairs further stabilises the structure. The two strands of the DNA double helix run anti-parallel to one another i.e. one strand runs 3' to 5' and the complementary strand runs 5' to 3'<sup>2</sup>.

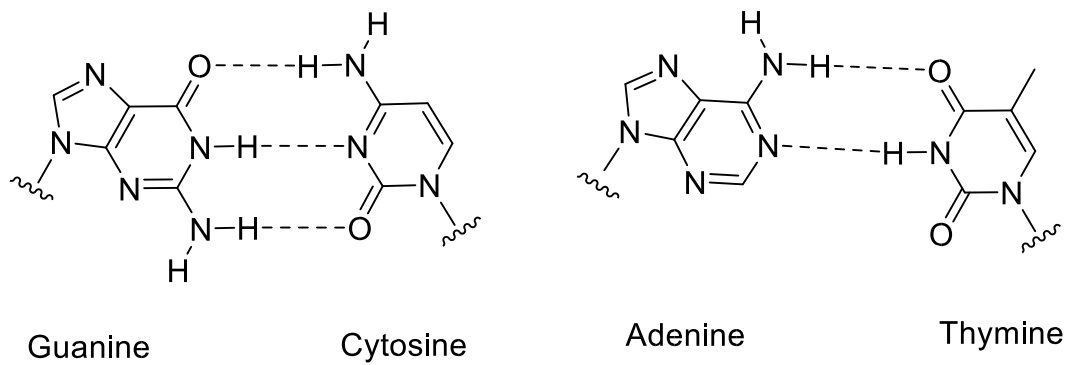


Figure 3. Base pairing through hydrogen bonds between DNA bases.

The double helical structure modelled by Watson and Crick is commonly referred to as B-DNA which is the most abundant conformation in organisms. This is a right-handed double helix which creates a major and a minor groove along the edges of the bases. Each helix has 10.5 base pairs per turn. Other secondary structures of DNA have been discovered, such as Z-DNA which arises preferentially in left-handed conformation sequences<sup>3</sup>. Sequences rich in alternating G and C nucleobases tend to prefer this conformation. A-DNA is broader and shorter in conformation, caused due to having 11 base pairs per turn of the helix<sup>3</sup>. This conformation is a preference for sequences rich in non-alternating G and C nucleobases<sup>3</sup>.

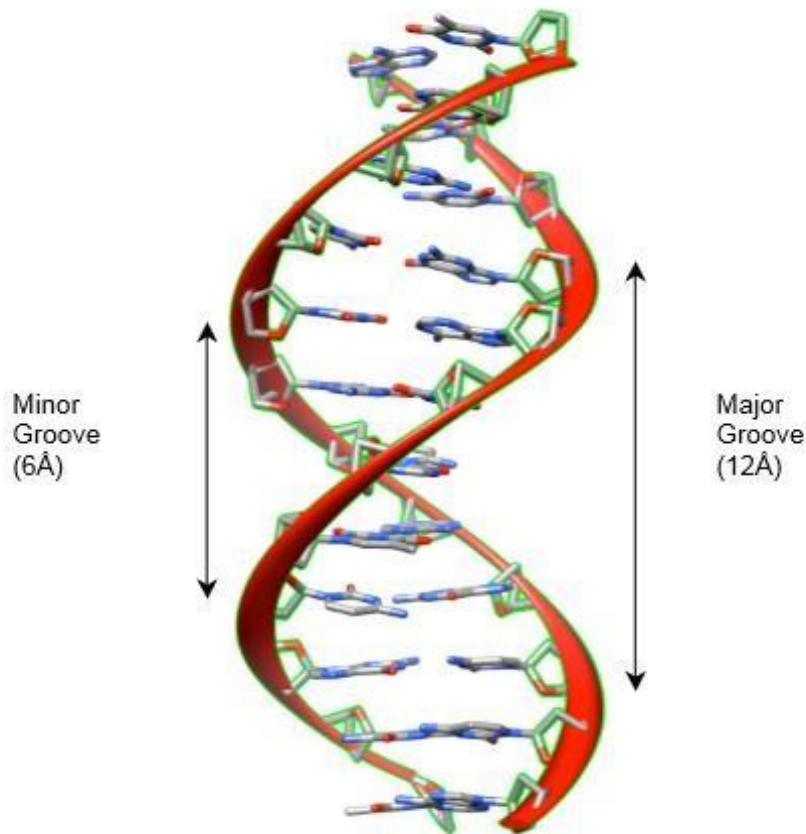


Figure 4. The DNA double helix and the positions of the major and minor groove of B-DNA (Wilkinson, 2012).

Watson-Crick base pairing is not the only base pairing found in DNA. X-ray analysis of sequences rich in A and T nucleobases has indicated alternative base pairing between A and T. Instead of hydrogen bonding of N6 and N1 on A with O4 and N3 on T respectively, a Hoogsteen base pair involves hydrogen bonding of N7 and N6 of A with N3 and O4 of T (Figure 5). Hoogsteen base pairing has also been noted between G and C when the N3 position of C is protonated<sup>2</sup>.

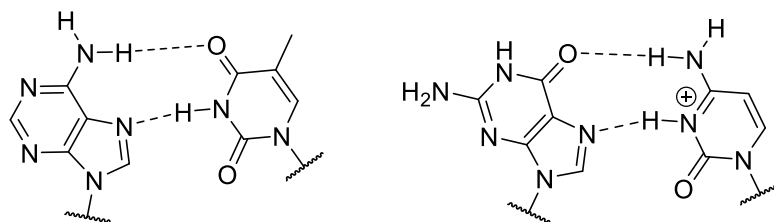


Figure 5. Hoogsteen base pairing between A: T and G: C.

Hoogsteen base pairing narrows the distance between nucleobases from 10.5 Å to 8.65 Å and subsequently is incompatible with B-DNA because the duplex is destabilised<sup>3</sup>. However, Hoogsteen base pairing does enable triple helical structures as well as quadruplex structures<sup>4</sup>. Triple helical structures form between B-DNA duplex and a third DNA strand that bonds to the duplex *via* Hoogsteen base pairing. These structures have been noted in gene regulation<sup>4-6</sup>. Quadruplex structures form between planes of four bonded guanine bases which then subsequently stack on top of each other. Guanine-quartets form through cyclic Hoogsteen base pairing and stabilised by presence of a monovalent cation, such as K<sup>+</sup> or Na<sup>+</sup> (Figure 6).

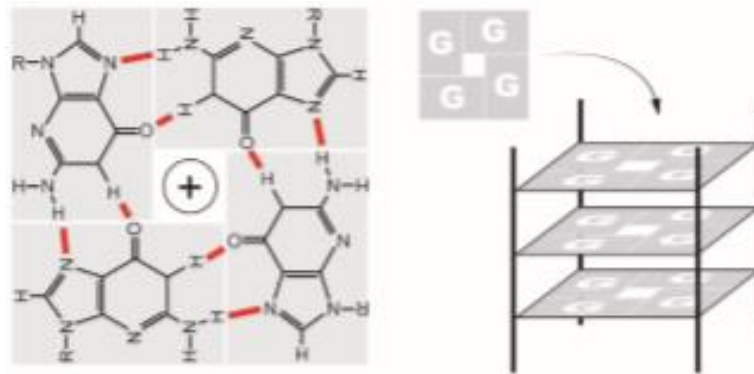


Figure 6. G-quartets formed through cyclic Hoogsteen base pairing and packing of these quartets to form G-quadruplexes (Rhodes 2015)<sup>4</sup>.

Ribonucleic acid (RNA) translates the genetic information in DNA into proteins.

RNA has an analogous primary structure to DNA but contains ribose instead of 2-deoxyribose whilst the nucleobase thymine is replaced by uracil<sup>2</sup>.



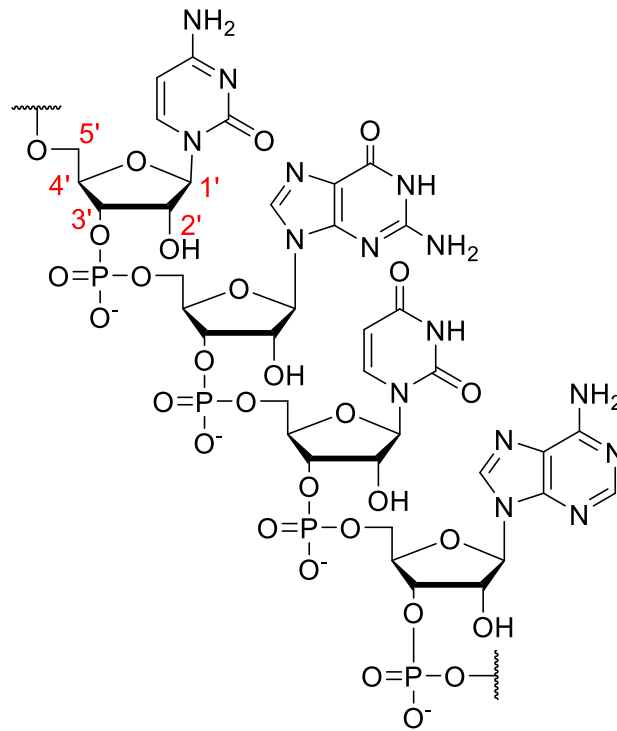
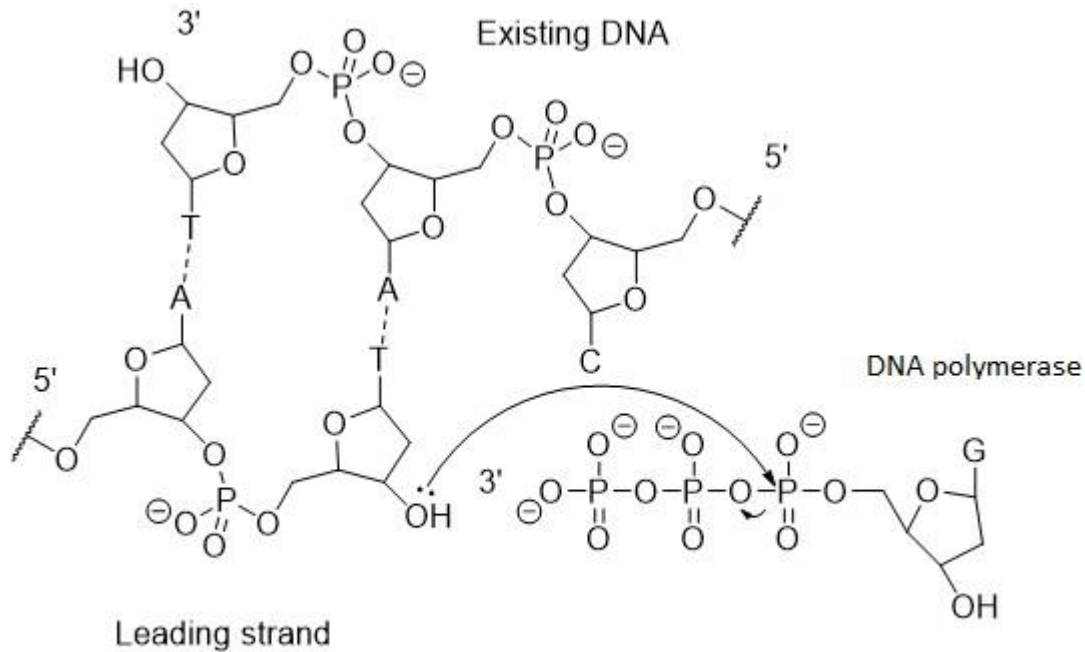


Figure 7. Sequence of RNA and RNA base uracil which is present instead of thymine.

## 1.2.DNA Replication

It is vital for a living organism to have the ability to replicate its DNA during cell division in order to pass on the exact same genetic coding to the new cells. Complementary base pairing ensures that the code is maintained because each strand of the double helix can act as a template to allow specific enzymes known as DNA polymerases to generate complementary copies<sup>2,3</sup>. This results in two identical copies of the DNA strands and thus each new cell contains the same genetic coding as the old cells. This process is said to be semi-conservative as each new DNA double helix is

comprised of one 'new' strand and one 'old' one. Scheme 1 shows DNA polymerase inserting bases onto an existing strand of DNA.



Scheme 1. Nucleotide insertion by DNA polymerase during the process of DNA replication.

The first step of the mechanism is the unzipping of the double helix. DNA gyrase initiates the separation of the double stranded DNA into two single strands. A helicase enzyme unwinds the DNA resulting in the formation of a replication fork. New nucleotides are added to the separated strands by DNA polymerase in the presence of  $Mg^{2+}$ . In this process 2'-deoxynucleoside-5'-triphosphates (dNTPs) are added sequentially to the 3'-end of the growing DNA strand such that A is incorporated opposite T and G inserted opposite C and *vice versa*<sup>7</sup>. After this step, the inserted nucleotides are "proofread" whereby the DNA polymerase checks to see if the correct base has been inserted. Due to this, the 5' to 3' strand is known as the leading strand.

The 3' to 5' (or lagging) strand is not replicated continuously but occurs *via* formation of a series of Okazaki fragments. Once the replication fork in DNA is reached, replication ends, and DNA ligase seals the fragments into one continuous strand. The DNA winds up again and the replication is complete<sup>1,2</sup>.

### **1.3.DNA Transcription and Translation**

Proteins are essential for an organism to function, catalysing metabolism to ensure organisms grow, reproduce and maintain their structure. Proteins are formed from polypeptides that are long-chains of single-strand amino acids, of which there are 20 that occur naturally. Amino acids are determined by the genetic code of codons, a sequence of three nucleotides that correspond to specific amino acids. There are 64 possible codons that can form from G, C, A and T which means amino acids can have more than one codon specific to them. This whole process is described as transcription and translation.

Initially, the genetic code of DNA is transcribed to a complementary strand of messenger RNA (mRNA). RNA polymerase unzips the DNA double helix and synthesise the complementary mRNA strand before releasing it. Synthesis is initiated at the site of a specific codon and stops once RNA polymerase reaches one of three codons that stop mRNA synthesis.

Once released, mRNA is transported to ribosomes where its own codons are “read”, bringing to the ribosome the amino acid corresponding to each codon. This occurs because amino acids are attached to transfer RNA (tRNA) that base pair to each mRNA codon. This allows polypeptides to form when amino acids on each

tRNA form amide bonds with adjacent amino acids that are attached to their own tRNA. Polypeptides fold into secondary, tertiary and quaternary structures to form functional proteins.

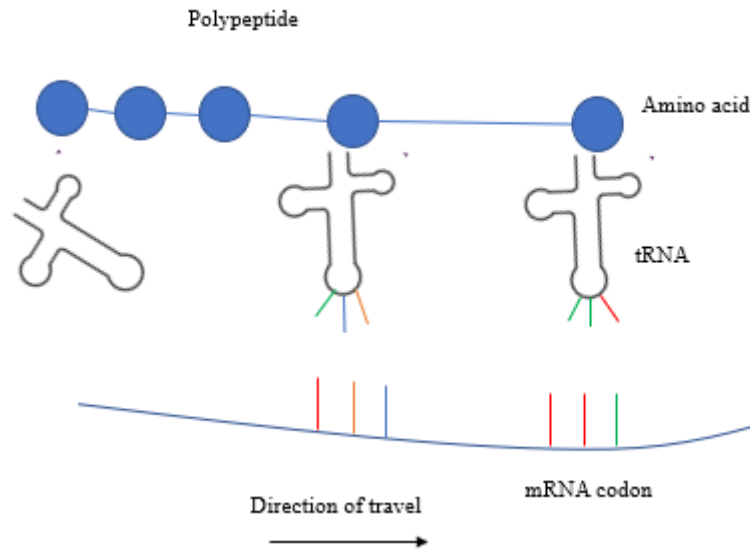


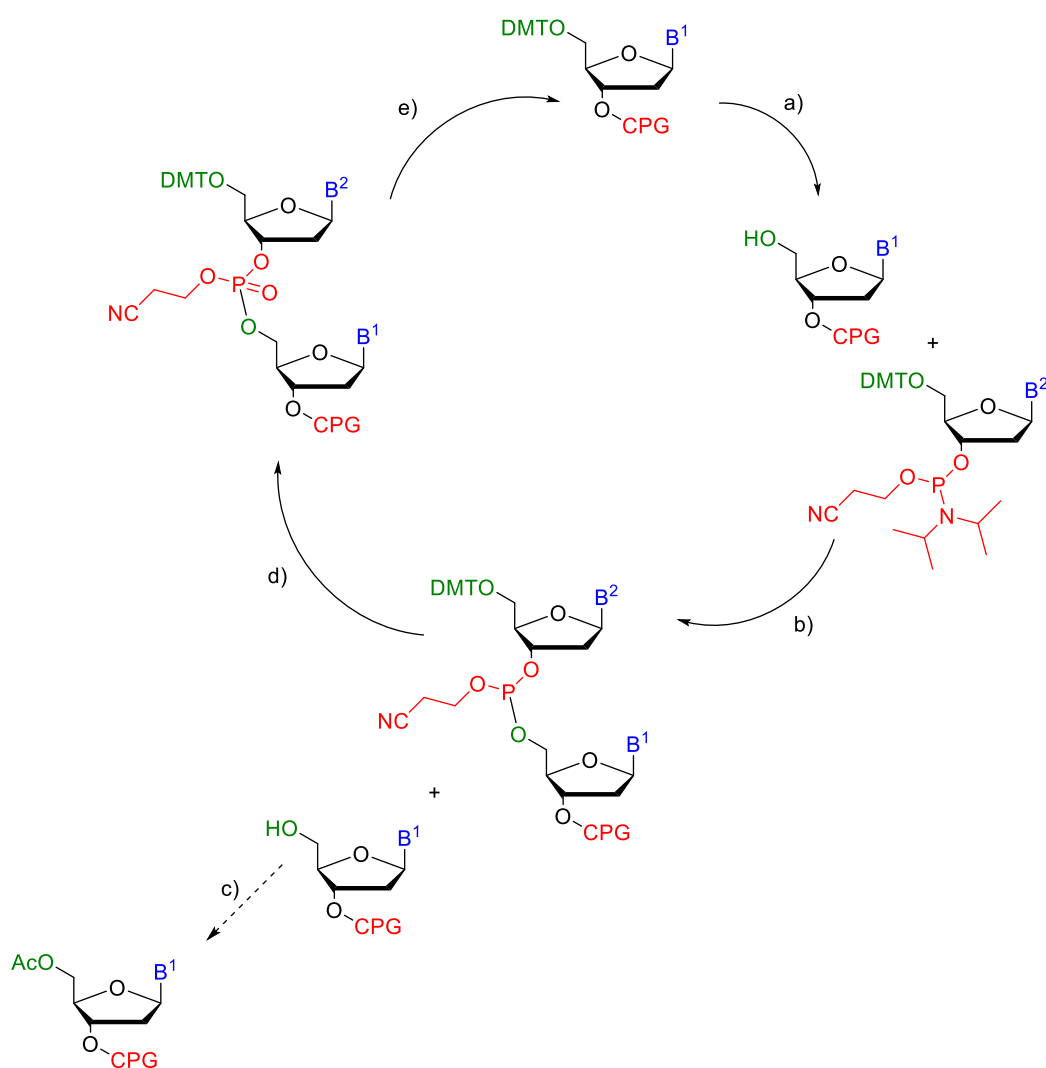
Figure 8. Translation of mRNA to form polypeptides.

This intricate system is dependent on DNA remaining unchanged. If damage occurs, this can affect the subsequent processes, culminating in the formation of an incorrect protein, thus affecting the functionality of the organism. For example, if the nucleotide codon is “read” incorrectly then the wrong amino acid is inserted. Proteins rely on the specific properties of the amino acids to enable correct folding and functionality therefore an incorrect amino acid can alter this.

#### 1.4. Chemical synthesis of DNA

The chemical synthesis of DNA for use as primers for DNA sequencing, the polymerase chain reaction and for structural studies is now routine. For a number of decades, the development of such a route proved hard to find<sup>2</sup> but, during the 1980s, a

highly efficient and automated DNA synthesis process was developed by Caruthers<sup>8</sup>. The 5'-OH group of a solid support-bound deoxyribonucleoside reacts with another deoxyribonucleoside containing a 3'-*O*-(*N,N*-diisopropyl)-*O*-alkyl phosphoramidite to build a sequence of up to 150 base pairs (Scheme 2)<sup>2</sup>. Being bound to a controlled pore glass (CPG) solid support enables reagents to be used in excess, ensuring efficient coupling and purification whilst phosphoramidites are highly reactive phosphorus (III) compounds<sup>8</sup>.



Scheme 2. Basic steps of the phosphoramidite method of synthesising DNA sequences<sup>2</sup>; a) Deprotection-TCA, DCM; b) Coupling condensation- tetrazole; c) Capping- acetic anhydride, *N*-methylimidazole; d) Oxidation- iodine, aq pyridine; e) repetition of cycle.

This method requires protection of nucleophilic hydroxyl and exocyclic amino groups of the G, C and A bases otherwise numerous side reactions would occur driving down the yield of reaction. The nucleobases are protected using acyl groups, benzoyl for C and A and isobutyryl for G. Thymine does not have an exocyclic amino group to protect. These protecting groups can be removed by heating in aq conc. ammonia solution at 55 °C for 6 hours at the end of the synthesis. Under these conditions the cyanoethyl group is removed from the phosphate triester groups and the oligodeoxyribonucleotide (ODN) is cleaved from the solid support.

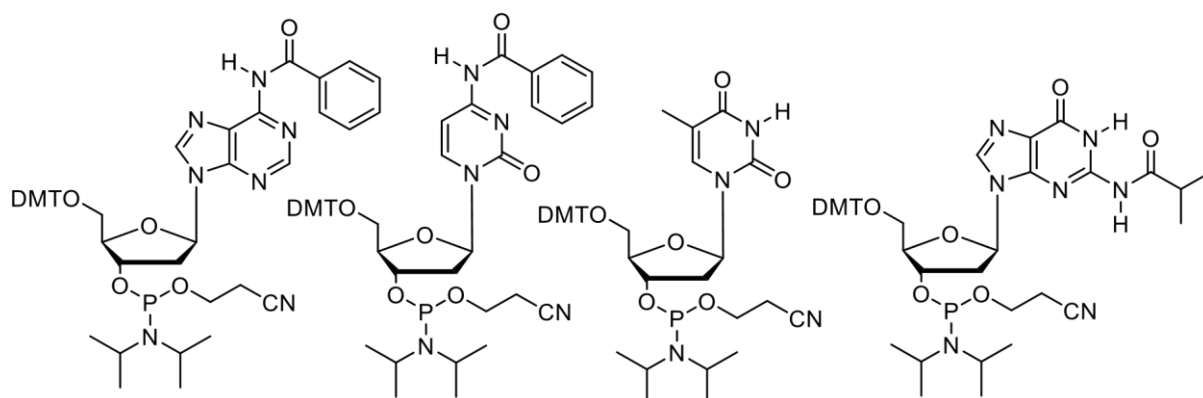


Figure 9. The phosphoramidites of fully protected A, C, T and G for DNA synthesis. The exocyclic amino groups on A and C are protected by a benzoyl group whilst G is protected by an isobutyryl group. T does not require protection because it has no exocyclic amino group.

Nowadays, thanks to the development of the chemistry, milder deprotection conditions can be used at the end of the synthesis. Thus, A and G can be protected with phenoxyacetyl (pac) and C by acetyl<sup>9</sup>. These protecting groups can be removed with aqueous ammonia solution in 4 hours at room temperature.

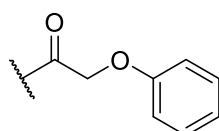


Figure 10. Phenoxyacetyl used for the exocyclic amino group protection of G and A.

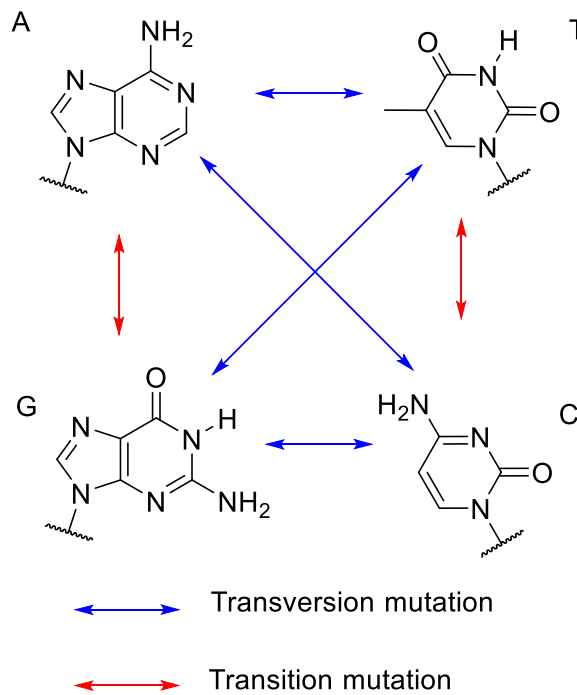
The 5'-OH of each sugar is protected with 4,4'-dimethoxytritylchloride (DMTCl) whilst the 3'-OH has 2-cyanoethyl-*N,N*-diisopropylaminophosphoramidite attached. Though not a protecting group, the phosphoramidite enables coupling to take place and leads to the formation of a phosphite triester.

The synthesis is initiated by the removal of the 5'-DMT group from the CPG-bound nucleoside by acid. The phosphoramidite of the next desired nucleoside in the sequence is introduced and is activated by tetrazole to form a phosphotetrazolide. This phosphotetrazolide reacts with the 5'-OH of the ODN on the solid support to form a phosphite triester. Capping follows the phosphite triester formation. Capping is done to ensure that any chains that are unreacted in the cycle do not react in subsequent cycles, leading to the synthesis of incorrect sequences. Capping is achieved using acetic anhydride and *N*-methylimidazole tetrahydrofuran<sup>8</sup>.

Oxidation of the phosphite triester to the phosphate is required and is achieved by using a mixture of iodine and aqueous pyridine. This is the last step of the cycle. After this is done, the cycle can be repeated to add a new nucleotide to the sequence or deprotected and removed from the solid support using concentrated aqueous ammonia.

### **1.5.DNA Damage and Mutations**

Damage and mutations to DNA are induced in various ways. Mutations are changes in the base sequence of the DNA and are designated as transition or transversion mutations. As shown in Scheme 3, a transition mutation is an interchange between purine (A↔G) or pyrimidine (C↔T) bases whilst a transversion mutation is an interchange between a purine base and a pyrimidine base or vice versa<sup>3</sup>. This means there is one possible transition mutation and two possible transversion mutations for each nucleobase.

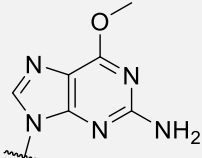
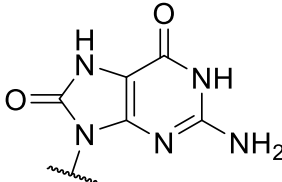
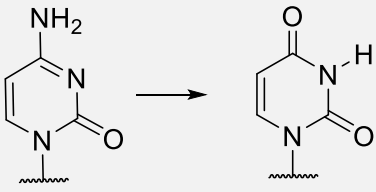
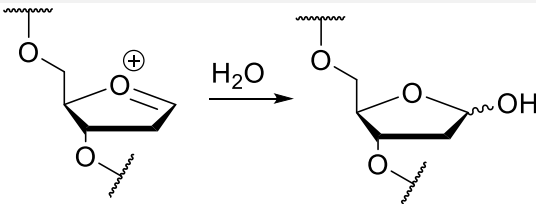


Scheme 3. Interchange of nucleobases that result in either transversion or transition mutations.

Once mutations occur, they are replicated into new cells and cannot be repaired. Mutational errors can occur during DNA replication when DNA polymerase enzymes misread the template strand and insert the wrong nucleobase. This occurs once per one million base pairs<sup>2, 10, 11</sup>. Alteration to the structure can cause decomposition which leaves an abasic site or change the base pairing properties that can cause the DNA polymerase enzymes to insert the wrong nucleobase during DNA replication. Table 1 summarises a few common ways that DNA is structurally damaged<sup>2</sup>.



Table 1. Origin of common mutations in DNA<sup>2</sup>

<i>Mutation</i>	<i>Cause</i>	<i>Example</i>
<b>Alkylation-</b>  <i>addition of alkyl group to nucleophilic, N or O)</i>	Alkylating agent	 $O^6$ -methylguanine
<b>Oxidation</b>	Reactive oxygen species	 8-oxoguanine
<b>Deamination –</b>  <i>hydrolysis of NH<sub>2</sub></i>	Hydrolysis	 Cytosine                      Uracil
<b>Depurination-</b>  <i>loss of purine base from sugar</i>	Heat, acid or alkylation	 Abasic site

It is thought that the majority of DNA is damaged by endogenous factors<sup>12</sup>. These chemicals are formed from biochemical reactions, including metabolism. Reactive oxygen species (ROS) cause oxidative damage on the nucleobases or cause cleavage of the double or single strands. 8-Oxoguanine is a common oxidised base which is

mutagenic. It is considered to be a major cause of cancer as well as degenerative diseases<sup>13</sup>. 8-Oxoguanine can mispair with adenine instead of cytosine during DNA replication (Figure 11) causing transversion mutations that are difficult to repair<sup>13</sup>.

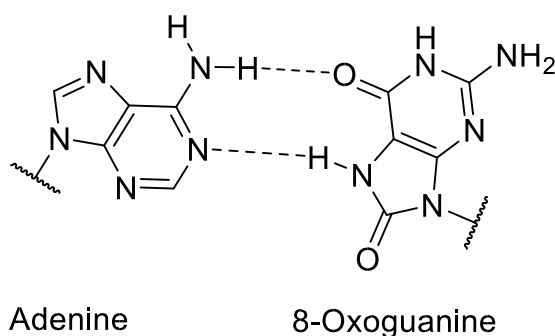
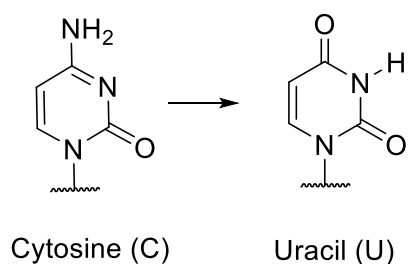


Figure 11. Base pairing between adenine and 8-oxoguanine arising from the oxidation of 8-guanine.

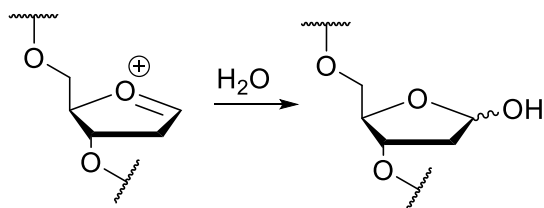
Deamination is another type of damage caused by endogenous reactions. Deamination is the spontaneous hydrolysis of an exocyclic amino- group to either a hydroxyl or a carbonyl group. This can occur on guanine, adenine and cytosine<sup>3</sup>. This type of damage is one of the most common types of DNA damage, occurring hundreds of times a day in each cell<sup>3</sup>.



Scheme 4. Deamination of cytosine to uracil

Purine-containing 2'-deoxyribonucleosides are prone to acid-catalysed hydrolysis. This results in the cleavage of the purine base (A or G) from the sugar leading to an oxocarbenium ion and an abasic site (Scheme 5). This type of damage can also affect purines in RNA but they depurinate more slowly due to the higher-energy transition state in the formation of the oxocarbenium ion and an abasic site<sup>14</sup>. Heating, alkylation

or acidic conditions encourage depurination to occur<sup>2</sup>. After depurination, the wrong base pair can be inserted by a DNA polymerase opposite the abasic site. Transition or transversion mutations can result because the DNA polymerase cannot recognise the base that has been incorrectly inserted<sup>3</sup>. Carcinogenic mutations have been known to occur from depurination<sup>14</sup>.



Scheme 5. 2-Deoxyribose oxocarbenium ion and abasic site after the cleavage of the base due to depurination.

Exogenous factors of DNA damage include UV light and X-rays. The radiative damage done by both affects DNA transcription and replication. Chemical reactions can occur in DNA that are initiated by UV light, such as the electrochemical coupling of two consecutive thymine bases. This leads to cyclobutane pyrimidine dimers forming which permanently alter the helix structure. X-rays cause single- and double-strand breaks to the backbone of the nucleic acid<sup>2</sup>.

### 1.6. Alkylation of Nucleobases

Alkylation of DNA can occur from both endogenous and exogenous sources with base alkylation a primary source of mutagenesis<sup>15, 16</sup>. *O*-Alkylation also occurs on the phosphodiester linkages which make up the backbone but there is little information to suggest that these are carcinogenic or mutagenic<sup>2</sup>. The nucleobases have several nucleophilic sites including amines, carbonyl groups and ring nitrogens. The most susceptible sites for addition of an alkyl group on DNA are shown in Figure 12<sup>17</sup>. The N7 position of guanine is by far the most common site of alkylation, accounting for

70-80 % of all alkylations. This is because it is the most nucleophilic site<sup>18-20</sup>. Alkylation here can lead to the depurination of the base and the formation of an abasic site. However, in terms of mutagenic potential, N7-alkylation is considered low because it does not appreciably affect the Watson-Crick base pairing and depurination leaves a repairable abasic site.

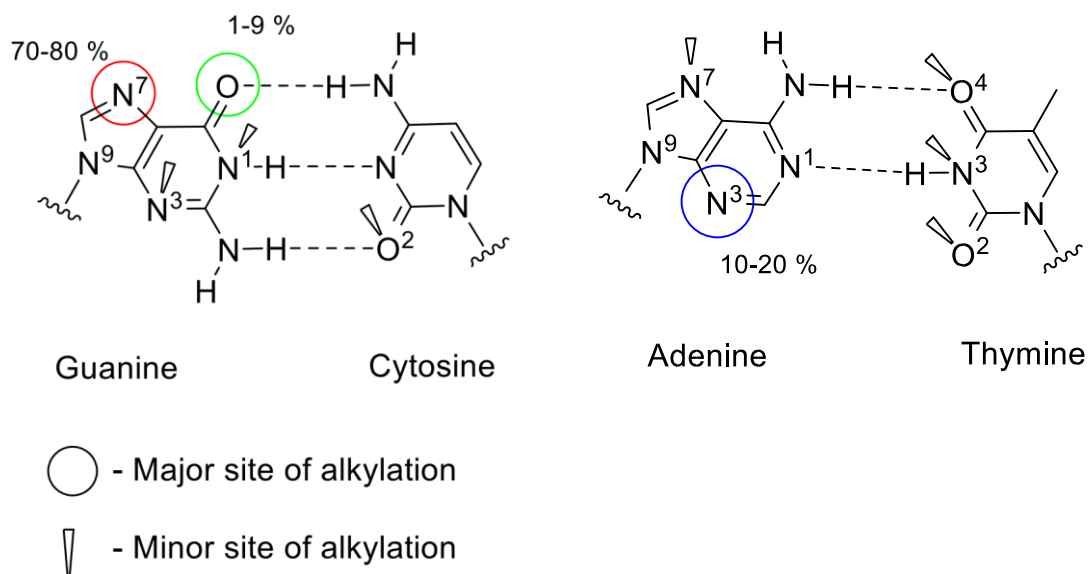


Figure 12. Sites of alkylation on the bases of DNA.

As seen in Figure 13, methylation of the N3 atom of thymine and cytosine inhibits hydrogen bonding with their respective complementary bases. This blocks DNA replication and can lead to apoptosis<sup>21,22</sup>.

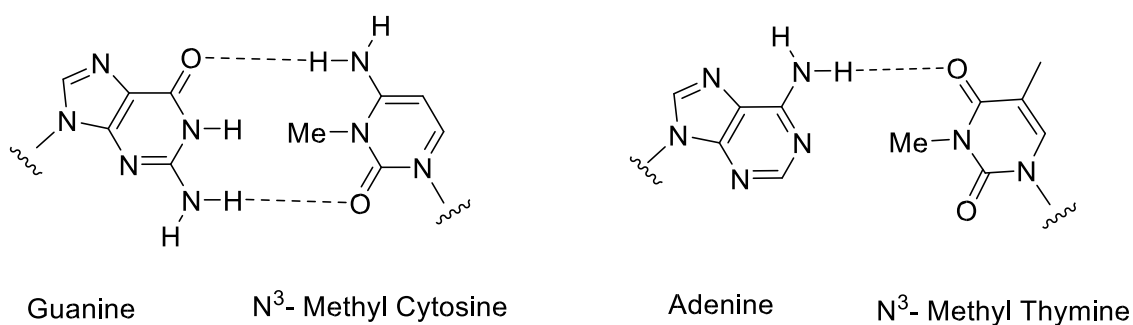


Figure 13. Hydrogen bonding between nucleobases with N<sup>3</sup>-methylation of the pyrimidine base.

Guanine has several sites susceptible to alkylation. The susceptible sites can be divided into those alkylated by ‘soft’ electrophiles *via* an S<sub>N</sub>2 mechanism and those alkylated by ‘hard’ electrophiles *via* an S<sub>N</sub>1 mechanism<sup>17</sup>.

#### 1.6.1. Alkylation at N-sites

Alkylating agents that are deemed soft acids, such as methyl halides and methyl methane sulfonates, are found to preferentially alkylate N7 sites of purine. This is due to nitrogen being a softer nucleophilic site than oxygen.

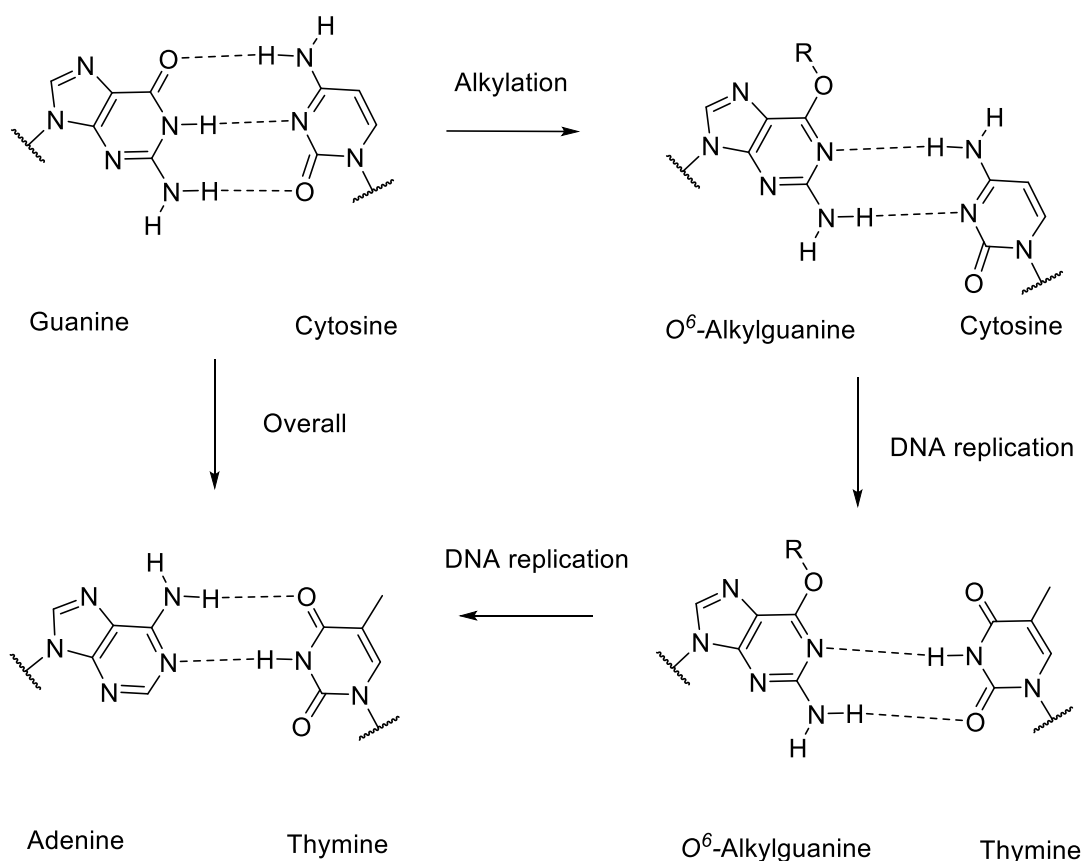
The second most common alkylation site is the N3 site on adenine. *N*<sup>3</sup>-Methyladenine is highly cytotoxic because it can block DNA polymerases and therefore hinder DNA synthesis<sup>23</sup>. *N*<sup>3</sup>-Methyladenine has been shown to inhibit DNA synthesis in yeast<sup>24</sup> but the full effects of *N*<sup>3</sup>-methyladenine are not fully established because it readily forms alongside *N*<sup>7</sup>-methylguanine and *O*<sup>6</sup>-methylguanine<sup>24</sup> both of which are toxic. A further problem is that *N*<sup>3</sup>-methyladenine readily undergoes depurination, especially when in a single stranded structure<sup>13</sup>.

The exocyclic N2 amino group of guanine and the N6 amino group of adenine are also known to react with aldehydes and enals, amongst other chemicals, to form, in the latter case, tricyclic N1, N2-guanine adducts. These adducts have an effect on DNA replication and cause mainly transition mutations.

#### 1.6.2. *O*<sup>6</sup>-Alkylguanines

Oxygen sites have ‘hard-base’ characteristics. This means ‘hard’ electrophiles such as carbocations can readily alkylate these sites<sup>17</sup>. Alkylation at the O6-position of guanine, though only accounting for 1-9 % of the total alkylation on DNA nucleobases, is considered to be the most mutagenic<sup>25, 26</sup> and its formation has been linked to colorectal cancer<sup>27</sup>.

An  $O^6$ -alkylguanine ( $O^6$ -alkylG) adduct is both neutral and stable which means it is more persistent and not easily detected by repair proteins. This enables it to disrupt DNA replication which is the cause of its mutagenicity. As shown in Scheme 6, G:C  $\rightarrow$  A:T transition mutations occur when thymine is incorrectly paired opposite  $O^6$ -alkylG during the initial step of the replication<sup>28</sup>. When this strand is then replicated, adenine is correctly inserted opposite thymine but, overall, there is a transition mutation from guanine to adenine. The mutations occur because of the alteration to the hydrogen bonding preference of guanine, the O6 of guanine is no longer able to hydrogen bond to the N4 atom of cytosine. Though it can still pair with cytosine, the structure is no longer a Watson-Crick-type but is instead considered to be a reverse-wobble pair structure<sup>28</sup>. Although there is evidence that this base pair is more stable than that with thymine<sup>29</sup>, this latter structure has the same geometry as a Watson-Crick-type base pair and DNA polymerases preferentially rely on such base pairs during replication.



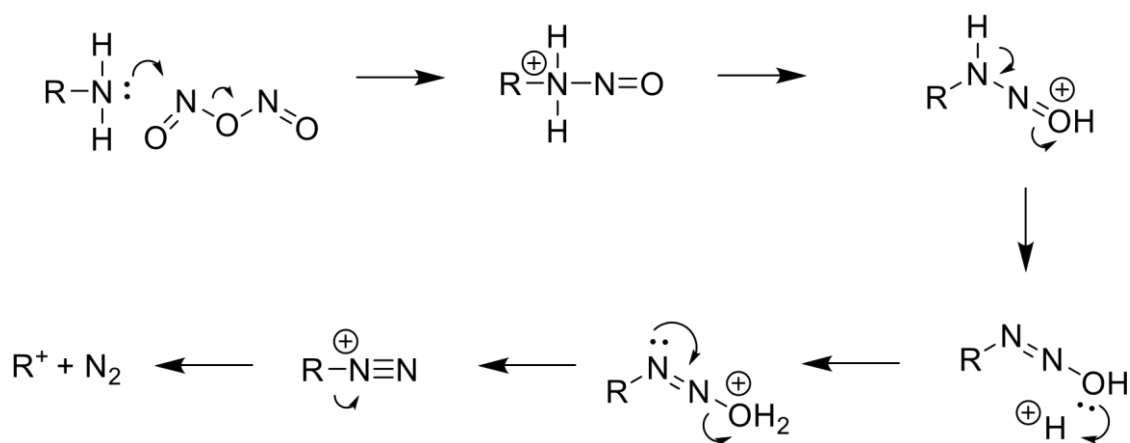
Scheme 6. Overall mutation of DNA replication due to  $O^6$ -alkylguanine base pairing with Thymine

### 1.7. Colorectal Cancer and *N*-Nitroso Compounds

G → A transition mutations are commonly found in colorectal cancer and the cause is associated with the ki-Ras protein<sup>29</sup>. Codons are sequences of three DNA nucleotides which correspond to a specific amino acid or stop signal during protein synthesis and mutations in either codon 12 or 13 of ki-Ras convert this gene into an active oncogene<sup>29</sup>. Colorectal cancer is the third most common cancer in the developed world though incidence rates vary around 20-fold globally. Lifestyle, especially diet, is a major factor determining risk and that as a result, colorectal cancer could be avoided in an estimated 70 % of cases<sup>30</sup>. There are strong links between CRC and an increase in red and processed meat in the diet whilst a diet high in fibre decreases the risk of CRC. In recent years it has been found that *N*-Nitroso compounds (NOCs), a

wide range of compounds including nitrosamides and nitrosamines, have a significant impact on CRC<sup>31</sup>.

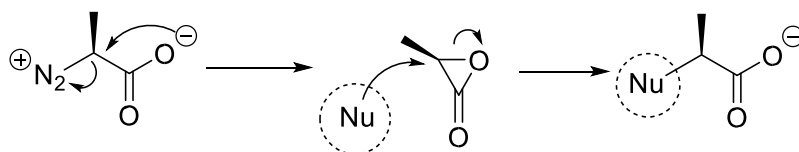
Besides red and processed meat, NOCs are present in tobacco smoke, some anticancer drugs, such as carmustine and dacarbazine, and many foods such as smoked fish and bacon and can be produced through the endogenous N-nitrosation of amines and amides<sup>27</sup>. Endogenous N-nitrosation occurs readily due to the abundance of proteins and amino acids as well as nitrosating agents, such as nitrite and nitrosated haem<sup>27</sup>. This process is summarised in Scheme 7 which shows the formation of the alkylating agent and the elimination of nitrogen. Reaction of nitrite or nitrate with dinitrogen trioxide has been noted in the stomach whilst exposure to nitrosyl haem means nitrosation has been detected in the colon. There is a high content of haem in red meat and studies looking at detection of NOCs in human faeces formed a link between red meat intake and faecal NOCs<sup>31</sup>.



Scheme 7. Formation of alkylating agents from amines under acidic conditions.

Nitrosation of amino acids leads to the formation of lactones of which there are three possible structures,  $\alpha$ ,  $\beta$  or  $\gamma$ . Studies by Garcia-Santos *et al.* indicated that  $\alpha$ -lactones are highly reactive due to the instability of its three-membered ring and small size compared to the other structures. Scheme 8 shows the formation of the  $\alpha$ -lactone<sup>32</sup>.





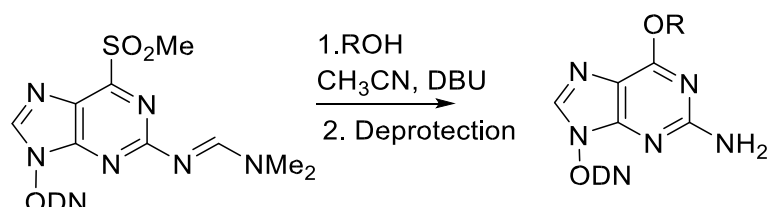
Scheme 8. Formation of  $\alpha$ -lactone of alanine with elimination of  $N_2$  and the subsequent reaction with a nucleophile.

Two important  $O^6$ -alkylG adducts,  $O^6$ -methyl- and  $O^6$ -(carboxymethyl)guanine, ( $O^6$ -MeG and  $O^6$ -CMG) have been detected in human colonic tissue and colonocytes<sup>31, 33</sup>. Both  $O^6$ -MeG and  $O^6$ -CMG adducts can potentially derive from the nitrosation of glycine and its derivatives. NOCs found in red meat and of which high levels have been detected in faeces, further supporting the hypothesis that CRC is linked to N-nitrosation and diet. The putative  $O^6$ -alkylG adduct  $O^6$ -(carboxyethyl)guanine ( $O^6$ -CEG) may also form though the adduct has not been recorded in DNA samples thus far. Shephard *et al.* have noted that meat contains high concentrations of alanine<sup>33</sup> which will produce the reactive  $\alpha$ -lactone upon N-nitrosation<sup>32</sup>. It can be therefore hypothesised that the  $\alpha$ -lactone of diazo alanine can be expected to react at the  $O^6$  position of guanine to produce  $O^6$ -CEG.

### 1.8.Synthesis of ODNs containing $O^6$ -Alkylguanine Adducts

Development of a synthesis enabling the incorporation of the 2'-deoxyriboside of any  $O^6$ -alkylG adduct into ODN sequences, using a specific phosphoramidite reagent of 2-amino-6-methylsulfonylpurine, has made it possible to quickly and efficiently synthesise these sequences. As outlined in Scheme 9, a modified nucleoside containing a reactive 2-amino-6-methylsulfonylpurine base is converted into the desired  $O^6$ -alkylG-containing ODN following reaction with the appropriate alcohol in the

presence of DBU. The high reactivity of the methylsulfonyl group enables facile nucleophilic aromatic substitution. The ODN is then fully deprotected to allow characterisation of the adduct<sup>34</sup>.



Scheme 9. Synthesis of *O*<sup>6</sup>-alkylguanine from 2-amino-6-methylsulfonylguanine in DNA<sup>34</sup>.

## 1.9. DNA Repair Mechanisms

DNA repair systems have developed for the repair of DNA damage and the maintenance of genome integrity<sup>2</sup>. There are four main repair mechanisms employed: base mismatch repair; base excision repair; nucleotide excision repair; and direct damage reversal repair.

Base mismatch repair (MMR) is the fundamental repair mechanism of errors caused by polymerases during DNA replication. Acting as a proofreading aid, proteins involved in MMR scan the daughter strand and remove any damage within a nucleotide chain. In *E. coli*, the mismatched base is recognised by a homodimer of the protein MutS. The endonuclease MutH then makes an incision in the strand containing the error. The helix is unwound by helicase II which allows exonucleases to digest the now single-strand. This leaves a gap of between 1000-3000 nucleotides which is then repaired fully by DNA polymerase and resealed by DNA ligase to restore the DNA to the correct sequence<sup>35</sup>.

Base excision repair (BER) cleaves the *N*-glycosidic bond in order to remove any damaged bases. The sugar is then removed and a new nucleotide is inserted. DNA glycosylase initiates the repair mechanism by recognising the damaged base and then it cuts the *N*-glycosidic bond (between the base and the sugar-phosphate backbone) to generate an AP site. This is then cleaved by an AP endonuclease to leave a gap which is repaired by DNA polymerase and DNA ligase. Many types of endogenous damage are repaired by BER including *N*-alkylated adducts of DNA such as *N*<sup>3</sup>-methyladenine and other such sites of damage such as 8-oxoguanine and uracil (cytosine deamination)<sup>2</sup>.

Nucleotide excision repair (NER) resolves more extensive damage than BER, dealing with damage that can distort the helix such as cyclobutane thymidine dimers. NER removes the daughter strand (up to 30 nucleotides) and DNA polymerase inserts a new, undamaged strand that is annealed by ligase. To start with, the endonuclease enzyme complex, such as Uvr ABC, scans the DNA and recognises any distortions after which it loads onto the damaged site. Part of the complex then cleaves the phosphodiester backbone either side of the damage and the damaged section is removed. The complex is then able to recruit DNA polymerase to insert the correct complementary base. DNA ligase then seals the nick to complete the repair<sup>36</sup>. NER repairs, amongst others, cyclobutane pyrimidine dimers and 6,4-photoproducts that are induced by exposure to UV radiation.

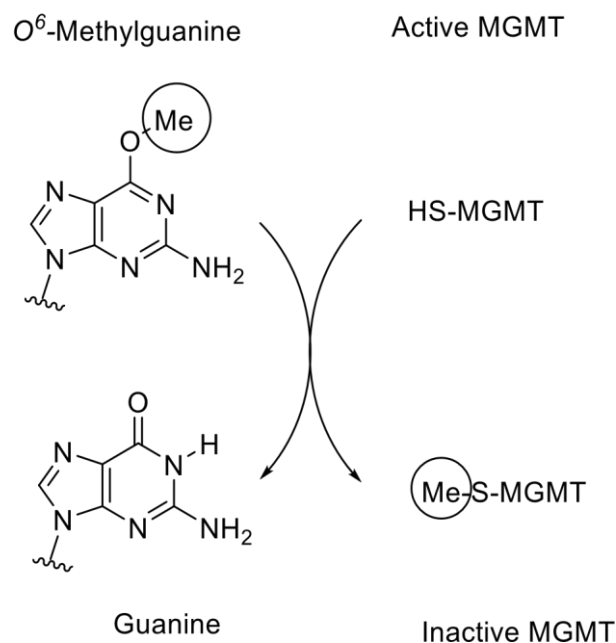
The direct damage reversal repair mechanism de-alkylates some bases that have undergone alkylation. The most abundant alkylated DNA adduct, *N*<sup>7</sup>-methylguanine, is not repaired directly but adducts such as *N*<sup>1</sup>-methyladenine, *N*<sup>3</sup>-methyladenine and *N*<sup>3</sup>-methylcytosine are. *O*<sup>6</sup>-Alkylguanine-DNA-alkyltransferase (AGT) proteins repair highly mutagenic *O*<sup>6</sup>-alkylG adducts *via* direct damage reversal and, therefore, are

important in protecting organisms from carcinogenic effects. The human protein is also known as *O*<sup>6</sup>-methylguanine-DNA methyltransferase (MGMT)<sup>37</sup>. The most well-characterised bacterial AGTs in *E.coli* are the constitutive Ogt protein and the inducible Ada protein<sup>18</sup>. Both proteins are known to repair *O*<sup>4</sup>-alkyl T adducts and *O*<sup>6</sup>-alkylG adducts whilst Ada can also repair phosphate triesters in the sugar-phosphate backbone of DNA<sup>18</sup>. Ogt shows a preference to *O*<sup>4</sup>-alkyl T adduct repair<sup>38</sup> and wild type *E.coli* contains an estimated 30 Ogt molecules<sup>39</sup>. Ada is part of the adaptive response of the bacteria<sup>40</sup> of which a full adapted cell contains 3000 molecules<sup>39</sup>. Ada does not have a repair preference<sup>41</sup> though it has previously been reported that it preferentially repair *O*<sup>6</sup>-alkylG adducts over *O*<sup>4</sup>-alkyl T adducts.

## **1.10. MGMT**

MGMT is a mammalian suicide enzyme that repairs *O*<sup>6</sup>-methylguanine and various other *O*<sup>6</sup>-alkylguanines<sup>42</sup> by direct, irreversible transfer of the alkyl group in a S<sub>N</sub>2 reaction to an active cysteine residue found in the catalytic pocket of MGMT<sup>43, 44</sup>. In *E.coli*, the protein Ada performs the same role. Importantly, both *O*<sup>6</sup>-MeG and *O*<sup>6</sup>-CMG adducts are repaired by MGMT<sup>43</sup>. MGMT is the only enzyme involved in the mechanism<sup>45</sup> and the de-alkylation occurs *via* a single stoichiometric bimolecular reaction. After the transfer of the alkyl group, MGMT becomes unstable and inactive and is broken down in the cell. This gives rise to MGMT being referred to as a suicide enzyme<sup>46</sup>. Production of the enzyme by humans depends both on environmental factors and on an intragenic single nucleotide polymorphism (a type of genetic variation amongst a population) therefore the repair of this mutation varies with each individual. This means some individuals will be at greater threat from *O*<sup>6</sup>-alkylG

adducts than others which places importance on the development of a method of detection for  $O^6$ -alkylG adducts.



Scheme 10. Dealkylation of  $O^6$ -methylguanine by the enzyme MGMT. SH represents side chain of Cys 145.

### 1.11. MGMT Radioactive Assay

MGMT-inactivation assays can determine whether specific  $O^6$ -alkylguanines are substrates of MGMT. Due to the removal of the alkyl group from the ODN to MGMT being an irreversible reaction, an  $IC_{50}$  value (the amount of substrate required to inactivate half the protein) can be calculated from an inactivation assay.

Commonly, this uses DNA containing radiolabelled  $O^6$ -[ $^3H$ ]-MeG to quantify the amount of repair<sup>34</sup>. A known amount of an ODN containing the  $O^6$ -alkylguanine of interest is incubated with MGMT after which the radiolabelled DNA is introduced for further incubation.  $O^6$ -MeG is a known substrate of MGMT therefore the radiolabelled methyl group will be transferred to any active MGMT present during

the second incubation. The MGMT is then isolated and measurement of the counts per second (CPS) by scintillation quantifies the amount of radioactivity transferred to the MGMT. A high CPS indicates that there was little inactivation of MGMT in the initial incubation which indicates that the  $O^6$ -alkylguanine is a poor substrate of MGMT.

After repeating for a range of concentrations of the  $O^6$ -alkylguanine, the  $IC_{50}$  can be calculated. The lower the figure, the greater amount of repair of the substrate. The maximum for this is the total transfer of [ $^3H$ ]-Me to MGMT whilst the minimum is the observed background radiation. The  $IC_{50}$  allows for comparison to other  $O^6$ -alkylguanines.

A range of  $O^6$ -alkylG adducts have been tested to see if they are substrates of MGMT<sup>28, 34, 47</sup>. There is no clear trend to determine the likelihood of an  $O^6$ -alkylG adduct being a substrate. For example, MGMT has been found to repair small adducts such as  $O^6$ -MeG and  $O^6$ -EtG<sup>34</sup> as well as larger adducts such as  $O^6$ -benzylguanine ( $O^6$ -BzG) and  $O^6$ -(4-bromothienyl)guanine ( $O^6$ -BThG)<sup>47</sup>, whilst adducts such as  $O^6$ -hydroxyethenylguanine ( $O^6$ -HOEtG)<sup>34</sup> and  $O^6$ -[4-oxo-4-(3-pyridyl) butyl]guanine ( $O^6$ -PobG) are very poor substrates.

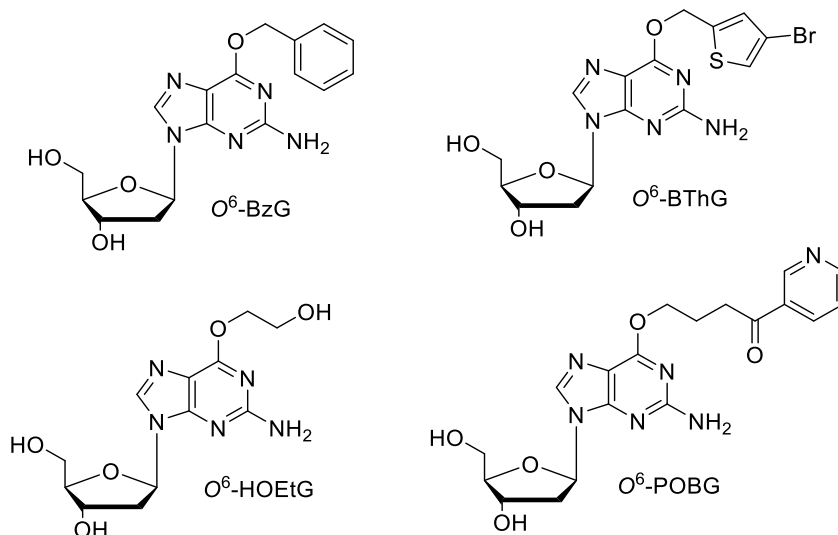


Figure 14. Structures of  $O^6$ -BzG,  $O^6$ -BThG,  $O^6$ -HOEtG and  $O^6$ -POBG which have been determined to be good or poor substrates of MGMT.

### 1.12. Alkyltransferase-Like (ATL) Proteins

Besides MGMT, there is another group of proteins that recognise  $O^6$ -alkylguanine adducts in DNA. ATL proteins are similar in structure to MGMT<sup>48, 49</sup> but instead of an active Cysteine, they have either Tryptophan (Trp) or Alanine (Ala) instead<sup>48</sup>. These amino acids are unable to directly de-alkylate the alkylation damage but do bind specifically to  $O^6$ -alkylG adducts in DNA and protect DNA from the adverse effects by enabling NER<sup>46</sup>.

*Schizosaccharomyces pombe* (Atl1) and *Thermus thermophilus* (TTHA1564) are two such ATL proteins and both have been found to show high affinity to ODNs containing  $O^6$ -alkylG adducts. Unlike MGMT, both proteins have a strong affinity for DNA containing a range of  $O^6$ -alkylG adducts irrespective of size, charge and polarity of the alkyl group<sup>49, 50</sup>. The binding site of ATL proteins has a larger active site cavity than that of MGMT which explains why this occurs.

Due to the high affinity of ATL proteins to  $O^6$ -alkylG adducts, it is possible to identify DNA samples containing such damage through characterisation methods such as non-denaturing polyacrylamide gel electrophoresis (PAGE), fluorescence-based binding assays or enzyme-linked immunosorbent assay (ELISA). AtI1 has a better ability to distinguish between  $O^6$ -alkylG and G<sup>50</sup> than TTHAI564 due to the use of an active site arginine that recognises the electrostatic potential surface of the base adduct.

### **1.13. Characterisation of DNA Containing Alkylated Base Adducts**

There are numerous techniques used to identify and characterise alkylated DNA base adducts. Non-denaturing PAGE has been used to map out damage done to DNA sequences such as strand breaks. Samples are loaded into wells in a polyacrylamide gel and, once a current is applied, will move through the gel. The gel is then stained to produce bands indicating the movement of each sample. Samples containing larger complexes have a retarded mobility. This allows for identification of  $O^6$ -alkylG adducts in DNA because a DNA sample's mobility is retarded in the gel due to the formation of a higher molecular weight protein-DNA complex. The mobility of this will be slower than unbound DNA which can be identified following staining of the gel. By titrating AtI1 protein with a DNA sample and analysing the products, an approximate dissociation constant ( $K_D$ ) can be determined for the complex.

$K_{DS}$  can also be determined using a fluorescent-binding assay in which concentration-dependent binding of AtI1 to a fluorescently labelled substrate is measured. This enables fluorescent quenching of the label and subsequent titration of AtI1 allows determination of  $K_D$  under equilibrium conditions.



An alternative method for detecting alkylated DNA bases uses ELISA. This technique has been used previously to detect  $O^6$ -CMG in DNA using  $O^6$ -CMG antibodies or At11 antibodies<sup>43</sup>. ELISA operates by immobilising a DNA sample to the surface of a microtiter plate and introducing At11 protein. The surface is washed in order to remove any unbound At11 protein before the antibody to At11 protein is introduced prior to a second wash to remove unbound antibody. Finally, a reagent that emits chemiluminescence when attached to the antibody is added. If chemiluminescence is detected then At11 protein has bound to the DNA sample, therefore indicating the presence of  $O^6$ -alkylG DNA damage. The technique can identify  $O^6$ -alkylG DNA in a gel and, similar to PAGE, this technique establishes whether the specific DNA damage is present in a sample but cannot establish the structure.

$^{32}\text{P}$ -postlabelling requires digestion of DNA into 3'-monophosphates before the transfer of  $\gamma$ - $^{32}\text{P}$  from adenosine 5'-[ $\gamma$ - $^{32}\text{P}$ ]-triphosphate to generate 3',5'-[5'- $^{32}\text{P}$ ]-diphosphates. This requires using T4 polynucleotide kinase, an enzyme that catalyses this transfer to produce ADP and 3',5'-[5'- $^{32}\text{P}$ ]-diphosphate. The labelled adducted nucleotides are then separated by TLC and quantified by scintillation counting. This platform can detect one adduct in  $10^{10}$  nucleotides using as little as 1  $\mu\text{g}$  of DNA, a limit of detection greater than other techniques<sup>51</sup>. Despite this advantage,  $^{32}\text{P}$ -postlabelling poses health risks due to the use of a radioactive isotope whilst the accuracy of quantification is dependent on no sample being lost during analysis<sup>51</sup>. It has been noted that this technique consistently underrepresents DNA damage adducts compared to other techniques. On top of this, no structural information is generated.

Quantification and structural determination of damaged DNA adducts can be determined using mass spectroscopy (MS). Liquid chromatography MS (LC-MS) is

the preferred method. Gas chromatography MS (GC-MS) can also be used but DNA adducts require extensive derivatisation steps prior to analysis. This is due to the adducts not being volatile enough to be analysed directly<sup>52</sup>. LC-MS is capable of direct separation of modified nucleotides, nucleosides, bases and ODNs therefore it does not require derivatisation.

Being able to quantify DNA adducts enables the increased knowledge on damage to tissue occurring at any given time which can be put to further use to determine repair potential as well as the susceptibility of an individual to cancer or other illnesses. Another application can be to set exposure limits to environmental and industrial chemicals, enabling safer working environments, for example.

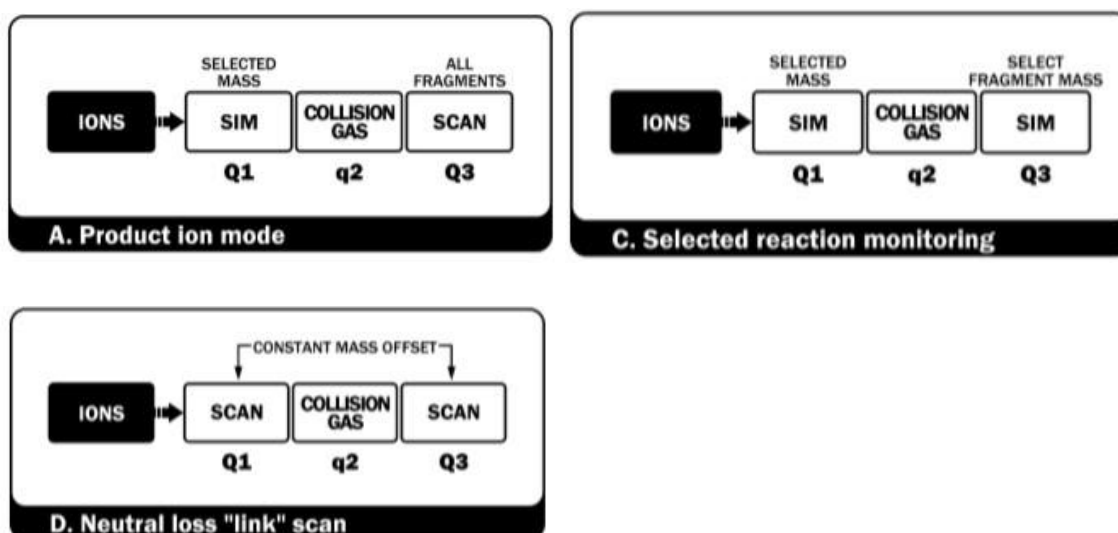
#### **1.14. Detection of DNA base adducts using Mass Spectrometry**

Electrospray Ionisation (ESI) and Matrix Assisted Laser Desorption Ionisation (MALDI) are the two most common ionisation techniques used to detect DNA modifications, especially when coupled with LC<sup>52</sup>. ESI enables direct analysis of polar molecules (ODNs, nucleotides/sides) where either the protonated or deprotonated ion is generated, dependent on the polarity of voltage applied. Sensitivity is enhanced with a slower flow rate in the HPLC column that fractionates the sample prior to MS. If ESI is used for analysis, the method has to consider ion suppression<sup>53, 54</sup>. This is the interference of other ionised compounds in the mixture with the analyte. This reduces sensitivity but can be solved by efficient HPLC separation. MALDI produces singly charged ions that are either positive or negative and is used to great proficiency for analysing ODNs or nucleic acids<sup>52</sup>.

Information on the analyte is gathered by ion scanning modes. Selected ion monitoring (SIM) and selected reaction monitoring (SRM) are highly sensitive so are

suitable to the quantification of the specific adduct<sup>52</sup>. The sensitivity is obtained because SIM only allows detection of a specific mass-to-charge ( $m/z$ ) ratio, therefore disregarding the rest of the spectrum. SRM utilises SIM across two analyses to optimise sensitivity (Scheme 11). Constant neutral loss (CNL) is used when multiple analytes or unknown adducts are being tested and enables the determination of the structure of the analyte<sup>55</sup>. For example, loss of the glycosidic bond between a nucleobase and deoxyribose ring produces a constant loss of 116  $m/z$ , the mass of deoxyribose. This is constant for all 2'-deoxyribonucleosides and provides further evidence of their presence (Figure 15).

Scheme 11 shows the process of these three scanning methods<sup>56</sup>.



Scheme 11. Tandem mass spectrometry scanning modes SIM; SRM and CNL. (Scheme taken from Tretyakova *et al.*)<sup>56</sup>.

Q1 refers to the first quadrupole mass analyser. Mass analysers have a filter that sets a range of  $m/z$  that enables certain ions to pass through whilst forbidding ions outside the range. q2 is a collision cell where ions collide with inert gas which results in the cleavage of weaker bonds, called collision induced dissociation (CID). The fragments produced from such collisions enable further identification because masses

corresponding to specific fragments would be observed, thus adding to the overall structure of the analyte (Figure 15).

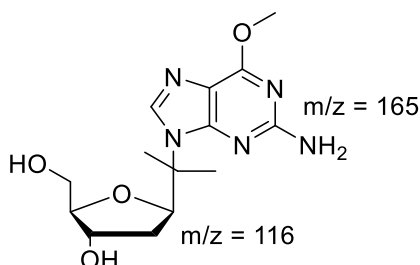


Figure 15. Fragmentation of N9 glycosidic bond of  $O^6$ -MeG, induced by CID, resulting in a CNL fragmentation of  $m/z = 116$  corresponding to 2'-deoxyribose.  $M/z = 165$  would also be detected, indicating  $O^6$ -MeG nucleobase.

These cleaved ions are finally analysed in Q3, another quadrupole mass analyser. Triple quadrupoles (Q1q2Q3) are very selective, sensitive, precise and have a large range therefore are commonly used.

### 1.15. Stable Isotope Dilution (SID)

Quantification of DNA adducts is possible because of isotope dilution mass spectrometry (IDMS). A stable isotopically labelled analogue of the adduct is used as an internal standard. Quantification requires high sensitivity and precision and IDMS enables a level of specificity that cannot be attained by any other analytical techniques. In stable isotope dilution (SID), the internal standard is spiked into a sample at a known concentration<sup>52</sup>. After validation of the method, whereby a calibration plot of the amount of labelled analyte vs MS signal is determined, unknown samples can be looked at. A ratio is established between the analyte and the labelled standard and, using the calibration plot, the absolute amount of the analyte in the sample can be deduced.

Having a stable isotope-labelled analogue is the most desirable standard because it will mimic exactly how the behaviour of the analyte during extraction and the processing of the sample as it is chemically identical<sup>52</sup>. It will also account for ion suppression. A structural analogue will have a different retention time and different ionisation properties which means the accuracy will not be absolute. <sup>13</sup>C or <sup>15</sup>N isotopically-labelled analytes will have near on identical retention times to the unlabelled analyte which again improves the precision of the technique<sup>52</sup>. Deuterated isotopes can be at risk of isotopic exchange under certain conditions and usually have a shift in retention time<sup>57</sup>.

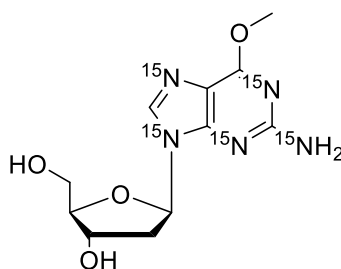


Figure 16. <sup>15</sup>N<sub>5</sub>-labelled *O*<sup>6</sup>-MeG that can be used to develop a SID method to quantify *O*<sup>6</sup>-MeG adducts in DNA.

### 1.16. Previous Studies Using SID

Many isotope-labelled biomarkers have been used to quantify various compounds that can lead to cardiovascular or neurodegenerative diseases or cancer<sup>58, 59</sup>. Biomarkers for lipids, steroids, homocysteine and glutathione adducts have all been studied. *O*<sup>6</sup>-CMG has been studied using a deuterated standard<sup>60</sup> and with tubercidin (Figure 17)<sup>61</sup>.

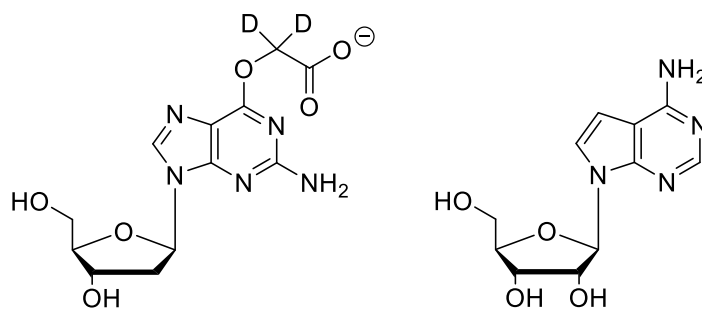


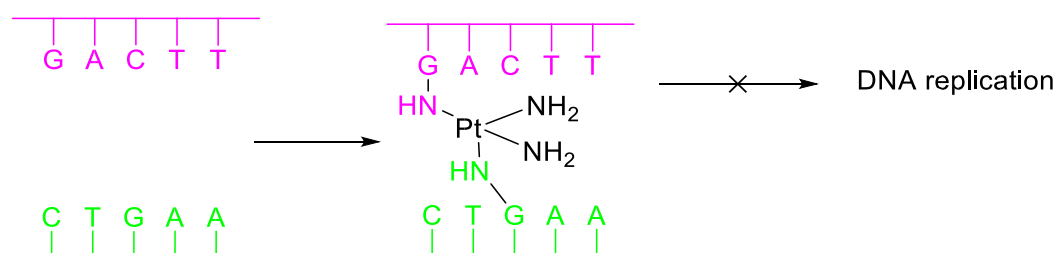
Figure 17. Structures of  $^2\text{H}_2$ -labelled  $O^6$ -CMG and tubercidin previously used in the development of a SID method to identify  $O^6$ -alkylG adducts.

The first method was able to quantify DNA taken from calf thymus but suffered from decomposition of the deuterated standard, around 12 %<sup>60</sup>. For the second method, tubercidin, though not isotopically-labelled, was used as the ISTD because its structure, ionisation pattern and sensitivity were all deemed appropriate. However, the sample DNA was taken from urine and there was a problem with the limit of quantification (LOQ)<sup>61</sup>. This may have been due to the fact that urine contains adducts that have been removed by NER and BER whilst  $O^6$ -CMG is most likely repaired by MGMT and it is unclear whether adducts removed by MGMT appear in urine. If a SID method could be developed that incorporated  $O^6$ -alkylG adducts, many doors would be opened in characterising the properties of the adducts. SID enables the study of DNA repair mechanisms<sup>62, 63</sup> whilst also being put to use as a way of determining the risk individuals were at of developing life-threatening diseases. Studies looking at  $O^6$ -MeG have determined the effect of the DNA sequence on the repair capabilities of MGMT<sup>63</sup> as well as ISC repair of cancer patients undergoing chemotherapy<sup>64</sup>.

### 1.17. Interstrand Crosslinks

Interstrand crosslinks (ICLs) block DNA replication and transcription. This can cause mutagenic and carcinogenic effects whilst also causing cytotoxic DNA adducts.

ICLs covalently link two complementary strands of a DNA duplex and can be caused through reaction with bifunctional electrophilic agents such as  $\alpha$ ,  $\beta$ -unsaturated aldehydes. Due to their cytotoxicity, ICL agents are used in numerous anticancer chemotherapies such as urea and platinum complexes<sup>65, 66</sup>. However, tumours are developing resistance to these agents and are able to remove ICLs from DNA, making the drugs obsolete<sup>67</sup>.



Scheme 12. Cisplatin induced formation of an interstrand crosslink in a DNA duplex.

The carcinogenic and mutagenic effects caused by enals lead to several transition and transversion mutations, predominantly G→T as well as G→C and G→A. Eukaryotic cells can be killed by a single interstrand crosslink. Malondialdehyde is a common enal that has been found to cause ICLs whilst formaldehyde, though not an enal, is a highly reactive species that also causes ICLs<sup>68</sup>. These endogenous and exogenous metabolites pose a threat to DNA but ICL repair pathways do occur. ICLs can occur between several different parts of the DNA duplex.

### 1.18. Formaldehyde

Formaldehyde is an abundant chemical in the body, a by-product of oxidative demethylation reactions caused by enzymes in the nucleus. It is also present in cigarette smoke and methanol (trace amounts that can arise in the diet). Formaldehyde readily forms ICLs with both proteins and nucleic acids<sup>69-71</sup> due to its high reactivity. The

exocyclic amino groups at the N2 position of guanine and N6 position of adenine are susceptible to reaction with formaldehyde and lead to ICLs, either across the minor groove forming dG-dG ICLs or across the major groove forming dA-dA ICLs (Figure 18).

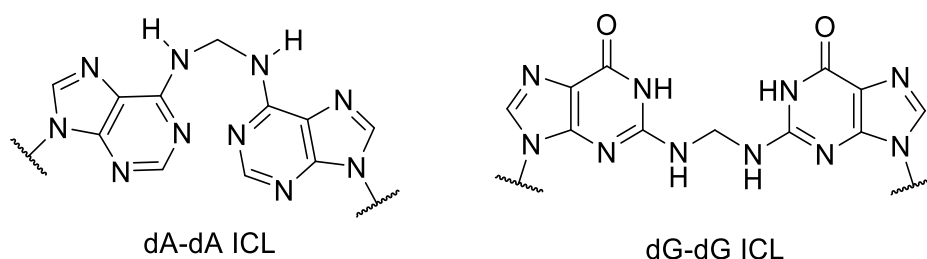
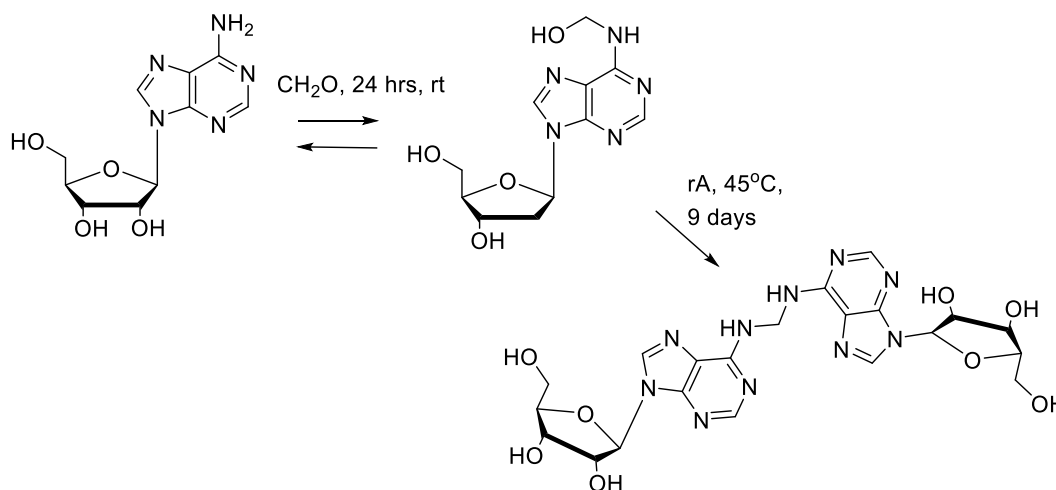


Figure 18. Formaldehyde-induced interstrand crosslinks between 2'-deoxyadenosine bases and 2'-deoxyguanosine bases.

The reaction has been studied by Chaw *et al.* who synthesised a methylene bridged di-adenosine unit by reacting adenosine with aqueous formaldehyde. The products were purified and characterised by reverse-phase HPLC and  $^1\text{H}$  NMR<sup>72</sup>.



Scheme 13. Reaction of adenosine with aqueous formaldehyde to form a methylene interstrand crosslink.

Huang and Hopkins extended these studies and synthesised a variety of ICL-containing DNA samples containing formaldehyde-induced dA-dA ICLs (the



predominant ICL) and found that the DNA sequence influences the chances of forming an ICL<sup>73</sup>. Importantly, the crosslink is noted to form in the same groove (due to its short bridging ability) and would be between adjacent base pairs (because of the need of the amino groups to be in the same groove)<sup>73</sup>. Table 2 shows three of the most efficient nucleotide sequences in producing an ICL between two adenines though it is important to take note that the extent of ICL formation is very low in products, less than 1 %<sup>73</sup>. These three sequences varied by four bases in the middle of the sequence, denoted by N<sub>4</sub> in the table. The ICLs were synthesised by reacting ODNs with aqueous formaldehyde in a sodium acetate buffer for 9 days and then characterised by denaturing polyacrylamide gel electrophoresis. The amounts of products obtained were determined using gel-scanning densitometry.

Table 2. Nucleotide sequences and percentage of interstrand crosslinking between adenines when reacted with formaldehyde<sup>73</sup>.

<i>Nucleotide sequence</i>	<i>% of interstrand crosslinked product</i>
<i>General sequence and complementary strand: 5' TACAACN<sub>4</sub>GTTGT 3'</i> <i>3'TGTTGN<sub>4</sub>CAACAT 5'</i>	
<b>N<sub>4</sub> = 5' AATT</b>	0.35
<b>N<sub>4</sub> = 5' ATAT</b>	0.31
<b>N<sub>4</sub> = 5' TATA</b>	0.18

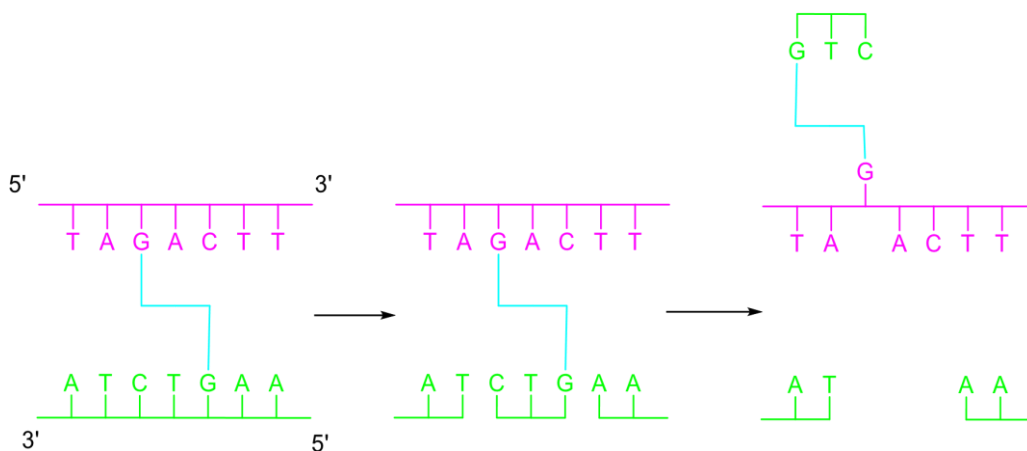
The products of such formaldehyde damage are removed from cells by the protein, aldehyde dehydrogenase gene ADH5<sup>74</sup>. ADH proteins operate to remove damage derived through reaction with specific aldehydes. Therefore, if the body is deficient in a specific ADH protein, it is more susceptible to damage from the relevant aldehyde.

For example, ADH2, which removes acetaldehyde, will not remove formaldehyde if the body is lacking ADH5<sup>75</sup>. Studies looking at the inhibition of ADH5 found that formaldehyde-induced ICLs lead to kidney, liver and bone marrow failure<sup>75</sup>. This is clear evidence that the build-up of formaldehyde in the body has carcinogenic effects.

### 1.19. Fanconi Anemia and ICL Repair

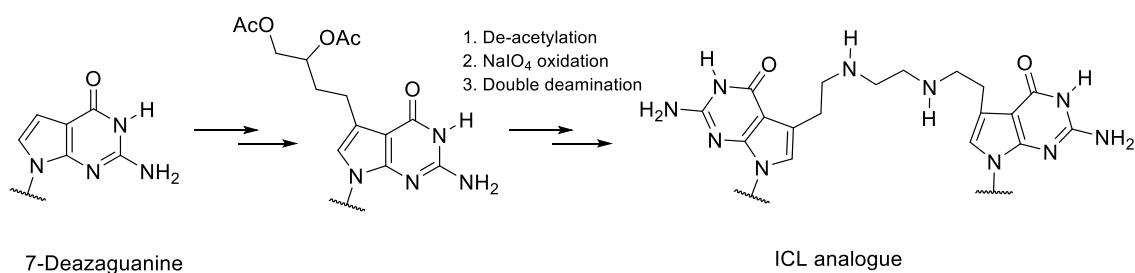
Fanconi anemia (FA) is a human cancer predisposition that can be characterised by sensitivity to DNA ICL-inducing agents and instability of the genome<sup>76</sup>. This means characterisation of the removal of ICLs is important for patients with FA. Previous research into the repair of ICLs has identified the FA proteins FANCI and FANCD2 as well as the nuclease XPF-ERCC1 as key components in repairing ICLs in DNA<sup>74, 76</sup>.

The mechanism of ICL repair is complicated. Two replication forks descend on the ICL, where one causes a dual incision either side of the adduct whilst the other stops 20 nucleotides away. The cleaved side is on the lagging strand which allows sister chromatid separation. This step is known as “unhooking”. The adduct is bypassed by DNA polymerase which allows the regeneration of an intact sister chromatid, enabling homologous recombination repair of the double strand break<sup>74</sup>.



Scheme 14. Process of ‘unhooking’ during the repair of an interstrand crosslink. The gap that is left on the complementary strand has new base pairs inserted in by DNA polymerase whilst the nucleotides that still contain the crosslink are removed by NER.

Research into the repair mechanism was enabled by the synthesis of complementary ODNs containing 7-deazaguanine residues bearing aldehyde groups<sup>77</sup>. Through double reductive amination, ICLs are formed<sup>77</sup>.



Scheme 15. Synthesis of interstrand crosslink-containing analogue from 7-deazaguanosine enabling characterisation of the ICL repair mechanism<sup>78</sup>.

These ICL analogues can be synthesised in high yield and purity<sup>79</sup>, an advantage over ICLs synthesised from direct reaction between DNA and ICL agents. The latter analogues are a mixture of monoadducts, intra- and interstrand crosslinks where only 1-5 % of the desired ICL is obtained<sup>80-82</sup>. Another advantage of the 7-deazaguanine analogue is that it enables the formation of a major-groove ICL whilst not disturbing the double helix, unlike cisplatin<sup>83</sup>. A disadvantage of using 7-deazaguanine, however, is that it is a synthetic, rather than a naturally occurring, ICL.

Figure 19 shows the ODN sequences containing 7-deazaguanine used to study the repair mechanism of ICLs. Mice deficient of *SLX4* were susceptible to FA symptoms and the repair mechanism studies showed the role of XPF-ERCC1<sup>74</sup>.

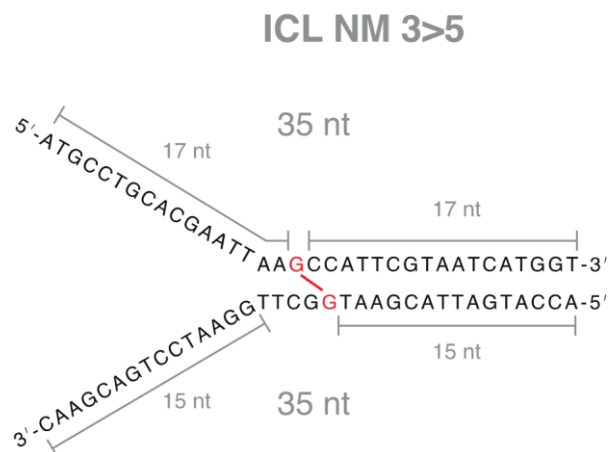
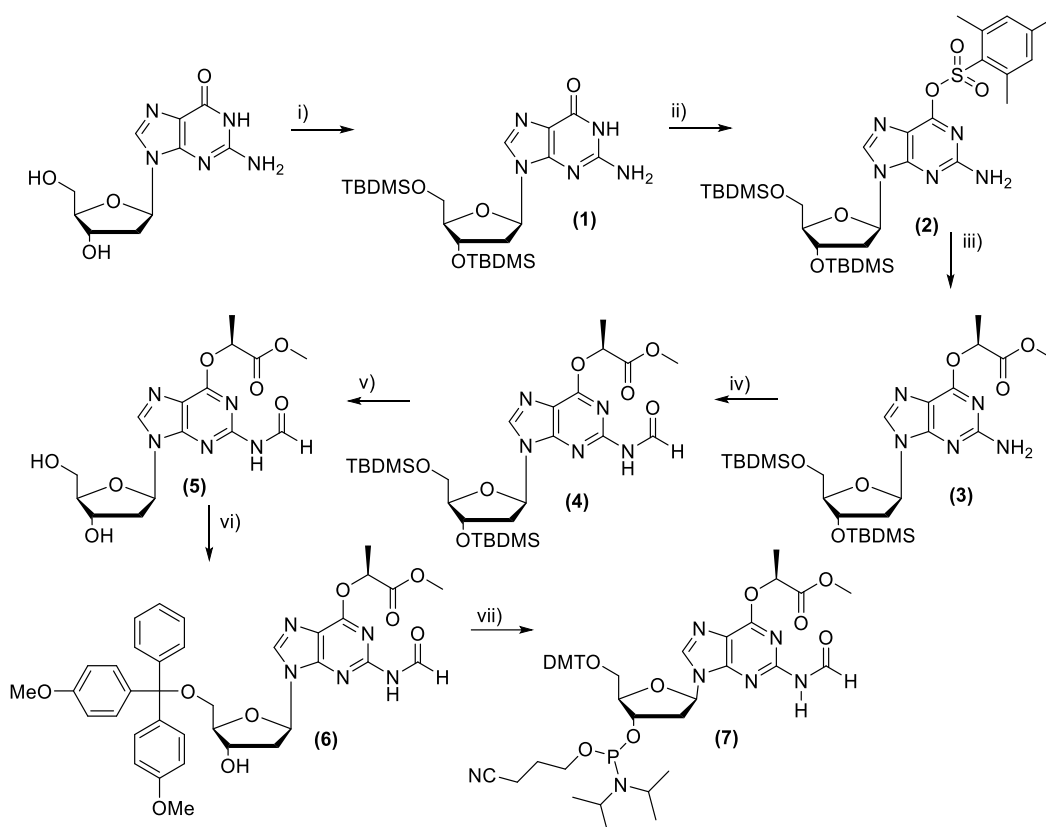


Figure 19. Duplex formed by the synthesis of an ICL between two guanine bases inserted used to study the ICL repair mechanism in mice<sup>74</sup>.

## 1.20. Project Aims

This project has three aspects, each looking at and characterising a specific type of DNA damage. The first aim is to study the recognition and repair of  $O^6$ -alkylguanines derived from nitrosated amino acids, specifically glycine and alanine. This will require synthesis of ODNs containing  $O^6$ -CMG and  $O^6$ -CEG, for which a novel phosphoramidite synthesis of the latter is required. The proposed route is shown in Scheme 16 and based on the synthesis of the phosphoramidite of  $O^6$ -CMG<sup>85</sup>.



Scheme 16. Proposed synthesis of  $O^6$ -CEG phosphoramidite from 2'-deoxyguanosine. i) TBDMSO, imidazole, anhydrous DMF; ii) mesitylenesulfonyl chloride, DMAP, Et<sub>3</sub>N, DCM; iii) methyl(-)-lactate, DABCO, DBU, 1, 2-DME; iv) *N,N*-dimethylformamide dimethyl acetal, anhydrous DMF; v) TBAF, THF; vi) dimethoxytrityl chloride, pyridine; vii) 2-cyanoethyl-*N*-diisopropylamine chlorophosphoramidite, DIPEA, anhydrous DCM.

Previous research in the Williams group has reported issues with the synthesis of ODNs containing carboxylates that require ODN deprotection using aq NaOH rather than aq conc. ammonia solution (that produces the carboxamide). There is uncertainty whether MGMT substrate data reported by Senthong *et al.*<sup>43</sup> is for  $O^6$ -CMG or  $O^6$ -CMG carboxamide, therefore, ODN deprotection conditions will be re-examined to distinguish between carboxylate and carboxamide formation. MGMT assays will then be performed to identify whether  $O^6$ -CMG or  $O^6$ -CEG are substrates of MGMT and to examine the repair of the corresponding carboxamide analogues.

The second aim of the project is to synthesise and develop SID methods for identifying  $O^6$ -alkylG adducts in DNA using mass spectrometry for four <sup>15</sup>N<sub>5</sub>-labelled

$O^6$ -alkylG internal standards. These are  $O^6$ -MeG,  $O^6$ -EtG,  $O^6$ -CMG and  $O^6$ -CEG (Figure 20).

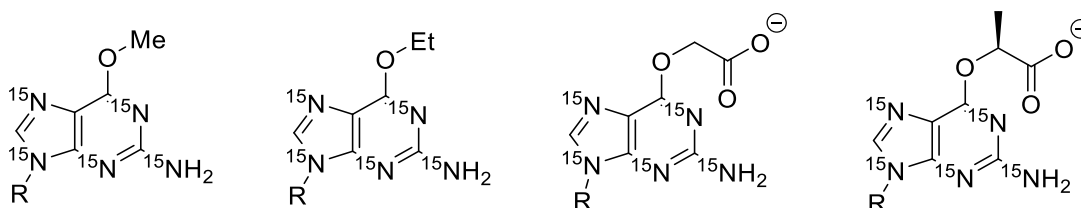
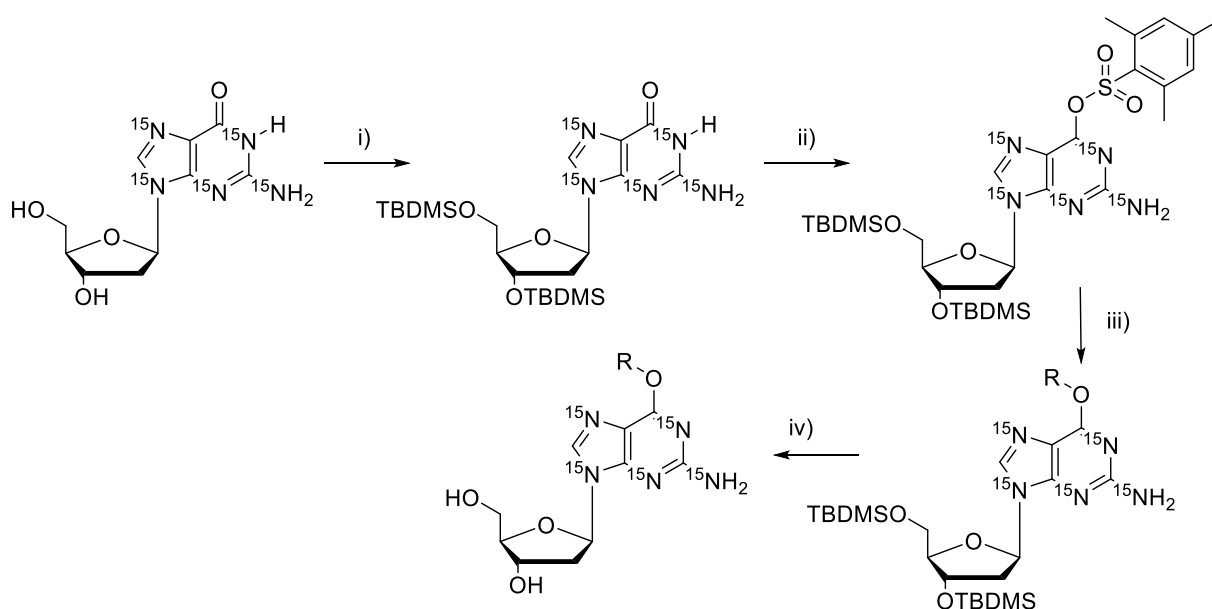


Figure 20.  $^{15}\text{N}_5$ -labelled structures of  $O^6$ -alkylguanine compounds  $O^6$ -MeG,  $O^6$ -EtG,  $O^6$ -CMG and  $O^6$ -CEG, left to right, to be synthesised for use as stable-isotopically labelled internal standards in SID methods. R = 2'-deoxyribose sugar.

The internal standards will be synthesised following a procedure by Millington *et al.* (Scheme 17)<sup>85</sup>.



Scheme 17. Proposed synthetic route to  $^{15}\text{N}_5$ -labelled  $O^6$ -alkyl-2'-deoxyguanosines, based on Millington *et al.*<sup>84</sup>, to be used as SIL internal standards starting from 2'-deoxyguanosine (100 mg scale); i) TBDMSCl, imidazole, DMF; ii)  $\text{Et}_3\text{N}$ , DMAP, 2-mesitylenesulfonyl chloride, DCM; iii) DABCO, ROH, DBU, 1, 2-DME; iv) TBAF in THF.

SID is the preferred approach in quantifying DNA adducts within an unknown sample because it offers high sensitivity and precision as well as quantitative and qualitative data. To develop each SID method, a calibration coefficient is needed to be calculated. To achieve this, an ODN containing a single modification of the required

*O*<sup>6</sup>-alkylG adduct will be digested using a nuclease mixture and spiked with a known amount of the <sup>15</sup>N<sub>5</sub>-labelled internal standard. The analysis of the relative abundance between the unlabelled and labelled adduct by LC-MS will enable calculation of the coefficient. A limit of quantification is also required. This is defined as lowest amount of analyte in a sample, which can be quantitatively determined with suitable precision and accuracy. The SID method will be developed using UHPLC which is deemed the best system.

The third aim of the project is to synthesise an ODN duplex containing a fICL that would be used by collaborators to characterise the repair mechanism of such DNA damage. Formaldehyde is found abundantly in the human body and will cause crosslinks at various sites. Formaldehyde-induced ICLs lead to kidney, liver and bone marrow failure<sup>47</sup>. Potential routes to this modified DNA will explore reaction of methylene diamine or an equivalent reagent with complementary ODN sequences containing 6-chloropurine following hydrolysis of amidine-protected adenine precursors or by developing methods using reaction with aqueous formaldehyde.





**Chapter 2- *Recognition and Repair of O<sup>6</sup>-Alkylguanines  
Derived From Nitrosated Amino Acids***

## **2. Recognition and Repair of $O^6$ -Alkylguanines Derived From Nitrosated Amino Acids**

### **2.1. Introduction**

$O^6$ -Alkylguanine adducts are significant risk factors for development of colorectal cancer (CRC). Colonic tumours have been found in regions expressing low MGMT activity whilst in normal colonic tissue, low activity results in a greater number of GC→AT transition mutations (that are most commonly derived from  $O^6$ -alkylguanines)<sup>18, 43, 85</sup>. Since MGMT removes  $O^6$ -alkylguanine adducts from DNA and these adducts result in transition mutations, it is clear that MGMT provides a valuable role in protecting against CRC.

Diets low in fibre are long known to increase the risk of developing CRC. In addition, diets rich in red and processed meats also increases CRC risk due to the greater exposure of colonic DNA to alkylating agents. These alkylating agents include *N*-nitroso compounds that are found in the diet or form from bacterial or chemical *N*-nitrosation of amines<sup>32, 33</sup>. As discussed in the introduction, *N*-nitroso compounds can be converted to alkylating agents that in turn produce mutagenic  $O^6$ -alkylguanine adducts. One such  $O^6$ -alkylguanine adduct is  $O^6$ -CMG (Figure 21) which can be attributed to *N*-nitrosation of glycocholic acid, a component of bile, of which several grams are produced daily<sup>31, 43</sup>. It has been shown that DNA in exfoliated colon cell DNA contains significant levels of  $O^6$ -CMG<sup>31, 43</sup>, providing a potential link between diet,  $O^6$ -CMG formation and CRC.

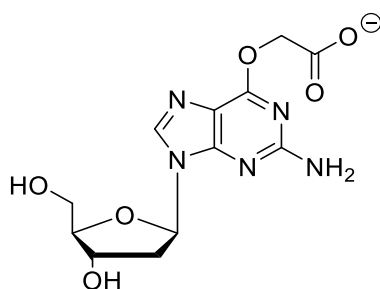
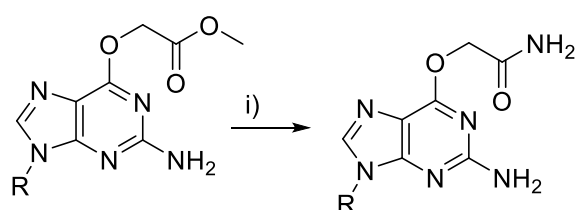


Figure 21. Structure of  $O^6$ -CMG.

The ability of MGMT to repair a given  $O^6$ -alkylG adduct will affect its persistence and therefore its toxicity. Evidence that the carboxymethyl group of  $O^6$ -CMG was removed by MGMT initially proved inconclusive. Using cell-free extracts from *E. coli*, first treated with a carboxymethylating agent, indicated that the adduct was not removed by alkyltransferases<sup>86</sup>. However, *E. coli* Ada, instead of mammalian MGMT, was used in the study and the two may act differently. Likewise, research by Pletsas *et al.* studied the inactivation of MGMT by an  $O^6$ -CMG-modified purine base that further suggested that it is not a substrate<sup>87</sup>. However, ODNs containing the modification were not studied. This is required because the mechanism of removal may rely on DNA-protein interactions that enable MGMT to identify and remove the adduct which are not replicated through purine-protein interactions.

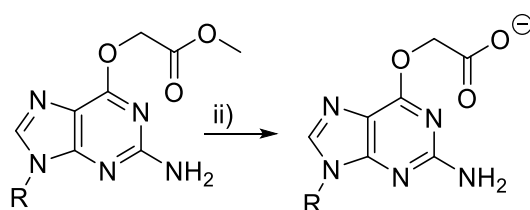
More recently, Senthong *et al.* analysed the inactivation of MGMT and Ogt proteins by ODNs containing  $O^6$ -CMG to calculate an  $IC_{50}$  to determine whether  $O^6$ -CMG is a substrate of MGMT. Analysis of the active site peptide of MGMT after alkyl transfer and treatment with trypsin using mass spectrometry indicated the transfer of the alkyl group from the ODN to MGMT<sup>43</sup>. The  $IC_{50}$  calculated indicated that  $O^6$ -CMG is a substrate of MGMT, with a comparable value to  $O^6$ -MeG. The  $IC_{50}$  for  $O^6$ -CMG was reported as 1.7 nM whilst the  $IC_{50}$  of  $O^6$ -MeG was calculated as 0.9 nM. However, the calculated mass of the alkylated active site peptide was one less than expected

suggesting that the ODN used in the study contained a different modification to carboxymethyl. This may have arisen during the deprotection of the ODN whilst preparing it for MGMT studies. As discussed in section 1.4, ODNs are treated directly with conc. aq ammonia to remove the base-labile protecting groups that are in place during the ODN synthesis. However, as noted by Xu<sup>88</sup>, ammonia can convert the methyl ester group into a carboxamide instead of the required carboxylate group (Scheme 18).



Scheme 18. Formation of ODN containing *O*<sup>6</sup>-CMG carboxamide after treatment of ODN containing *O*<sup>6</sup>-methylglycolate with conc. aq ammonia (i).

To avoid this, initial hydrolysis of the ester by aq NaOH is required, followed by treatment with conc. aq ammonia (Scheme 19)<sup>84</sup>. The latter treatment ensures the complete removal of base protecting groups. The initial hydrolysis ensures full conversion of the methyl ester to the carboxylate, therefore forbidding any carboxamide formation, whilst the base-labile protecting groups that are in place during the ODN synthesis are still removed upon treatment with conc. aq ammonia. Importantly, the carboxamide will have a mass one less than the carboxylate. This would explain the incorrect mass analysis of MGMT by Senthong *et al.*<sup>43</sup>.



Scheme 19. Formation of ODN containing *O*<sup>6</sup>-CMG after treatment of ODN containing *O*<sup>6</sup>-methylglycolate with aq 0.5 M NaOH followed by conc. aq ammonia (ii).

Synthesising an ODN containing  $O^6$ -CMG and ensuring the method of deprotection was correct would enable clear and conclusive evidence as to whether  $O^6$ -CMG is effectively repaired by MGMT and that the work reported by Senthong *et al.*<sup>43</sup> was carried out using  $O^6$ -(carboxamidylmethyl) guanine ( $O^6$ -CMG carboxamide). To ensure this, the ODN will be deprotected by both described methods and the samples analysed by MS and RP-HPLC. We expect the ODN sample treated solely with conc. aq ammonia to have a mass one less than the sample treated with aq NaOH and conc. aq ammonia because the first sample will contain the carboxamide. The substrate properties of both ODNs will then be tested with MGMT.

A putative adduct related to  $O^6$ -CMG is  $O^6$ -CEG. Although there has been no research confirming the presence of  $O^6$ -CEG (Figure 22) in human DNA, since alanine, like glycine, is abundant in red meat,<sup>33</sup> *N*-nitrosation of alanine is likely to occur. *N*-nitrosation of  $\alpha$ -amino acids produces the  $\alpha$ -lactones, which are known to be highly reactive alkylating agents<sup>32</sup>.

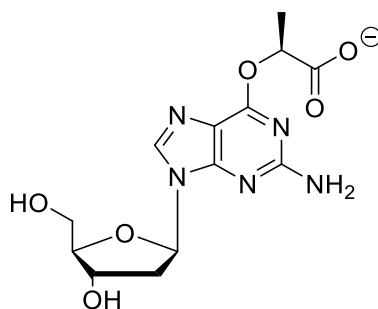


Figure 22. Structure of  $O^6$ -CEG.

It is hypothesised that  $O^6$ -CEG occurs following reaction of guanine with from the alkylation of 2'-deoxyguanosine by the alkylating agent derived from *N*-nitroso alanine (Figure 23). Studying the repair of  $O^6$ -CEG by MGMT will give an insight into the potential mutagenic risk posed by the putative adduct: adducts that are poor MGMT substrates are likely to persist and be more problematic.

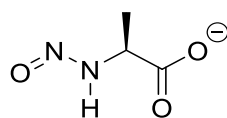


Figure 23. *N*-nitroso alanine.

Synthesising ODNs containing *O*<sup>6</sup>-CEG would require a novel synthesis of the *O*<sup>6</sup>-CEG phosphoramidite (Figure 24). Since there is a potential problem in inadvertent preparation of *O*<sup>6</sup>-cboxamide adducts, we decided to prepare ODNs containing both the carboxylate and carboxamide forms of *O*<sup>6</sup>-CMG and *O*<sup>6</sup>-CEG.

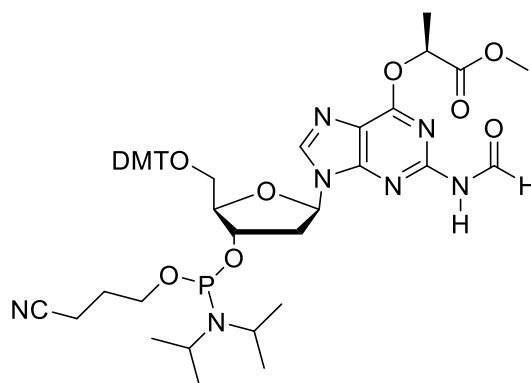
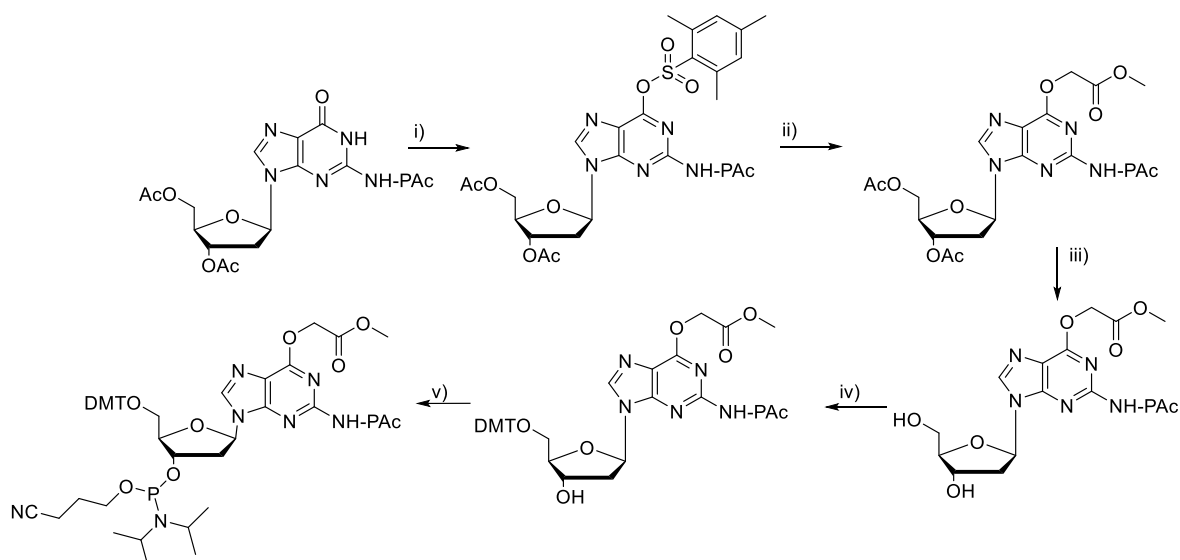


Figure 24. Phosphoramidite of *O*<sup>6</sup>-CEG used to synthesise an ODN containing *O*<sup>6</sup>-CEG to study MGMT repair of the adduct.

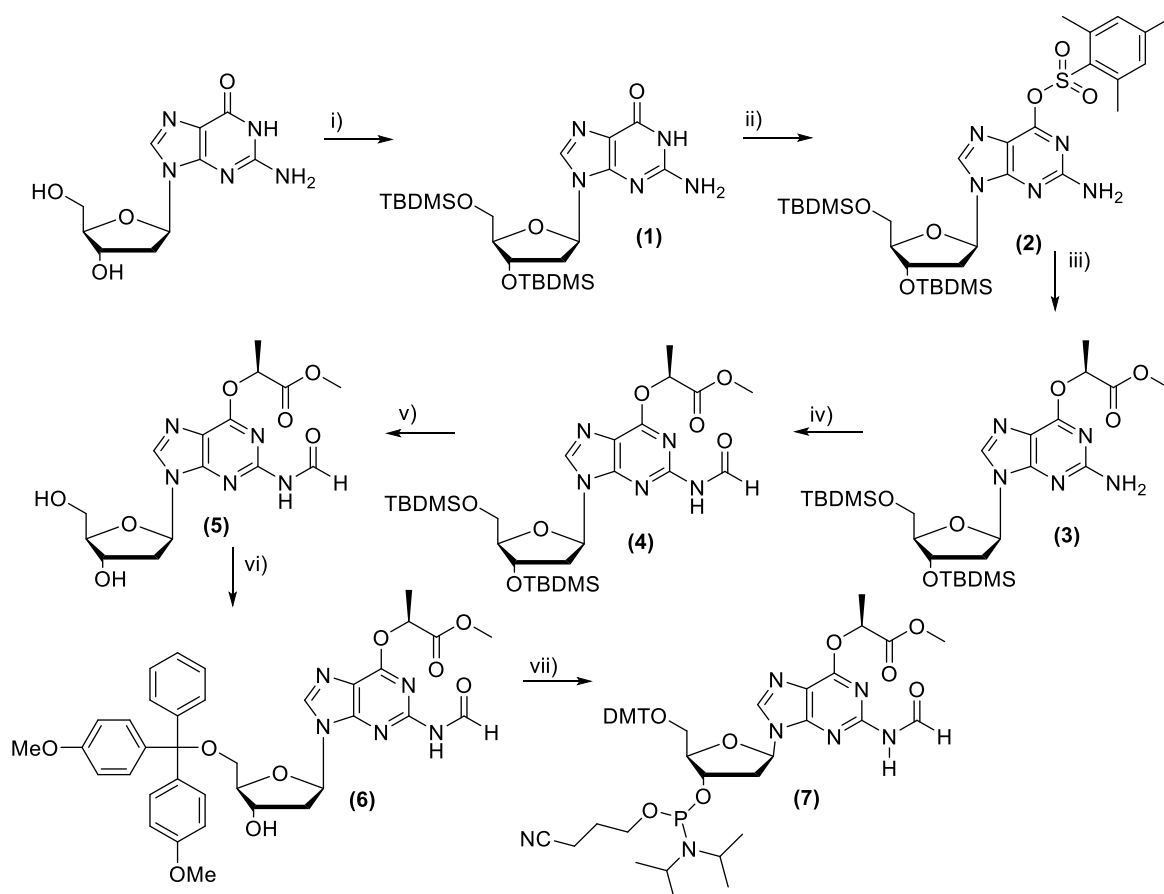
## 2.2.Synthesis of *O*<sup>6</sup>-CEG Phosphoramidite

The synthesis of the *O*<sup>6</sup>-CEG phosphoramidite followed the general synthetic route that the Williams group had previously developed for the synthesis of the phosphoramidite for the incorporation of *O*<sup>6</sup>-CMG<sup>84</sup>. ODNs containing *O*<sup>6</sup>-CMG have previously been prepared by Xu (Scheme 20)<sup>88</sup>.



Scheme 20. Synthetic route of phosphoramidite of  $O^6$ -CMG derived by Xu<sup>88</sup>. i) 2-Mesitylenesulfonyl chloride, DMAP, Et<sub>3</sub>N, DCM; ii) methyl glycolate, Quinuclidine, DBU, 1, 2-DME; iii) 0.5 M Et<sub>3</sub>N/ MeOH; iv) dimethoxytrityl chloride, pyridine; v) 2-cyanoethyl-*N*-diisopropylamine chlorophosphoramidite, DIPEA, anhydrous DCM.

However, two issues arise. Firstly, Xu protected the N2 position with the phenylacetyl group. This requires relatively harsh conditions for full removal following ODN synthesis that can lead to deamination<sup>2</sup>. Thus, protection by the more labile formamidine group was deemed more suitable<sup>89</sup>. Secondly, the displacement of the mesitylene sulfonate ester is achieved using quinuclidine and methyl glycolate. Quinuclidine is a catalyst that activates the site by displacement of the sulfonate ester, but it is very expensive. However, work reported by Lakshman indicated that activation of the site for the displacement was able to be achieved using DABCO, a more affordable substitute<sup>90</sup>. The full synthetic route to the  $O^6$ -CEG phosphoramidite was based on the previous synthesis of the  $O^6$ -CMG phosphoramidite<sup>84</sup> and can be seen below (Scheme 21).



Scheme 21. Synthesis of  $O^6$ -CEG phosphoramidite from 2'-deoxyguanosine. i) TBDMSCl, imidazole, anhydrous DMF; ii) mesitylenesulfonyl chloride, DMAP, Et<sub>3</sub>N, DCM; iii) methyl(-)-lactate, DABCO, DBU, 1, 2-DME; iv) *N,N*-dimethylformamide dimethyl acetal, anhydrous DMF; v) TBAF, THF; vi) dimethoxytrityl chloride, pyridine; vii) 2-cyanoethyl-*N*-diisopropylamine chlorophosphoramidite, DIPEA, anhydrous DCM.

Unlike the previous synthesis routes, the stereochemistry of the alcohol had to be taken into consideration during step iii), in order to synthesise the correct enantiomer.  $O^6$ -CEG may occur naturally upon reaction of the  $\alpha$ -lactone of *N*-nitrosated alanine with the O6 position of guanine and this gives the *S*-enantiomer with no racemisation. Previous work in the Williams group (unpublished) displaced mesitylene sulfonate ester with methyl(-)-lactate and the reaction was deemed to replicate the naturally occurring alkylation by *N*-nitroso alanine. This meant that during the synthesis, methyl(-)-lactate was used as a reagent in step iii).

The initial step was the protection of the 3' and 5' hydroxyl groups of 2'-deoxyguanosine by TBDMSCl in anhydrous DMF. After removal of the solvent, and



addition of MeOH, **1** was isolated in quantitative yield. Subsequently, reaction with 2-mesitylenesulfonyl chloride in anhydrous DCM for 18 hours, gave **2** with an 81 % yield after silica column chromatography. This was followed by displacement of the mesitylene sulfonate group with methyl(-)-lactate, *via* a DABCO salt intermediate, to give **3** which was purified by silica column chromatography with an 80 % yield.

The methoxycarbonylethyl group was left as its methyl ester form to act as a protecting group for the carboxylate during the rest of the phosphoramidite synthesis and the ODN synthesis that followed. Hydrolysis of the ester during the ODN deprotection (section 1.4) would enable formation of the desired modification.

Another alteration to the synthetic route<sup>84</sup> was the conversion of **3** to **4**. Previously, the formamidine group was introduced as a protecting group at the N2 position of dG and the product purified by silica gel chromatography with Et<sub>3</sub>N included in the eluent. This was to neutralise the acidic silica and ensure that no protonation of the formamidine occurred<sup>84</sup>, which partially converted the formamidine to the *N*<sup>2</sup>-formyl compound. We expected that the formamidine could then be fully converted to the *N*<sup>2</sup>-formyl group after removal of the silyl protecting groups in the next step. Instead, during purification of **4** by silica gel chromatography, the crude solution was adsorbed onto the silica and eluted after 1 hour to allow for the full conversion of the formamidine to the *N*<sup>2</sup>-formyl group. No Et<sub>3</sub>N was used in the eluent. Following elution from the column, a yield of 90 % was obtained. Converting the formamidine to the formyl during this step meant that the focus of the following step would only be on the removal of the silyl protecting groups. Treatment of **4** with TBAF in THF removed the TBDMS groups to produce **5** which was purified by silica column chromatography and obtained in a 72 % yield. **5** was then reacted with DMTCl and, after elution from silica column chromatography, produced **6** with a 78 % yield. The phosphoramidite,

**7**, was synthesised upon reaction with 2-cyanoethyl-*N*-diisopropylamine chlorophosphoramidite and DIPEA. The crude product was purified by silica column chromatography and **7** was obtained with a 72 % yield.

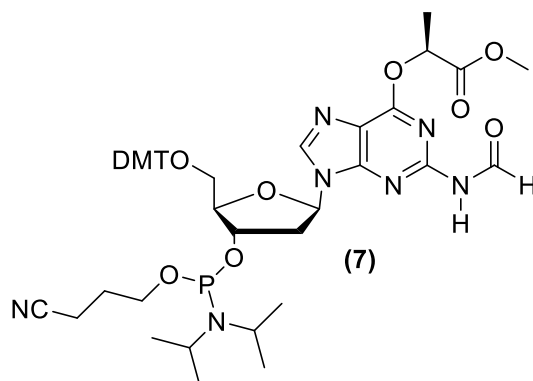


Figure 25. Phosphoramidite of  $O^6$ -CEG synthesised starting from 2'-deoxyguanosine.

ODNs containing the phosphoramidite of  $O^6$ -CEG and  $O^6$ -CMG were synthesised using standard techniques: phenoxyacetyl protection of G and A and acetyl protection of C (Table 3).

Table 3. ODN codes and ODN sequences synthesised containing  $O^6$ -CMG phosphoramidite (**AJ03**) and  $O^6$ -CEG phosphoramidite (**AJ04**) to study MGMT repair of DNA adducts  $O^6$ -CMG and  $O^6$ -CEG.

<i>ODN Code</i>	<i>ODN Sequence</i>
<b>ODN1</b>	5' GAA CTX CAG CTC CGT GCT GGC CC 3' X = $O^6$ -CMG phosphoramidite
<b>ODN2</b>	5' GAA CTX CAG CTC CGT GCT GGC CC 3' X = $O^6$ -CEG phosphoramidite

### 2.3.ODN deprotection and purification

Table 4. ODN codes and ODN sequences containing  $O^6$ -CMG,  $O^6$ -CEG,  $O^6$ -CMG carboxamide and  $O^6$ -CEG carboxamide.

<i>ODN Code</i>	<i>Sequence</i>
<b>AJ03</b>	5' GAA CTX CAG CTC CGT GCT GGC CC 3'
<b>AJ04</b>	5' GAA CTX CAG CTC CGT GCT GGC CC 3'
<b>AJ05</b>	5' GAA CTX CAG CTC CGT GCT GGC CC 3'
<b>AJ06</b>	5' GAA CTX CAG CTC CGT GCT GGC CC 3'

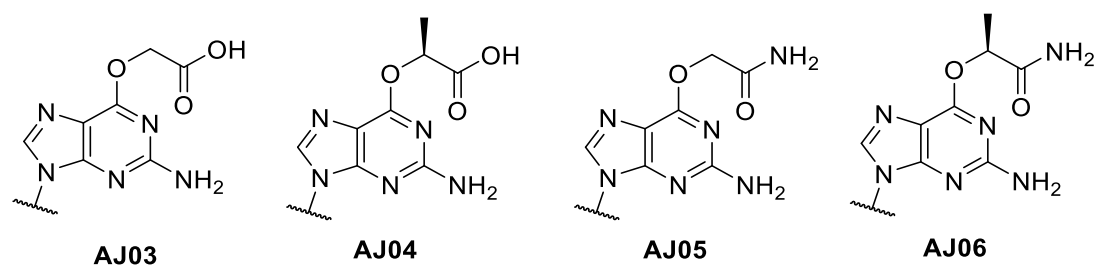


Figure 26. Modifications denoted by **X** in the ODN sequences **AJ03**, **AJ04**, **AJ05**, **AJ06** in Table 4.

In order to cleave the ODN from the CPG support, remove the nucleobase protecting groups and hydrolyse the ester protecting group of the alkyl group at the  $O^6$  position, a literature procedure developed by Millington was followed<sup>84</sup>. Thus, the solid support was treated with aq 0.5 M NaOH for two days followed by the addition of conc. aq ammonia and left for a further three days. Both Millington and Xu<sup>88</sup> had noted that treatment of solid supports containing an ester modification solely by conc. aq ammonia produced the carboxamide on the modified guanine base whilst treatment solely by aq 0.5 M NaOH did not fully remove the nucleobase protecting groups. This meant it was necessary to treat the CPG with aq 0.5 M NaOH followed by conc. aq ammonia.

After the removal of the supernatant from the CPG and evaporation of the ammonia solution, the ODN was desalted using a NAP<sup>TM</sup>-10 column (gel filtration). Purification of the ODNs by RP-HPLC (buffer A- 0.1 M TEAB, buffer B- 1:1 0.1 M TEAB: MeCN; 5- 40 % B over 30 mins) allowed the collection of one, clean peak.

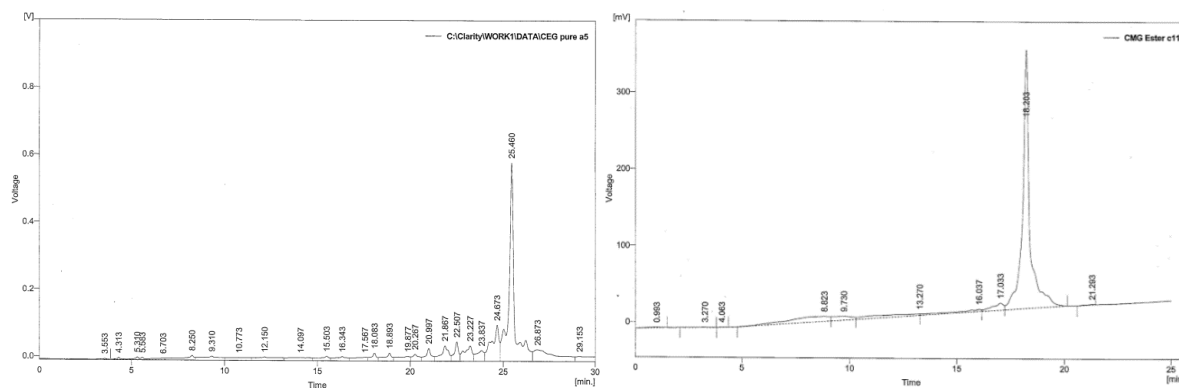


Figure 27. RP-HPLC traces of **AJ03** (left, 25.46 mins) and **AJ04** (right, 18.20 mins). Buffer A- 0.1 M TEAB, buffer B- 1:1 0.1 M TEAB: MeCN; 5- 40 % B over 30 mins.

Analysis by ESI- mass spectroscopy revealed that the mass obtained for the ODN containing *O*<sup>6</sup>-CMG was 7059 (**AJ03**) and the mass obtained for the ODN containing *O*<sup>6</sup>-CEG was 7073 (**AJ04**). Both are the expected masses and meant that the ODNs were suitable for use in the MGMT assays.

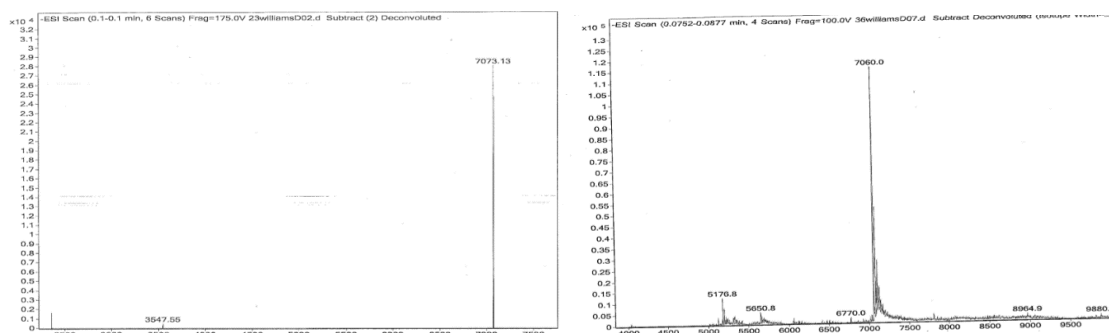


Figure 28. Mass spectroscopy analysis of **AJ03** (left, 7073.13) and **AJ04** (right, 7060.0) indicating expected masses for each ODN.

The ODN deprotection method was verified by treating a CPG of **ODN2** with only conc. aq ammonia for three days to produce the ODN containing  $O^6$ -CEG carboxamide (**AJ06**) and by treating a CPG of **ODN1** with conc. aq ammonia for three days to produce the ODN containing  $O^6$ -CMG carboxamide (**AJ05**). After the respective CPGs had been treated with ammonia for three days at room temperature, **AJ06** was purified by RP-HPLC (buffer A- 0.1 M TEAB, buffer B- 1:1 0.1 M TEAB: MeCN; 5- 40 % B over 30 mins) which allowed the collection of one, clean peak. **AJ05** was purified by RP-HPLC under the same conditions.

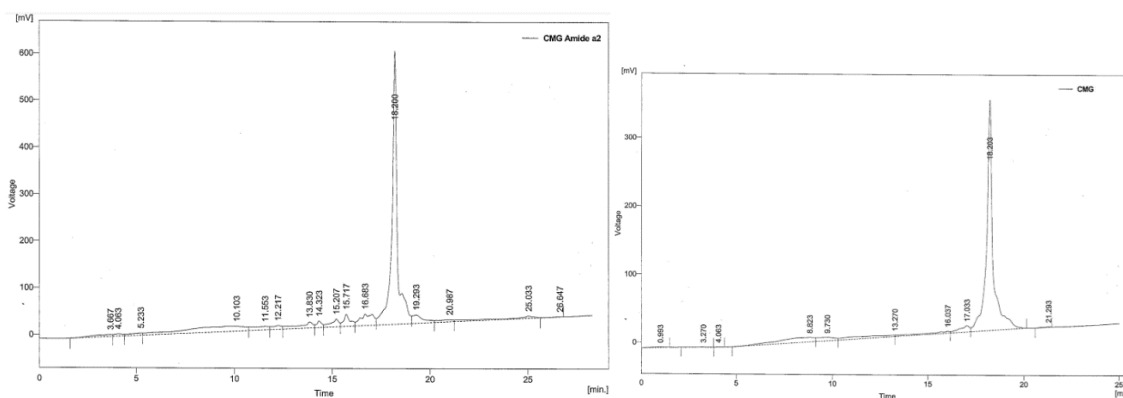


Figure 29. RP-HPLC traces of **AJ05** (left, 18.20 mins) and **AJ06** (right, 18.20 mins). Buffer A- 0.1 M TEAB, buffer B- 1:1 0.1 M TEAB: MeCN; 5- 40 % B over 30 mins.

Analysis by mass spectroscopy gave a mass of 7058 for **AJ05** and a mass of 7072 for **AJ06**. These are the expected masses, both of which are 1 mass unit less than the corresponding carboxylate-containing ODNs, **AJ03** and **AJ04**. This verifies the ODN deprotection procedure required to synthesise ODNs containing carboxylate modifications.

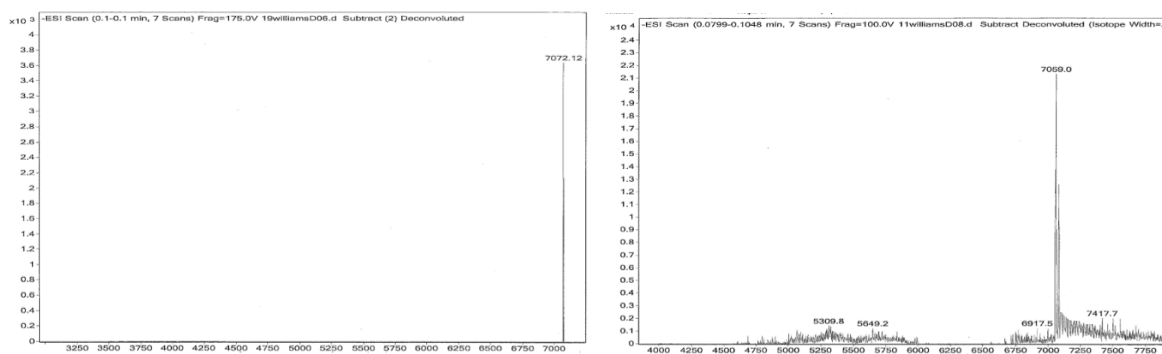


Figure 30. Mass spectroscopy analysis of **AJ05** (left, 7072.12) and **AJ06** (right, 7059.0) indicating expected masses for each ODN.

As further characterisation, a sample containing equal concentrations of ODNs **AJ04** and **AJ06** was analysed by RP-HPLC (Figure 31). This showed two distinct peaks of equal area. The concentration of **AJ06** was doubled and the sample analysed again. As seen in Figure 31, the peak at 13.6 minutes increased and the calculated area was twice that of the peak at 13.1 minutes.

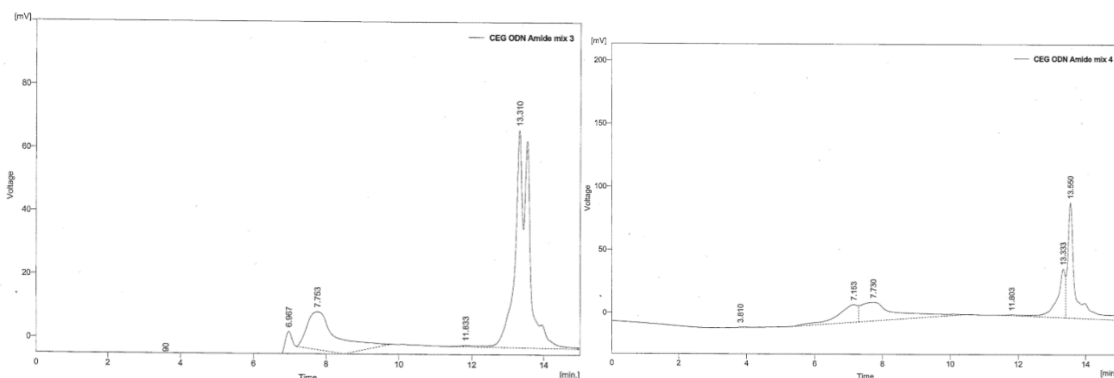


Figure 31. RP-HPLC traces of a sample containing **AJ04** (13.31 and 13.33 mins) and **AJ06** (13.55 and 13.56 mins). On the left,  $[\text{AJ04}] = [\text{AJ06}]$ ; on the right,  $[\text{AJ04}] = \frac{1}{2} [\text{AJ06}]$ . Buffer A- 0.1 M TEAB, buffer B- 1:1 0.1 M TEAB: MeCN; 5- 40 % B over 20 mins.

This result indicates that the retention time of the carboxamide is longer than the retention time of the carboxylate. This is to be expected because carboxamides are less polar than carboxylates therefore should be retained longer on RP-HPLC. This further suggests that the method for deprotecting ODNs containing carboxylic acids is reliable and furthermore that they can be resolved by RP-HPLC.

#### **2.4. Nucleoside composition analysis of ODNs containing *O*<sup>6</sup>-CMG, *O*<sup>6</sup>-CEG, *O*<sup>6</sup>-CMG carboxamide and *O*<sup>6</sup>-CEG carboxamide**

It was deemed necessary to provide further evidence to distinguish the carboxamide from the carboxylate-containing ODNs and this could be provided by LC-MS analysis following enzymatic digestion of each ODN. Digesting each ODN down to its individual nucleosides would mean the mass of the single modified 2'-deoxyguanosine base in each ODN would be obtained and, due to the mass difference of one between *O*<sup>6</sup>-CEG carboxylate and *O*<sup>6</sup>-CEG carboxamide, there would be clear evidence that the ODNs contained different modifications.

Digestion of ODNs to their nucleoside components requires a phosphodiesterase enzyme and the alkaline phosphatase. The phosphodiesterase enzyme cleaves the phosphodiester linkages of the ODN, which forms the individual nucleotides, and then the monophosphate groups are removed by alkaline phosphatase, resulting in the individual nucleosides. Upon analysis of the digestion mixture by LC-MS, the individual masses of the four nucleosides can be obtained alongside the masses of any modified nucleosides.

DNA Degradase Plus™ is a nuclease mixture that digests ODNs down to their individual nucleosides in one step. This is advantageous over other methods that require a two-step procedure and is cheaper than the separate acquisition of a phosphodiesterase enzyme and the calf-intestinal alkaline phosphatase. Following the protocol provided by Zymo Research<sup>91</sup>, ODN **AJ06** was treated with DNA Degradase Plus™ for 2 hours at 37 °C and analysed by LC-MS (Figure 32).

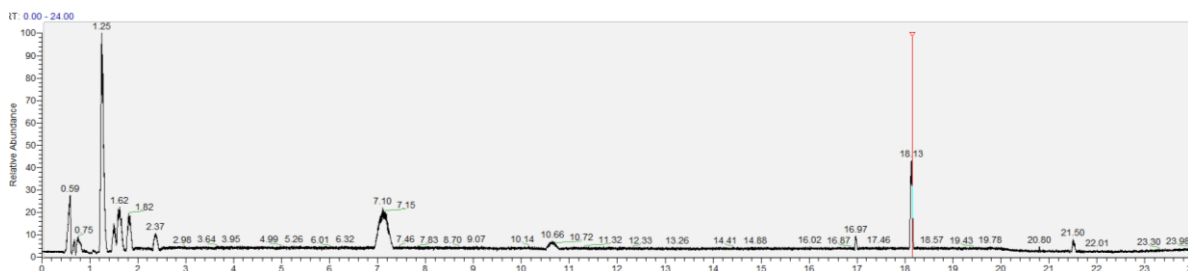


Figure 32. LC-MS analysis of digested **ODN AJ06** showing complete digestion. dC (1.25 mins), dG (7.10 mins), T (10.66 mins), *O*<sup>6</sup>-CEG carboxamide (18.13 mins).

The LC-MS trace shows peaks corresponding to the masses of dC, dG, T and *O*<sup>6</sup>-CEG carboxamide at 1.25, 7.10, 10.66 and 18.13 minutes respectively. However, there is no peak accounting for dA. Decomposition of dA may have occurred whilst the digested sample was waiting to be analysed because the DNA Degradase Plus™ had not been fully inactivated by heating for 20 minutes at 70 °C. The presence of adenosine deaminase could have potentially hydrolysed 2'-deoxyadenosine to 2'-deoxyinosine.

Besides ensuring full inactivation of DNA Degradase Plus™, the solution to the issue required the use of Nanosep® centrifugal device that separates the digested nucleosides from the nuclease mix by membrane ultrafiltration. The digest solution is injected into the Nanosep® centrifugal device and is centrifuged for 5 minutes at 1500 x g. During this time, the smaller nucleosides pass through the semipermeable membrane whilst the larger proteins are retained enabling separation of the two.



#### 2.4.1. Nucleoside composition analysis of $O^6$ -CEG

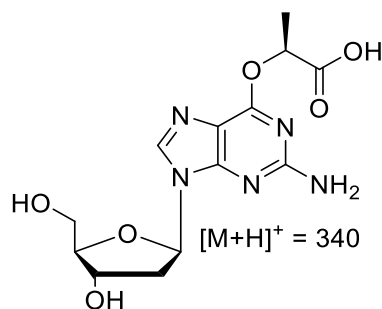


Figure 33. Structure and [M+H]<sup>+</sup> of  $O^6$ -CEG.

**AJ04** was digested by DNA Degradase Plus™ for 2 hours at 37 °C. The sample was then heated at 70 °C for 20 minutes to inactivate the DNA Degradase Plus™ and then injected into a Nanosep® centrifugal device. This was centrifuged at 1500 x g for 5 minutes to separate the nuclease mix from the digested ODN following which, the digested ODN was analysed by LC-MS (Figure 34).

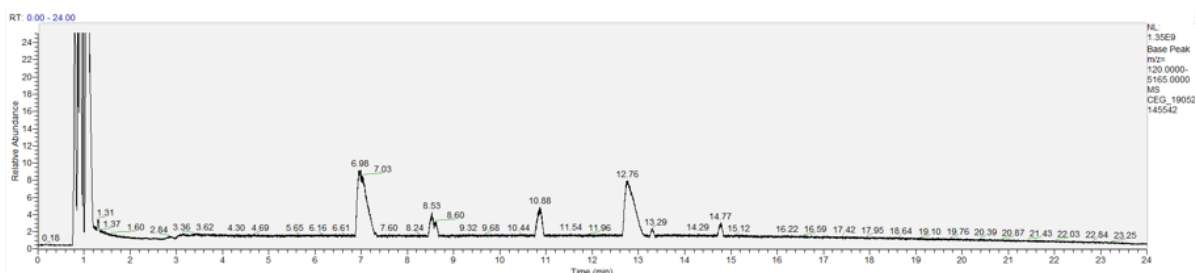


Figure 34. LC-MS analysis of digested **ODN AJ04** showing complete digestion. dC (6.98 mins), T (8.53 mins), dA (10.88 mins), dG (12.76 mins) and  $O^6$ -CEG (13.29 mins). Buffer A- 5 mM ammonium acetate, buffer B- 5 mM ammonium acetate/ 80 % MeCN; 20- 90 % B over 20 mins.

Nucleoside composition analysis of **AJ04** indicated complete digestion of the ODN and that a [M+H]<sup>+</sup> = 340.1 was observed at 13.29 minutes (Figure 35). This mass corresponds to  $O^6$ -CEG indicating that the ODN contained the correct modification. This verifies the method of hydrolysis of the methyl ester group to the carboxylate, indicating that treatment of the CPG initially with aq 0.5 M NaOH is required.

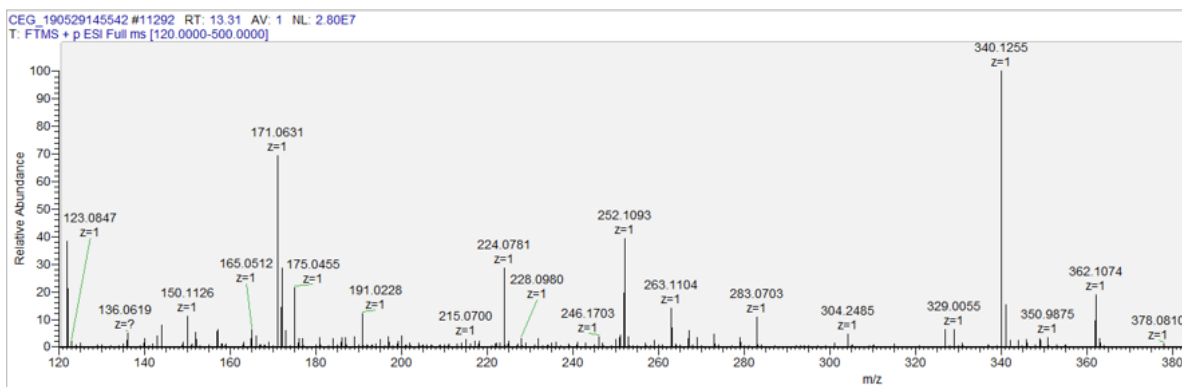


Figure 35. MS analysis of peak at 13.29 minutes of Figure 34 indicating  $m/z = 340.1$  which corresponds to  $O^6$ -CEG.

Five other peaks are observed. Peaks at 6.98 mins, 8.53 mins, 10.88 mins and 12.76 mins correspond to dC, T, dA and dG respectively. The peak at 14.77 minutes relates to nuclease mix that was small enough to pass through the Nanosep® centrifugal device and is observed in all following nucleoside composition analyses (sections 2.4.2, 2.4.3 and 2.4.4).

The ODN composition was calculated after analysing the sample by UHPLC with UV detection (Figure 36).

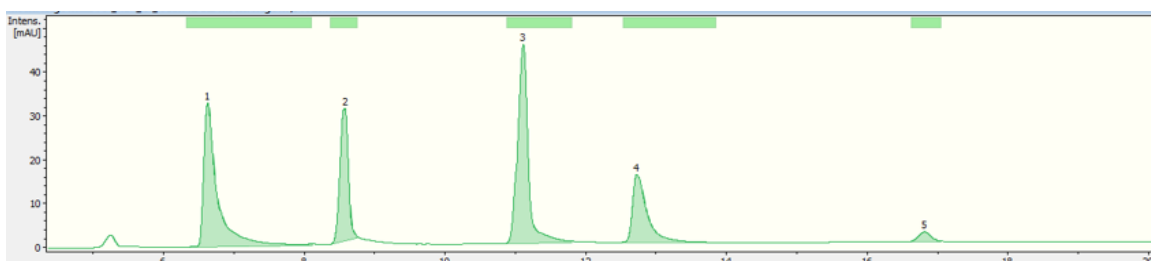


Figure 36. UHPLC UV analysis of **AJ04** indicating dC (6.6 mins), T (8.6 mins), dG (11.1 mins), dA (12.7 mins) and  $O^6$ -CEG (16.8 mins). Buffer A- 5 mM ammonium acetate, buffer B- 5 mM ammonium acetate/ 80 % MeCN; 20- 90 % B over 20 mins.

Using  $\epsilon_{260}$  values calculated by Cavaluzzi and Borer<sup>92</sup> and a literature value for  $O^6$ -MeG<sup>34</sup> (there is no recorded value for  $O^6$ -CEG or  $O^6$ -CMG), the compositions were calculated (Table 5).

Table 5. Calculation of nucleoside composition of **AJ04** by UHPLC.

<i>Nucleoside</i>	<i>Area</i>	<i>Area/ <math>\epsilon_{260}</math></i>	<i>Composition</i>	<i>Actual Composition</i>
<i>dC</i>	445.9	62.8	9.3	9.0
<i>T</i>	252.8	29.5	4.3	4.0
<i>dG</i>	484.4	39.7	5.9	6.0
<i>dA</i>	224.8	15.0	2.4	3.0
<i>O<sup>6</sup>-CEG</i>	27.0	8.0	1.2	1.0

#### 2.4.2. Nucleoside composition analysis of *O<sup>6</sup>*-CEG carboxamide

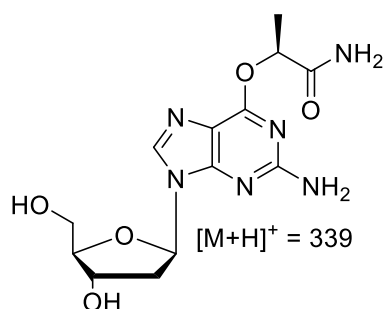


Figure 37. Structure and  $[M+H]^+$  of *O<sup>6</sup>*-CEG carboxamide.

**AJ06** was digested and the nucleoside composition analysed following the procedure mentioned in Section 2.4.1. Analysis indicated that the ODN contained the intended modification.

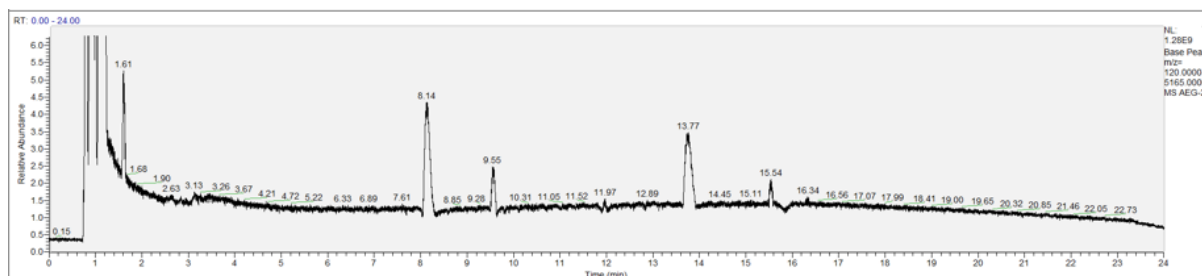


Figure 38. LC-MS analysis of digested **ODN AJ06** showing complete digestion. dC (8.14 mins), T (9.55 mins), dA (11.97 mins), dG (13.77 mins) and *O<sup>6</sup>*-CEG carboxamide (15.85 mins). Buffer A- 5 mM ammonium acetate, buffer B- 5 mM ammonium acetate/ 80 % MeCN; 20- 90 % B over 20 mins.

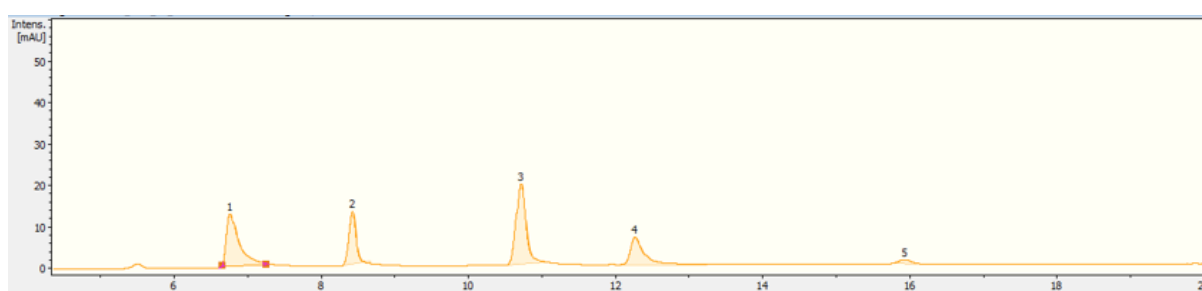
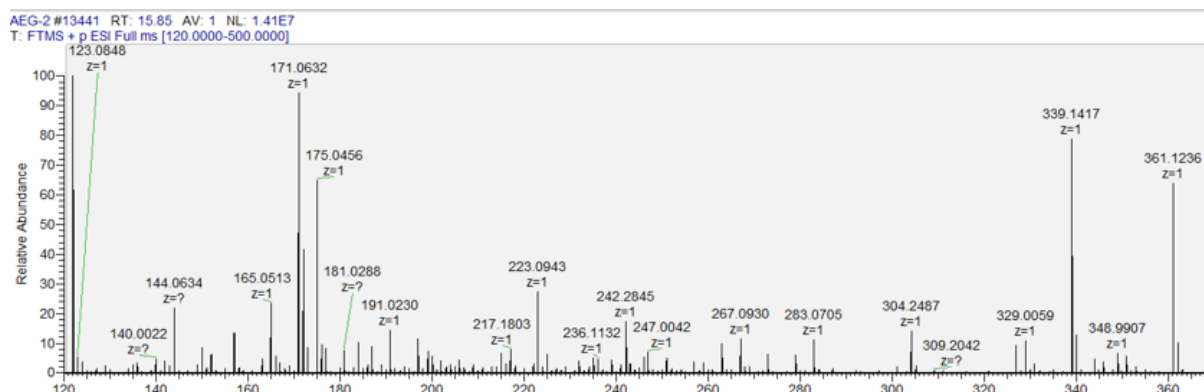


Table 6. Calculation of nucleoside composition of **AJ06** by UHPLC. Actual area calculated by (Calculated area/  $\epsilon_{260}$ ).

Nucleoside	Area	Area/ $\epsilon_{260}$	Composition	Actual composition
dC	255.2	35.9	8.8	9.0
T	150.3	17.6	4.3	4.0
dG	295.4	24.3	5.9	6.0
dA	138.9	9.3	2.5	3.0
$O^6$ -CEG carboxamide	23.8	7.0	1.5	1.0

### 2.4.3. Nucleoside composition analysis of $O^6$ -CMG

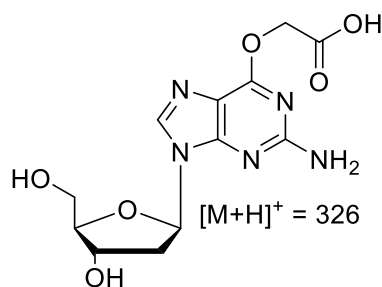


Figure 41. Structure and [M+H]<sup>+</sup> of  $O^6$ -CMG.

**AJ03** was digested and the nucleoside composition analysed following the procedure mentioned in Section 2.4.1. Analysis indicated that the ODN contained the intended modification.

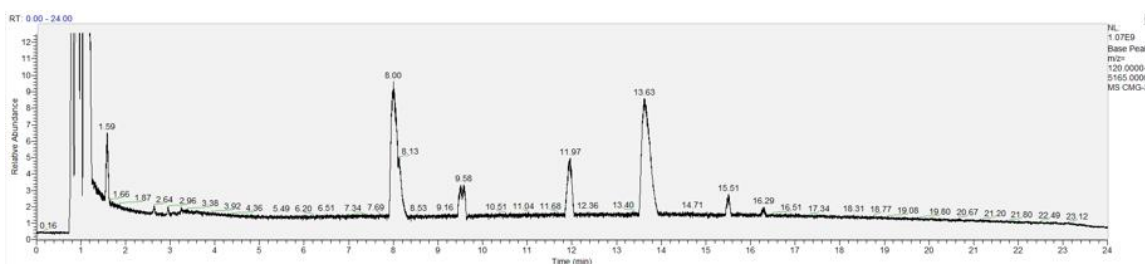


Figure 42. LC-MS analysis of digested **ODN AJ03** showing complete digestion. dC (8.00 mins), T (9.58 mins), dA (11.97 mins), dG (13.63 mins) and  $O^6$ -CMG (16.29 mins). Buffer A- 5 mM ammonium acetate, buffer B- 5 mM ammonium acetate/ 80 % MeCN; 20- 90 % B over 20 mins.

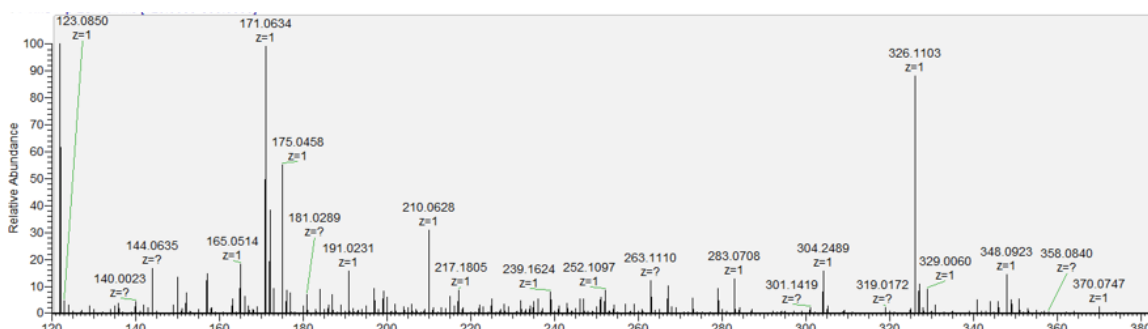


Figure 43. MS analysis of peak at 16.29 minutes of Figure 42 indicating  $m/z = 326.1$  which corresponds to  $O^6$ -CMG.

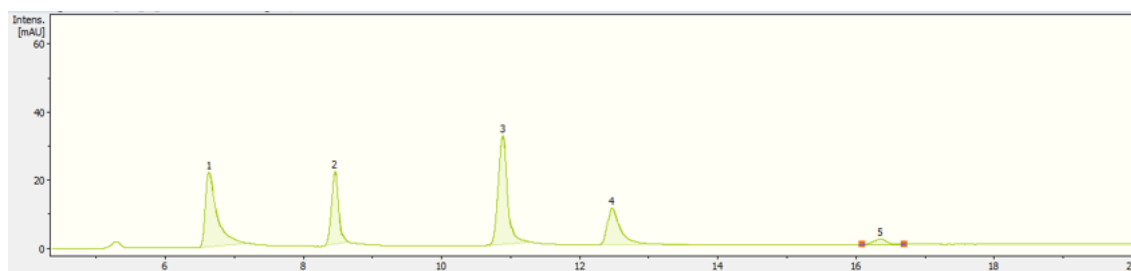


Figure 44. UHPLC UV analysis of **AJ03** indicating dC (6.6 mins), T (8.5 mins), dG (10.9 mins), dA (12.5 mins) and *O*<sup>6</sup>-CMG (16.4 mins). Buffer A- 5 mM ammonium acetate, buffer B- 5 mM ammonium acetate/ 80 % MeCN; 20- 90 % B over 20 mins.

Table 7. Calculation of nucleoside composition of **AJ03** by UHPLC. Actual area calculated by (Calculated area/  $\epsilon_{260}$ ).

<i>Nucleoside</i>	<i>Area</i>	<i>Area/ <math>\epsilon_{260}</math></i>	<i>Composition</i>	<i>Actual composition</i>
<i>dC</i>	117.5	16.5	8.6	9.0
<i>T</i>	75.0	8.8	4.5	4.0
<i>dG</i>	136.6	11.2	5.8	6.0
<i>dA</i>	71.4	4.7	2.5	3.0
<i>O</i> <sup>6</sup> - <i>CMG</i>	10.9	3.2	1.6	1.0

#### 2.4.4. Nucleoside composition analysis of *O*<sup>6</sup>-CMG carboxamide

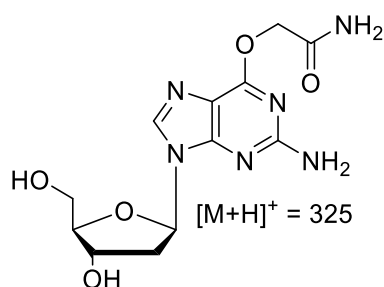


Figure 45. Structure and  $[M+H]^+$  of *O*<sup>6</sup>-CMG carboxamide.

**AJ05** was digested and the nucleoside composition analysed following the procedure mentioned in Section 2.4.1. Analysis indicated that the ODN contained the intended modification.

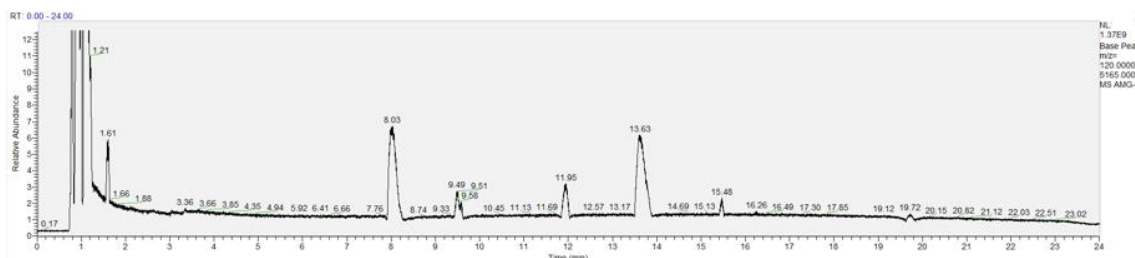


Figure 46. LC-MS analysis of digested **ODN AJ05** showing complete digestion. dC (8.03 mins), T (9.49 mins), dA (11.95 mins), dG (13.63 mins) and *O*<sup>6</sup>-CMG carboxamide (19.72 mins). Buffer A- 5 mM ammonium acetate, buffer B- 5 mM ammonium acetate/ 80 % MeCN; 20- 90 % B over 20 mins.

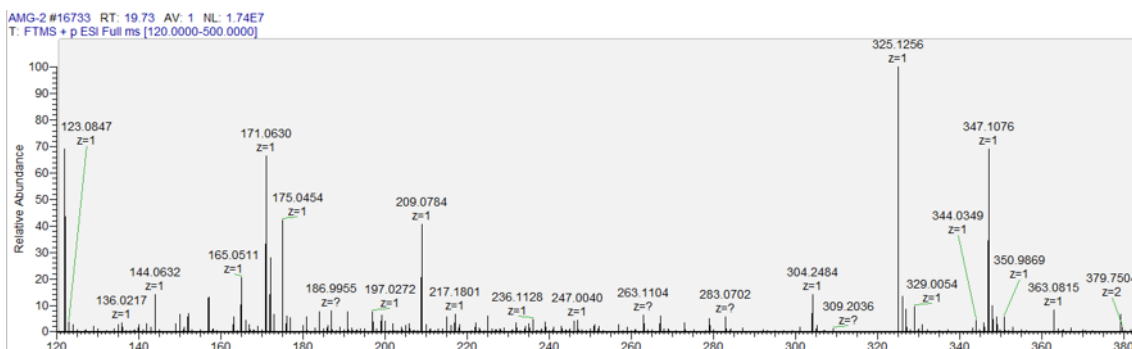


Figure 47. MS analysis of peak at 19.73 minutes of Figure 46 indicating  $m/z = 325.1$  which corresponds to *O*<sup>6</sup>-CMG carboxamide.

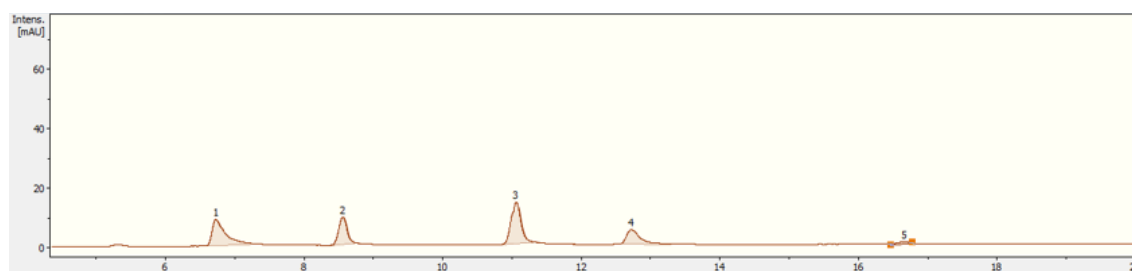


Figure 48 UHPLC UV analysis of **AJ05** indicating dC (6.7 mins), T (8.6 mins), dG (11.1 mins), dA (12.7 mins) and *O*<sup>6</sup>-CMG (16.7 mins). Buffer A- 5 mM ammonium acetate, buffer B- 5 mM ammonium acetate/ 80 % MeCN; 20- 90 % B over 20 mins.

Table 8. Calculation of nucleoside composition of **AJ05** by UHPLC. Actual area calculated by (Calculated area/  $\epsilon_{260}$ ).

<i>Nucleoside</i>	<i>Area</i>	<i>Area/ <math>\epsilon_{260}</math></i>	<i>Composition</i>	<i>Actual composition</i>
<i>dC</i>	155.8	21.9	8.9	9.0
<i>T</i>	82.1	9.6	3.9	4.0
<i>dG</i>	182.4	15	6.1	6.0
<i>dA</i>	92.3	6.2	2.5	3.0
<i>O<sup>6</sup>-CMG carboxamide</i>	11.4	3.3	1.3	1.0

#### 2.4.5. Analysis of nucleoside composition spectra

The analyses of the nucleoside compositions indicated that UHPLC UV detection can be used to determine the composition of an ODN sample and the m/z of individual nucleosides identified by LC-MS. The calculated ODN compositions are coherent with the actual composition of each ODN even though the  $\epsilon_{260}$  used for the *O<sup>6</sup>-alkylG* modifications is unknown. This explains why it has a greater calculated value across the four spectra. Throughout the analysis, the calculated amount of dA was lower than expected. This is due to its decomposition in the nuclease mixture used to digest the ODN samples. LC-MS is unable to quantify the composition alone because the ionisation of each nucleoside is independent of its amount. That is why a separate UV analysis was required.

LC-MS analysis confirmed the difference in modifications between the carboxylate and carboxamide samples. Both the analyses of **AJ03** and **AJ05** and between **AJ04** and **AJ06** indicated a m/z difference of 1 between the *O<sup>6</sup>-alkylG* modifications. This is the expected m/z difference between the carboxylate and carboxamide. Therefore,



ODNs containing carboxylate modifications require treatment with aq 0.5 M NaOH initially, followed by treatment with aq conc. ammonia.

## 2.5. MGMT assays

As discussed in section 1.11, repair of  $O^6$ -alkylG modifications by MGMT can be quantified and an  $IC_{50}$  (the amount of substrate required to inactivate 50 % of protein) generated through an inactivation assay using  $O^6$ - $[^3H]$ -MeG. The ODN samples, at a known concentration, were individually incubated with MGMT for 90 minutes after which DNA containing  $O^6$ - $[^3H]$ -MeG was introduced, and the samples incubated for a further 120 minutes. The MGMT was isolated and the amount of radioactivity transferred to the MGMT from  $O^6$ - $[^3H]$ -MeG was recorded. This was repeated across a range of concentrations until the  $IC_{50}$  for each ODN sample could be calculated. The results in Table 9 were obtained by our collaborators, the Margison group at the University of Manchester.

Table 9. Reported  $IC_{50}$ s for ODNs **AJ03-06** obtained by the Margison group at the University of Manchester.

<i>ODN</i>	<i>IC<sub>50</sub>/nM</i>
<b>AJ03- <math>O^6</math>-CMG</b>	133
<b>AJ04- <math>O^6</math>-CEG</b>	72.1
<b>AJ05- <math>O^6</math>-CMG carboxamide</b>	20.8
<b>AJ06- <math>O^6</math>-CEG carboxamide</b>	56.4
<b><math>O^6</math>-MeG</b>	1.15

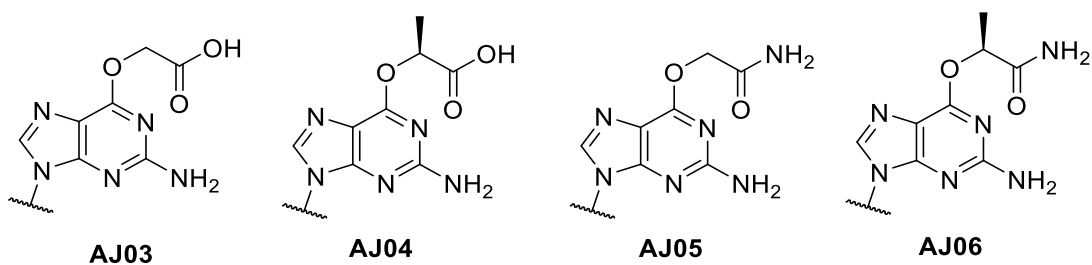


Figure 49. Modifications denoted by **X** in the ODN sequences **AJ03**, **AJ04**, **AJ05**, **AJ06** in Table 9.

The results obtained for the four ODNs provide clear answers to our initial aims of determining: i) whether the  $IC_{50}$  result obtained by Senthong *et al.*<sup>43</sup> was for  $O^6$ -CMG or  $O^6$ -CMG carboxamide and ii) whether both  $O^6$ -CEG and  $O^6$ -CMG are substrates of MGMT. The steps taken during the synthesis and purification and subsequent characterisation of each ODN ensured that we were absolutely certain that each ODN contained the correct modification.

The  $IC_{50}$ s calculated for the four ODNs indicate a clear difference in MGMT affinity to the carboxamide and the carboxylate, with MGMT having a greater affinity towards carboxamides. The  $IC_{50}$  for **AJ03** (133 nM) was over six-fold more than the  $IC_{50}$  for **AJ05** (20.8 nM) whilst the  $IC_{50}$  for **AJ04** (72.1 nM) was greater than for **AJ06** (56.4 nM). Pletsas *et al.* reported that the carboxylate and carboxamide inhibit Atase, an  $O^6$ -alkyltransferase, at proportionally the same rate<sup>87</sup>. However, these assays were performed using simple  $N^9$ -alkylated purines instead of ODNs which are known to have significantly lower  $IC_{50}$ s<sup>34</sup>. Moreover, crystal structures of MGMT interactions with  $O^6$ -MeG and  $O^6$ -benzyl G recorded by Daniels *et al.*<sup>93,94</sup> indicate that MGMT affinity is greater for hydrophobic  $O^6$ -alkyl groups over hydrophilic  $O^6$ -alkyl groups. This explains why  $O^6$ -benzyl G has an affinity 2000-fold greater than  $O^6$ -MeG<sup>93,94</sup>. This is due to the Tyr158 side chain of MGMT which packs close to the  $O^6$ -alkyl group during the repair process. Tyr158 provides a hydrophobic environment therefore a more hydrophobic  $O^6$ -alkyl group would pack closer to the

protein and enable greater speed of alkyl transfer from guanine to Cysteine.

However, there is no crystal structures available to determine whether this is a factor in the results obtained by the Margison group.

There is clearly a greater affinity towards carboxamide adducts than carboxylate adducts. This may be due to the carboxylate being negatively charged which would repel the active site protein. The  $O^6$ -alkyl group is removed by the negatively charged thiolate group of cysteine therefore a repulsion between the thiolate group and the carboxylate group can be expected.

The  $IC_{50}$  data acquired for **AJ03** and **AJ05** strongly suggests that the data collected by Senthong *et al.* in their paper and represented as an ODN containing  $O^6$ -CMG was actually data collected for an ODN containing  $O^6$ -CMG carboxamide. Senthong *et al.* reported that the  $IC_{50}$  for the radioactive MGMT assay of their ODN was 1.7 nM whilst we reported an  $IC_{50}$  of 133 nM for  $O^6$ -CMG, 78 times greater.

During the deprotection of the ODN prepared by Senthong *et al.*<sup>43</sup>, the ODN presumably provided the carboxamide, highlighting the need to undertake the aq NaOH deprotection step for at least two days first before adding aq conc. ammonia.

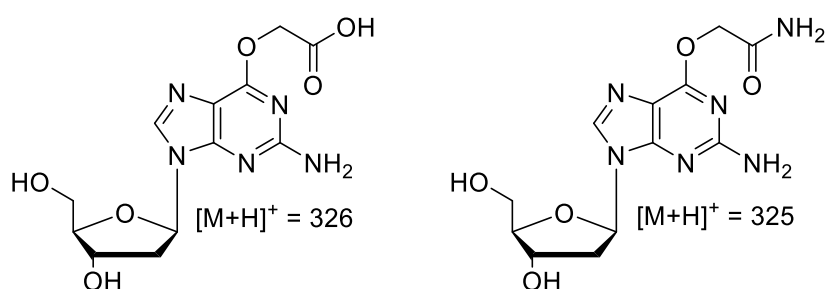


Figure 50. Structures and  $[M+H]^+$  of  $O^6$ -CMG, left, and  $O^6$ -CMG carboxamide.

Importantly, these analyses provide structural data of the  $O^6$ -alkyl groups, indicating the masses of the  $O^6$ -alkyl groups. Senthong *et al.* performed MS analysis

of the alkyl group transfer to MGMT and recorded a transfer of 57. They associated this to a transfer of the alkyl group  $-\text{CH}_2\text{CO}_2^-$  to MGMT. However, our nucleoside component analyses indicate that, during MGMT repair of  $O^6$ -CMG, it would be expected that the alkyl group  $-\text{CH}_2\text{CO}_2\text{H}$  would be transferred resulting in a mass transfer of 58. The repair of  $O^6$ -CMG carboxamide would result in the transfer of the alkyl group  $-\text{CH}_2\text{C}(\text{O})\text{NH}_2$ , resulting in a mass transfer of 57. The expected mass transfer for  $O^6$ -CMG carboxamide is the same mass observed by Senthong *et al.*

However, the  $\text{IC}_{50}$  data produced by the Margison group does confirm that both  $O^6$ -CMG and  $O^6$ -CEG are substrates of MGMT. This means that MGMT may protect against human CRC by repairing such  $O^6$ -alkylG adducts although clearly the repair of these types of adducts is less efficient than others such as  $O^6$ -MeG.

**Chapter 3 – Use of  $^{15}\text{N}_5$ -Labelled  $\text{O}^6$ -Alkyl-2'-Deoxyguanosine For  
Quantification of Alkylation Damage Using MS**

### 3. Use of $^{15}\text{N}_5$ -labelled $O^6$ -Alkyl-2'-Deoxyguanosine For Quantification of Alkylation Damage Using MS

#### 3.1. Synthesis of $O^6$ -alkylguanine adducts

##### 3.1.1. Introduction

Developing a SID method using  $^{15}\text{N}_5$ -labelled internal standards of  $O^6$ -alkyl-2'-deoxyguanosine adducts would be a substantial step to identifying and quantifying them in human DNA and allowing for the development of techniques to study the aetiology of cancer.  $O^6$ -alkylguanine adducts occur in 4 in  $10^8$  nucleobases (which equates to 60 fmol of  $O^6$ -alkylguanine adducts in 100  $\mu\text{g}$  double stranded DNA) therefore an accurate and sensitive quantification method is paramount to detect such levels. This should be achievable by using  $^{15}\text{N}_5$ -labelled internal standards.

A number of  $O^6$ -alkylguanine SIL internal standards have previously been synthesised<sup>95, 104–106</sup>. Yang *et al.* synthesised  $^2\text{H}_3$ -labelled  $O^6$ -MeG whilst Moore and Shuker synthesised  $^2\text{H}_2$  and  $^{13}\text{C}$ -labelled  $O^6$ -CMG (Figure 51).

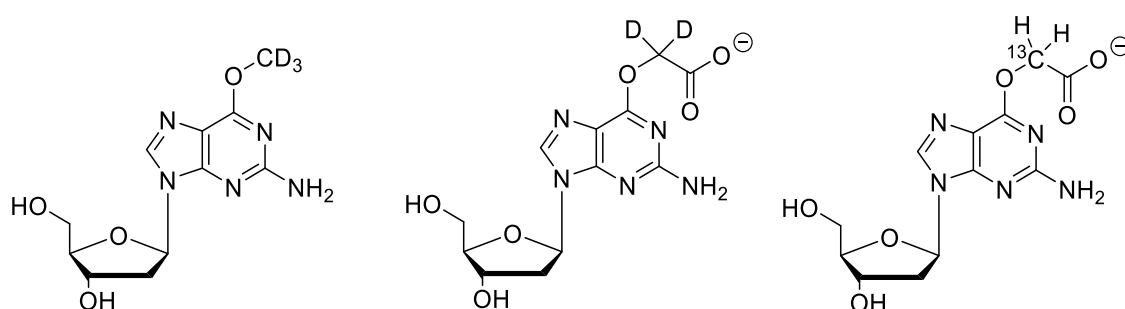


Figure 51. SIL internal standards  $^2\text{H}_3$ -labelled  $O^6$ -MeG,  $^2\text{H}_2$ -labelled  $O^6$ -CMG and  $^{13}\text{C}$ -labelled  $O^6$ -CMG.

However, all three synthetic routes encounter issues that limit their application to a wide range of  $O^6$ -alkylguanine adducts. As discussed in section 1.15, deuterated SIL internal standards are affected by deuterium-proton exchange<sup>57</sup>. Also, this

modification alters the lipophilicity of the internal standard (which can lead to an alteration of its retention time relative to the analyte during LC-MS) and leads to mono- or *bis*-deuterated  $O^6$ -MeG compounds, as noted by Moore and Shuker (Figure 52)<sup>105</sup>. Accurate quantification of the analyte is reduced because multiple masses will be observed whilst reducing the overall synthesis yield.

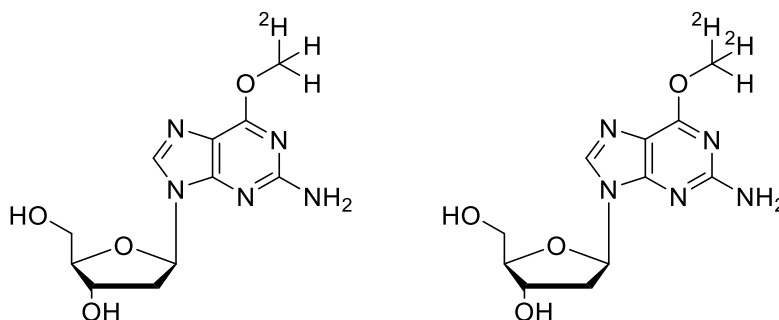


Figure 52. Mono- and *bis*-deuterated  $O^6$ -MeG SIL internal standard formed due to deuterium-proton exchange during synthesis.

Moreover, neither the purine ring nor ribose sugar were stable-isotopically labelled in synthetic routes devised by Yang *et al.* and Moore and Shuker. This means that the reagent required to introduce the modification at the  $O^6$  position needs to be stable-isotopically labelled. This is an infeasible method to build a catalogue of SIL internal standards of  $O^6$ -alkylguanine adducts. Furthermore, overall yields of the synthetic routes were below 10 %.

Churchwell *et al.* synthesised  $^{15}N_5$ -labelled  $O^6$ -MeG and  $O^6$ -EtG by reacting  $^{15}N_5$ -labelled dG directly with diazomethane and diazoethane, respectively, with no yields reported<sup>95</sup>. Again, this is not a feasible method to synthesise a range of  $O^6$ -alkylguanine adducts because alkyl groups of interest are not all readily available in their diazo form.

The 2'-deoxyribosides of four  $O^6$ -alkylguanine adducts were identified as being suitable for the synthesis of their  $^{15}\text{N}_5$ -labelled equivalents.  $O^6$ -MeG,  $O^6$ -EtG and  $O^6$ -CMG are adducts found in human DNA<sup>28, 30, 95, 96</sup>.  $O^6$ -MeG and  $O^6$ -EtG are both known to cause GC→AT transition mutations and form from *in vitro* exposure to *N*-methyl-*N*-nitrosourea and *N*-ethyl-*N*-nitrosourea respectively.  $O^6$ -CEG is an adduct of interest to our group due to its potential link to CRC in connection with  $O^6$ -CMG.

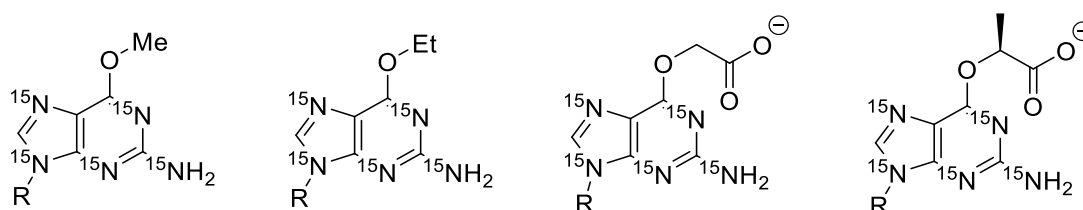


Figure 53.  $^{15}\text{N}_5$ -labelled structures of  $O^6$ -alkylguanine compounds  $O^6$ -MeG,  $O^6$ -EtG,  $O^6$ -CMG and  $O^6$ -CEG, left to right, to be synthesised for use as stable-isotopically labelled internal standards in SID methods. R = 2-deoxyribose sugar.

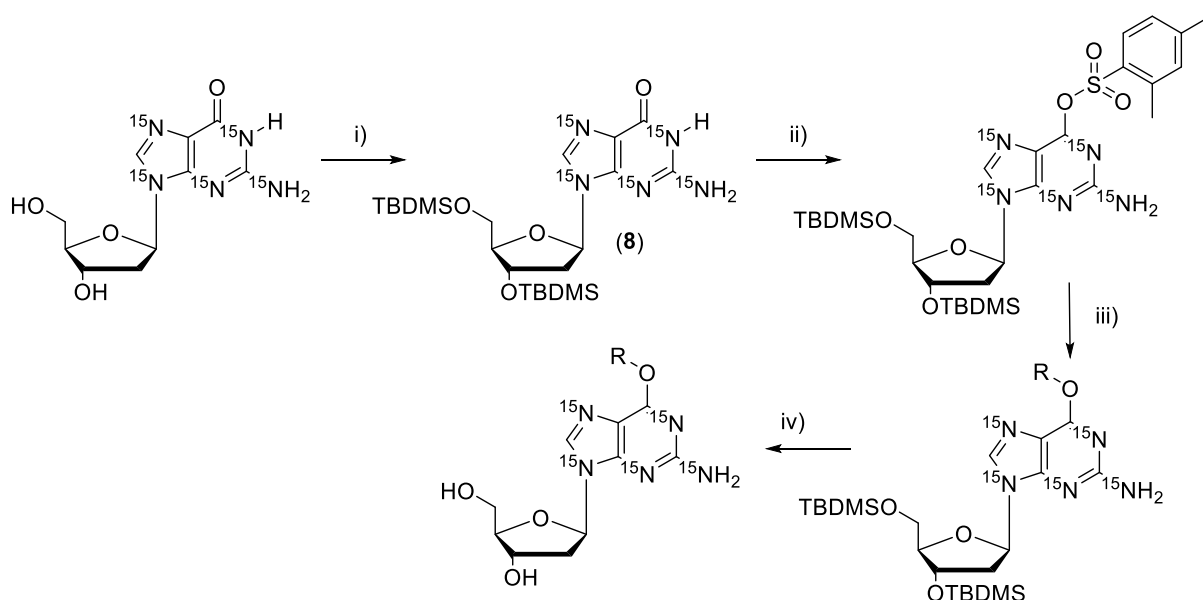
Obtaining the starting material to synthesise these standards is expensive, but, due to the small amounts of material needed to detect  $O^6$ -alkylguanine adducts, a range of adducts can be synthesised in abundance starting from 100 mg of starting material. In order to synthesise the  $^{15}\text{N}_5$ -labelled  $O^6$ -alkylguanine adducts, it was therefore necessary to optimise the yields for their synthesis initially using unlabelled compounds on a suitable scale, for repeating with the expensive  $^{15}\text{N}_5$ -labelled 2'-deoxyguanosine.

### 3.1.2. Synthesis of labelled standards using 3', 5' silyl protecting groups

The initial envisaged route to synthesise the  $O^6$ -alkyl-2'-deoxyguanosines followed the synthesis route of the  $O^6$ -CMG phosphoramidite developed by Millington *et al.* (Scheme 22)<sup>84</sup>. This involved silyl protection of the 3' and 5' hydroxy groups of 2'-deoxyguanosine and subsequent mesitylation of the  $O^6$ -position, using 2-



mesitylenesulfonyl chloride. Following this, we attempted displacement of the mesitylene sulfonate ester sequentially by DABCO then the required alcohol for the respective modification, followed by removal of the silyl groups by TBAF.



Scheme 22. Attempted efficient synthesis of  $O^6$ -alkyl-2'-deoxyguanosines, based on Millington *et al.*<sup>84</sup>, to be used as SIL internal standards starting from 2'-deoxyguanosine (100 mg scale); i) TBDMSO, imidazole, DMF; ii)  $\text{Et}_3\text{N}$ , DMAP, 2-mesitylenesulfonyl chloride, DCM; iii) DABCO, ROH, DBU, 1, 2-DME; iv) TBAF in THF.

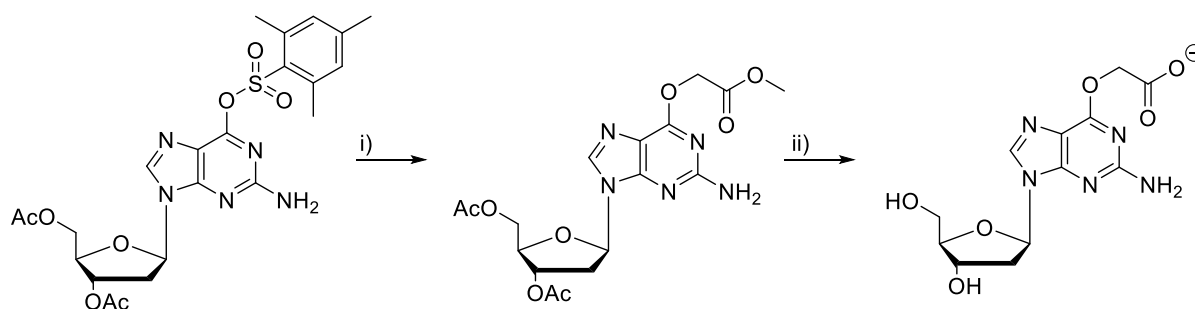
When the initial silylation was repeated, attempted collection of the product by precipitation in methanol followed by sonication and vacuum filtration as described gave only 9 mg (5 %) of the desired product. The problem here was assumed to be due to the scale of the reaction. Previous work used 33 times more starting material and any product lost during the sonication and vacuum filtration was offset by the large scale of the reaction.

The purpose of the purification in the first step was to remove imidazole which was expected to disrupt the next two steps, namely formation of the  $O^6$ -sulfonate ester and its subsequent displacement. The reaction was attempted again on a 100 mg scale and, once analysis by TLC showed completion of the reaction, the solvent was removed. The crude material was subsequently dissolved in DCM and washed with 0.1 M HCl

to remove the imidazole as its hydrochloride salt. TBDMS groups can be removed at  $\text{pH} < 4$  therefore dilute HCl was used. After extraction of the organic layer and purification by silica column chromatography, **8** was obtained in 37 % yield. This yield was considered too low to be viable for synthesising the  $^{15}\text{N}_5$ -labelled compound due to the expense of the starting material. Consequently, attention turned to exploring alternative hydroxyl-protecting groups.

### 3.1.3. Synthesis of labelled standards using 3', 5' acetyl protecting groups

High yields have been reported for the acetylation of the 3' and 5' hydroxy groups of 2'-deoxyguanosine<sup>97</sup> whilst Janeba *et al.* have reported the mesitylation of 3',5'-bis-O-acetyl-2'-deoxyguanosine<sup>98</sup>. Moreover, following this route we expected that during incorporation of the respective alcohols at the  $O^6$  position, the removal of acetyl protecting groups under basic conditions might also be achieved at the same time or after brief treatment with aq NaOH. In the case of the  $O^6$ -CMG compound, after the insertion of methyl glycolate, hydrolysis of the methyl ester could be taken simultaneously (Scheme 23).

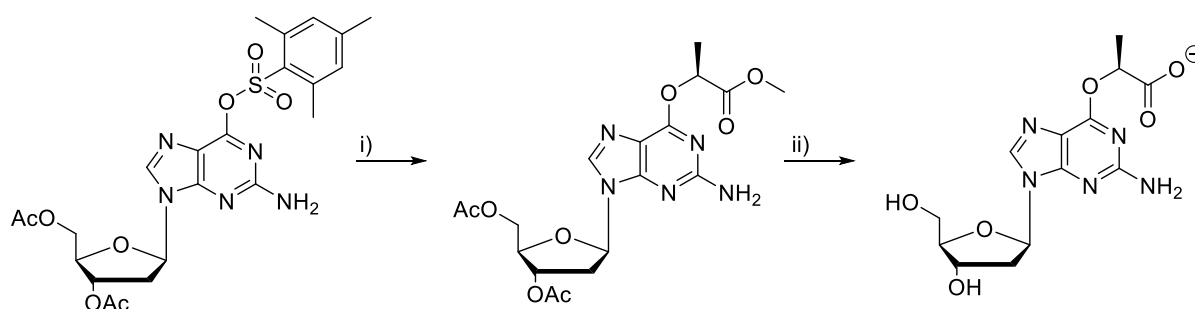


Scheme 23. Displacement of 6-mesitylenesulfonyl-3',5'-bis-O-acetyl-2'-deoxyguanosine by methyl glycolate followed by hydrolysis of  $O^6$ -methylglycolate alkyl group to  $O^6$ -CMG; i) methyl glycolate, DABCO, DBU, 1, 2-DME; ii) 0.5 M aq NaOH.

Likewise, an analogous route in which methyl(-)-lactate displaces mesitylenesulfonyl to form  $O^6$ -methyl(-)-lactate followed by hydrolysis of the

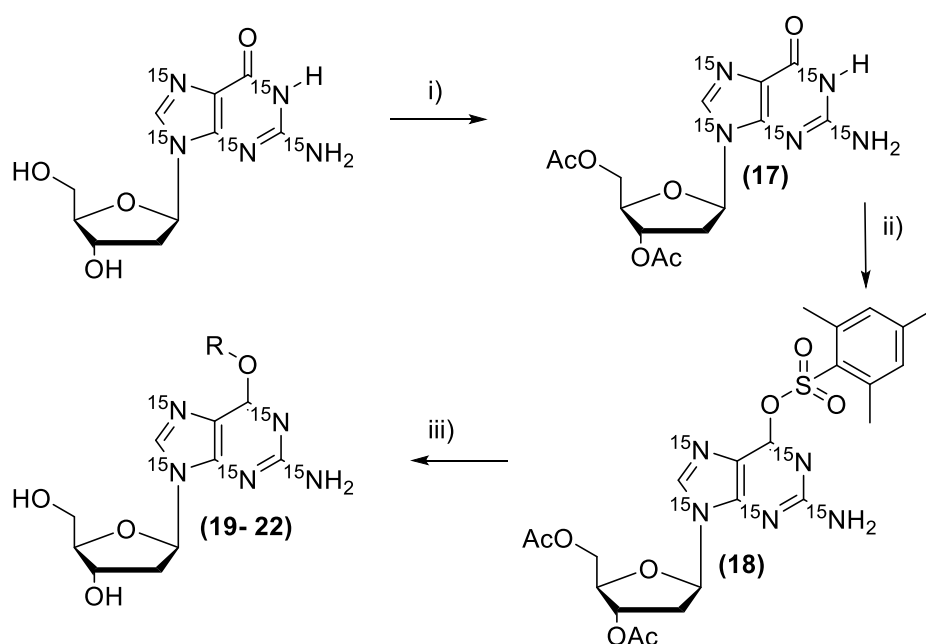
methyl ester would complete the synthesis of the labelled  $O^6$ -CEG nucleoside

(Scheme 24).



Scheme 24. Displacement of 6-mesitylenesulfonyl-3',5'-bis-*O*-acetyl-2'-deoxyguanosine by methyl(-)-lactate followed by hydrolysis of  $O^6$ -methyl(-)-lactate alkyl group to  $O^6$ -CEG; i) methyl(-)-lactate, DABCO, DBU, 1, 2-DME; ii) 0.5 M aq NaOH.

This can be undertaken during the removal of the acetyl groups therefore replacing silyl groups with acetyl groups was deemed appropriate and a new synthetic route designed (Scheme 25).



Scheme 25. Synthetic route to  $^{15}\text{N}_5$ -labelled  $O^6$ -alkylguanine compounds to be used as SIL internal standards starting from  $^{15}\text{N}_5$ -labelled 2'-deoxyguanosine; i)  $\text{Ac}_2\text{O}$ ,  $\text{Et}_3\text{N}$ , DMAP, DMF; ii)  $\text{Et}_3\text{N}$ , DMAP, 2-mesitylenesulfonyl chloride, DCM; iii) DABCO, ROH, DBU, 1, 2-DME; iv) 0.5 M aq NaOH.

Acetylation of 2'-deoxyguanosine followed the method of Tintoré *et al.*<sup>97</sup>. Thus, 2'-deoxyguanosine was dissolved in DMF and reacted with acetic anhydride (2.5 eq.) in

the presence of Et<sub>3</sub>N for four hours. However, analysis by ESI MS showed that there was formation of a tri-acetyl compound ( $[M+H]^+ = 394$ ) and the <sup>1</sup>H NMR spectrum indicated a third acetyl group.

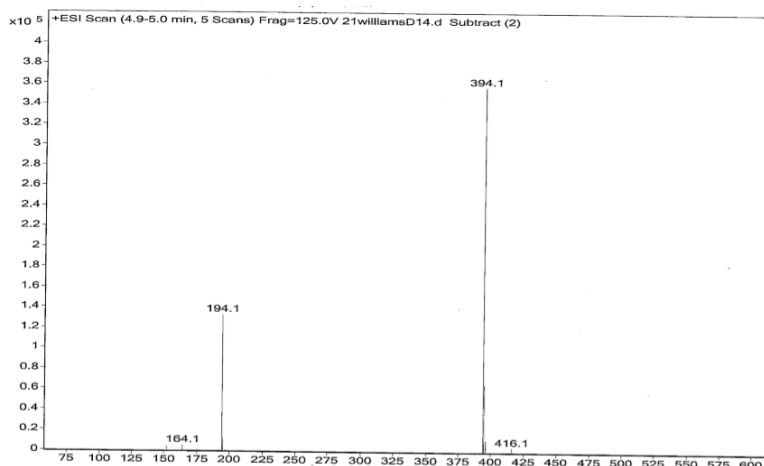


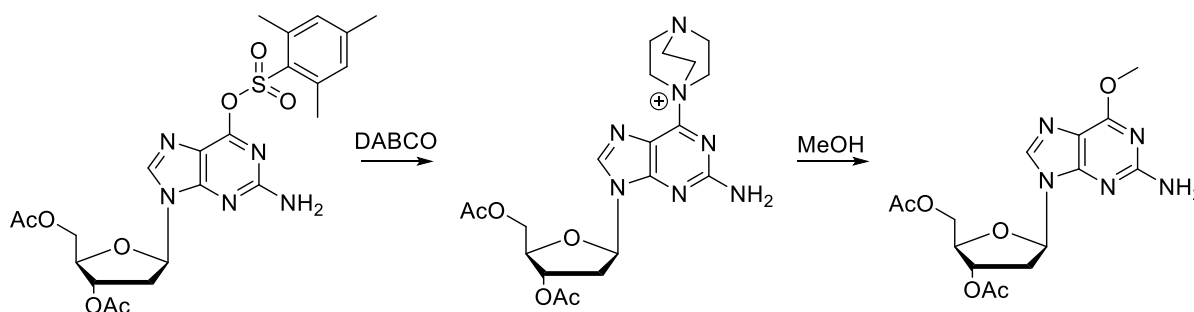
Figure 54. MS analysis showing mass corresponding to 3', 5'-bis-O-acetyl-2-N-acetyl-2'-deoxyguanosine.

This can be assigned to acetylation of both the sugar hydroxy groups and the exocyclic N<sup>2</sup>-amino group. To combat this, the method was modified by stopping the reaction after two hours instead of four hours. The diacetyl compound was collected, following removal of the solvent, by precipitation in DCM and vacuum filtration of the mixture. The diacetylated compound is insoluble in DCM therefore enabling its purification without the need of silica column chromatography and subsequently achieving a greater yield. **9** was synthesised with 98 % yield. This was then converted to the sulfonate ester without further purification using mesitylenesulfonyl chloride, triethylamine and DMAP, adapting the procedure used by Janeba *et al.*<sup>98</sup>. Analysis by TLC indicated completion of the reaction and subsequent purification of **10** by silica column chromatography resulted in a 62 % yield. This was deemed insufficient, but purification was necessary to remove DMAP. The presence of DMAP in future synthetic steps would result in side reactions and reduce the overall efficiency. Similar to the removal of imidazole mentioned previously, it was decided that DMAP could

potentially be removed using an acid extraction into an aqueous layer. Thus, mesitylation of **9** followed by a dilute aq HCl wash resulted in **10** with a 90 % yield.

### 3.1.4. Synthesis of *O*<sup>6</sup>-methyl-2'-deoxyguanosine

Displacement of the mesitylene sulfonate ester group by DABCO activates the *O*<sup>6</sup> position by formation of an *O*<sup>6</sup>-DABCO salt that can be displaced by a nucleophile. This enables the synthesis of *O*<sup>6</sup> modifications. *O*<sup>6</sup>-MeG is introduced through the displacement of DABCO by methanol (Scheme 26).



Scheme 26. Activation of the *O*<sup>6</sup> position of 6-mesitylenesulfonyl-3',5'-bis-*O*-acetyl-2'-deoxyguanosine by DABCO to enhance conversion to *O*<sup>6</sup>-MeG<sup>90</sup>.

**10** was reacted with DABCO, methanol (5 eq.) and DBU in 1,2-DME. After 18 hours, analysis by thin layer chromatography (TLC) indicated three products and the reaction mixture was analysed by ESI MS. The masses of the three products were  $[M+H]^+ = 282$  (the major peak),  $[M+H]^+ = 324$  and  $[M+H]^+ = 366$ .

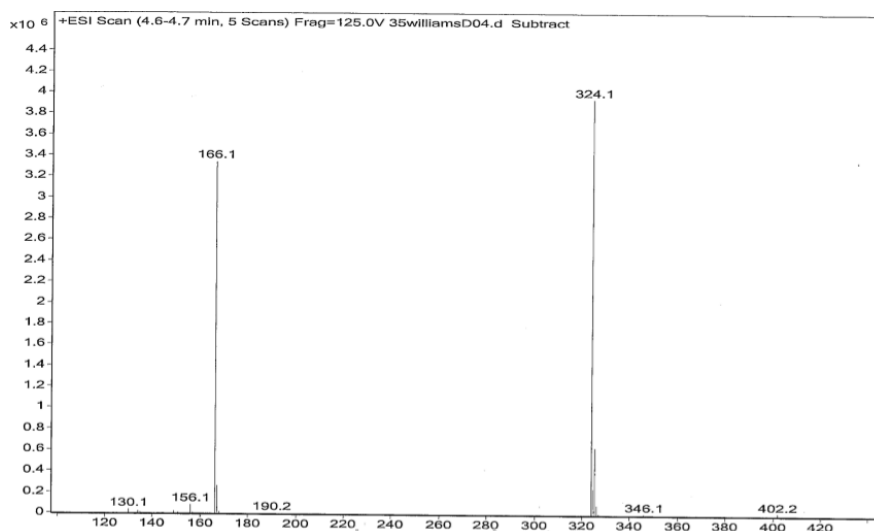


Figure 55. MS analysis indicating mass corresponding to monoacetylated **11**.

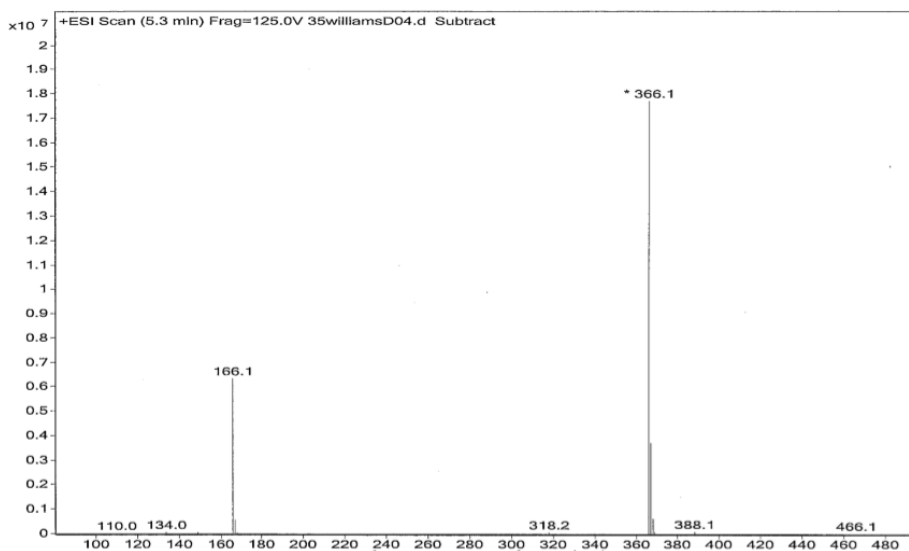


Figure 56. MS analysis indicating mass corresponding to diacetylated **12**.

The expected mass of  $O^6$ -MeG is  $[M+H]^+ = 282$  whilst the mass difference of 42 between the first two products corresponds to the addition of an acetyl group. This indicates that the second product is the mono-acetylated compound (**11**).  $[M+H]^+ = 366$  corresponds to the *bis*-acetylated compound (**12**).

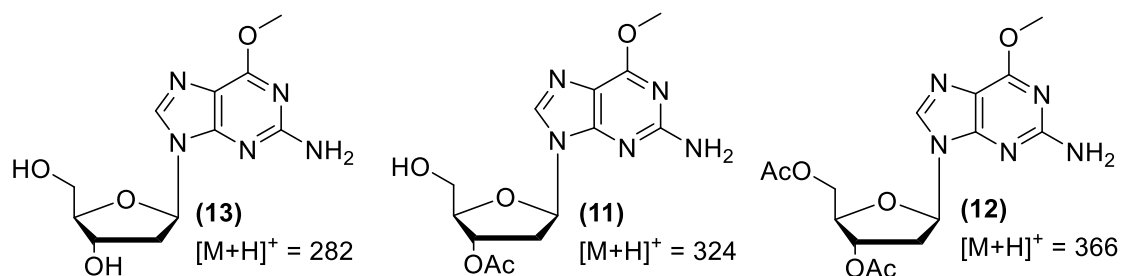


Figure 57. Structures of  $O^6$ -MeG, 3'-*O*-acetyl-2'-deoxy-6-methoxyguanosine and 6-methoxy-3', 5'-*bis*-*O*-acetyl-2'-deoxyguanosine formed during the reaction of 6-mesitylenesulfonyl-3', 5'-*bis*-*O*-acetyl-2'-deoxyguanosine with methanol, DBU and DABCO.

Formation of the desired  $O^6$ -MeG compound indicated that the acetyl groups could be removed under the basic conditions of the reaction without the addition of aq NaOH. The reaction was continued and analysed by TLC after 48 and 72 hours. After 48 hours, only two products were present, indicating no *bis*-acetyl compound, whilst after 72 hours, there was only one product corresponding to an unprotected nucleoside based on its  $R_f$  by silica TLC. This meant full deprotection had occurred. Due to the polarity of unprotected nucleosides, it was necessary to purify by RP-HPLC (5-20 % acetonitrile: water). **13** was purified with an 83 % yield.

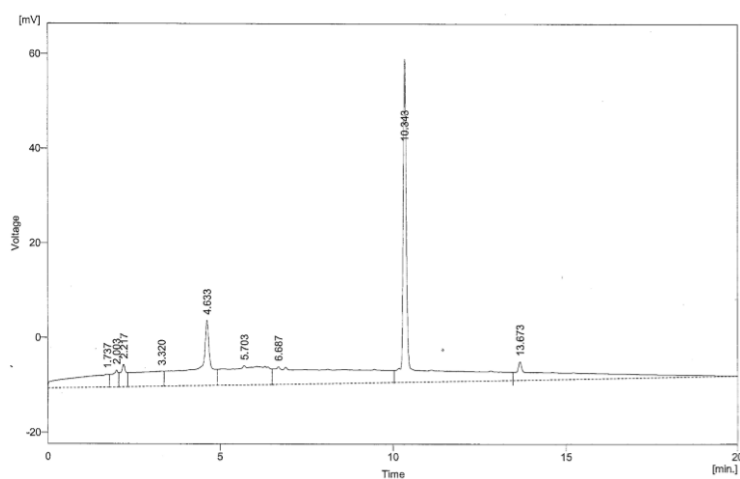


Figure 58. RP-HPLC trace of purification of  $O^6$ -MeG (10.34 mins). Peak present at 4.63 mins corresponds to DABCO reagent.

### 3.1.5. Synthesis of $O^6$ -EtG

$O^6$ -EtG was synthesised from **10** following the same method as for  $O^6$ -MeG: treating **10** with DABCO, DBU and ethanol in 1,2-DME and reacting for three days. Purification by RP-HPLC (5-20 % acetonitrile: water) resulted in **14** which was obtained with a 62 % yield.

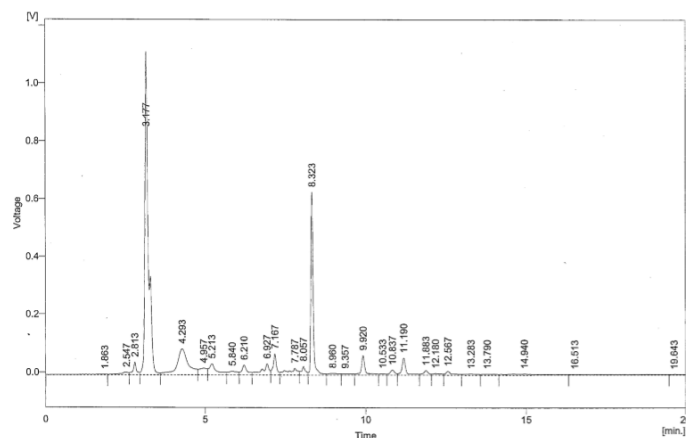


Figure 59. RP-HPLC trace of purification of  $O^6$ -EtG (8.32 mins). Peak present at 3.18 mins corresponds to DABCO reagent.

### 3.1.6. Synthesis of $O^6$ -CMG

**10** was reacted with DABCO, methyl glycolate and DBU in 1,2-DME for one day to convert the mesitylenesulfonyl group to the methylglycolyloxy group and partially remove the acetyl groups on the 3' and 5' hydroxy positions. 0.5 M aq NaOH was added and the reaction continued for a further two days. This was necessary because the methylglycolyloxy group needed to be hydrolysed to the carboxymethyl group. The addition of 0.5 M aq NaOH also resulted in the removal of the 3' and 5' acetyl groups. After purification by RP-HPLC (5-20 % acetonitrile: 0.1 M TEAB), **15** was obtained with a 67 % yield.



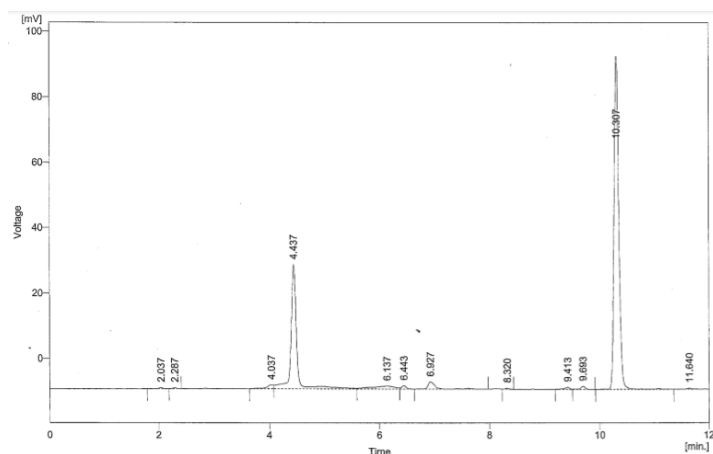


Figure 60. RP-HPLC trace of purification of *O*<sup>6</sup>-CMG (10.31 mins). Peak present at 4.44 mins corresponds to DABCO reagent.

### 3.1.7. Synthesis of *O*<sup>6</sup>-CEG

**10** was reacted with DABCO, methyl(-)-lactate and DBU in 1,2-DME for one day to convert the mesitylenesulfonyl group to the methyl lactyloxy group and partially remove the acetyl protecting groups on the 3' and 5' hydroxy positions. 0.5 M aq NaOH was added and the reaction continued for a further two days. This was necessary because the methyl lactyloxy group needed to be hydrolysed to the carboxyethyl group. The addition of 0.5 M aq NaOH also resulted in the removal of the acetyl groups. After purification by RP-HPLC (5-20 % acetonitrile: 0.1 M TEAB), **16** was synthesised with a 58 % yield.

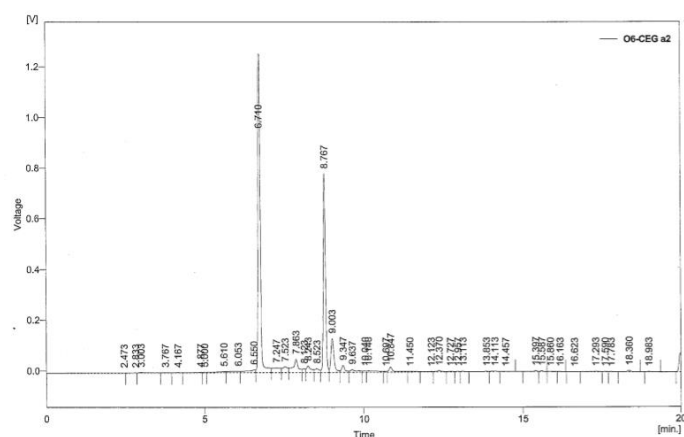


Figure 61. RP-HPLC trace of purification of *O*<sup>6</sup>-CEG (8.77 mins). Peak present at 6.71 mins corresponds to DABCO reagent.

## 3.2. Synthesis of $^{15}\text{N}_5$ -labelled $O^6$ -alkylguanine adducts

### 3.2.1. Introduction

With the successful synthesis of the four unlabelled  $O^6$ -alkylguanine adducts, a complete, three-step synthetic route was found that allowed the synthesis of these adducts in high yields starting from 2'-deoxyguanosine. The overall yield for the first two steps in the synthetic route was 88 % whilst the yields for the conversion of **10** into the respective  $O^6$ -alkylguanine adducts ranged from 58- 83 %. This was deemed acceptable and efficient enough for the synthesis of the respective  $^{15}\text{N}_5$ -labelled adducts.

### 3.2.2. Synthesis of $^{15}\text{N}_5$ -labelled 6-mesitylenesulfonyl-3',5'-bis-O-acetyl-2'-deoxyguanosine

Acetylation of  $^{15}\text{N}_5$ -labelled 2'-deoxyguanosine resulted in **17** (97 %) and its subsequent mesitylation resulted in **18**. As discussed in section 3.1.3, **18** was not purified, rather the catalytic amount of DMAP required for the mesitylation was removed by acid extraction.

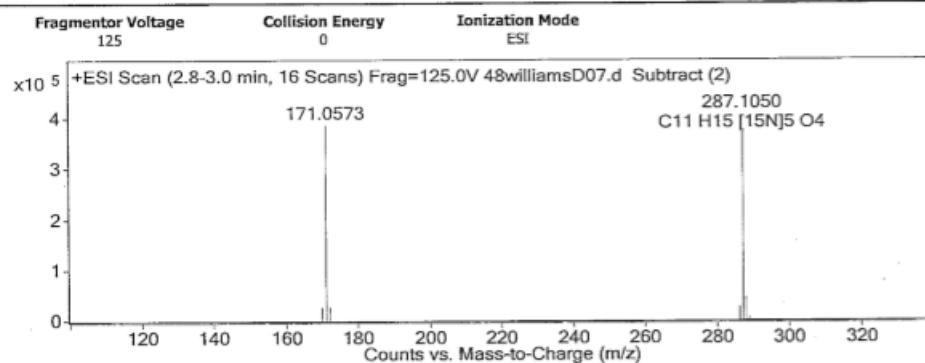
### 3.2.3. Synthesis of $^{15}\text{N}_5$ -labelled $O^6$ -MeG

$^{15}\text{N}_5$ -labelled  $O^6$ -MeG was synthesised by reacting **18** with DABCO, DBU and methanol (5 eq.) in 1,2-DME and reacting for three days. **19** was synthesised with an 88 % yield after purification by RP-HPLC (5-20 % acetonitrile:  $\text{H}_2\text{O}$ ).

## Qualitative Analysis Report

Data Filename	48williamsD07.d	Sample Name	A. Johns
Position	Vial 79	Acq Method	sheffield.m
Acquired Time	30/11/2017 18:49:22	DA Method	sheffield.m
Comment	O6-MeG N15		
Sample Group		Info.	
Acquisition SW	6200 series TOF/6500 series		
Version	Q-TOF B.05.01 (B5125.1)		

### User Spectra



#### Peak List

<i>m/z</i>	<i>z</i>	Abund	Formula	Ion	PPM Difference	Calculated <i>m/z</i>
171.0573	1	387661.72				
287.105	1	373395.56	C11 H15 [15N]5 O4	(M+H)+	-0.62	287.1049

Figure 62. Qualitative analysis report of  $^{15}\text{N}_5$ -labelled  $O^6$ -MeG indicating pure sample.  $M/z = 171$  indicates loss of 2'-deoxyribose during CNL analysis.

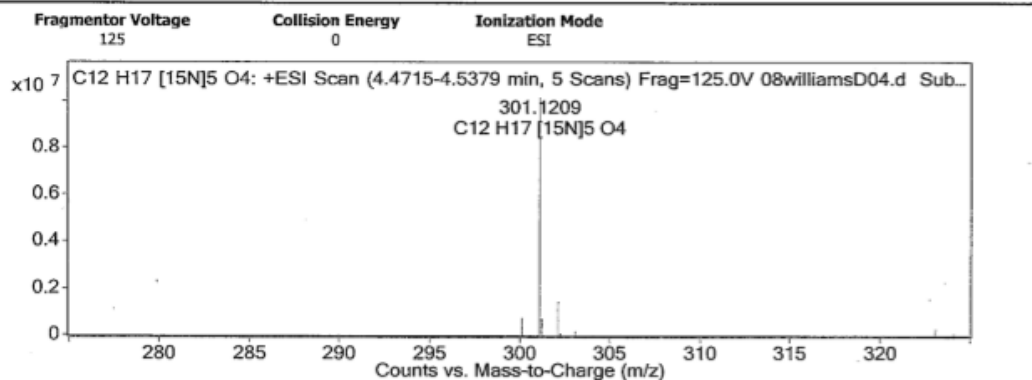
### 3.2.4. Synthesis of $^{15}\text{N}_5$ -labelled $O^6$ -EtG

$^{15}\text{N}_5$ -labelled  $O^6$ -EtG was synthesised by reacting **18** with DABCO, DBU and ethanol (5 eq.) in 1,2-DME and reacting for three days. **20** was synthesised with a 75 % yield after purification by RP-HPLC (5-20 % acetonitrile:  $\text{H}_2\text{O}$ ).

## Qualitative Analysis Report

<b>Data Filename</b>	08williamsD04.d	<b>Sample Name</b>	A. Johns
<b>Position</b>	Vial 57	<b>Acq Method</b>	sheffield_A2B2.m
<b>Acquired Time</b>	19/02/2018 18:56:51	<b>DA Method</b>	sheffield.m
<b>Comment</b>	15N06-EtG		
<b>Sample Group</b>		<b>Info.</b>	
<b>Acquisition SW</b>	6200 series TOF/6500 series		
<b>Version</b>	Q-TOF B.05.01 (BS125.1)		

### User Spectra



#### Peak List

<i>m/z</i>	<i>z</i>	Abund	Formula	Ion	PPM Difference	Calculated <i>m/z</i>
301.1209	1	10134663	C12 H17 [15N]5 O4	(M+H) <sup>+</sup>	-1.17	301.1205

Figure 63. Qualitative analysis report of <sup>15</sup>N<sub>5</sub>-labelled *O*<sup>6</sup>-EtG indicating pure sample.

### 3.2.5. Synthesis of <sup>15</sup>N<sub>5</sub>-labelled *O*<sup>6</sup>-CMG

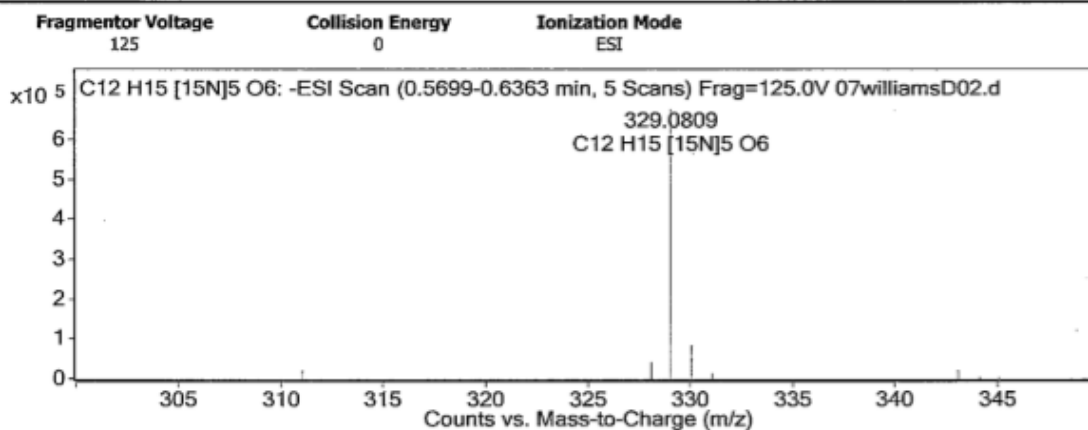
**18** was reacted with DABCO, methyl glycolate and DBU in 1,2-DME for one day to convert the mesitylenesulfonyl group to the methylglycolyloxy group and partially remove the acetyl protecting groups on the 3' and 5' hydroxy positions. 0.5 M aq NaOH was added and the reaction continued for a further two days. **21** was synthesised with a 67 % yield after purification by RP-HPLC (5-20 % acetonitrile: 0.1 M TEAB).

## Qualitative Analysis Report

<b>Data Filename</b>	07williamsD02.d	<b>Sample Name</b>	A. Johns
<b>Position</b>	Vial 22	<b>Acq Method</b>	sheffieldneg.m
<b>Acquired Time</b>	12/02/2018 16:38:28	<b>DA Method</b>	sheffield.m
<b>Comment</b>	15NO6-CMG		

<b>Sample Group</b>	<b>Info.</b>
<b>Acquisition SW</b>	6200 series TOF/6500 series
<b>Version</b>	Q-TOF B.05.01 (B5125.1)

### User Spectra



#### Peak List

<i>m/z</i>	<i>z</i>	Abund	Formula	Ion	PPM Difference	Calculated <i>m/z</i>
329.0809	1	676664.75	C12 H15 [15N]5 O6	(M-H) <sup>-</sup>	-2.48	329.0801

Figure 64. Qualitative analysis report of <sup>15</sup>N<sub>5</sub>-labelled O<sup>6</sup>-CMG indicating pure sample.

### 3.2.6. Synthesis of <sup>15</sup>N<sub>5</sub>-labelled O<sup>6</sup>-CEG

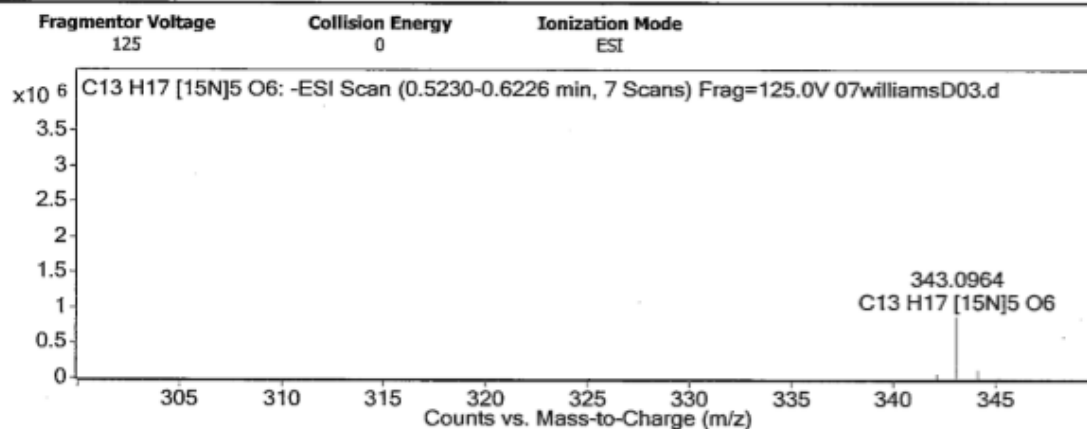
**18** was reacted with DABCO, methyl(-)-lactate and DBU in 1,2-DME for one day to convert the mesitylenesulfonyl group to the methyl lactyloxy group and partially remove the acetyl protecting groups on the 3' and 5' hydroxy positions. 0.5 M aq NaOH was added and the reaction continued for a further two days. **22** was synthesised with a 67 % yield after purification by RP-HPLC (5-20 % acetonitrile: 0.1 M TEAB).

## Qualitative Analysis Report

Data Filename	07williamsD03.d	Sample Name	A. Johns
Position	Vial 23	Acq Method	sheffieldneg.m
Acquired Time	12/02/2018 16:54:14	DA Method	sheffield.m
Comment	15N06-CEG		

Sample Group	Info.
Acquisition SW	6200 series TOF/6500 series
Version	Q-TOF B.05.01 (B5125.1)

### User Spectra



#### Peak List

m/z	Abund	Formula	Ion	PPM Difference	Calculated m/z
343.0964	887406.2	C13 H17 [15N]5 O6	(M-H)-	-1.85	343.0958

Figure 65. Qualitative analysis report of <sup>15</sup>N<sub>5</sub>-labelled O<sup>6</sup>-CEG indicating pure sample.

The synthetic routes presented here overcome the limitations of the aforementioned methods and offers a facile and efficient route to a range of SIL O<sup>6</sup>-alkylguanine adducts for use as SIL internal standards for SID methods. <sup>15</sup>N<sub>5</sub>-labelled O<sup>6</sup>-MeG, O<sup>6</sup>-EtG, O<sup>6</sup>-CMG and O<sup>6</sup>-CEG were synthesised following the three-step synthetic route with overall yields between 57- 75 %. This indicates that the route is suitable for a range of O<sup>6</sup>-alkylguanine adducts and that the procedure can be adapted according to the required modification. Acetylation of 3' and 5' hydroxy groups is fast, efficient and high yielding, likewise the mesitylation of the O6 position. The alkyl modification is introduced in the third step, after activation of the O6 position by formation of DABCO salt, *via* nucleophilic displacement. During this final step, the acetyl groups

are removed either by the basic nature of the nucleophile or by addition of 0.5 M aq NaOH.

Using  $^{15}\text{N}_5$ -labelled dG as starting material is expensive but SID methods require SIL internal standard amounts as low as  $1 \text{ fmol } \mu\text{L}^{-1}$ . This enables synthesis of numerous  $O^6$ -alkylguanine adducts in abundance therefore enabling examination of numerous genotoxic agents. No other synthesis of stable-isotopically labelled  $O^6$ -alkylguanines is as effective or precise.

### **3.3. Validation of SID method**

#### 3.3.1. Introduction

Validation of a SID method requires determining the limit of detection (LOD, specified as a 3: 1 signal-to-noise ratio) of the stable isotopically-labelled (SIL) internal standard as well as ensuring reproducible and accurate quantification by producing a calibration curve plotting the amount of internal standard injected against the amount detected. This requires spiking the analyte (the non-isotopically labelled compound of interest) with various concentrations of the internal standard, processing and then analysing the sample by LC-MS. The accuracy and reproducibility are determined by the correlation coefficient ( $r^2$ ) of the resulting plot, where a greater value of  $r^2$  corresponds to greater accuracy and reproducibility ( $r^2 \leq 1$ ). As discussed earlier in section 1.15, the SIL internal standard is structurally identical to the analyte, therefore, if there is any loss of the sample during the processing, it equates to an equal loss of both the analyte and the internal standard. Any ionisation suppression or matrix effects during LC-MS analysis will also affect the analyte and internal standard equally thus preserving the ratio between the two. This achieves greater certainty and accuracy during the quantification.

### 3.3.2. Calibration of SID method

Calibration of the method was performed using the  $^{15}\text{N}_5$ -labelled  $O^6$ -MeG compound as the SIL internal standard and  $O^6$ -MeG as the analyte. Accurate concentrations of the sample were calculated because the molar extinction coefficient of the unlabelled compound has been reported in literature<sup>34</sup>. However, the literature reports various values therefore the molar extinction coefficient of **13** was calculated. The purity of **13** was verified by RP-HPLC and  $^1\text{H}$  NMR and the extinction coefficient value calculated by UV analysis. The previous literature value reported was  $\epsilon(260 \text{ nm}) = 3400 \text{ M}^{-1} \text{ cm}^{-1}$  whilst for **13**  $\epsilon(260 \text{ nm}) = 3390 \text{ M}^{-1} \text{ cm}^{-1}$ , a discrepancy of 0.29 %. This can be attributed to discrepancies established during the preparation and analysis of the sample. Concentrations of the internal standard were calculated using an extinction coefficient of  $3390 \text{ M}^{-1} \text{ cm}^{-1}$ .

In order to test the quantification method, we used a 23-mer ODN containing a single  $O^6$ -MeG nucleobase for use as the analyte during the calibration (**ODN AJ07**) (Table 10). The concentration of the modified nucleobase (and analyte) could be calculated, therefore, once the analyte had been spiked with a known amount of the internal standard and processed. Analysis by LC-MS would establish a peak-abundance ratio between the  $^{14}\text{N}$  and  $^{15}\text{N}_5$   $O^6$ -MeG. This is possible since the  $^{15}\text{N}_5$ -labelled compound has a mass five units more than the unlabelled compound. Plotting the results of various spiked analytes would establish the value of  $r^2$  which is in direct proportion to the accuracy and reliability of the method.

Table 10. ODN **AJ07** synthesised for SID validation method containing a single  $O^6$ -MeG modification.

<i>Code</i>	<i>Sequence</i>
<i>AJ07</i>	5' GCC CTX CAG CTC CTG GCT GGC CC 3' X = $O^6$ -MeG



The preparation of the analyte required digestion of **ODN AJ07**, resulting in the individual nucleosides. As discussed in 2.4.1, DNA Degradase Plus™ was deemed suitable to achieve this. Analysis of the digested ODN by LC-MS showed efficient separation of the five nucleosides (Figure 66). The elution times of dC, dA, dG, T and *O*<sup>6</sup>-MeG were 1.43, 5.52, 6.52, 10.37 and 17.79 mins respectively. A prewash occurred between 0- 1 min and the peak at 13.39 mins possibly corresponds to the K<sup>+</sup> adduct of T, formed due to the presence of K<sup>+</sup> ions in the eluent, whilst DNA Degradase Plus™ is attributed to the peak eluted at 16.62 mins.

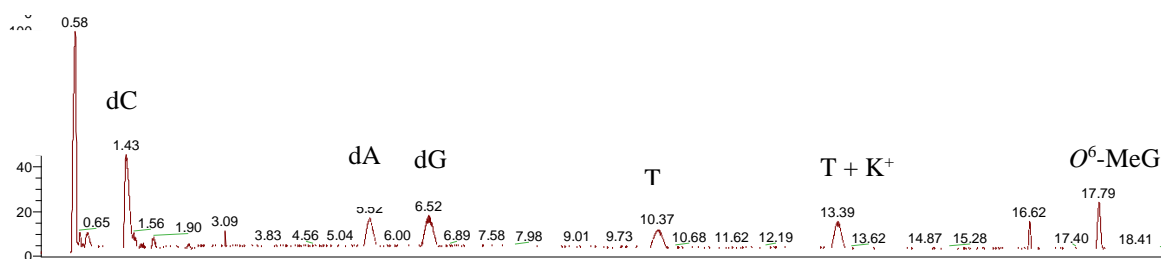


Figure 66. LC-MS analysis of digested **ODN AJ07** showing complete digestion and separation of nucleosides. dC (1.43 mins), dA (5.52 mins), dG (6.52 mins), T (10.37 mins), T+K<sup>+</sup> (13.39 mins), DNA Degradase Plus™ buffer (16.62 mins), *O*<sup>6</sup>-MeG (17.79 mins).

LC-MS analysis of digested **ODN AJ07** indicated successful digestion and clean separation of nucleosides whilst all peaks were accounted for. This meant the requirements needed for calibration were met because the analysis ensured complete digestion of **ODN AJ07** and no interference to the elution of the analyte, allowing accurate quantification. To maximise sensitivity and mass accuracy, UHPLC (Ultra HPLC) was coupled with a Q Exactive™ HF Orbitrap mass analyser. UHPLC increases sample throughput and enables fast analysis<sup>99</sup> whilst the orbitrap produces the greatest sensitivity, mass accuracy and resolving power of any currently available mass analyser<sup>52</sup>.

After initial analysis of the digested sample, the analyte (8.6 nmol  $\mu\text{L}^{-1}$ ) was spiked with various amounts of the SIL internal standard (0- 34.4 nmol  $\mu\text{L}^{-1}$ ) and analysed by LC-MS.

The analyte and the internal standard eluted at the exact same time due to the identical structure of the two. MS analysis of the eluted peak indicated the relative abundance of the masses of the two compounds from which the peak-abundance ratio was obtained (Figure 67).

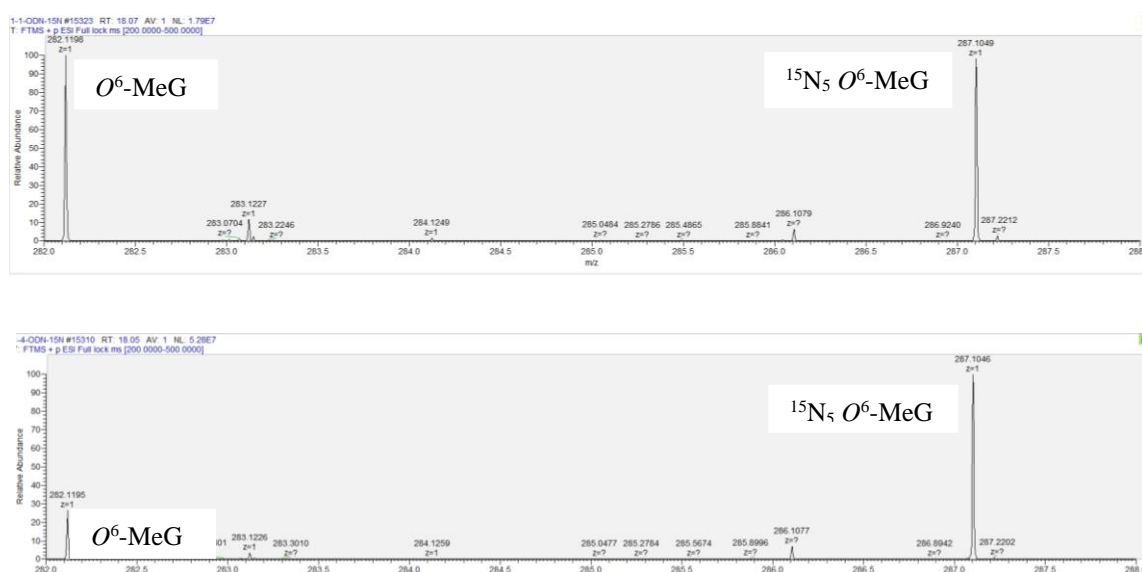


Figure 67. MS relative abundance of peak eluting  $O^6$ -MeG ( $[M+H]^+ = 282.1$ ) and SIL internal standard  $^{15}\text{N}_5$ -labelled  $O^6$ -MeG ( $[M+H]^+ = 287.1$ ). **ODN AJ07** digested by DNA Degradase Plus™ and spiked with various concentrations of SIL internal standard  $^{15}\text{N}_5$ -labelled  $O^6$ -MeG.  $[O^6\text{-MeG}] = [^{15}\text{N}_5\text{-labelled } O^6\text{-MeG}]$  (top) and  $[O^6\text{-MeG}] = \frac{1}{4} [^{15}\text{N}_5\text{-labelled } O^6\text{-MeG}]$  (bottom).

Analysis by LC-MS was undertaken after the analyte had been spiked with 8.6 nmol  $\mu\text{L}^{-1}$  and 34.4 nmol  $\mu\text{L}^{-1}$  of the SIL internal standard. Comparison of the peak-abundancies for the initial spike showed a 1: 1 ratio between the analyte ( $[M+H]^+ = 282.1$ ) and the internal standard ( $[M+H]^+ = 287.1$ ). This is to be expected because equal amounts of the compounds were present in the analysed sample. Furthermore, the second analysis produced a peak-abundance ratio corresponding identically to the

expected result. In this case, the analyte was spiked with four times the amount of internal standard, resulting in a peak-abundance ratio of 1: 4.

MS analysis solely of the analyte (0 nmol  $\mu\text{L}^{-1}$  internal standard) indicated no presence of a mass relating to the internal standard (Figure 68). This ensured that any detection of the internal standard mass is only from the synthesised compound.

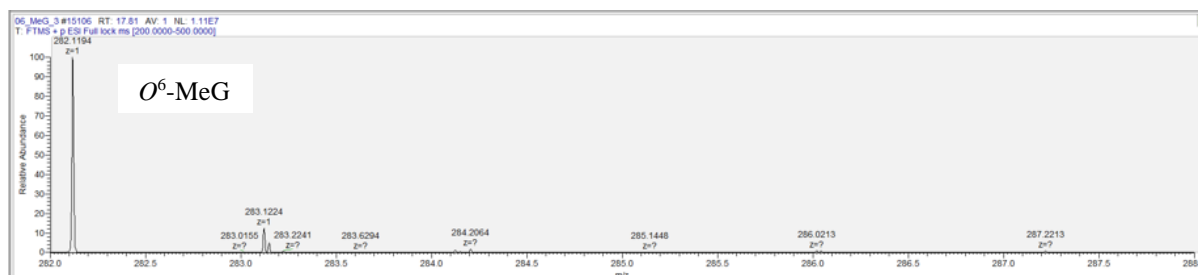


Figure 68. MS relative abundance of peak eluting  $O^6\text{-MeG}$  ( $[\text{M}+\text{H}]^+ = 282.1$ ) and SIL internal standard  $^{15}\text{N}_5$ -labelled  $O^6\text{-MeG}$  ( $[\text{M}+\text{H}]^+ = 287.1$ ). **ODN AJ07** digested by DNA Degradase Plus <sup>TM</sup> and spiked with 0 nmol  $\mu\text{L}^{-1}$  of SIL internal standard  $^{15}\text{N}_5$ -labelled  $O^6\text{-MeG}$ .

Likewise, the internal standard was analysed alone by MS to ensure that the compound was solely isotopically-labelled and no unlabelled compound was present.

MS analysis established both cases to be true (Figure 69).

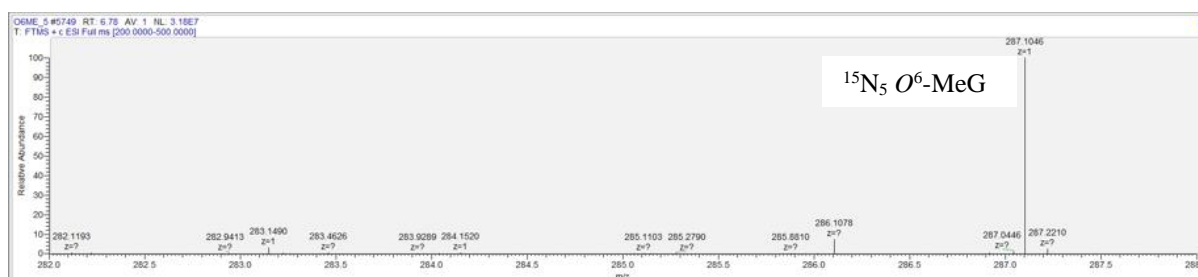


Figure 69. SIL internal standard  $^{15}\text{N}_5$ -labelled  $O^6\text{-MeG}$  ( $[\text{M}+\text{H}]^+ = 287.1$ ).

The accuracy and reliability of quantification was measured by plotting the ratio of peak-abundance (analyte/ internal standard) against  $[\text{analyte}]/ [\text{internal standard}]$  and measuring  $r^2$  of the plot (Figure 70).

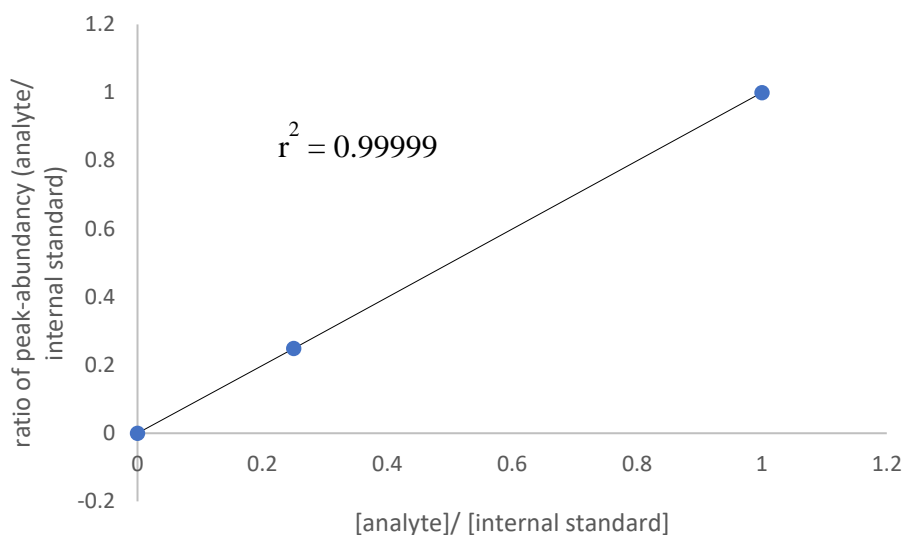


Figure 70. Chart plotting ratio of peak-abundance ( $O^6$ -MeG/  $^{15}N_5$ labelled  $O^6$ -MeG) against [ $O^6$ -MeG/  $^{15}N_5$ labelled  $O^6$ -MeG].

The calculated value of  $r^2 = 0.999$  indicates that the calibration of the method was successful.

### 3.3.3. Limit of Quantification

Limit of detection (LOD) is defined as “lowest amount of analyte in the sample, which can be detected but not necessarily quantitated under stated experimental conditions”<sup>100</sup> whilst limit of quantification (LOQ) is defined as “lowest amount of analyte in a sample, which can be quantitatively determined with suitable precision and accuracy”<sup>100</sup>. Both are determined by calculating the signal-to-noise ratio (S/N) of a spectrum:  $S/N = 2H/h$  and is visualised in Figure 71<sup>100</sup>. A S/N value of 3 is an accepted value for LOD whilst a S/N value of 10 is an acceptable value for LOQ<sup>100</sup>.

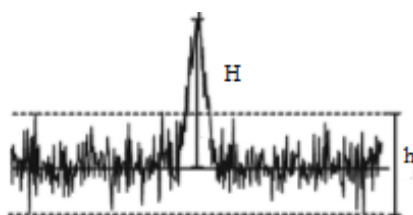


Figure 71. Determination of signal-to-noise ratio used to calculate limit of detection, where  $S/N = 2H/h$ , Shrivastava (2011)<sup>100</sup>.

The importance of LOQ is that it indicates that the SIL standard can be detected at similar levels to that expected of the analyte. This ensures accurate quantification of the analyte in a sample. If the concentration of the SIL standard at its LOQ is a lot greater than the expected concentration of analyte, calculating the abundance ratio between the two compounds becomes difficult because the difference in abundance of the two is so great.

1  $\mu\text{L}$  of  $^{15}\text{N}_5$ -labelled  $O^6$ -MeG (0.1  $\mu\text{M}$ ) was injected into the UHPLC and the S/N calculated from analysis of the resulting spectrum (Figure 72). It was calculated as 10.60 ( $H = 13.25$ ,  $h = 2.50$ ) which is an acceptable value for LOQ.

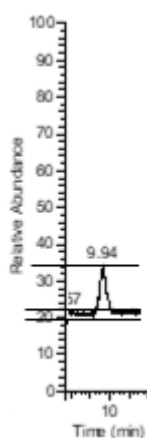


Figure 72. LC-MS spectrum of 1  $\mu\text{M}$   $^{15}\text{N}_5$ -labelled  $O^6$ -MeG to determine S/N of sample.

LC-MS calibration of the initial SID method using  $^{15}\text{N}_5$ -labelled  $O^6$ -MeG as the SIL internal standard determined the correlation coefficient,  $r^2$ , to be 0.99999. This is comparable to other successful SID methods<sup>61, 95, 104, 106</sup> and indicates that our method is reproducible and accurate. This value was obtained by analysing a digested ODN containing 8.6  $\text{nmol } \mu\text{L}^{-1}$   $O^6$ -MeG that had been spiked with  $^{15}\text{N}_5$ -labelled  $O^6$ -MeG (0-34.4  $\text{nmol } \mu\text{L}^{-1}$ ) and observing the relative abundances of the unlabelled analyte and SIL standard. The ratio (analyte/ SIL standard) obtained was plotted against [analyte]/[SIL standard] to generate  $r^2$ .

Despite the success of synthesis of the SIL standards and calibration of  $^{15}\text{N}_5$ -labelled  $O^6$ -MeG, there is one limitation to our proposed method. Approximately 4  $O^6$ -alkylG adducts occur in every  $10^8$  nucleotides which equates to 6 fmol in 100  $\mu\text{g}$  DNA. Previous SID methods developed for  $O^6$ -MeG have measured LOQ ranging between 24- 130 fmol<sup>61, 106</sup>. Our calculated LOQ of 100 fmol is therefore greater than existing methods and needs to be reduced to make it suitable to measure  $O^6$ -MeG adducts in DNA samples. However, there were time and cost restraints which limited LC-MS analysis for  $^{15}\text{N}_5$ -labelled  $O^6$ -MeG. This explains why previous SID methods acquired lower LOQs because, even though the LC-MS systems were the same, our system was unable to be calibrated specifically for our needs. This initial LOQ value indicates that the SID method has the potential to be developed further.

Further work assessing these criteria in more detail will ensure a clearer idea of the suitability of the SID method for quantifying  $O^6$ -MeG adducts in DNA samples whilst SID methods can be developed for the SIL standards  $^{15}\text{N}_5$ -labelled  $O^6$ -EtG,  $O^6$ -CMG and  $O^6$ -CEG using UHPLC.

**Chapter 4 – *Synthesis of Formaldehyde-Induced Interstrand  
Crosslinks Between Adenine Nucleobases in Double-Stranded DNA***

## **4. Synthesis of Formaldehyde-Induced Interstrand Crosslinks Between Adenine Nucleobases in Double-Stranded DNA**

### **4.1. Introduction**

Huang and Hopkins first synthesised DNA containing a single formaldehyde-induced dA-dA ICL by reacting formaldehyde directly with an ODN, but the route was limited due to a yield of less than 1 % and found to be highly dependent on sequence<sup>73</sup>. Adapting protocols that enable the synthesis of longer DNA duplexes containing a formaldehyde-induced ICL in a single defined position is highly desirable because it will enable studies on its repair mechanism. This chapter examines the use of formaldehyde and other methods to achieve this goal.

### **4.2. Synthesis of *bis*-(*N*<sup>6</sup>-2'-deoxyadenosyl) methane using methylene diamine**

Alkylene diamines have been used to prepare DNA containing *N*<sup>6</sup>-alkyladenine interstrand crosslinks ((-CH<sub>2</sub>)<sub>n</sub> where n > 1) by reacting 6-chloropurine-2'-deoxyriboside with the respective diamine<sup>107</sup>. Reacting an ODN containing 6-chloropurine-2'-deoxyriboside in adjacent strands with methylene diamine would in principle generate a DNA complex with a formaldehyde-induced ICL (fICL) between two adenine bases.

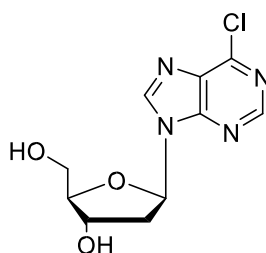


Figure 73. 6-Chloropurine 2'-deoxyriboside used for the synthesis of fICL between adjacent adenines.



Harris *et al.* synthesised and characterised a DNA duplex containing a four-carbon tether between two adenine bases on adjacent strands by reacting two complementary ODNs containing 6-chloropurine with diaminobutane in DIPEA and DMSO at 60 °C<sup>107</sup>. This was achieved using the phosphoramidite building block of 6-chloropurine 2'-deoxyribose during DNA synthesis followed by deprotection and purification of the 6-chloropurine-containing ODNs.

Enabling the investigation into the characterisation of FANCD2 and its involvement in the repair of fICLs was envisaged by combining the work of Harris *et al.* and Hodkinson *et al.* (unpublished data). For this, two ODN sequences, AJ-01 and AJ-02, were obtained to react with methylene diamine. The two sequences are complementary therefore they will anneal together and enable reaction of methylene diamine with the respective 6-chloropurine bases on each strand to form the fICL. The structure of the duplex is shown in Table 11.

Table 11. ODN sequences containing 6-chloropurine deoxyribose used to react with methylene diamine to synthesise an fICL-containing duplex.

<i>Code</i>	<i>ODN Sequence</i>	<i>Molecular Weight/ g mol<sup>-1</sup></i>
<i>AJ01</i>	5' ATG CCT GCA ATT ACX TAT GTC GTA ATC ATG GT 3'	10764
<i>AJ02</i>	3' C AAG CAG TCC TAA GGT XTA CAG CAT TAG TAC CA 5'	10767

ODN sequences AJ-01 and AJ-02 were fully deprotected by treatment of the CPG with aq 0.1 M NaOH for two days at room temperature and desalted by gel filtration using a NAP<sup>TM</sup>-10 column. Purification was achieved by reverse-phase HPLC.

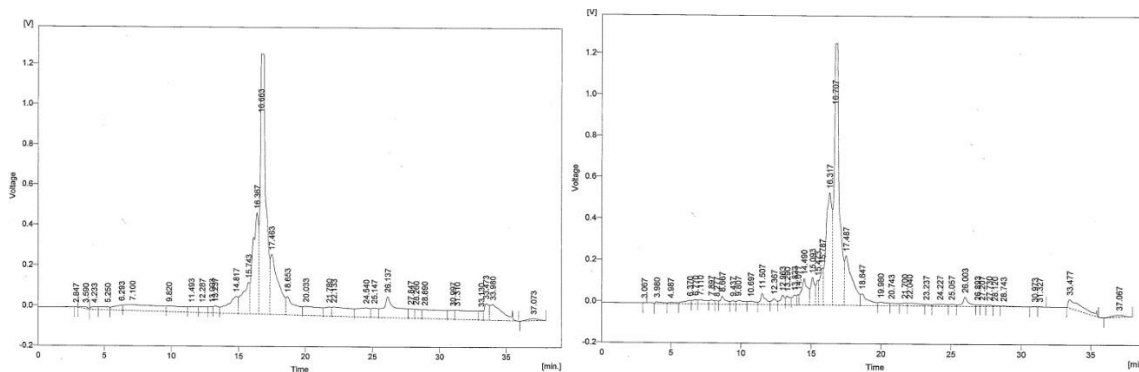


Figure 74. Crude analytical HPLC traces of the AJ-01 (left, 16.66 mins) and AJ-02 (right, 16.21 mins). Buffer A: 0.1 M TEAB, Buffer B: MeCN. 5– 20 % B over 30 minutes.

Analysis by ESI MS showed masses of 10764 and 10767 respectively indicating successful deprotection and purification of each ODN.

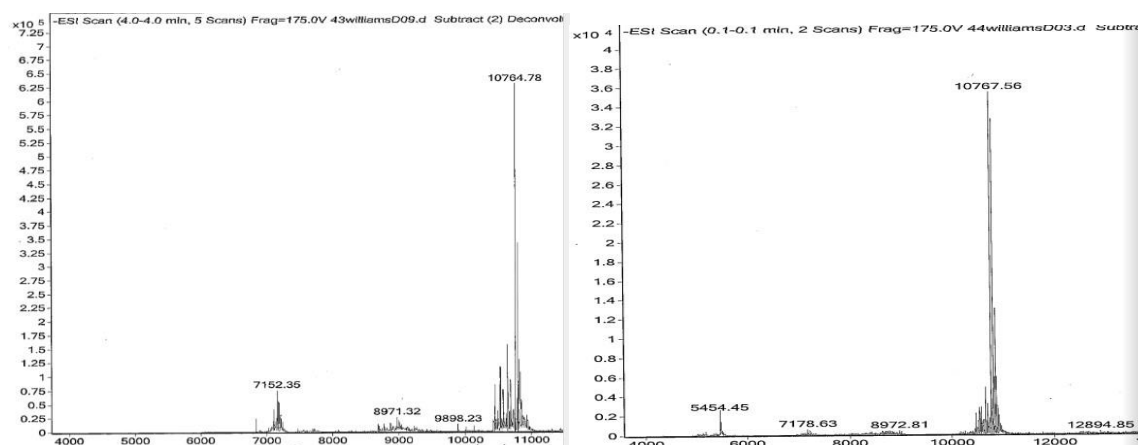
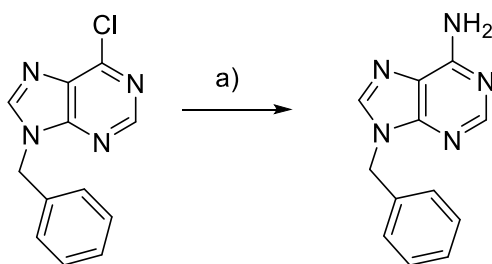


Figure 75. ESI MS analysis of purified ODN sequences AJ-01 (left, 10764.78) and AJ-02 (right, 10767.56).

Before reaction of purified AJ-01 and AJ-02 with methylene diamine, a model reaction was run by attempting a reaction between 9-benzyl-6-chloropurine and methylene diamine. This heterocycle was an available compound previously synthesised by the Williams group and was reacted with methylene diamine dihydrochloride, DIPEA and acetonitrile at 55 °C for 2 days. The reaction formed a white precipitate. Analysis by ESI MS showed a peak of mass 226 which corresponds to 9-benzyladenine suggesting that the methylene diamine dihydrochloride had decomposed to ammonia and methylamine prior to the displacement reaction.



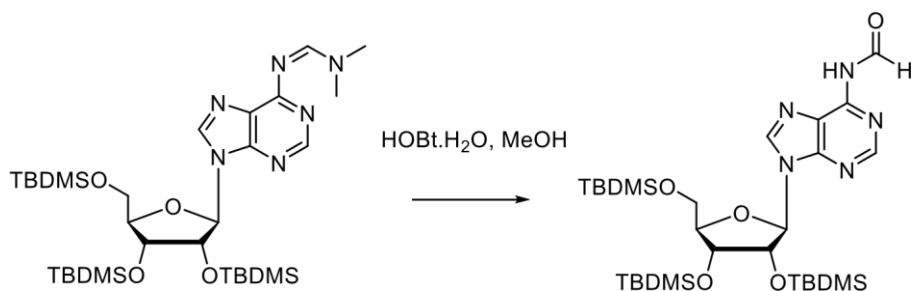
Scheme 27. Reaction of 9-benzyl-6-chloropurine with methylene diamine dihydrochloride leading to formation of 9-benzyladenine; a) methylene diamine dihydrochloride, DIPEA, MeCN, 55 °C.

Protecting one end of methylene diamine would stabilise the compound and enable it to react with 9-benzyl-6-chloropurine instead of decomposing. To this end, we attempted to react methylene diamine dihydrochloride with DMTCl in the presence of DMAP and triethylamine in pyridine. After 18 hours, silica TLC only showed the starting material, so the reaction was heated to 50 °C. After a further 18 hours, silica TLC again only showed the starting material. Analysis by ESI MS showed a peak of mass 120 which corresponds to methylene diamine dihydrochloride. It was deemed that using methylene diamine dihydrochloride to form the ICL was unsuitable due to its instability therefore another method to synthesise an ODN containing a defined fICL was sought.

#### 4.3. Synthetic route to formaldehyde-induced interstrand crosslinks using *N*<sup>6</sup>-[(dibutylamino)methylene]-2'-deoxyadenosine

In a second approach that was considered to synthesise a methylene ICL between two adenine bases on complementary ODNs strands, we envisaged synthesising an ODN containing *N*<sup>6</sup>-formyl-2'-deoxyadenosine. Annealing a complementary ODN strand, which would have a specifically placed adenine base, would, after reduction of the formyl group, allow formation the ICL *via* Schiff base formation from the *N*<sup>6</sup>-

hydroxymethyl adenine unit. Fu *et al.* have described the synthesis of  $N^6$ -formyladenosine in RNA and synthesised the nucleoside,  $N^6$ -formyladenosine, by treating the formamidine-protected nucleoside with *N*-hydroxy benzotriazole (HOBt.H<sub>2</sub>O) in methanol<sup>108</sup>.



Scheme 28. Synthesis of  $N^6$ -formyl-[2',3',5'-*tris*-O-(TBDMS)] adenosine<sup>108</sup>.

However, they noted that the compound is only stable in dry organic solvent because the *N*-formyl compound can readily be hydrolysed to the amine<sup>108</sup>.

$N^6$ -[(Dimethylamino)methylene]-2'-deoxyadenosine (dA<sup>dmf</sup>) (**23**) was synthesised in 97 % yield following the procedure set out by McBride *et al.*<sup>89</sup>.

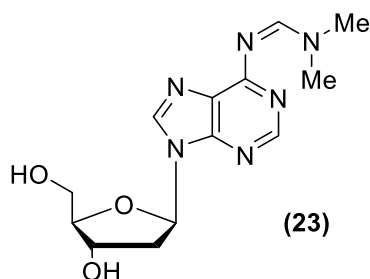


Figure 76.  $N^6$ -[(Dimethylamino)methylene]-2'-deoxyadenosine.

Synthesis of  $N^6$ -formyl-2'-deoxyadenosine was attempted following the procedure of Fu *et al.*<sup>108</sup> by reacting **23** with hydroxybenzotriazole in methanol at 0 °C for 1 hour. Analysis by ESI MS showed a peak of mass 306 which corresponds to the unreacted amidine starting material. With no available literature to refer to that would indicate

the expected reaction time of  $N^6$ -formyl-2'-deoxyadenosine, the reaction was repeated but this time it was followed by  $^1\text{H}$  NMR at room temperature. Analysing the peak shifts at regular intervals allowed us to gain an idea of how quickly the formyl-group was forming and the amount of time it would take before full decomposition of the amidine to the unprotected nucleoside.

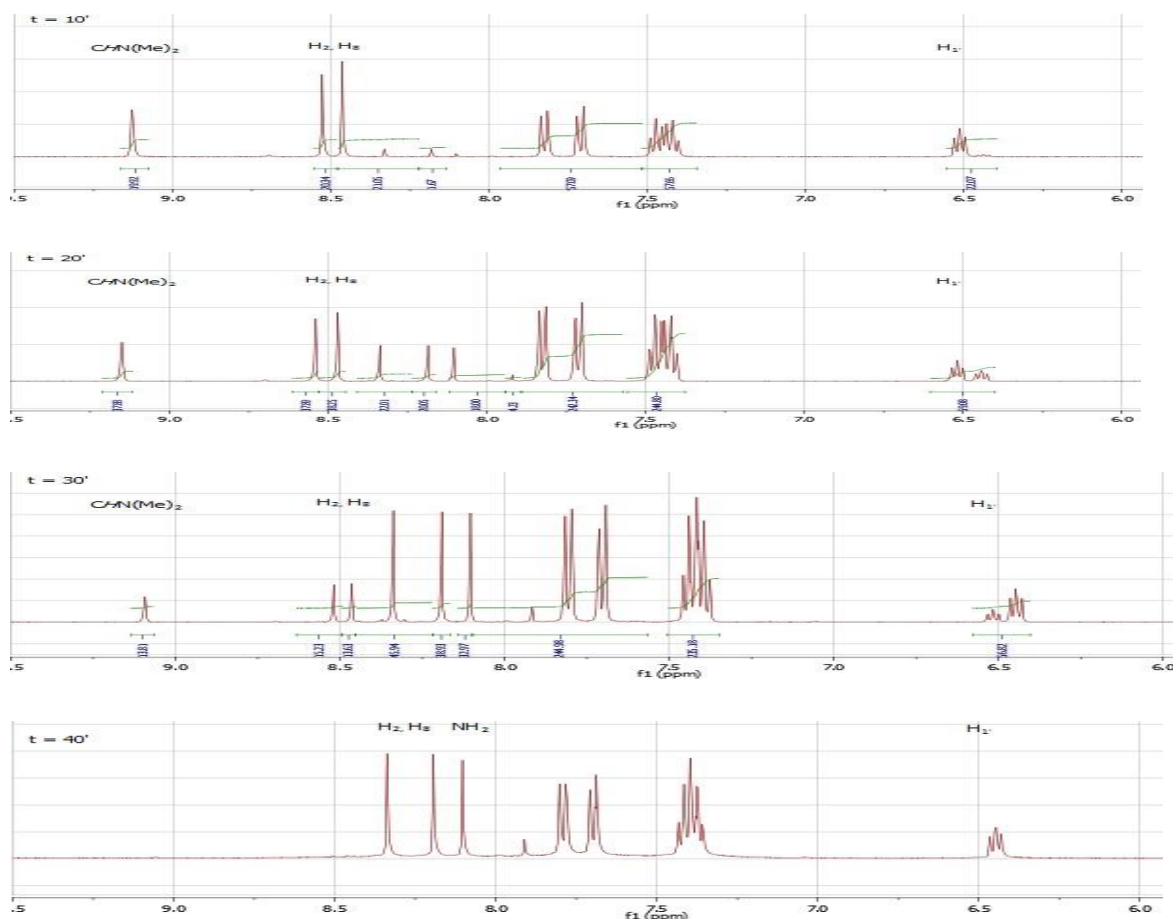


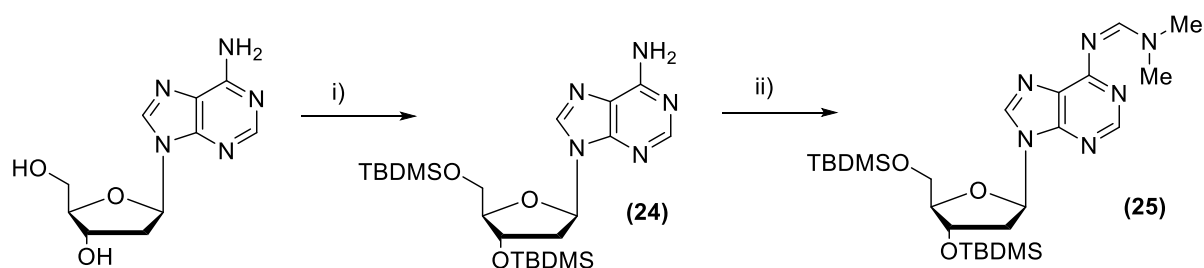
Figure 77.  $^1\text{H}$  NMR experiment of the reaction of  $N^6$ -[(dimethyl amino)methylene]-2'-deoxyadenosine with  $\text{HOBt}\cdot\text{H}_2\text{O}$  (3 eq.) in  $\text{D}_2\text{O}$ . Spectrum was taken at 10-minute intervals and the experiment was run for 40 minutes. The last spectrum indicates full deprotection of the amidine group leading to the formation of 2'-deoxyadenosine.

Following the reaction by  $^1\text{H}$  NMR at 10-minute intervals proved to be informative. After 10 minutes, the amidine CH signal at 9.18 ppm showed that the starting material was still present. After 20 minutes, the appearance of peaks at 6.42 ppm and between 8.0-8.35 ppm corresponded to 2'-deoxyadenosine. Analysing at 30 minutes the integration indicated that amidine group was being deprotected which was confirmed

by the last spectrum at 40 minutes. This spectrum matches that of 2'-deoxyadenosine indicating the full deprotection of the amidine group was complete.

The experiment followed by  $^1\text{H}$  NMR indicated that the full deprotection of the amidine to the amine occurred in just 40 minutes whilst there was no clear evidence of the intermediate formation of  $N^6$ -formyl-2'-deoxyadenosine. If the intermediate had formed, a new signal would be expected around 9.0 ppm corresponding to the formyl group whilst a shift would be seen in the  $\text{H}_8$  and  $\text{H}_2$  signals. This means isolating  $N^6$ -formyl-2'-deoxyadenosine would prove difficult.

3',5'-bis-*O*-(TBDMS)-2'-deoxyadenosine (**24**) was synthesised with a yield of 71 % in order to synthesise  $N^6$ -[(dimethylamino)methylene]-3',5'-bis-*O*-(*t*-butyldimethylsilyl)-2'-deoxyadenosine (**25**) as described<sup>108</sup>. **25** was obtained with a yield of 80 % after treating **24** with dimethylformamide dimethyl acetal. These compounds were synthesised because Fu *et al.* isolated  $N^6$ -formyl-2',3',5'-tri-*O*-(TBDMS) adenosine and we hypothesised that the TBDMS groups afforded to the compound may enable purification of the  $N^6$ -formyl compound by silica chromatography.



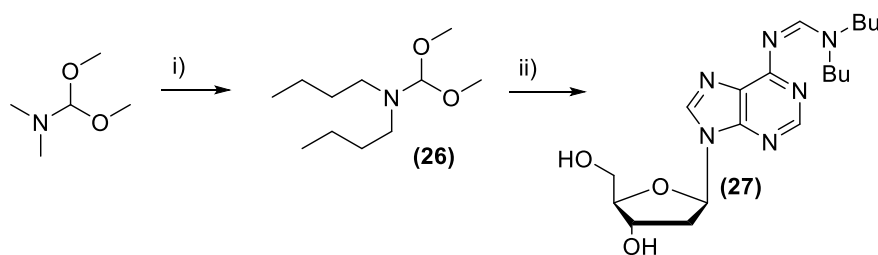
Scheme 29. Synthesis of  $N^6$ -[(dimethylamino)methylene]-3',5'-bis-*O*-(TBDMS)-2'-deoxyadenosine from 2'-deoxyadenosine; i) TBDMSO, DMF, imidazole (71 %); ii) *N,N'*-dimethylformamide dimethyl acetal, MeOH (80 %).

However, when compound **25** was reacted with HOBt in methanol at room temperature, analysis by ESI MS after 1 hour showed a peak of mass 480 which corresponds to the deprotected nucleoside **24**. Monitoring of  $\text{H}_8$  and  $\text{H}_2$  proton signals

of compound **25** by  $^1\text{H}$  NMR experiment indicated full hydrolysis to **24** within 30 minutes.

The two experiments using compounds **23** and **25** indicated that the formamidine group is completely hydrolysed in a short amount of time to give the 2'-deoxyadenosine. This was not completely unexpected since dimethylformamidine-protected dA is known to be relatively easily hydrolysed in acidic, basic and neutral conditions<sup>18, 89</sup>. Consequently, we turned our attention a more sterically-hindered protecting group dialkylformamidines, namely dibutylformamidine (dbf). Stability studies by McBride *et al.* found that 6-*N*-((di-*n*-butylamino)-methylene)-2'-deoxyadenosine ( $\text{dA}^{\text{dbf}}$ ) decomposed to dA 24 times slower than  $\text{dA}^{\text{dmf}}$  in concentrated aqueous ammonia solution<sup>89</sup>. Using dbf as a protecting group would therefore enhance the possibility of isolating the *N*<sup>6</sup>-formyl-2'-deoxyadenosine compound. Moreover, Ohkubo *et al.* have reported formation of *N*<sup>6</sup>-formyl-2'-deoxyadenosine by treating *N*<sup>6</sup>-[(dibutylamino)methylene]-2'-deoxyadenosine with acid though they did not isolate it<sup>109</sup>.

Synthesis of  $\text{dA}^{\text{dbf}}$  required the initial synthesis of *N*, *N*'-di-*n*-butylformamide dimethyl acetal. Thus compound (**26**) was obtained in 52 % yield following the procedure by Frohler and Matteucci<sup>18</sup>. This involved reacting dimethylformamide dimethyl acetal with di-*n*-butylamine.  $\text{dA}^{\text{dbf}}$  (**27**) was then synthesised in 67 % yield.



Scheme 30. Synthesis of *N*<sup>6</sup>-formyl-[(dibutylamino)methylene]-2'-deoxyadenosine from *N*, *N*'-dimethylformamide dimethyl acetal; i) di-*n*-butylamine, 100 °C, 52 %; ii) 2'-deoxyadenosine, anhydrous DMF, 67 %.

Initially, a reaction of **27** with concentrated aqueous ammonia solution was run. This was to assess the stability of the protecting group under basic conditions. If the ICL was to be synthesised *via*  $N^6$ -formyl adenine, we reasoned it would be advantageous to remove all the other nucleobase protecting groups on the ODN whilst leaving only the single amidine-protected unit, thereby facilitating purification and duplex formation.

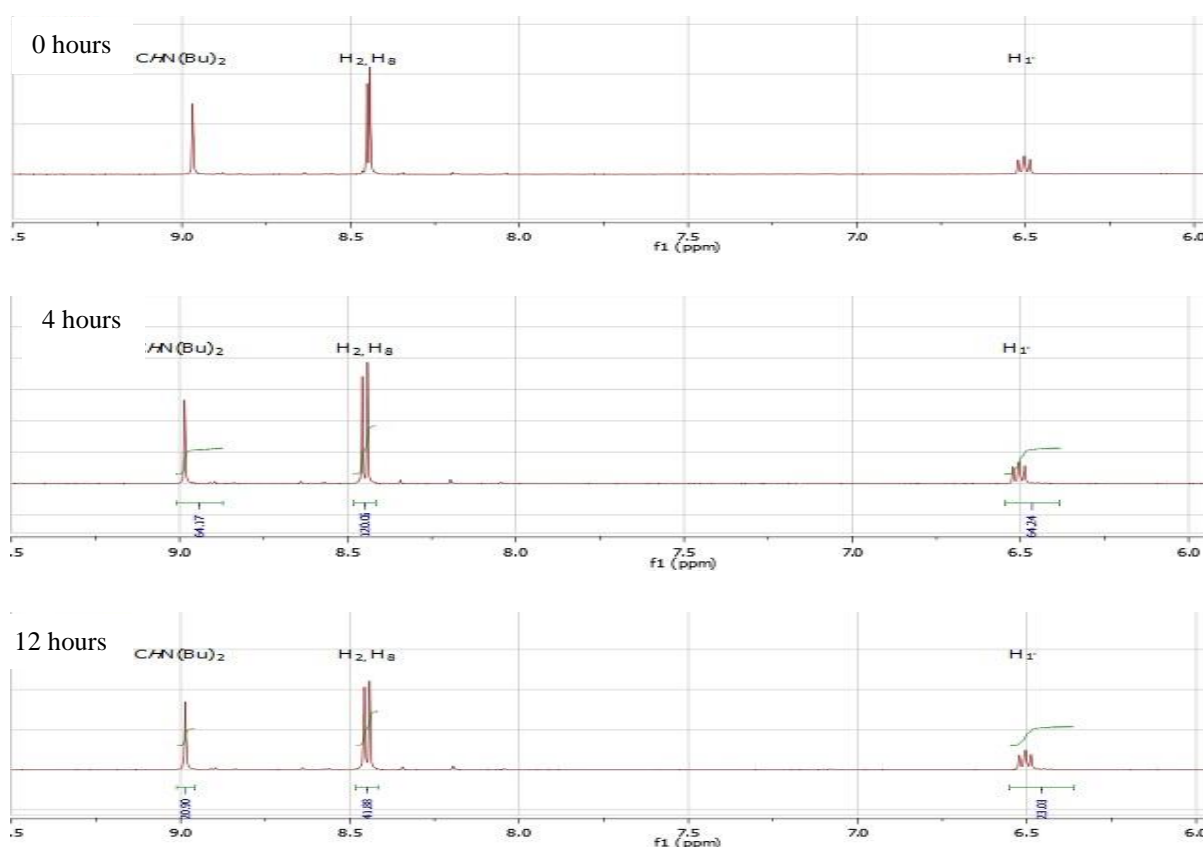


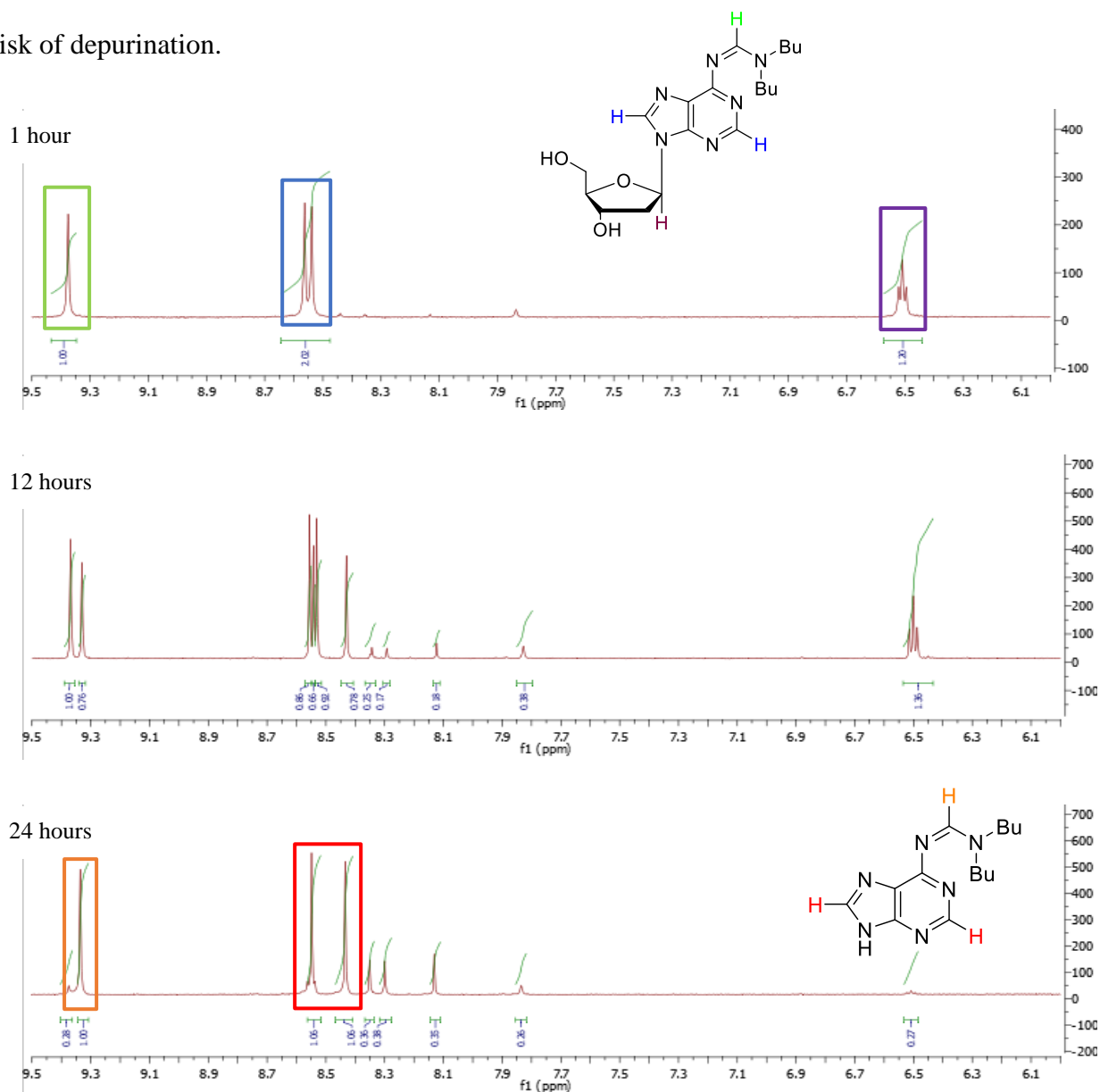
Figure 78.  $^1\text{H}$  NMR experiment of the reaction of  $N^6$ -[(dibutylamino)methylene]-2'-deoxyadenosine with concentrated aqueous ammonia solution. Spectrum was taken at 0, 4 and 12 hours.

A  $^1\text{H}$  NMR spectrum was taken on initial addition of the concentrated ammonia solution, at 4 hours and finally at 12 hours. No further spectra were taken because the removal of other ODN nucleobase protecting groups, such as pac, is complete within 12 hours<sup>8</sup>. All spectra in Figure 78 show a signal at 8.94 ppm which corresponds to the amidine proton of  $\text{CHN}(\text{Bu})_2$ . Thus, the amidine protecting group appears stable over



12 hours at room temp. and therefore it is a suitable group for the synthesis of the  $N^6$ -formyl adenine containing ODN.

Thus, we then investigated using  $^1\text{H}$  NMR to follow formation of  $N^6$ -formyl-2'-deoxyadenosine by monitoring the disappearance of the -NCHN- proton peak as well as the  $\text{H}_8$  and  $\text{H}_2$  of the nucleobase. New signals for the N-formyl group as well as new  $\text{H}_8$  and  $\text{H}_2$  signals were expected. Hydrolysis was performed using  $10\ \mu\text{M}$  DCl in  $\text{D}_2\text{O}$  at pH 5. Using stronger acid would not be compatible with ODNs due to the increased risk of depurination.



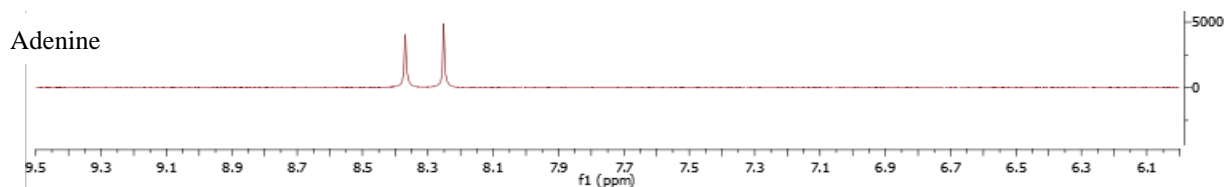


Figure 79.  $^1\text{H}$  NMR experiment of the reaction of  $N^6$ -[(dibutylamino)methylene]-2'-deoxyadenosine with 10  $\mu\text{M}$  DCl in  $\text{D}_2\text{O}$ . Spectrum was taken at 1, 12 and 24 hours. The bottom spectrum is adenine in 10  $\mu\text{M}$  DCl in  $\text{D}_2\text{O}$ .

Initial analysis of  $N^6$ -[(dibutylamino)methylene]-2'-deoxyadenosine shows the amidine signal ( $\text{CHN}(\text{Bu})_2$ ) at 9.40 ppm whilst  $\text{H}_8$  and  $\text{H}_2$  are between 8.5- 8.6 ppm. After 12 hours, there has been clear hydrolysis of the compound due to formation of a new signal at 9.34 ppm as well as signals at 8.54 and 8.43 ppm. Analysis of the  $\text{H}_{1'}$  signal also indicates hydrolysis of the compound but, importantly, there is no formation of a new  $\text{H}_{1'}$  environment in the range expected for nucleosides. An explanation for this is that depurination of  $N^6$ -[(dibutylamino)methylene]-2'-deoxyadenosine to  $N^6$ -[(dibutylamino)methylene] adenine has occurred and the signals at 9.34, 8.54 and 8.43 ppm correspond to  $N^6$ -[(dibutylamino)methylene] adenine. Integration of the respective purine ring protons indicated that roughly half of  $N^6$ -[(dibutylamino)methylene]-2'-deoxyadenosine has been depurinated at this point. The spectrum obtained after 24 hours indicated complete depurination of the formamidine compound, as seen with the almost disappearance of the  $\text{H}_{1'}$  signal. Comparison to the  $^1\text{H}$  NMR spectrum of adenine suggests that over 24 hours, a quarter of the dibutyl formamidine group had completely hydrolysed to adenine. MS analysis of the reaction supports the explanation that depurination had occurred. Analysis of the reaction after 12 hours showed  $[\text{M}+\text{H}]^+ = 275$  as well as  $[\text{M}+\text{H}]^+ = 391$ .  $[\text{M}+\text{H}]^+ = 391$  corresponds to  $\text{dA}^{\text{dbf}}$  whilst  $[\text{M}+\text{H}]^+ = 275$  indicates the loss of the 2'-deoxyribose resulting in  $N^6$ -[(dibutylamino)methylene] adenine. After 24 hours,  $[\text{M}+\text{H}]^+ = 136$  was present, which corresponds to adenine.

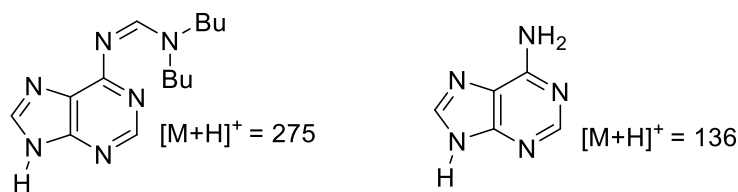


Figure 80. Structures and masses of  $N^6$ -[(dibutylamino)methylene] adenine and adenine.

The depurination of  $dA^{dbf}$  to  $N^6$ -[(dibutylamino)methylene] adenine in 24 hours was a surprise. McBride *et al.* studied the depurination of  $dA^{dbf}$  under mild acidic conditions (DCA/  $CH_2Cl_2$ , 1: 50 v/v) and recorded a  $t_{1/2}$  of 30 hours<sup>89</sup>. However, there was no recorded pH of the sample therefore our reaction may have been under stronger acidic conditions. Moreover, neither MS nor  $^1H$  NMR analysis indicated formation of  $N^6$ -formyl-2'-deoxyadenosine. Instead, both indicated partial hydrolysis of the dibutyl formamidinium group to the amine.

Despite issues with characterising the sample and no clear indication of the formation of  $N^6$ -formyl-2'-deoxyadenosine, a phosphoramidite containing dibutyl formamidinium protected  $dA$  was synthesised by Chris Millington, a fellow researcher in the Williams group. The phosphoramidite was then incorporated into an ODN using standard conditions and phenoxyacetyl protection of dG and acetyl protection of dC (Table 12).

Table 12. ODN sequence containing  $N^6$ -[(dibutylamino)methylene]-2'-deoxyadenosine.

<i>Code</i>	<i>ODN Sequence</i>
<b>AJ10</b>	5' CCC TCC AXT TGT TTG 3' <b>X</b> = $N^6$ -[(dibutylamino)methylene]-2'-deoxyadenosine

Once synthesised, the ODN was deprotected and purified to ensure the dA<sup>dbf</sup> modification did not decompose during the removal of base labile protecting groups. Thus, ODN *AJ10* (0.33  $\mu$ mol), whilst still attached to CPG solid support, was treated with aq ammonia solution (33 %) for 12 hours at room temperature that previous research indicated would not decompose the formamidine group. However, analysis by RP-HPLC showed that decomposition had occurred (Figure 81).

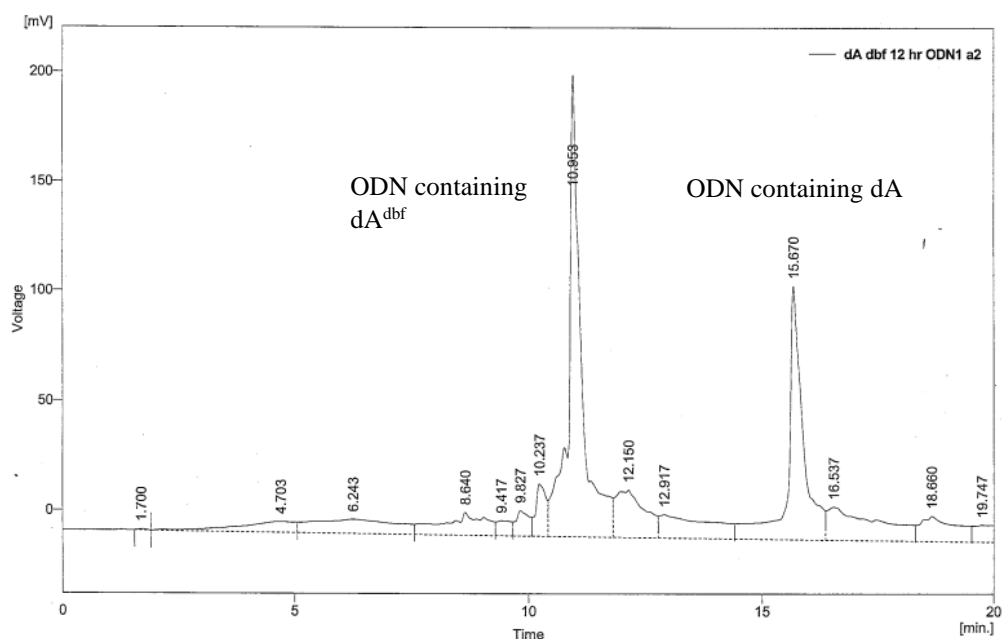


Figure 81. RP-HPLC analysis of ODN *AJ10* after treatment with ammonia after 12 hours. Peak at 10.95 mins corresponds to ODN containing dA<sup>dbf</sup> whilst peak at 15.67 mins corresponds to ODN containing dA. Buffer A: 0.1 M TEAB, Buffer B: MeCN. 5 – 18.5 % B over 20 minutes.

ESI MS analysis of the sample gave a mass of 4632. This is the expected mass of sequence containing the formamidine. However, the analysis also gave a mass of 4493. This is the mass of sequence 5' CCC TCC AAT TGT TTG 3', indicating that the dibutyl formamidine group had been fully hydrolysed.

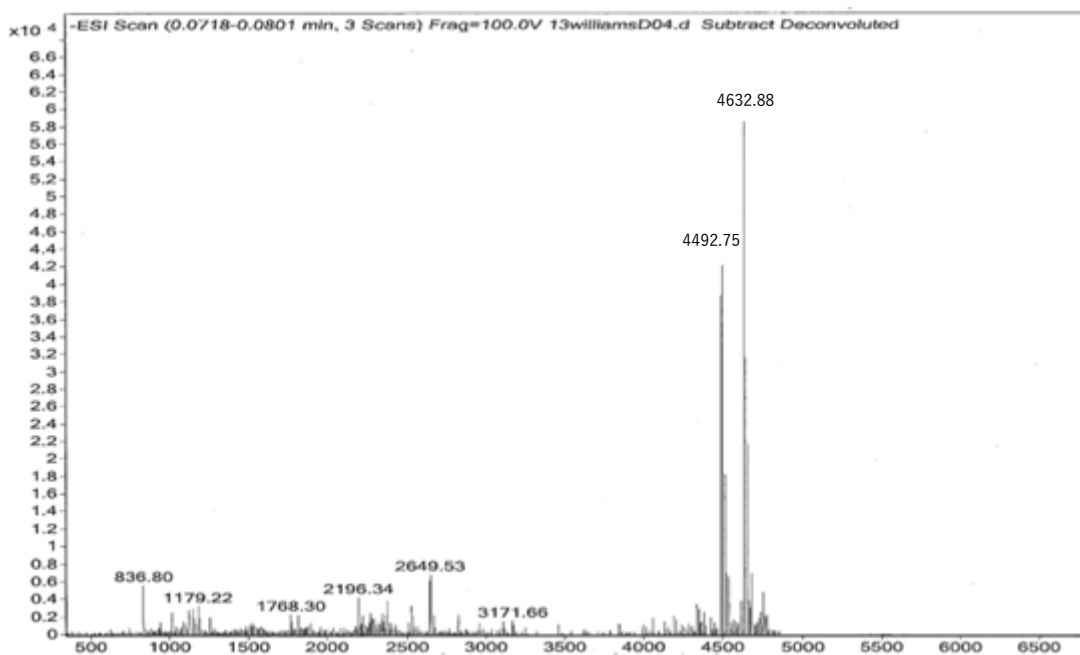


Figure 82. MS analysis of ODN *AJ10*. 4632.88 corresponds to sequence 5' CCC TCC AXT TGT TTG 3' containing the desired modification. However, mass 4492 corresponds to 5' CCC TCC AAT TGT TTG 3'.

In order to avoid decomposition, CPG of *AJ10* (0.33  $\mu$ mol) was treated with aq 0.1 M NaOH for four hours at room temperature. This provided milder basic conditions that were expected to remove the base labile groups without affecting the dibutyl formamidinium. The sample was desalted and analysed by RP-HPLC. However, RP-HPLC analysis again indicated decomposition had occurred (Figure 83).

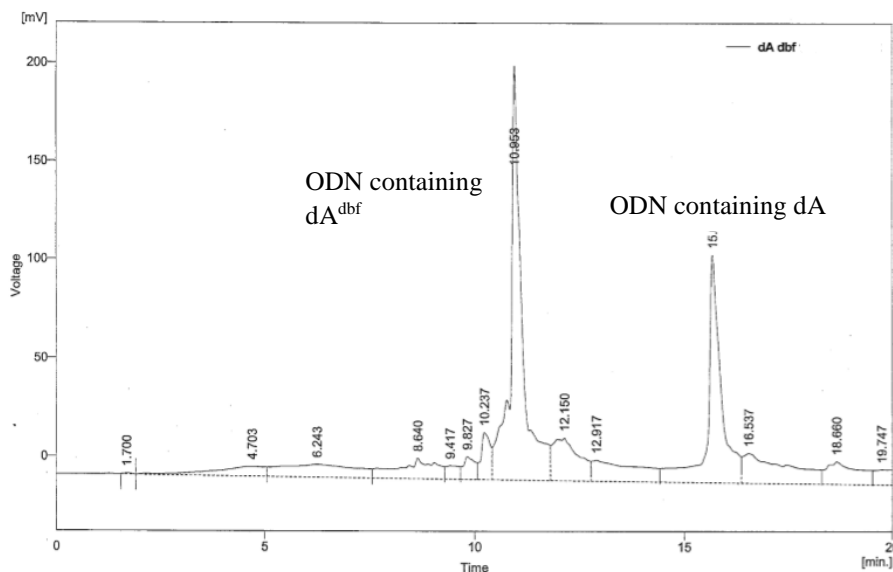


Figure 83. RP-HPLC analysis of ODN *AJ10* after treatment with 0.1 M aq NaOH for four hours. Peak at 10.95 mins corresponds to ODN containing dA<sup>dbf</sup> whilst peak at 15.67 mins corresponds to ODN containing dA. Buffer A: 0.1 M TEAB, Buffer B: MeCN. 5 – 18.5 % B over 20 minutes.

A third way was attempted. This time CPG of *AJ10* (0.33  $\mu$ mol) was treated with 0.05 M  $K_2CO_3$ / MeOH. This made no improvement (Figure 84). There were two possible reasons for the result: either the formamidine group decomposed during deprotection of the ODN or the decomposition occurred during the ODN synthesis on the CPG and therefore contained the mixture anyway. A greater understanding of the deprotection process was required in order to rule out one of the above without further wastage of ODN *AJ10*.

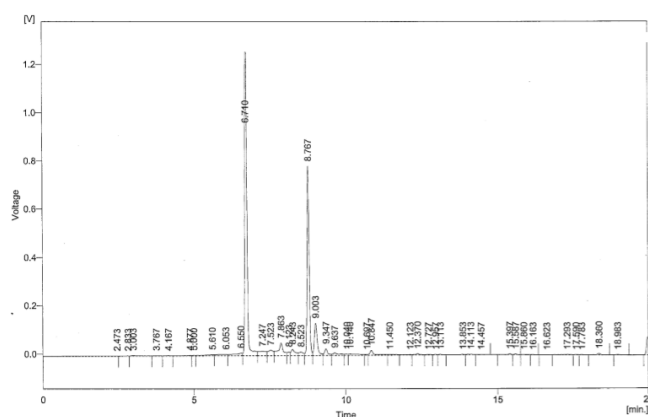


Figure 84. RP-HPLC analysis of ODN *AJ10* after treatment with 0.1 M aq NaOH for four hours. Peak at 6.71 mins corresponds to ODN containing dA<sup>dbf</sup> whilst peak at 8.77 mins corresponds to ODN containing dA. Buffer A: 0.1 M TEAB, Buffer B: MeCN. 5 – 47 % B over 20 minutes.

#### 4.4. Synthesis of d(GpA<sup>dbf</sup>)

To better understand the deprotection process, it was envisaged that synthesis of a dimer containing pac-protected dG (to replicate the protected nucleobase in *AJ10*) and dA<sup>dbf</sup> would allow analysis into the deprotection step without wasting CPG of *AJ10* (Figure 85). If the removal of pac from dG occurred without the decomposition of dA<sup>dbf</sup> under any of the three conditions mentioned above, then it would be ruled that decomposition of the phosphoramidite occurred. This would mean that deprotection and purification of CPG containing *AJ10* could continue. However, if decomposition of the dimer dA<sup>dbf</sup> occurred then conditions would need to be sought in order to achieve deprotection and purification without loss of material.

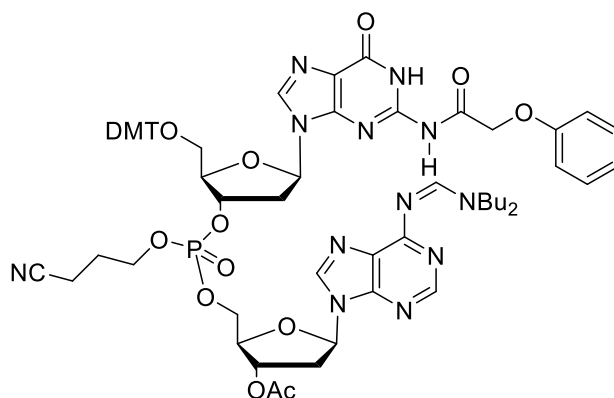
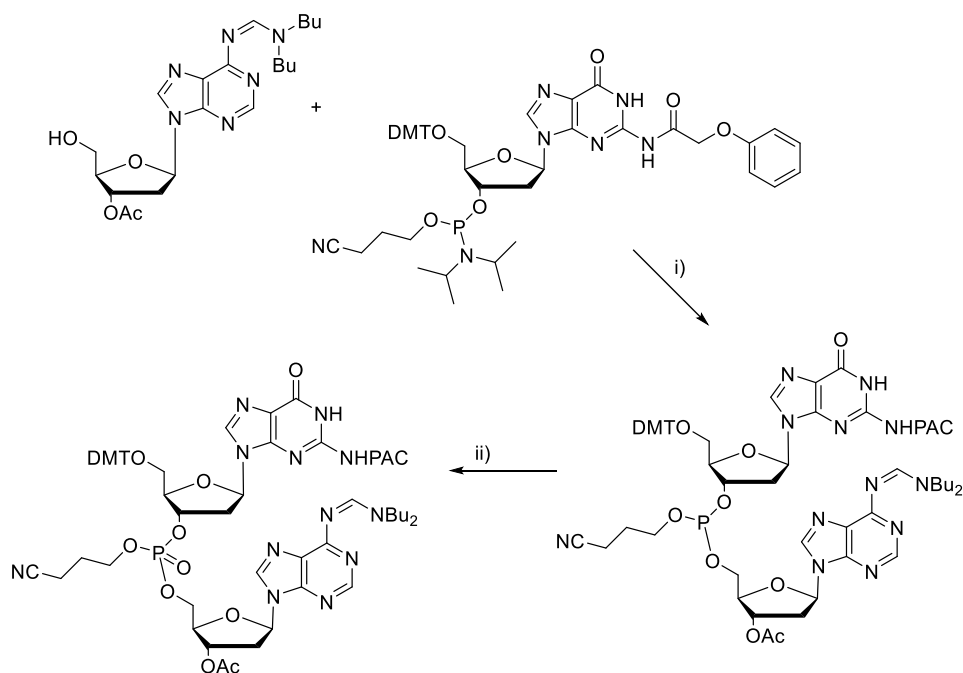


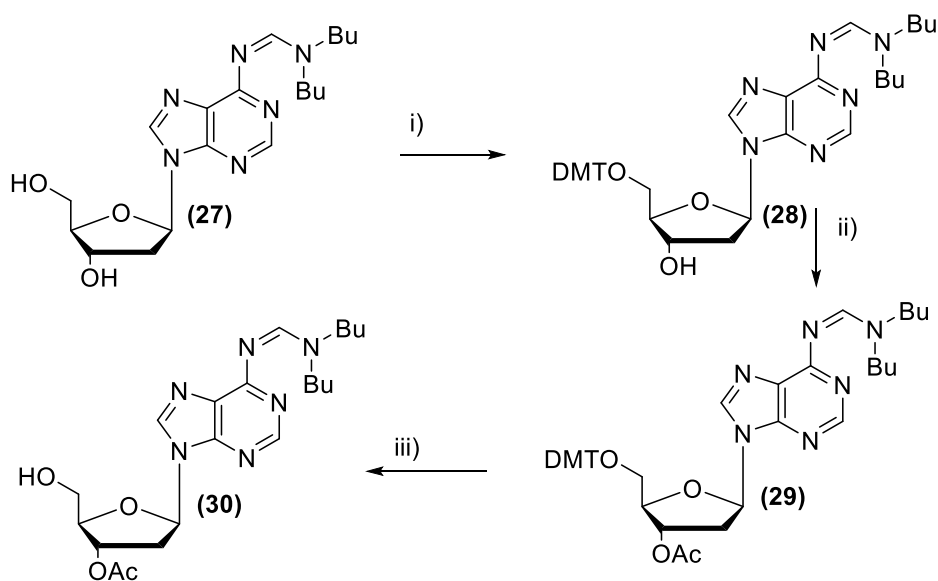
Figure 85. Structure of d(G<sup>pac</sup>pA<sup>dbf</sup>).

Synthesis of the dimer could be achieved by reacting 3'-O-acetyl-6-N',N'-[(dibutylamino)methylene]-2'-deoxyadenosine with pac-dG using standard ODN synthesis protocol (Scheme 31.)<sup>110</sup>.



Scheme 31. Proposed synthesis of  $d(G^{\text{pac}}pA^{\text{dbf}})$ ; i) 0.4 M (ethylthio) tetrazole, MeCN; ii) 0.2 M iodine, THF.

The 3'-phosphoramidite reagent of pac-dG is available to purchase directly whilst a three-step synthesis of 3'-*O*-acetyl-6-*N'*, *N'*-[(dibutylamino)methylene]-2'-deoxyadenosine was devised starting from  $dA^{\text{dbf}}$  (Scheme 32).



Scheme 32. Synthesis of 3'-*O*-acetyl-6-*N'*, *N'*-[(dibutylamino)methylene]-2'-deoxyadenosine from  $N^6$ -*N'*, *N'*-[(dibutylamino)methylene]-2'-deoxyadenosine; i) DMTCl, pyridine, DMAP; ii) acetic anhydride, DMAP,  $\text{Et}_3\text{N}$ , anhydrous DMF; iii) 40 % AcOH



The synthesis required consideration of which hydroxy protecting groups to use that would enable protection of the 3'-position whilst leaving the 5'-position unprotected. This was necessary for the synthesis of the desired dimer. Also, addition and removal conditions of the groups would have to be compatible with the dibutyl formamidine group as not to decompose it. This meant mild conditions during synthesis.

Initial protection of the 5'-position with DMTCI was deemed suitable because DMT is a bulky group which would allow controlled protection of 5'-position without risk of protecting 3'-position. This is due to 5'-position being a primary alcohol whilst the 3'-position is secondary and is also closer in proximity to the deoxyribose ring. DMT is removed quickly under mild acidic conditions therefore neutralisation of the reaction once the DMT group was removed would stop decomposition of the formamidine group that takes longer to occur. Acetylation of 3'-position was compatible with both groups because acetylation occurs under mild basic conditions whilst acetyl groups are stable under mild acidic conditions that are required to remove DMT.

**27** was reacted with DMTCI (1.2 eq.) for 18 hours. Analysis by silica TLC indicated complete reaction of starting material and formation of a new product. Purification by silica column chromatography eluted **28** (93 %).

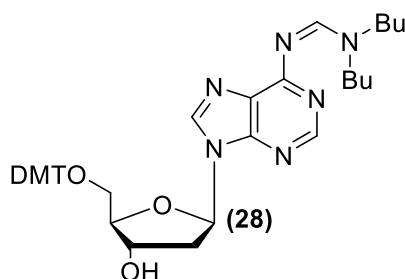


Figure 86. 5'-O-dimethoxytrityl-6-N', N'-[(dibutylamino)methylene]-2'-deoxyadenosine.

TLC analysis of the reaction of **28** with acetic anhydride after 30 minutes indicated the reaction was complete and **29** was eluted after purification by silica column chromatography (94 %).

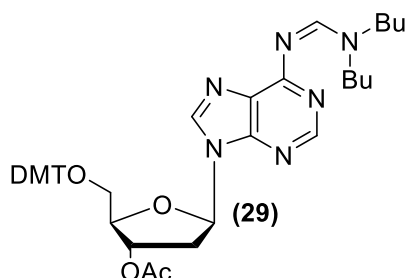


Figure 87. 5'-O-dimethoxytrityl-3'-O-acetyl-6-N', N'-[(dibutylamino)methylene]-2'-deoxyadenosine.

The final step required removal of DMT under mild acidic conditions. 40 % acetic acid was deemed suitable to achieve this whilst weak enough to not hydrolyse the formamidine. Previous studies showed that the formamidine group was stable under weak acidic conditions for at least three hours, therefore **29** was treated with 40 % acetic acid. After 1 hour, a sample was taken and neutralised with Et<sub>3</sub>N. TLC analysis of the sample indicated the reaction was complete. Neutralisation of the reaction, followed by silica column chromatography, gave **30** in 78 % yield. However, <sup>1</sup>H NMR in DMSO indicated that hydrolysis of the formamidine group to the formyl group had occurred (Figure 88).

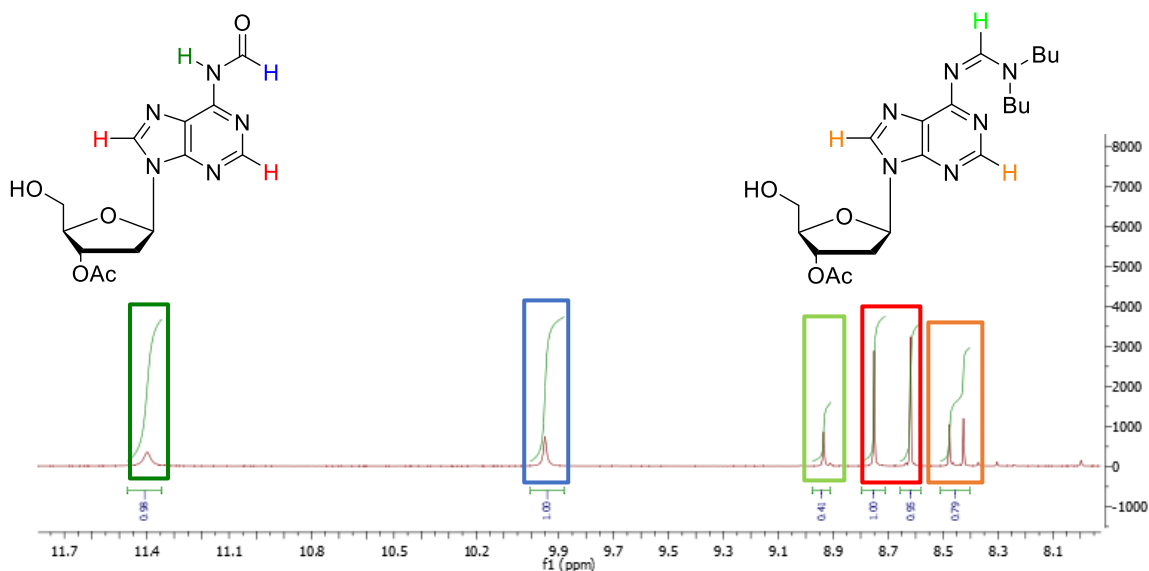


Figure 88.  $^1\text{H}$  NMR analysis of 3'-*O*-acetyl-6-*N'*, *N'*-[(dibutylamino)methylene]-2'-deoxyadenosine indicating presence of 3'-*O*-acetyl-6-formyl-2'-deoxyadenosine in DMSO.

Analysis of the spectrum indicates peaks at 8.42, 8.48 and 8.94 ppm correspond to the formamidine whilst peaks at 8.62, 8.75, 9.95 and 11.4 ppm indicate formyl formation. The formyl peaks correspond directly with  $^1\text{H}$  NMR data for  $N^6$ -formyladenosine<sup>108</sup>, as does the integration. The sample was analysed by TLC using the same conditions as before and analysis showed two compounds present. The sample was repurified by silica column chromatography and TLC analysis indicated two separate compounds were collected. However,  $^1\text{H}$  NMR analysis again showed decomposition of the formamidine.

The sample was purified again by silica column chromatography but this time  $\text{Et}_3\text{N}$  was not added to the solvent system. It had previously been added to neutralise the acidic silica that could hydrolyse the formamidine group. This alteration had no effect though because again  $^1\text{H}$  NMR analysis indicated hydrolysis (Figure 89).

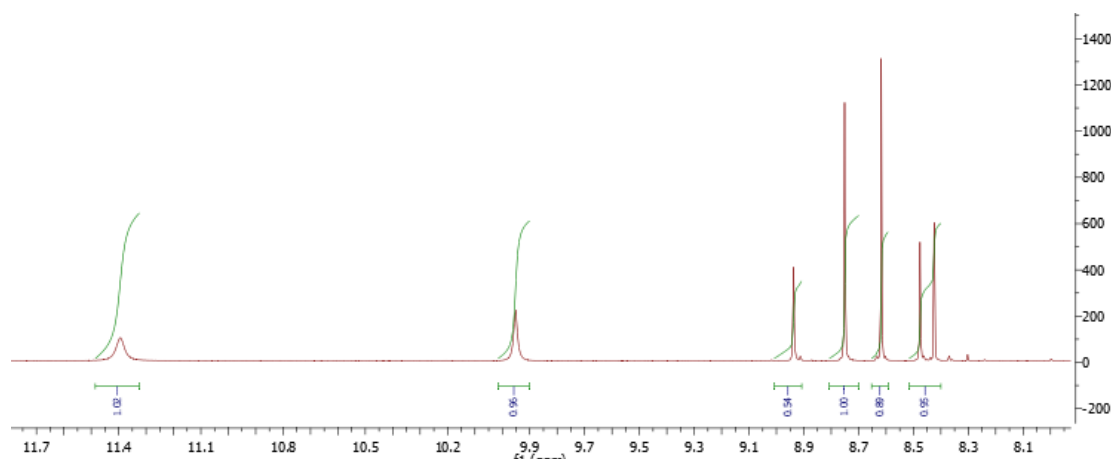


Figure 89.  $^1\text{H}$  NMR analysis of 3'-*O*-acetyl-6-*N'*, *N'*-[(dibutylamino)methylene]-2'-deoxyadenosine indicating presence of 3'-*O*-acetyl-6-formyl-2'-deoxyadenosine in  $\text{d}_6$ -DMSO.

It is unclear why this continued to occur. Firstly, if hydrolysis occurred due to the acidic conditions used to remove DMT, TLC analysis of the neutralised sample would indicate formation of two compounds. It only showed one. After neutralisation, hydrolysis could not occur because it required acidic conditions. Secondly, the deuterated solvent used for  $^1\text{H}$  NMR analysis was the same as for the analysis of **28** and **29**. Decomposition of the formamidine would have occurred analysing these two compounds if it was due to the deuterated solvent. Likewise, purification by silica column chromatography of **28** and **29** also added  $\text{Et}_3\text{N}$  to the solvent system. This was required to ensure the DMT group was not removed due to the acidity of silica.

The project was unable to move forward because of the uncertainty of whether the dimer containing the formamidine group could be synthesised and purified as a result of the complications accounted above. It would not have been cost effective to use the mixed sample for the dimer synthesis whilst time restraints meant restarting the synthesis of **30** was not feasible.

It was decided to purify the remaining ODN *AJ10* by treating it with 0.05 M K<sub>2</sub>CO<sub>3</sub>/ MeOH. Again RP-HPLC analysis indicated decomposition had occurred but the desired ODN was collected and confirmed by MS analysis.

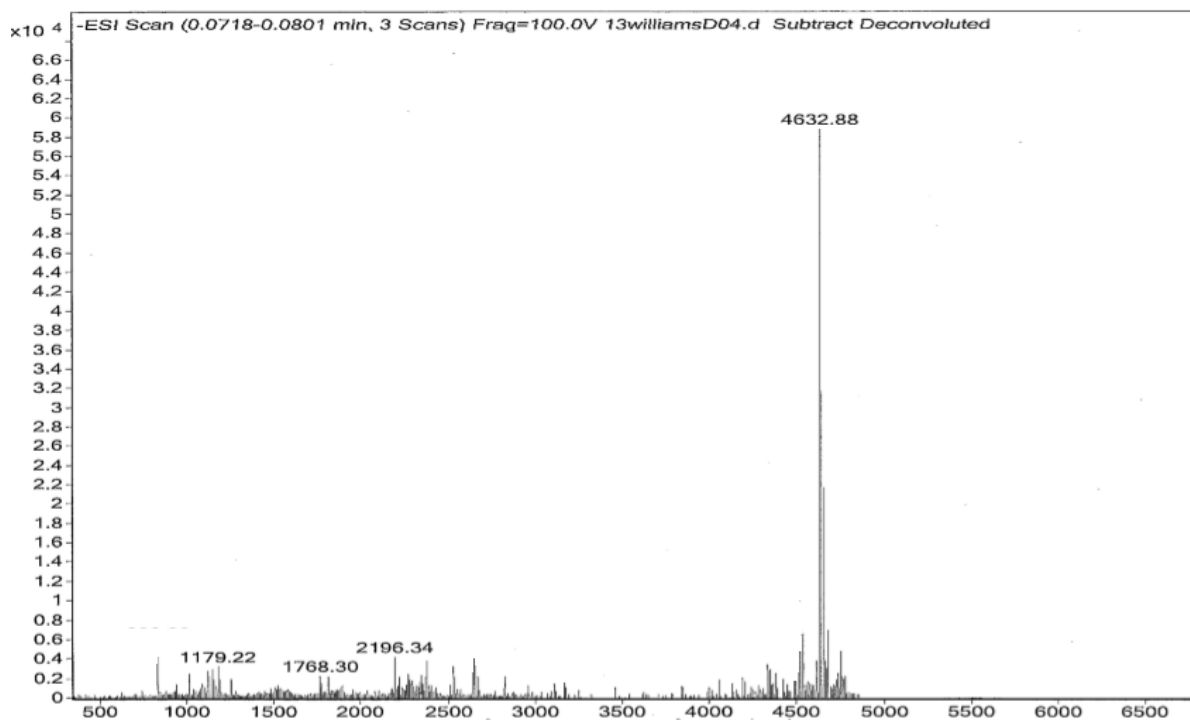


Figure 90. MS analysis of ODN *AJ10*. 4632.88 corresponds to sequence 5' CCC TCC AXT TGT TTG 3' containing the desired A<sup>dbf</sup> modification.

#### 4.5. Direct synthesis of formaldehyde-induced interstrand crosslink

Previous attempts to prepare ODNs containing formaldehyde-induced ICL did not come to fruition. Therefore, it was decided to attempt to synthesise the required ODN by reacting it directly with formaldehyde and hoping that a single crosslinked ODN could be isolated or purified from a mixture.

##### 4.5.1. Initial synthesis attempt

The synthesis of an ICL induced by aqueous formaldehyde followed the procedure laid out by Huang and Hopkins<sup>73</sup>. Two single strand ODNs, as outlined in the Table

13, were treated with aqueous formaldehyde (25 mM, 25 °C, pH 6.2, 50 mM sodium phosphate buffer, 25 mM sodium chloride) for 9 days. Both ODNs had 5' monophosphate to enable annealing to a larger ODN for repair mechanism studies.

Table 13. ODN sequences and their molecular weights used to react with aqueous formaldehyde in order to synthesise a formaldehyde-induced crosslink between two adenine bases plus the molecular weight of the desired ICL complex.

<i>Code</i>	<i>ODN Sequence</i>	<i>Molecular Weight/ gmol<sup>-1</sup></i>
<i>PK_FAI_JACS_V_Top</i>	5' pGCA CCA ACA ATT GTT G 3'	4945.2
<i>PK_FAI_JACS_V_Bot</i>	3' GTT GTT AAC AAC TCC Cp 5'	4896.2
<i>ICL Complex</i>	5' pGCAC CAACAATTGTTG 3' 3' GTTGTTAAACAAC TCCCp 5'	9853.4

The sequences were chosen for two reasons. Huang and Hopkins found that higher yields of a single fICL between two adenine bases were achieved when the sequence contained 5'-d(AATT) (73). This is due to fICLs forming between adjacent bases in the same groove which is because the methylene bridge is too short to span from one groove to another whilst no base contains two amino groups in the same groove. 5'-d(AATT) is highlighted in blue (Table 13). Secondly, to study the repair mechanism, the duplex is required to be ligated into a plasmid, therefore the sequences are not completely self-complementary but contain sticky ends to enable ligation.

Analysis for formation of an ICL was undertaken by using denaturing polyacrylamide gel electrophoresis (PAGE). Since PAGE separates nucleic acids principally on the basis of their size a covalently linked duplex is expected to have a

lower mobility (run slower) through the gel. The reaction mixture had two separate bands (Figure 91) which was a clear indication that there was formation of DNA structures with higher molecular weight than the single stranded ODNs. However, analysis by ESI MS showed a peak of mass 4816 which corresponds to the single strand ODN sequence 5' CCC TCA ACA ATT GTT G 3'. The literature yield for the ICL was < 1 %. Therefore, assuming the reaction yield here was similar, it is possible that preferential ionisation of the starting material, present in large excess, occurred in preference to the molecule with a greater mass.

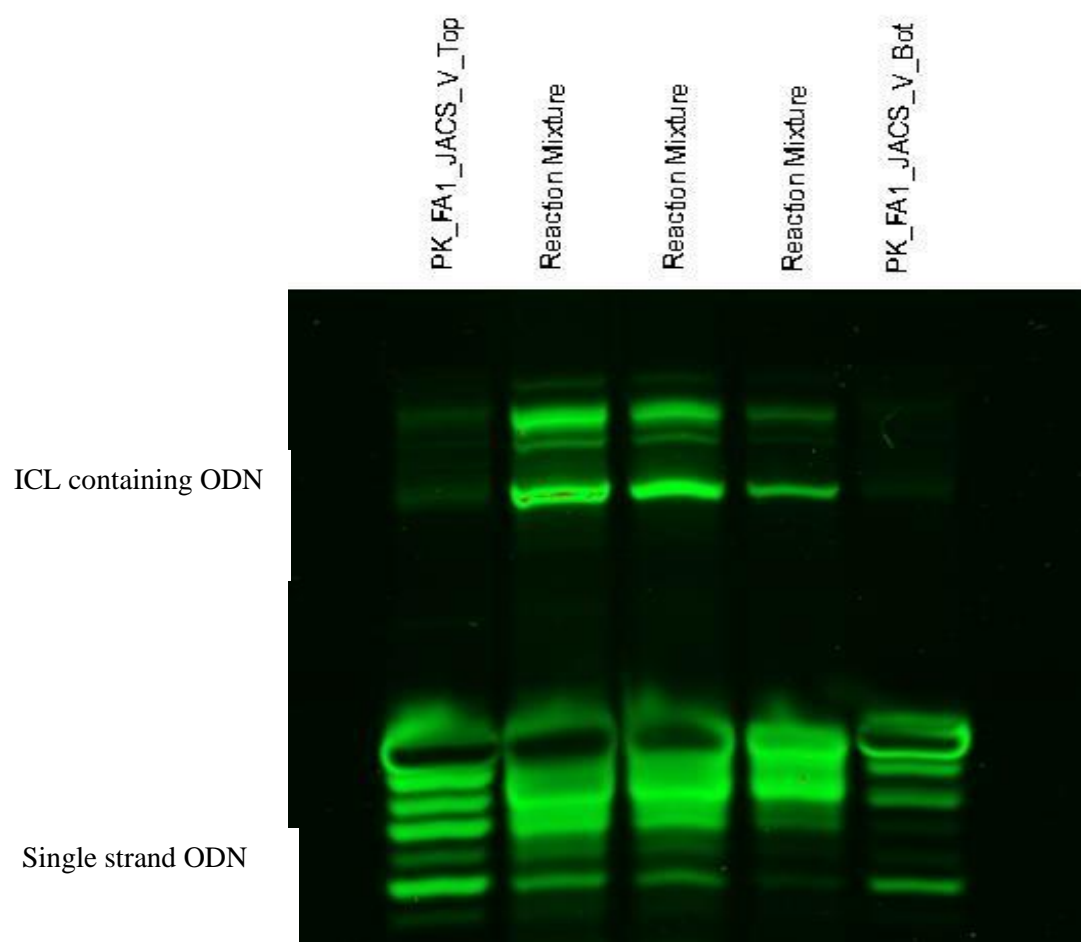


Figure 91. Denaturing PAGE analysis of the reaction between oligos 5' CCC TCA ACA ATT GTT 3' and 5' GCA CCA ACA ATT GTT G 3' with aqueous formaldehyde at pH 6.2. The reaction mixture was loaded in different amounts: Lane 2- 20 pmol; Lane 3- 10 pmol; Lane 4- 5 pmol. The two single strand ODNs were run as well. The slower running band is an oligo of higher molecular weight. 25 % polyacrylamide, 8 M urea.

In an attempt to separate the duplex from the single stranded ODN, the sample was analysed by RP-HPLC.

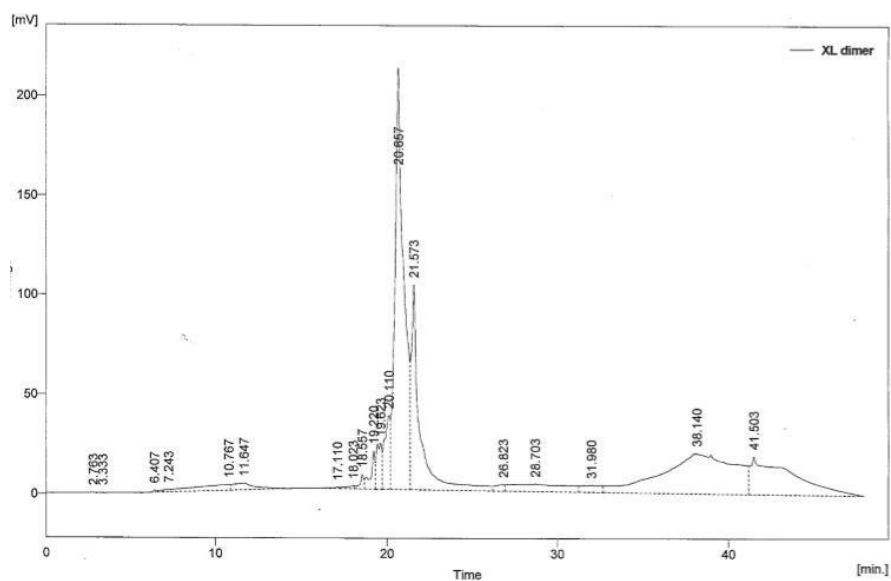


Figure 92. Reverse-phase HPLC trace of ODN duplex treated with aqueous formaldehyde. Four peaks were collected: 5-18 minutes, 18-20 minutes, 20-21.5 minutes and 21.5-25 minutes.

Four peaks were collected and analysed by denaturing PAGE. As seen in the Figure 93, one of the collected fractions had a slower running band and was therefore sent for analysis by ESI MS. Analysis by ESI MS showed a peak with mass 4877. This corresponds to a formaldehyde-induced intrastrand crosslink formed on single stranded sequence 5' GCA CCA ACA ATT GTT G 3'.



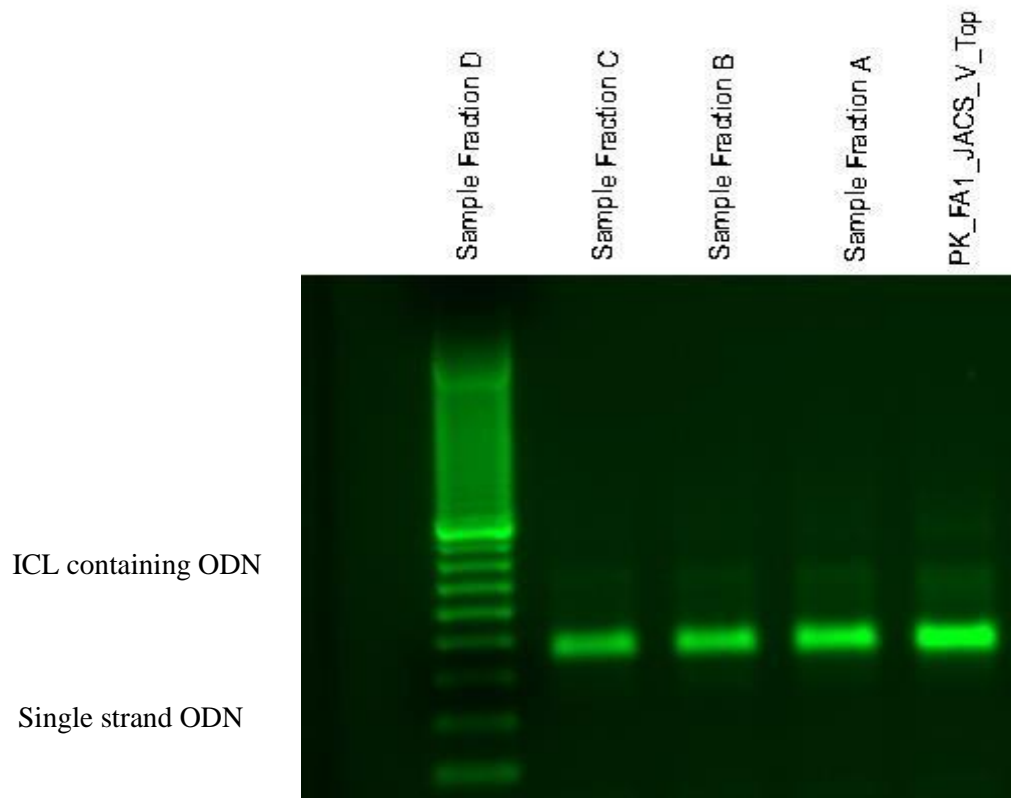


Figure 93. Denaturing PAGE analysis of the fractions collected from reverse-phase HPLC analysis of the reaction between ODNs 5' CCC TCA ACA ATT GTT 3' and 5' GCA CCA ACA ATT GTT G 3' with aqueous formaldehyde at pH 6.2 as well as the single stranded ODN PK\_FA1\_JACS\_V\_Top. The slower running band is an ODN of higher molecular weight. 25 % polyacrylamide, 8 M urea.

Purification of the slower running band using PAGE by our collaborator Michael Hodkinson (MRC Laboratory of Molecular Biology Cambridge) allowed analysis by MALDI MS.

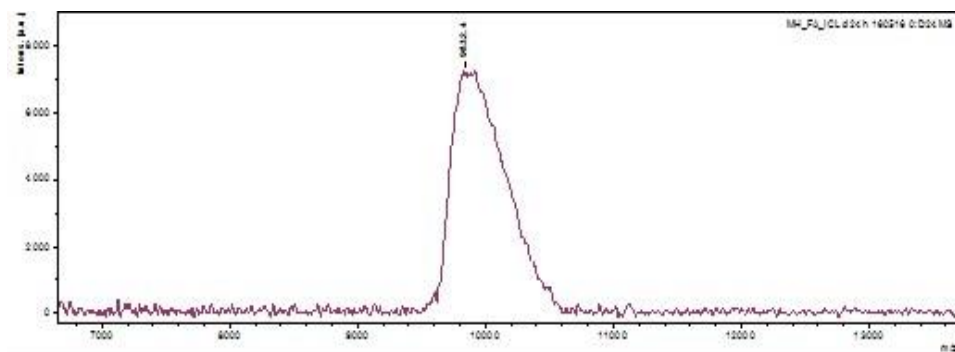


Figure 94. MALDI MS analysis of the gel purified crosslinked product showing mass of 9832.4 which corresponds to the self-annealed ODN sequence 5' CCC TCA ACA ATT GTT 3'.

The trace above shows a peak of mass 9832.4 which corresponds to the self-annealment of the ODN sequence PK\_FA2\_JACS\_V\_Bot and formation of  $N^6$ -CH<sub>2</sub>OH on an adenine base. Before the addition of aqueous formaldehyde to the ODNs, they were annealed at 95 °C which is where the self-annealment occurred. The sequences were adopted from those used by Huang and Hopkins<sup>73</sup> therefore it was not considered whether the ODNs would self-anneal. The Table 14 shows the possibility of both strands annealing to themselves.

Table 14. ODN complexes of the two ODN sequences showing that they could self-anneal and the respective molecular weights plus the molecular weight of the desired ICL complex.

<i>Code</i>	<i>Annealed ODN Sequence</i>	<i>Molecular Weight / gmol<sup>-1</sup></i>
<i>PK_FA1_JACS_V_Top</i>	5' GCA CCA ACA ATT GTT G 3' 3' GT TGT TTT AAC AAC CAC G 5'	9902.4
<i>PK_FA1_JACS_V_Bot</i>	5' CCC TCA ACA ATT GTT G 3' 3' GT TGT TAA CAA CTC CC 5'	9804.4
<i>ICL Complex</i>	5' GCA CCA ACA ATT GTT G 3' 3' GT TGT TAA CAA C TC CC 5'	9853.4

#### 4.5.2. Secondary synthesis route

Due to the formation of the crosslink between two strands of the ODN sequence PK\_FA1\_JACS\_V\_Bot, two new sequences were designed and treated with aqueous formaldehyde (25 mM, 25 °C, pH 6.2, 50 mM sodium phosphate buffer, 25 mM sodium chloride) for 9 days. The new sequences were modified from the previous sequences to prevent the ODNs self-annealing. Again, both ODNs had 5' monophosphate to enable annealment to a larger ODN for repair mechanism studies.

Table 15. ODN sequences and their molecular weights used to react with aqueous formaldehyde in order to synthesise a formaldehyde-induced crosslink between two adenosine bases plus the molecular weight of the desired ICL complex.

<i>Code</i>	<i>ODN Sequence</i>	<i>Molecular Weight / gmol<sup>-1</sup></i>
<i>PK_FA2_JACS_Vb_Top</i>	5' pGCA CCA AAC AAT TGG 3'	4650.0
<i>PK_FA2_JACS_Vb_Bot</i>	3' GTT TGT TAA CCT CCCp 5'	4573.9
<i>ICL Complex</i>	5' pGCA CCA AAC AAT TGG 3' 3' GT TTG TTA ACC TCCCp 5'	9235.9

As well as the reaction at pH 6.2, two further reactions were run, one at pH 7.0 and the other at pH 4.5. This was done to see whether the pH influenced the formation of an ICL. After 9 days, all three reactions were analysed by denaturing PAGE and all showed formation of a crosslinked, slower running species. However, the reaction that was run at pH 4.5 showed that some of the crosslink had decomposed whilst the reaction that was run at pH 7.0 appeared to not produce a significantly greater amount of crosslinked product than that formed at pH 6.2.

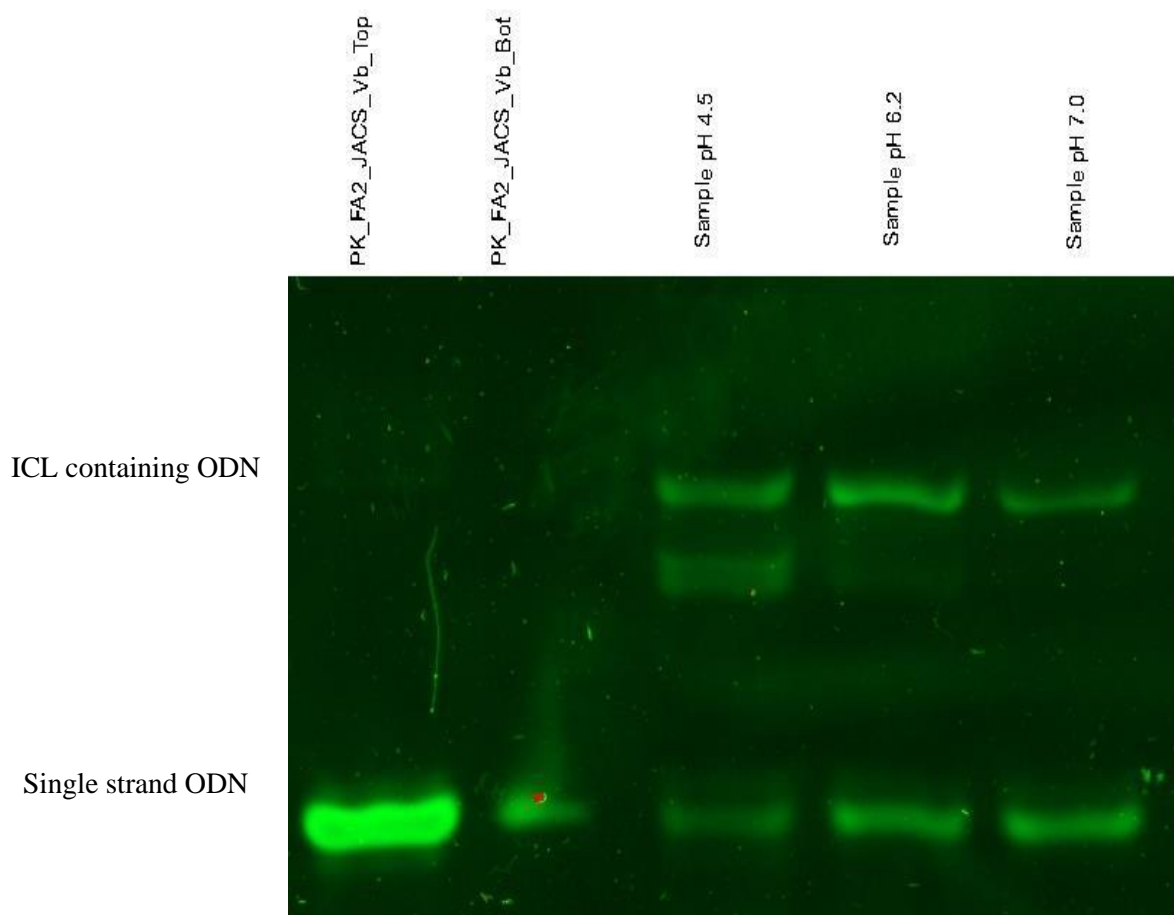


Figure 95. Denaturing PAGE analysis of reaction between ODNs 5' GCA CCA AAC AAT TGG 3' and 5' CCC TCC AAT TGT TTG 3' with aqueous formaldehyde at pH 4.5, 6.2 and 7.0 as well as the two single strands. The slower running band is an ODN of higher molecular weight.

Having been satisfied with the formation of a crosslinked species, a large-scale reaction (300 nmols) was run at pH 6.2 for 9 days and analysed by denaturing PAGE. Again, there was evidence of a crosslinked, slower running species.

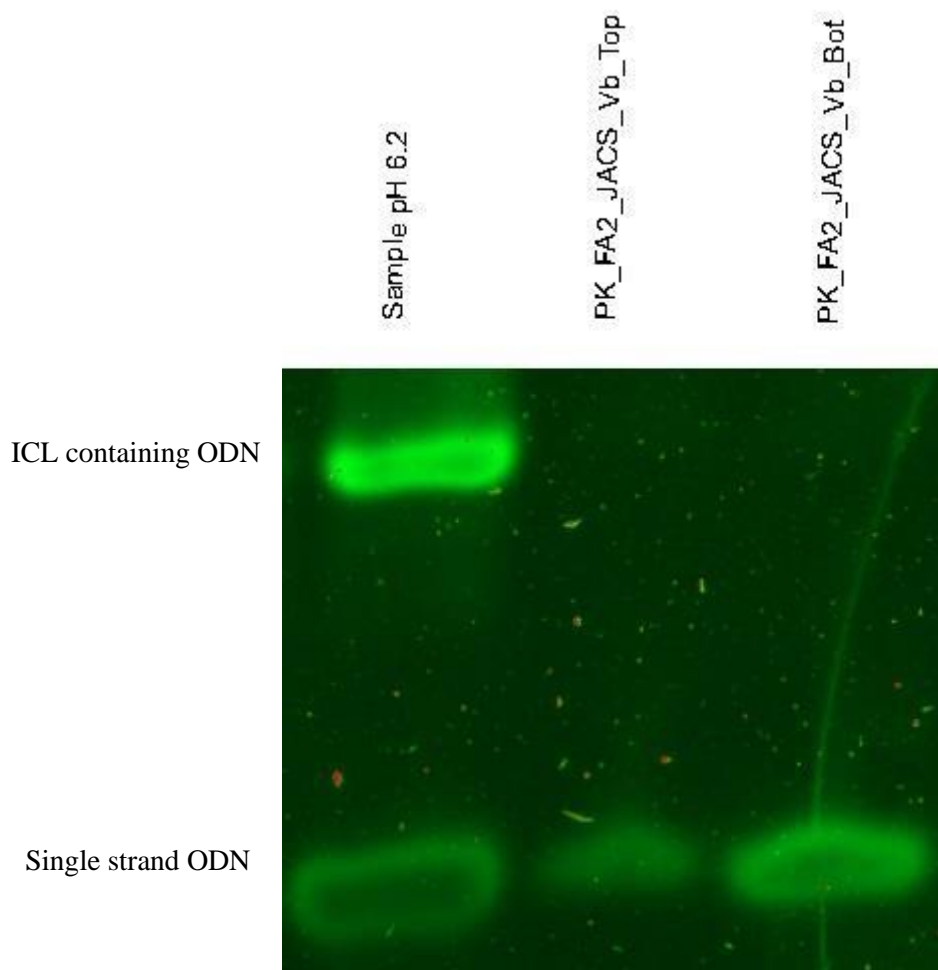


Figure 96. Denaturing PAGE analysis of reaction between ODNs 5' GCA CCA AAC AAT TGG 3' and 5' CCC TCC AAT TGT TTG 3' with aqueous formaldehyde at pH 6.2 as well as the two single strands. The slower running band is an ODN of higher molecular weight.

Due to the lack of success in analysing previous samples by ESI MS, it was decided to purify the new sample by reverse-phase HPLC prior to further ESI MS analysis. Isolation of the ICL-containing duplex would enable better ionisation during MS analysis. Using an AdvanceBio Oligonucleotide column, the ICL sample crosslinked at pH 6.2 was analysed using a gradient of acetonitrile in TEAA. The trace below shows the presence of the single strand oligo at 12.113 min and the larger molecule at 14.697, which is to be expected.

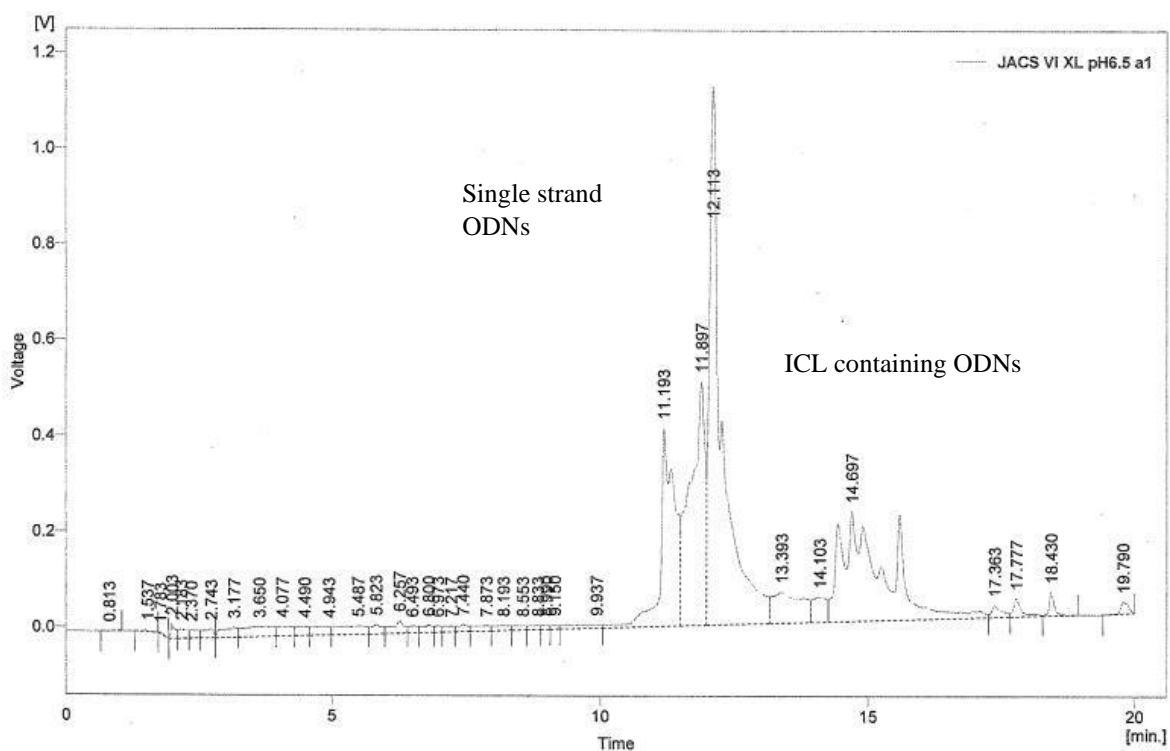


Figure 97. Analytical HPLC trace of large scale ICL reaction showing single stranded ODN and the ICL complex. Buffer A: 0.1 M TEAA, Buffer B: MeCN. 5 – 18.2 % B over 20 minutes.

Our collaborator Michael Hodkinson analysed the reaction mixture by MALDI MS and obtained a mass of 9232.8 which corresponds to  $[M-H]^-$  of the expected ICL complex.

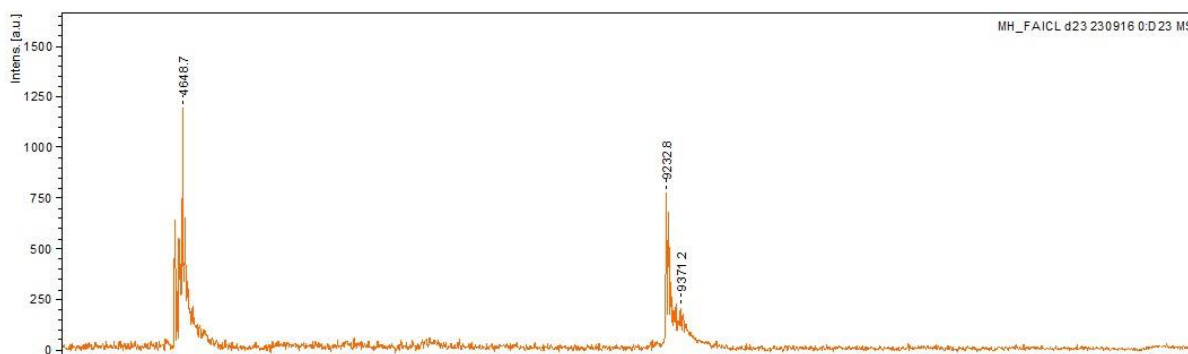


Figure 98. ESI MALDI analysis of the ICL complex by Michael Hodkinson. The mass of 4648 corresponds to the ODN PK\_FA2\_JACS\_Vb\_Top whilst the mass of 9232 corresponds to the desired ICL complex.

Purification of the sample was performed by RP-HPLC (Figure 99).

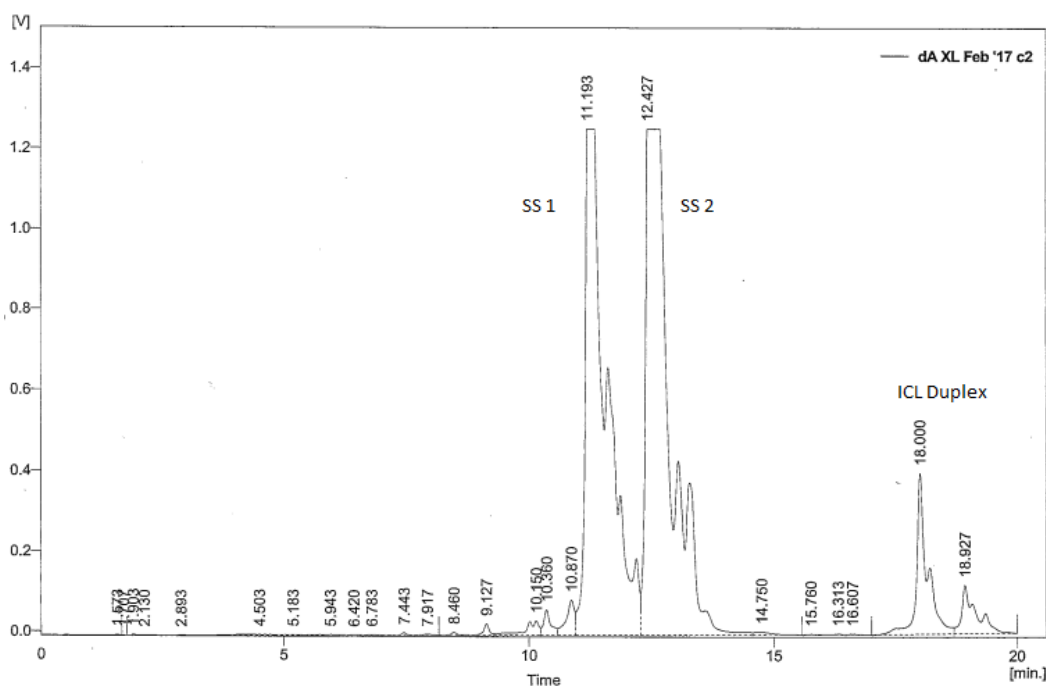


Figure 99. Analytical HPLC trace of large scale ICL reaction showing single stranded ODN and the ICL complex. Buffer A: 0.1 M TEAA, Buffer B: MeCN. 5 – 18.2 % B over 20 minutes at 60 °C.

Three peaks were collected and analysed. The first two peaks were unreacted single strands whilst ESI MS analysis of the third indicated formation of a methylene bridge between the two ODNs. This meant a successful synthesis but faced the same issues as Huang and Hopkins: the yield was 1 % and was dependent on sequence-specificity.

#### 4.5.3. Nucleoside composition analysis

Nucleoside composition analysis of the ODN was performed to confirm the formation of the fICL across two adenine bases. The ODN duplex was treated with DNA Degradase Plus™ for 1 hour at 37 °C. The sample was deactivated by heating at 70 °C for 20 minutes.

LC-MS analysis indicated that the dinucleoside *bis*-( $N^6$ -2'-deoxyadenosyl) methane ( $[M+H]^+ = 515.1$ ) was present at 16.30 minutes. The nucleosides dC, dG, T and dA are also detected whilst the other peaks are assigned to the digest mixture as well as fragments from incomplete digestion. The result confirms that the ODN duplex contained the fICL between two adenine bases on complementary strands.

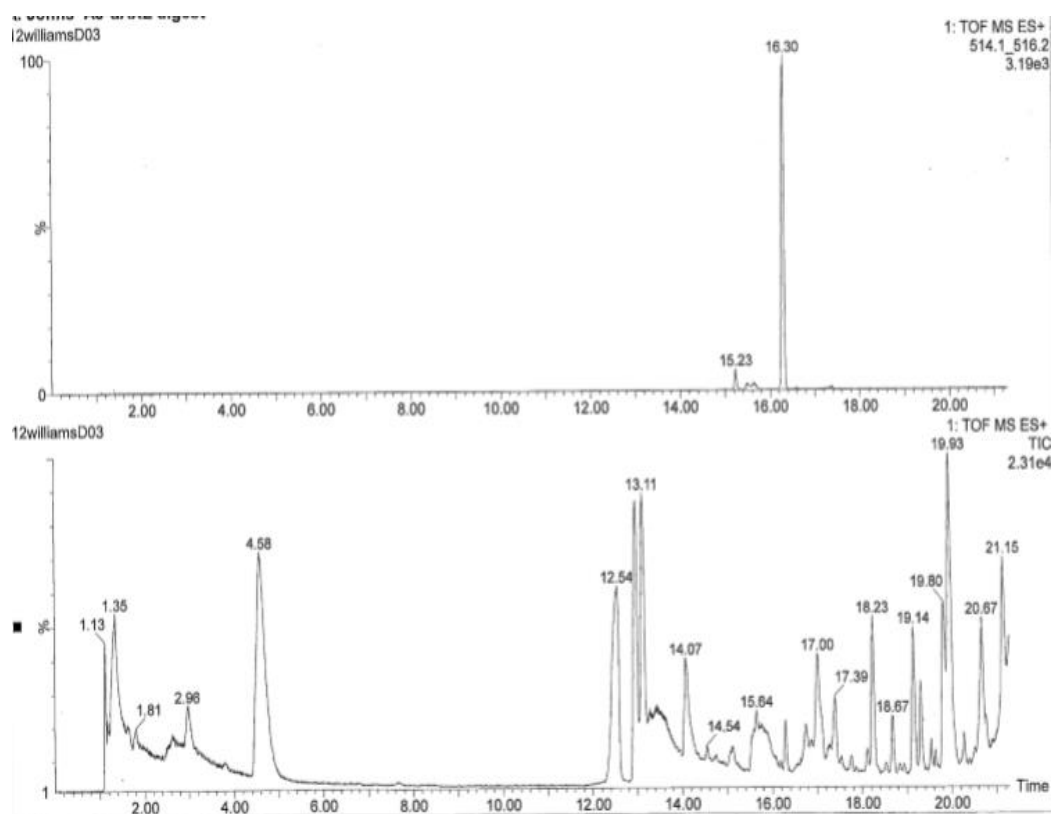


Figure 100. LC-MS analysis of the digest of the ODN duplex containing fICL. The top spectrum indicates the peak containing  $[M+H]^+ = 515.1$  which is attributed to *bis*-( $N^6$ -2'-deoxyadenosyl) methane. The bottom spectrum is the trace of the whole analysis indicating dC (4.58 mins), dA (12.54 mins), dG (12.96 mins) and T (13.11 mins). Buffer A-  $H_2O$ , buffer B- MeOH. 5- 80 % B over 25 mins.



#### 4.6. Formaldehyde-induced Interstrand Crosslink Within Drew-Dickerson Dodecamer

With the successful synthesis of an ODN containing a formaldehyde-induced ICL, it was envisaged that the crystal structure of the DNA modification could be acquired. This would give an insight into the effect such a modification has on the structure of  $\beta$ -DNA because it has not been reported in literature before.

The Drew-Dickerson dodecamer is a self-complementary ODN of sequence 5' CGC GAA TTC GCG 3' that acts as a model for B-DNA. It was first crystallised by Drew and Dickerson in 1981<sup>111</sup> and extensively in literature since then, providing very good details of crystallisation requirements and structural details. It was hypothesised that treatment of this ODN with formaldehyde would produce a single ICL between two central adenine bases and thus allow the crystal structure to be obtained. The ODN was also deemed suitable to use because it contains the AATT motif that Huang and Hopkins noted produced the greatest amount of crosslinked material when treated directly with formaldehyde (Figure 101).

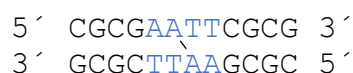


Figure 101. Drew-Dickerson duplex containing fICL between two adenine bases. The duplex contains the 5'-d(AATT) motif that Huang and Hopkins identified that encourages ICL formation, highlighted in blue.

Having acquired the Drew- Dickerson ODN, it was treated with formaldehyde (25 mM, 25 °C, pH 6.2, 50 mM sodium phosphate buffer, 25 mM sodium chloride) for 9 days. Denaturing PAGE analysis indicated the formation of a crosslinked, slower running species.

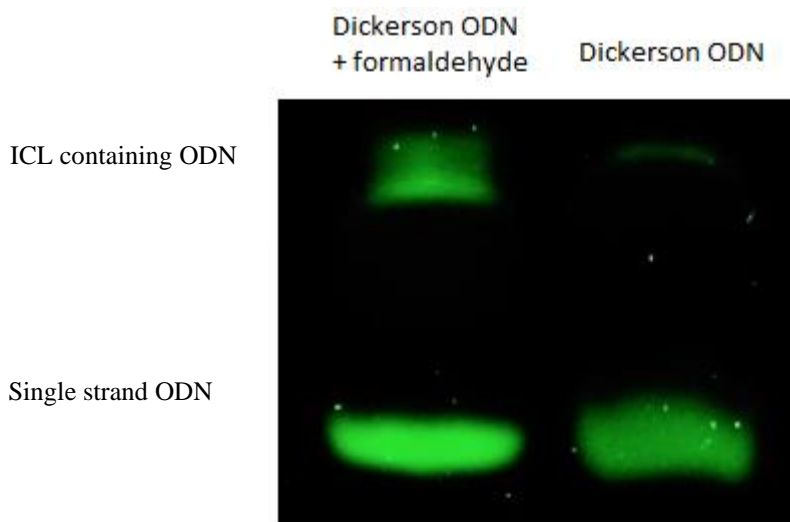


Figure 102. Denaturing PAGE analysis of reaction of ODN 5' CGC GAA TTC GCG 3' with aqueous formaldehyde at pH 6.2 as well as unreacted ODN 5' CGC GAA TTC GCG. The slower running band is an ODN of higher molecular weight.

The sample was analysed by RP-HPLC and four peaks were collected (22 mins, 29 mins, 32 mins and 42 mins) and analysed by MS.

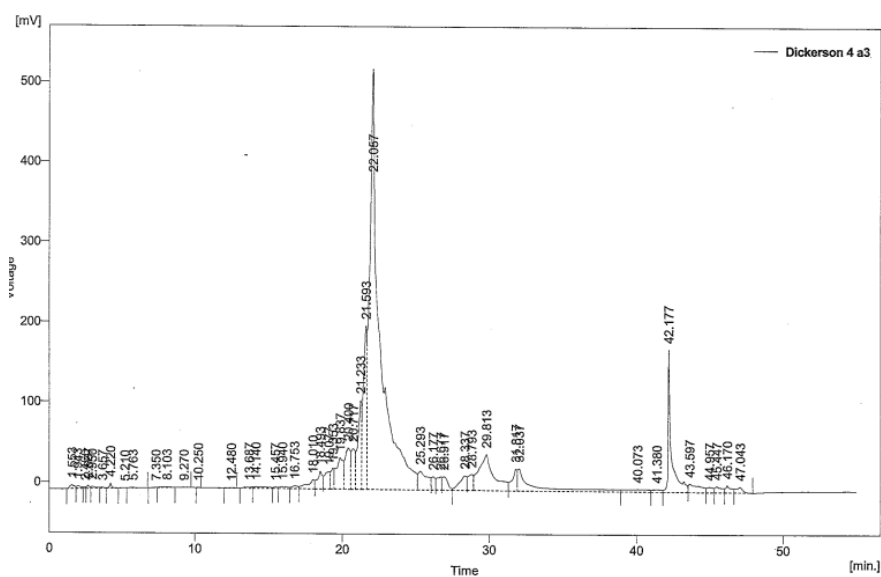


Figure 103. RP-HPLC trace of reaction of ODN 5' CGC GAA TTC GCG and aqueous formaldehyde. Buffer A: 0.1 M TEAA, Buffer B: MeCN, 5 – 30 % B over 50 minutes.

The major peak was identified as unmodified starting material whilst no other peak contained the desired mass of 7304.

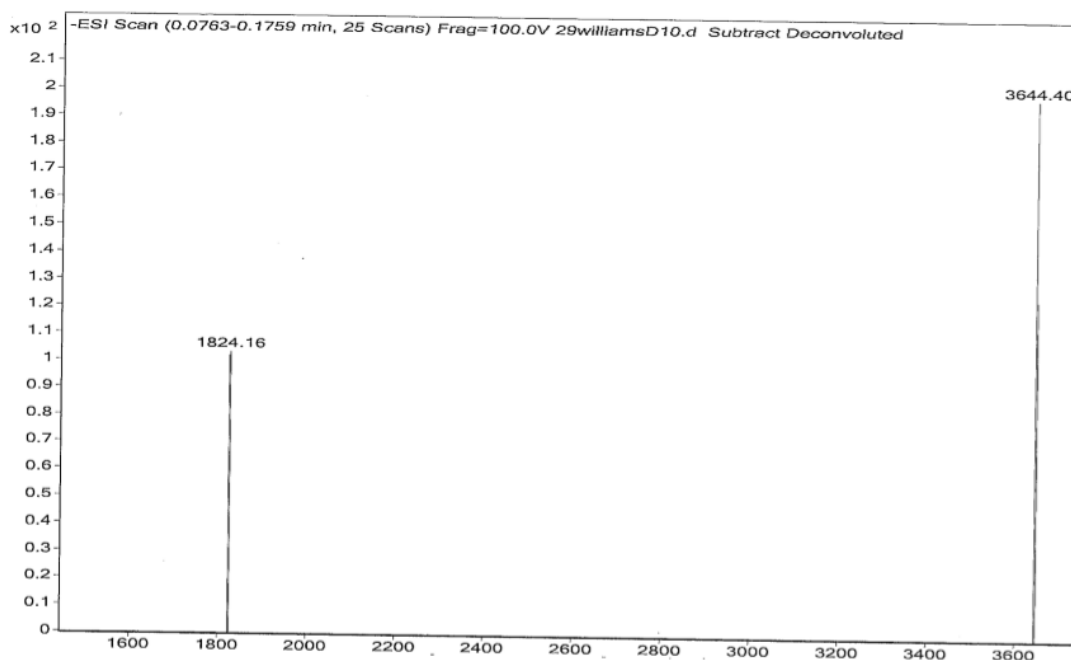


Figure 104. MS analysis of peak at 22.07 minutes from Figure 103. 3644.4 corresponds to sequence 5' CGC GAA TTC GCG 3'.

The sample was re-analysed and four samples collected. Samples were collected between 0- 16 mins; 16-25 mins; 25-35 mins and 35- 50 mins and submitted for MS analysis. The yield of the desired product was expected to be < 1 % therefore analysis of the whole spectrum at separate intervals would identify the region where the product was and enable its peak to be collected through a narrowing-down process. However, none of the four samples provided the expected mass nor masses in the expected range. In order to rule out ionisation issues experienced previously with MS, more in-depth analysis was carried out in the regions between 25-50 mins. This was because the crosslinked ODN is expected to elute later than the single-strand ODN, as seen previously with the synthesis of *AJ09*. Unfortunately, again no desired mass was obtained.

Numerous studies have proven that the mechanism of the reaction between an exocyclic amino group and formaldehyde in double-stranded DNA requires initial base unstacking, a fluctuation in the DNA<sup>71, 112</sup>. The very nature of the Drew-

Dickerson dodecamer, self-complementary B-DNA with Watson-Crick base-pairing, makes base unstacking very unlikely. Therefore, an explanation for the above result is that minimal fICL formation occurred, at levels undetectable by RP-HPLC, due to the nature of the ODN used. Moreover, McGhee and von Hippel reported that a reaction temperature below the melting temperature ( $T_m$ ) of the B-DNA duplex was required to initiate any potential reaction with formaldehyde<sup>70</sup>. The calculated  $T_m$  of the ODN was 40 °C whilst the reaction was undertaken at room temperature. However, this did not improve the yield of the reaction.

For future work, after analysis by denaturing PAGE, cutting out the band relating to the slower running species will enable analysis of it which will determine whether a single-sited fICL between two adenine bases has been synthesised and therefore enabling the determination of the crystal structure.

#### **4.7. <sup>1</sup>H NMR stability study of *bis*-(*N*<sup>6</sup>-2'-deoxyadenosyl) methane and *N*<sup>6</sup>-hydroxymethyl-2'-deoxyadenosine**

##### 4.7.1. Introduction

Stability studies of formaldehyde-induced DNA-protein crosslinks have been carried out<sup>113</sup> but none have studied the stability of either *bis*-(*N*<sup>6</sup>-2'-deoxyadenosyl) methane or *N*<sup>6</sup>-hydroxymethyl-2'-deoxyadenosine (*N*<sup>6</sup>-hydroxymethyl dA) (Figure 105). Acquiring such results would give valuable insight into the risk posed by these modifications and the necessity of their repair.

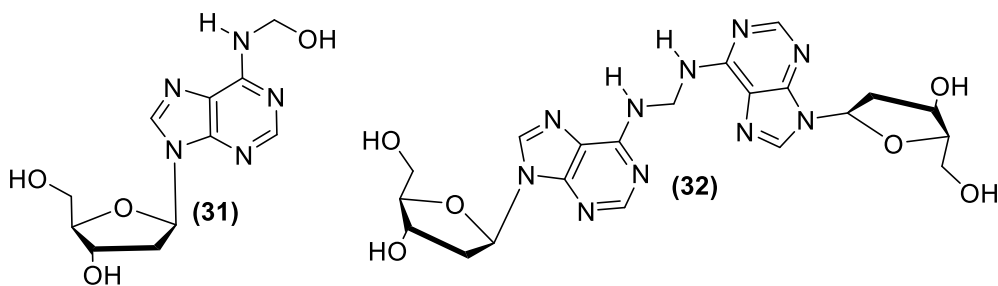


Figure 105. Structures of  $N^6$ -hydroxymethyl-2'-deoxyadenosine and *bis*-( $N^6$ -2'-deoxyadenosyl) methane.

Formation of  $N^6$ -hydroxymethyl dA upon treatment of 2'-deoxyadenosine with formaldehyde is known to be reversible<sup>72, 114</sup>, however, it has been detected in leukocyte DNA in 29 out of 32 heavy smokers<sup>115</sup> indicating its potential danger. *Bis*-( $N^6$ -2'-deoxyadenosyl) methane is known to irreversibly form when  $N^6$ -hydroxymethyl dA is heated in the presence of dA<sup>72</sup>, however, its stability under biological conditions has not been established. <sup>1</sup>H NMR analysis of the compounds under biological conditions would indicate their stability by analysing whether the compounds decomposed to dA. This would be obtained by studying the <sup>1</sup>H NMR peaks.

#### 4.7.2. Synthesis of $N^6$ -hydroxymethyl-2'-deoxyadenosine and *bis*-( $N^6$ -2'-deoxyadenosyl) methane

Chaw *et al.* reacted adenosine with 3.7 % aqueous formaldehyde solution for one day, forming  $N^6$ -hydroxymethyladenosine, then heated the solution to 50 °C and stirred for a further nine days, resulting in its conversion to *bis*-( $N^6$ -2'-adenosyl) methane<sup>72</sup>. Adapting this protocol, 2'-deoxyadenosine was reacted with 3.7 % aqueous formaldehyde solution for one day and then heated to 50 °C and stirred for a further nine days. After removal of formaldehyde by lyophilisation, the crude sample was purified by RP-HPLC (Figure 106).

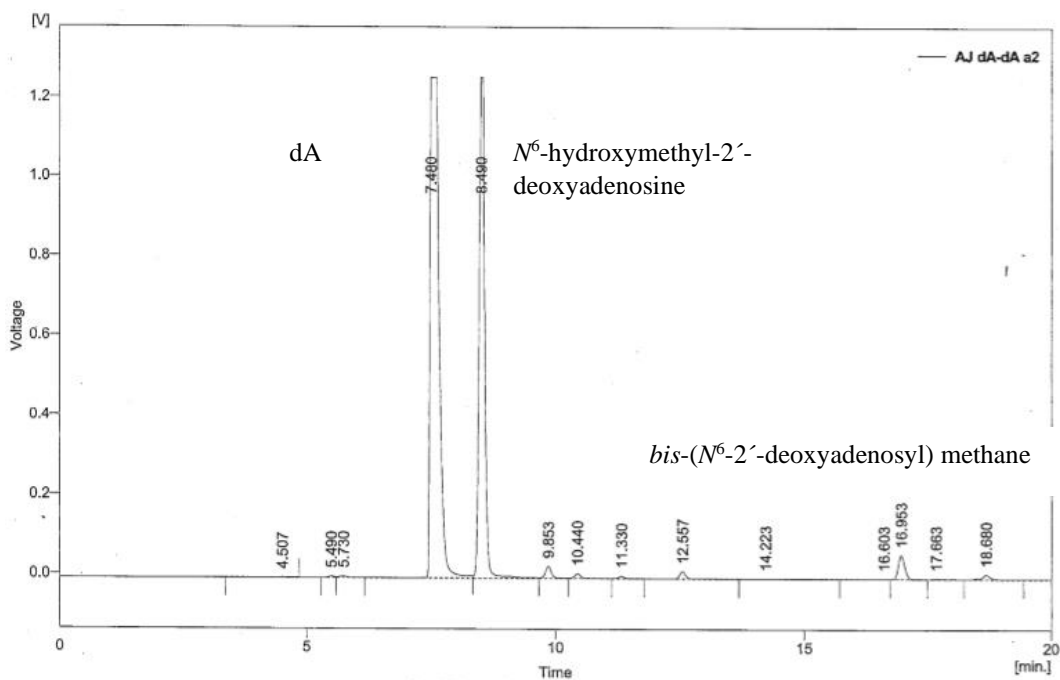


Figure 106. RP-HPLC trace of reaction 2'-deoxyadenosine and aqueous formaldehyde at 50 °C for nine days. Peak at 7.48 mins corresponds to dA, peak at 8.49 mins corresponds to *N*<sup>6</sup>-hydroxymethyl-2'-deoxyadenosine and peak at 16.95 mins corresponds to *bis*-(*N*<sup>6</sup>-2'-deoxyadenosyl) methane. Buffer A: H<sub>2</sub>O, Buffer B: MeCN. 5 – 30 % B over 20 minutes.

The RP-HPLC trace shows that three compounds were present in the crude mixture: 2'-deoxyadenosine, *N*<sup>6</sup>-hydroxymethyl dA and *bis*-(*N*<sup>6</sup>-2'-deoxyadenosyl) methane. The two desired compounds were collected separately, and the solvent removed by lyophilisation. ESI<sup>+</sup> MS analysis gave masses of 282 (*N*<sup>6</sup>-hydroxymethyl dA) and 515 (*bis*-(*N*<sup>6</sup>-2'-deoxyadenosyl) methane) indicating successful purification. Both samples were then dissolved in aqueous phosphate buffer (pH 7.1), in order to replicate biological conditions, which was then removed by lyophilisation and, finally, the resulting solids were dissolved in D<sub>2</sub>O, ready for analysis by <sup>1</sup>H NMR.

#### 4.7.3. <sup>1</sup>H NMR analysis of *N*<sup>6</sup>-hydroxymethyl-2'-deoxyadenosine

The sample of *N*<sup>6</sup>-hydroxymethyl dA contained residue which was assigned as phosphate buffer. The initial analysis indicated that decomposition to dA was already occurring (Figure 107). Signals at 8.36, 8.34 and 6.51 ppm correspond, respectively to

H8, H2 and H1' protons of *N*<sup>6</sup>-hydroxymethyl-2'-deoxyadenosine whilst signals at 8.32, 8.24 and 6.49 ppm correspond to these protons on 2'-deoxyadenosine.

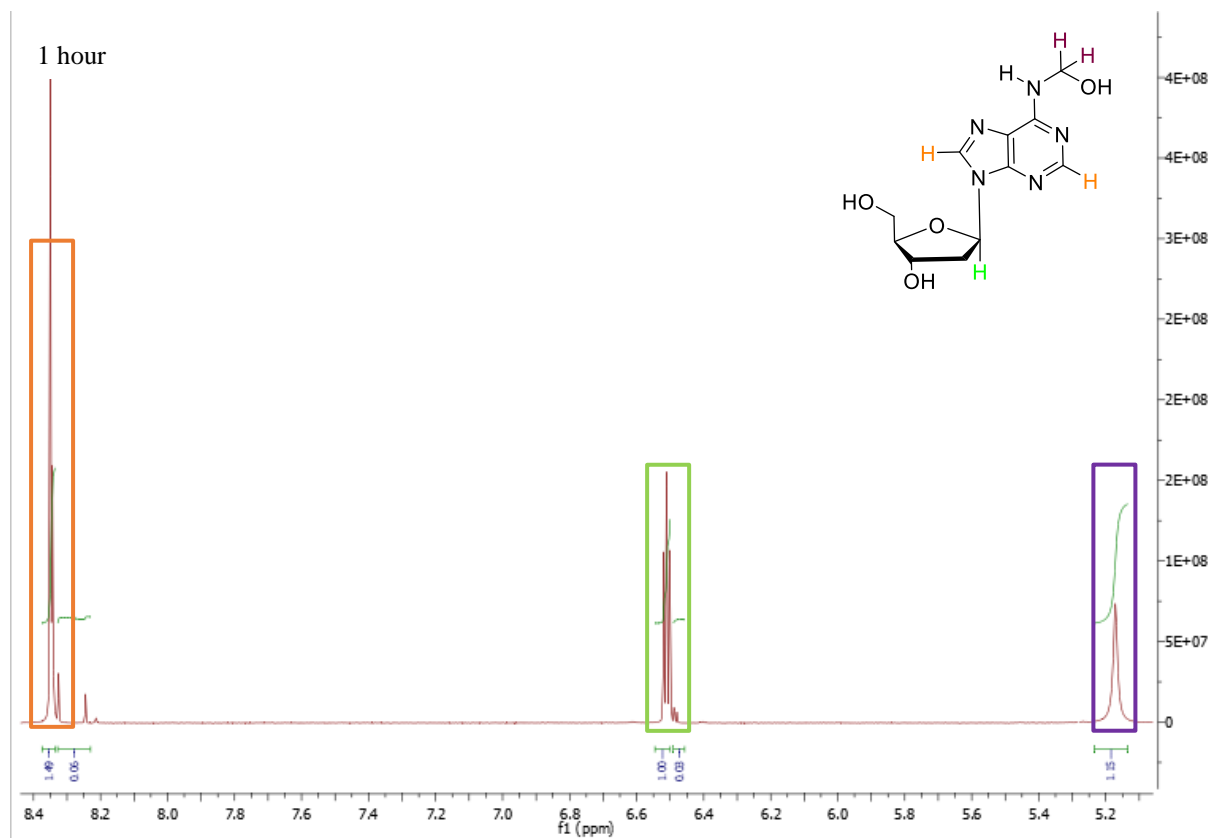


Figure 107. <sup>1</sup>H NMR analysis of *N*<sup>6</sup>-hydroxymethyl-2'-deoxyadenosine in phosphate buffered D<sub>2</sub>O, pH 7.1, after 1 hour.

Further analysis indicated that after a week, full decomposition to dA was almost complete (Figure 108). Integration of the signals corresponding to 2'-deoxyadenosine against signals corresponding to *N*<sup>6</sup>-hydroxymethyl-2'-adenosine indicate 3: 1 ratio of compounds.

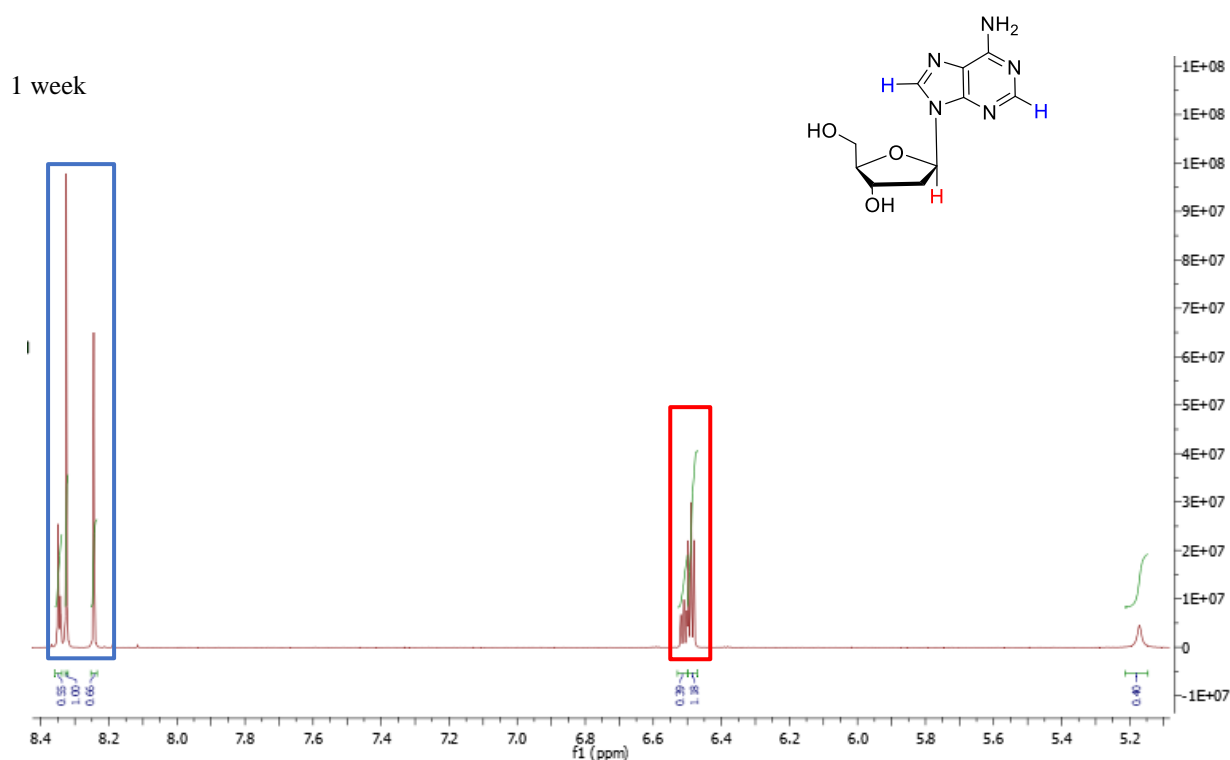
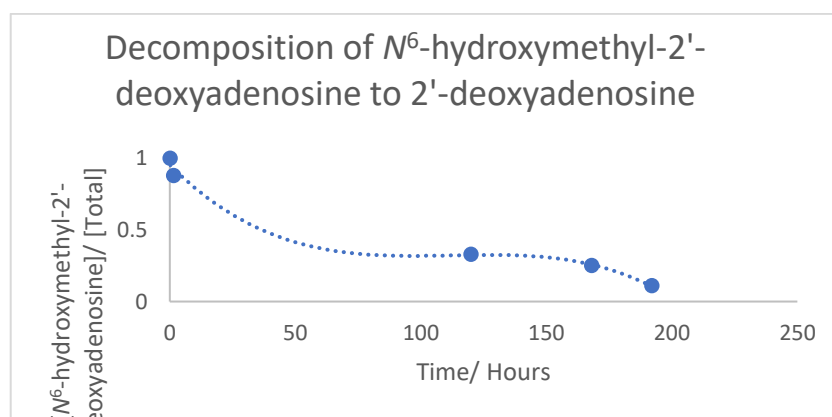


Figure 108.  $^1\text{H}$  NMR analysis of  $N^6$ -hydroxymethyl-2'-deoxyadenosine in phosphate buffered  $\text{D}_2\text{O}$ , pH 7.1 after seven days. The compound has decomposed to 2'-deoxyadenosine.

The half-life ( $t_{1/2}$ ) of the decomposition of  $N^6$ -hydroxymethyl-2'-deoxyadenosine was calculated from the  $^1\text{H}$  NMR data collected by plotting  $[\text{N}^6\text{-hydroxymethyl-2'-deoxyadenosine}] / [\text{total}]$  against time (graph 1). This was calculated as  $t_{1/2} = 36.5$  hours.



Graph 1. Data plotting  $[\text{N}^6\text{-hydroxymethyl-2'-deoxyadenosine}] / [\text{total}]$  against time (hours) to calculate  $t_{1/2}$  of the decomposition of  $N^6$ -hydroxy-2'-deoxyadeonsine to 2'-deoxyadenosine.



This finding indicates that unrepaired  $N^6$ -hydroxymethyl dA is a potential risk because its stability means there is a greater chance of ICLs forming.  $N^6$ -hydroxymethyl dA is the intermediate between non-crosslinked and crosslinked dA therefore the longer it exists, the greater chance there is of crosslink formation.

$^1\text{H}$  NMR analysis did record the proton shift of  $-\text{NHCH}_2\text{OH}$  that enabled differentiation from the methylene bridge of *bis*-( $N^6$ -2'-deoxyadenosyl) methane. The former appears at 5.17 ppm whilst the latter appears at 5.46 ppm. This was important in order to identify the presence of these compounds by NMR.

#### 4.7.4. $^1\text{H}$ NMR analysis of *bis*-( $N^6$ -2'-deoxyadenosyl) methane

Initial  $^1\text{H}$  NMR analysis of *bis*-( $N^6$ -2'-deoxyadenosyl) methane indicated that the compound was stable under biological conditions (Figure 109). The data obtained replicated previous  $^1\text{H}$  NMR data of the compound<sup>72</sup> whilst  $^1\text{H}$  COSY analysis confirmed the identification of the methylene bridge at 5.46 ppm (Figure 110), confirming a pure sample suitable for stability analysis. There is no recorded literature on the stability of *bis*-( $N^6$ -2'-deoxyadenosyl) methane.

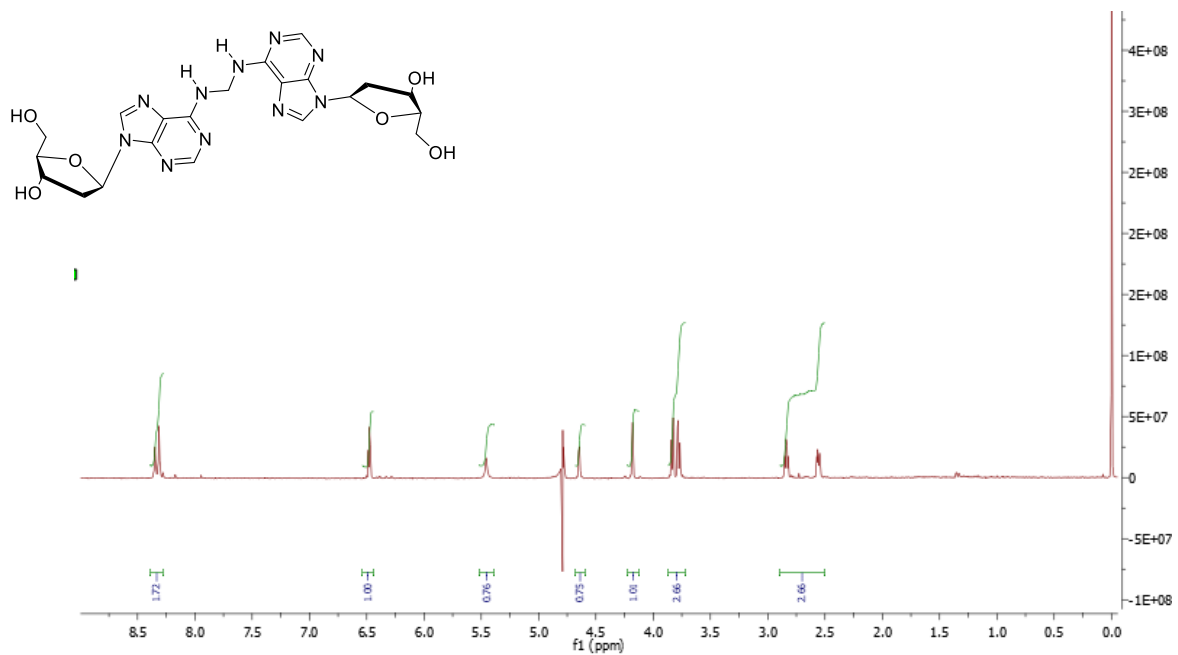


Figure 109.  $^1\text{H}$  NMR analysis of *bis*-( $N^6$ -2'-deoxyadenosyl) methane in phosphate buffer, pH 7.1.

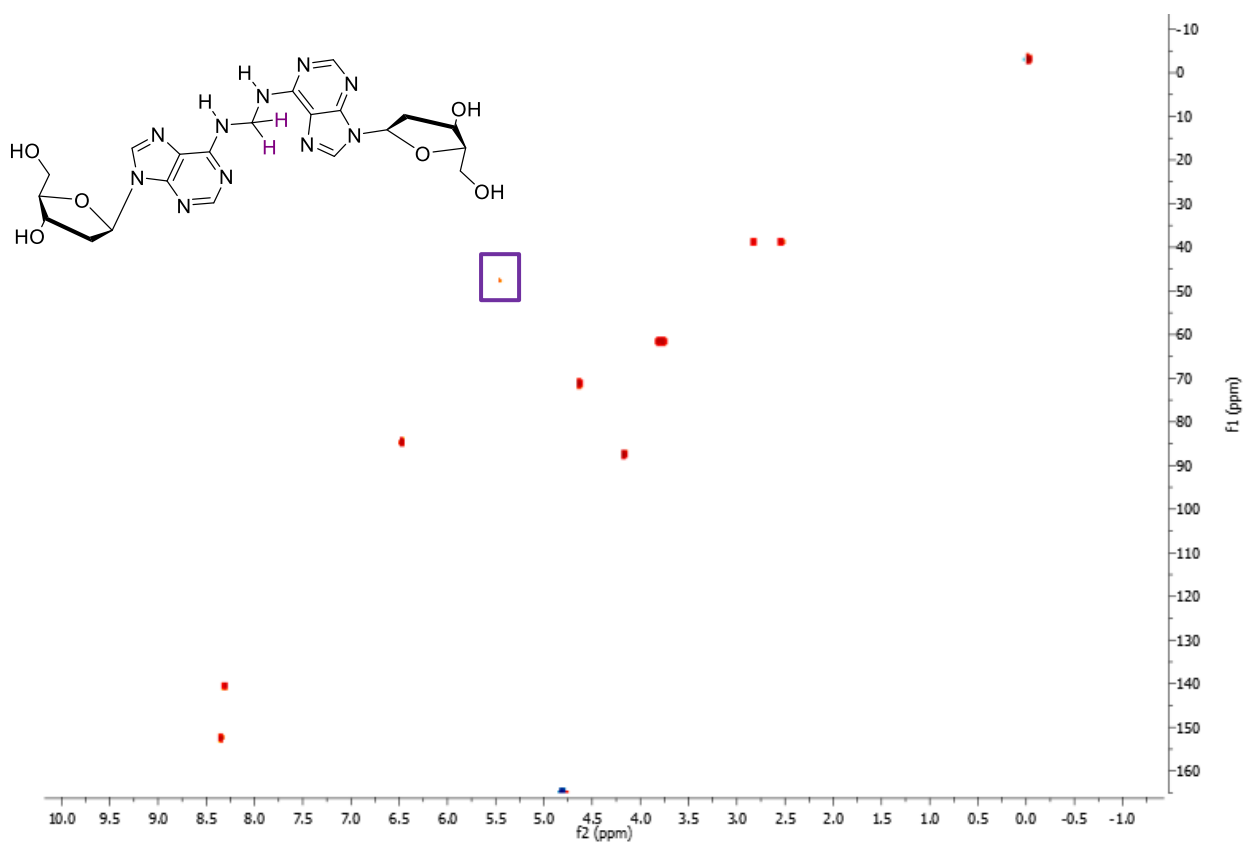


Figure 110.  $^1\text{H}$  COSY of *bis*-( $N^6$ -2'-deoxyadenosyl) methane.

However, residue was noted in the sample. This is thought to be phosphate buffer that had not gone into solution. 2'-Deoxyadenosine is soluble in aqueous conditions so would not precipitate and would therefore be identified by <sup>1</sup>H NMR. Also, MS analysis of the sample indicated no decomposition had occurred (Figure 111).

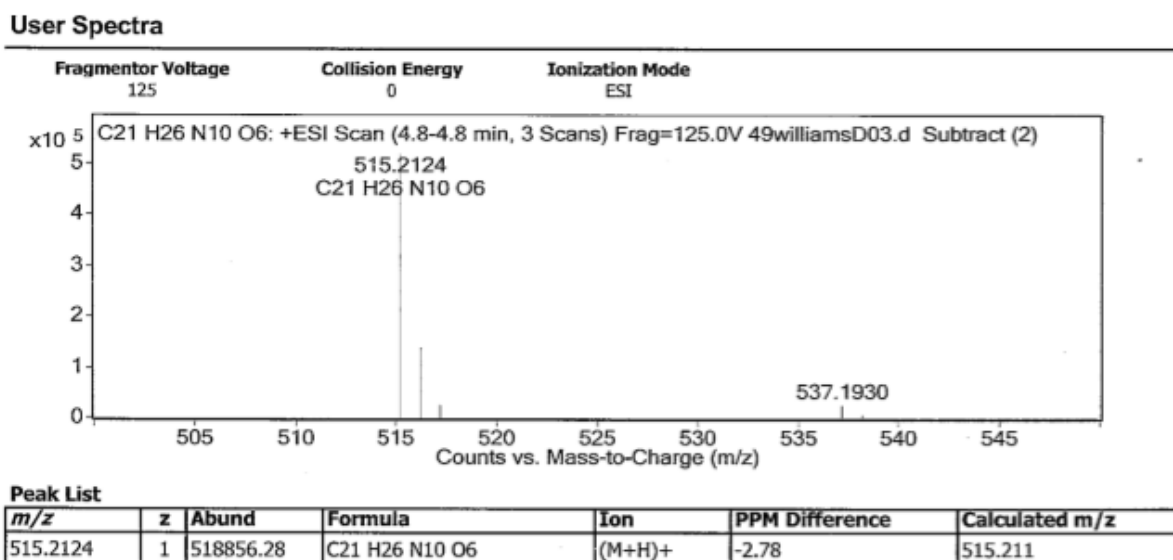


Figure 111. Accurate mass analysis of *bis*-(*N*<sup>6</sup>-2'-deoxyadeonsyl) methane in phosphate buffer, pH 7.1.

Further stability analyses were carried out over six weeks (Figure 112).

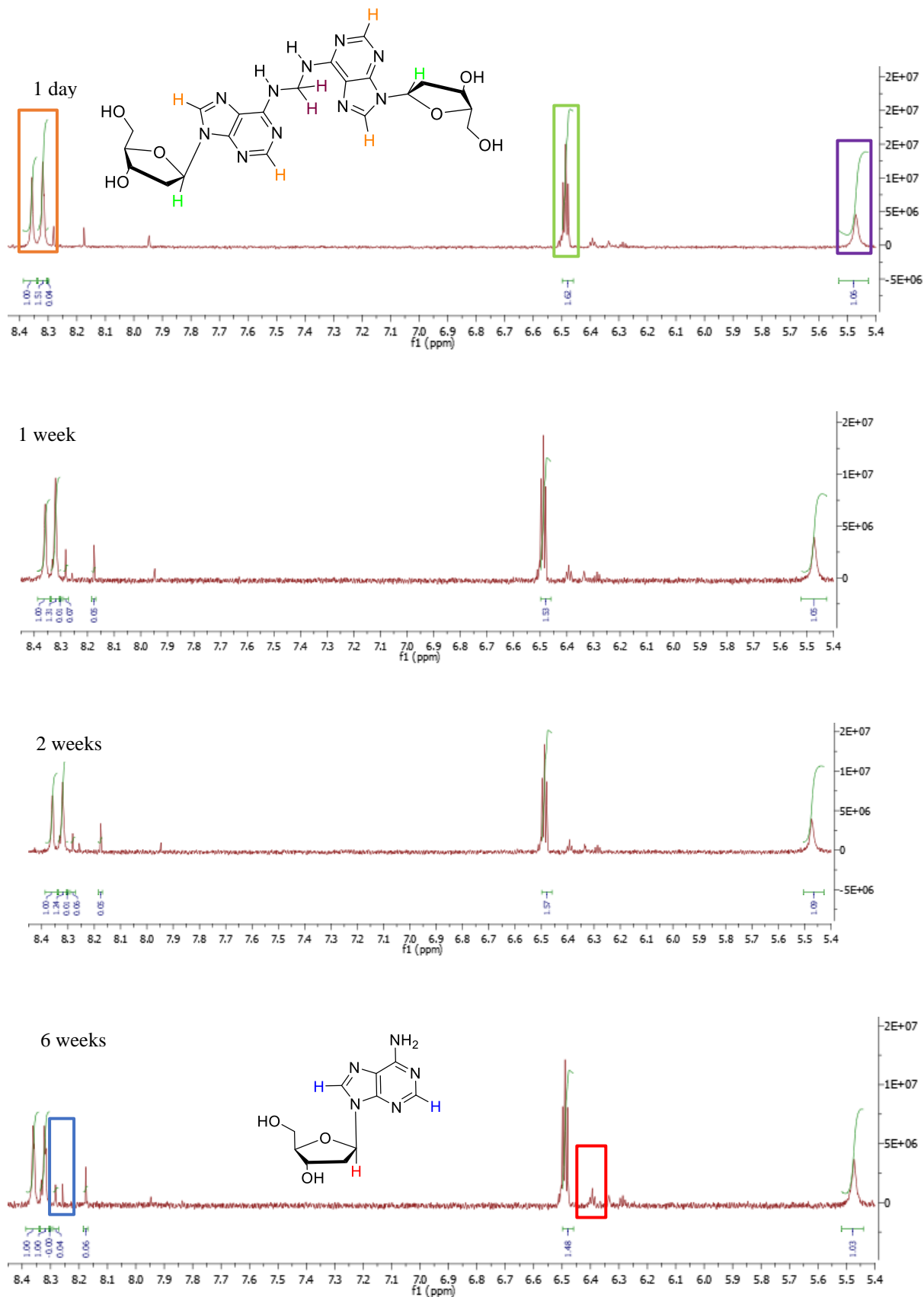


Figure 112.  $^1\text{H}$  NMR analysis of bis-( $N^6$ -2'-deoxyadeosyl) methane in phosphate buffer, pH 7.1. Spectra were collected at 1 day, 1 week, 2 weeks and 6 weeks and indicate integration of  $\text{H}_8$ ,  $\text{H}_2$  and  $\text{H}_{1'}$  protons between bis-( $N^6$ -2'-deoxyadeosyl) methane and 2'-deoxyadenosine.

Integration of peaks corresponding to H8, H2 and H1' indicated that, over six weeks, around 10 % decomposition occurred. There was no indication of formation of *N*<sup>6</sup>-hydroxymethyl dA, which would have been noted by the appearance of -NHCH<sub>2</sub>OH at 5.17 ppm. This was expected because our previous study showed rapid decomposition of *N*<sup>6</sup>-hydroxymethyl dA to dA. It was noted that over the six weeks, precipitation of the sample increased. Again, this is not attributed to formation of dA though explains the deterioration of the quality of <sup>1</sup>H NMR spectrum because less sample was in solution.

The <sup>1</sup>H NMR study of *bis*-(*N*<sup>6</sup>-2'-deoxydeonsyl) methane indicated that the dinucleoside was stable and underwent little decomposition under biological conditions. This means that such crosslinks pose a significant threat to the health of cells if unrepaired because the adducts will remain in place, hence hindering DNA replication and resulting in cell apoptosis. It further confirms the importance of synthesising ODNs containing formaldehyde-induced crosslinks to better understand the repair process and identify ways to resolve problems faced by Fanconi Anemia patients.



## **Chapter 5- *Conclusions and Future Work***

## 5. Conclusions and Future Work

### 5.1. Conclusions

ODNs containing  $O^6$ -CMG and  $O^6$ -CEG were synthesised to investigate the repair of these adducts by MGMT. The ODN containing  $O^6$ -CEG was synthesised after the initial, novel synthesis of  $O^6$ -CEG phosphoramidite which was achieved by adapting the procedure of Millington *et al.*

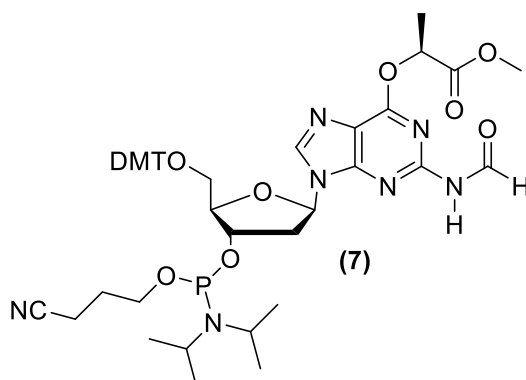


Figure 113. Synthesised phosphoramidite of  $O^6$ -CEG.

TBDMS protection of the 3' and 5' hydroxyl groups (1) was followed by mesitylation at the O6 position (2). Methyl(-)-lactate displaced the mesitylene group *via* a DABCO salt (3). Methyl(-)-lactate ensured that the correct, S enantiomer of  $O^6$ -CEG was introduced to the compound to replicate the stereochemistry of the naturally occurring  $O^6$ -alkylG adduct. Formyl protection of the N2 position followed *via* reaction with *N,N*-dimethylformamide dimethyl acetal and the subsequent hydrolysis of the formamidine group to the formyl group (4). This conversion ensured cleaner purifications for later steps because it has been noted that incomplete conversion of the formamidine to the formyl can occur, lowering the overall yield of the synthesis. The hydroxyl protecting groups were removed by TBAF in THF (5) before the dimethoxytrityl group was introduced at the 5'-hydroxyl position (6). The synthesis of



the phosphoramidite was completed with the introduction of 2-cyanoethyl-*N*-diisopropylamine chlorophosphoramidite at the 3'-hydroxyl position.

The conditions for the deprotection of ODNs containing *O*<sup>6</sup>-CMG and *O*<sup>6</sup>-CEG were verified.

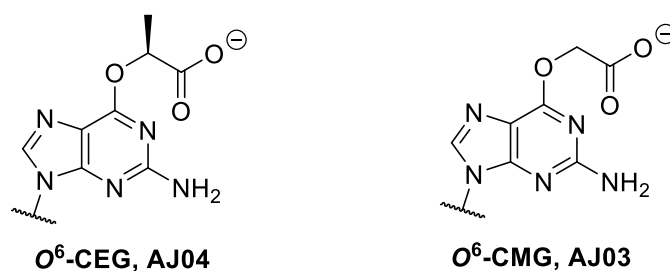


Figure 114. Structures of *O*<sup>6</sup>-CEG and *O*<sup>6</sup>-CMG and their anointed ODN code.

Following DNA synthesis, the ODNs containing *O*<sup>6</sup>-CMG and *O*<sup>6</sup>-CEG were treated with aq 0.5 M NaOH for two days to ensure hydrolysis of the methyl ester group to the carboxylate and then, subsequently, treated with conc. aq ammonia to remove the base-labile protecting groups. MS analysis of the purified ODNs gave the expected masses for the carboxylate adducts. Further characterisation was achieved by the direct treatment of the ODNs containing *O*<sup>6</sup>-CMG and *O*<sup>6</sup>-CEG with conc. aq ammonia. MS analysis gave the expected masses for the corresponding carboxamide adducts, a mass unit one fewer than the carboxylate-containing adducts.

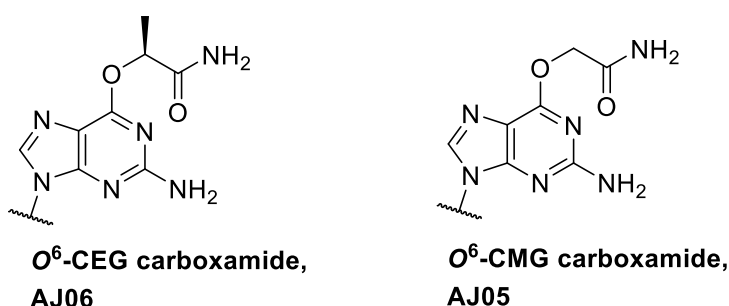


Figure 115. Structures of *O*<sup>6</sup>-CEG carboxamide and *O*<sup>6</sup>-CMG carboxamide and their anointed ODN code.

Further evidence for characterising these adducts in each of the ODNs was achieved by nucleoside composition analysis by LC-MS following enzymatic digestion.

The  $IC_{50}$  calculated from radioactive MGMT assays for **AJ03** and **AJ04** indicated that both  $O^6$ -CMG and  $O^6$ -CEG are substrates of MGMT. The  $IC_{50}$  for  $O^6$ -CMG is 133 nM and 72.1 nM for  $O^6$ -CEG. The ODNs containing the carboxamide groups showed greater affinity to MGMT than their carboxylate counterparts which may be explained due to the negative charge of the carboxylate. This may repel the negatively charged thiolate group of Cysteine that removes the  $O^6$ -alkyl group from guanine.

Through the results gathered above, it is indicated that previous data published by Senthong *et al.* was incorrectly assigned to an ODN containing  $O^6$ -CMG carboxamide rather than the desired  $O^6$ -CMG. However, consistent with this work is the finding that  $O^6$ -CMG is a substrate for MGMT.

This project also developed a high-yielding synthesis of four  $^{15}N_5$ -labelled  $O^6$ -alkylG SIL internal standards for the development of SID methods to identify and quantify these adducts in DNA samples. The synthesis was achieved by adapting protocols by Tintoré *et al.* and Janeba *et al.*

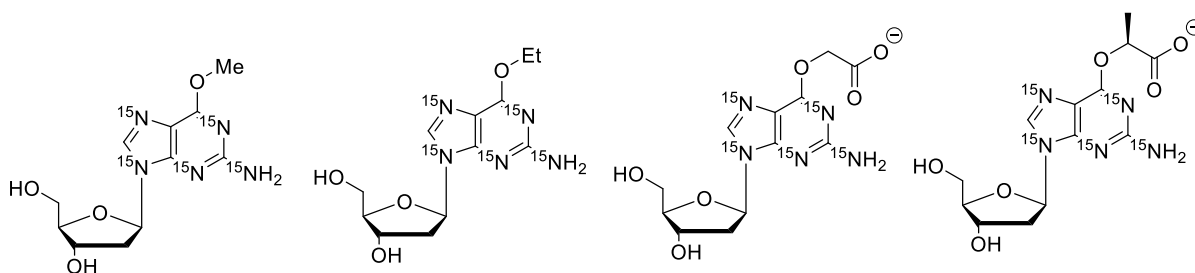


Figure 116. Structures of  $^{15}N_5$ -labelled  $O^6$ -MeG,  $O^6$ -EtG,  $O^6$ -CMG and  $O^6$ -CEG.

Acetylation of 3' and 5' hydroxy groups was fast, efficient and high yielding, likewise the mesitylation of the O6 position. The alkyl modification was introduced in the third step, after activation of the O6 position by formation of DABCO salt, *via*

nucleophilic displacement. During this final step, the acetyl groups were removed either by the basic nature of the nucleophile or by addition of aq 0.5 M NaOH. These steps enabled a simplified route without the need for chromatography that ensured minimal loss of yield.

Using  $^{15}\text{N}_5$ -labelled dG as starting material was expensive but SID methods require SIL internal standard amounts as low as  $1 \text{ fmol } \mu\text{L}^{-1}$ . This enabled the synthesis of numerous  $O^6$ -alkylguanine adducts in abundance therefore enabling future examination of numerous genotoxic agents. No other synthesis of stable-isotopically labelled  $O^6$ -alkylguanines is as effective or precise.

A SID method using  $^{15}\text{N}_5$ -labelled  $O^6$ -MeG has been developed with a calibration coefficient of 0.99999. The LOQ was calculated as 100 fmol which is greater than the LOQ calculated by Brink *et al.* for their  $O^6$ -MeG SIL internal standard SID method. However, our result was an initial result without optimisation of the UHPLC system.

In the final aspect of this project, an ODN duplex containing a single methylene bridge between two adenine bases on opposite strands was achieved by reacting a sequence-specific duplex directly with aqueous formaldehyde. This was done in order to better understand the repair mechanism. After synthesis of the fICL-containing duplex, nucleoside composition analysis, following enzymatic digest, confirmed the presence of the *bis*-( $N^6$ -2'-deoxyadenosyl) methane dinucleoside.

However, the yield of the synthesis was 1 % which meant it was not an efficient manner of synthesising fICLs between adenine bases. ODNs containing 6-chloropurine were purified and reacted with methylene diamine. However, the instability of methylene diamine voided the reaction because it decomposed to methylamine.

The Drew-Dickerson dodecamer ODN was reacted with aqueous formaldehyde in order to obtain a crystal structure of the fICL between two adenine bases. However, the desired ODN was not isolated. This is due to the dodecamer being self-complementary which inhibits the necessary base unstacking that initiates the reaction between an exocyclic amino group and formaldehyde in double-stranded DNA.

Attempts to hydrolyse a formamidine group at the N6 position of adenine to a formyl group in aq acid were unsuccessful. Initially, hydrolysis of *N*<sup>6</sup>-[(dimethylamino)methylene]-2'-deoxyadenosine resulted in 2'-deoxyadenosine. Hydrolysis of *N*<sup>6</sup>-[(dibutylamino)methylene]-2'-deoxyadenosine (**27**) under aq acid conditions resulted in depurination.

An ODN containing **27** was deprotected and purified. However, hydrolysis of the formamidine group to the amino group occurred during the deprotection resulting in low yields.

<sup>1</sup>H NMR analysis determined the stability *t*<sub>1/2</sub> of *N*<sup>6</sup>-hydroxymethyl-2'-deoxyadenosine (**31**) to be 36.5 hours under biological conditions. **31** is the intermediate species prior to fICL formation therefore the result indicates that fICL formation is likely. The stability of *bis*-(*N*<sup>6</sup>-2'-deoxyadenosyl) methane dinucleoside (**32**) under biological conditions was also analysed. After six weeks, around 10 % decomposition to dA had occurred. This indicates the importance of the repair of this DNA damage.

## 5.2. Future Work

### 5.2.1. Detection of *O*<sup>6</sup>-CEG in human DNA

Despite the abundance of dietary alanine, *O*<sup>6</sup>-CEG has yet to be detected in human DNA. This means that the risk posed by large amounts of dietary alanine is unclear, as

well as the importance behind it being a substrate of MGMT. It may be that its repair by the enzyme enhances protection for human CRC. Analysing the reaction between 2'-deoxyguanosine and diazo alanine, formed from *N*-nitrosation of alanine, by using *O*<sup>6</sup>-CEG as an internal standard would indicate whether a reaction occurs.

### 5.2.2. Development of SID methods for <sup>15</sup>N<sub>5</sub>-labelled SID internal standards

The initial development of a SID method using <sup>15</sup>N<sub>5</sub>-labelled *O*<sup>6</sup>-MeG as an internal standard is encouraging. However, for the method to be viable, the LOQ needs to be lowered. This requires extensive time developing and perfecting the UHPLC system using techniques such as column switching and lowering the flow rate *via* nanoflow LC.

With the successful calibration of the <sup>15</sup>N<sub>5</sub>-labelled *O*<sup>6</sup>-MeG internal standard, quantification of unknown ODN samples can be initiated. Though the current LOQ hinders quantification of large DNA samples, blind-testing short-chained ODN samples would enable further evidence of the application of the SID method and ensure greater accuracy.

SID methods can also be developed for the three other SIL *O*<sup>6</sup>-alkylG internal standards: *O*<sup>6</sup>-CMG, *O*<sup>6</sup>-CEG and *O*<sup>6</sup>-EtG. Initially, determination of their molar extinction coefficients is required in order to produce accurate concentrations required for method calibration.

### 5.2.3. Determination of the crystal structure of ODN containing fICL

Initial attempts to purify the Drew-Dickerson ODN containing a fICL by RP-HPLC were unsuccessful. Obtaining a crystal structure of a fICL between two adenine bases would provide vital structural information but requires significant amounts of ODN to

analyse. Another alternative would be to synthesise a Drew-Dickerson ODN with overhangs which would encourage reaction between the exocyclic amino group and formaldehyde due to the ability of the ODN to fluctuate.

#### 5.2.4. Development of antibodies for *bis*-( $N^6$ -2'-deoxyadenosyl) methane

With the successful synthesis and characterisation of *bis*-( $N^6$ -2'-deoxyadenosyl) methane, its antibodies can be synthesised. Having antibodies of this compound would enable quantification and characterisation in DNA samples using methods such as ELISA.

## **Chapter 6- *Experimental***

## 6. Experimental

All reagents were obtained from commercial suppliers and used without further purification. Dry solvents were obtained from the University of Sheffield Grubbs apparatus and all anhydrous reactions carried out in flame dried apparatus under N<sub>2</sub> using standard Schlenk techniques unless otherwise stated.

NMR spectra were recorded on a Bruker AV400/AV600 spectrometer (stated for each data set) and chemical shifts are reported in  $\delta$  values relative to tetramethylsilane as an external standard. J values are given in Hz.

Mass spectrometry was performed by the University of Sheffield Mass Spectrometry Service using the method of electrospray ionisation on a Water LCT Mass Spectrometer. Fragmentation voltage was 125.0 V, unless otherwise stated, and the concentration of samples was 1 mg/ mL.

Column chromatography was performed on silica gel for flash chromatography (30 – 70  $\mu$ m). Thin layer chromatography (TLC) was performed on pre-coated Merck silica gel 60 F254 aluminium backed plates. TLCs were visualised under UV (254 nm).

All modified ODNs were synthesised DMT-off using either 2-amino-6-methylsulfonyl-purine phosphoramidite, 2-amino-6-(methylglycolyloxy)purine phosphoramidite, 2-amino-6-(methyllactyloxy)purine phosphoramidite or 6-dibutylformamidinepurine phosphoramidite by LGC BioTechnology, Denmark. All natural bases used 'base labile' protecting groups (dA and dG with pac and dC with acetyl).



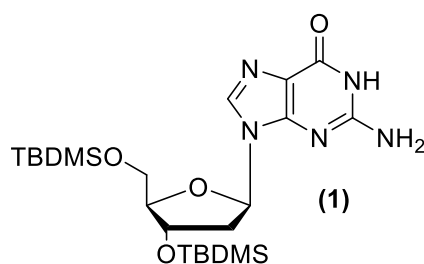
Analytical RP-HPLC performed on Walters 2695 or 2690 instrumentation using a Phenomenex Gemini C18 5  $\mu\text{m}$  4.6 x 250 mm column or AdvanceBio Oligonucleotides 2.7  $\mu\text{m}$  4.6 x 150 mm, flow rate 1 mL/min, UV detection was recorded at 260 nm unless specified otherwise. Preparative RP- HPLC performed using a Phenomenex Gemini C18 5  $\mu\text{m}$  110Å 21.2 x 250 mm column at a flow rate of 17 mL / min. UV detection was recorded at 260 nm unless specified otherwise.

Nucleoside composition analysis performed using UHPLC Ultimate 3000 instrumentation with a Thermo Fisher Scientific Hypercarb™ 5  $\mu\text{m}$  250 Å 100 x 4.6 mm, flow rate 0.08 mL / min. UV detection was recorded at 260 nm.

Limit of detection analysis performed using UHPLC Ultimate 3000 instrumentation with a Q Exactive HF Orbitrap, flow rate 0.1 mL / min. UV detection was recorded at 260 nm.

Denaturing PAGE gels were 25 % polyacrylamide, 8 M urea, 1 X TBE and were visualised by soaking in a 1 X TBE solution containing the SYBR Gold dye. Imaging was done using a BIO-RAD ChemiDOC MP Imaging system.

**3', 5'-Bis-O-tertbutyldimethylsilyl-2'-deoxyguanosine.**



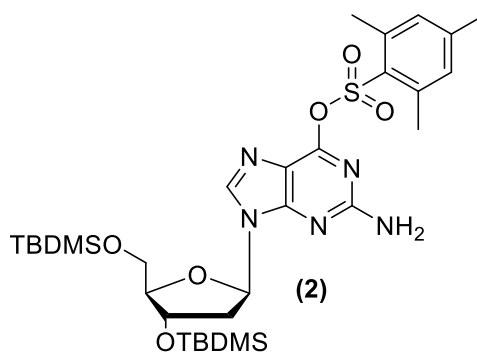
Prepared according to the literature<sup>84</sup>.

2'-Deoxyguanosine monohydrate (5.0 g, 17.5 mmol) was dried by evaporation with pyridine (3 x 25 mL) then suspended in anhydrous DMF (25 mL, 0.7 M). To this suspension was added imidazole (11.9 g, 0.18 mol) followed by TBDMSCl (6.6 g, 43.8 mmol) and stirred for 4 hrs at rt after which the solvent was evaporated. The resulting solid was then suspended and sonicated in MeOH and filtered to give the product as a white solid (8.88 g, quantitative yield).

$R_f$  (10% MeOH:DCM) = 0.26.

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>22</sub>H<sub>42</sub>N<sub>5</sub>O<sub>4</sub>Si<sub>2</sub> [M+H]<sup>+</sup>: 496.2775 Observed: 496.1893.

**2-Amino-9-[2-deoxy-3,5-bis-O-(*tert*butyldimethylsilyl)- $\beta$ -D-erythro-pentofuranosyl]-6-mesitylenesulfonyl-purine (2)**



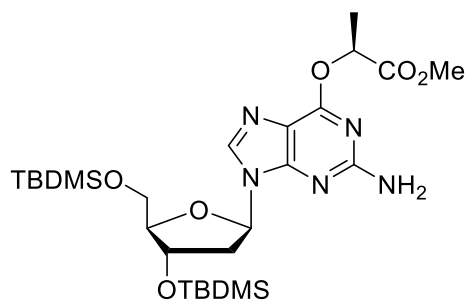
Prepared according to the literature procedure<sup>84</sup>.

Compound **1** (6.95 g, 14 mmol) was dissolved in DCM (56 mL, 0.25 M) and cooled to 0 °C. To this Et<sub>3</sub>N (11.7 mL, 84 mmol), 2-mesitylenesulfonyl chloride (6.1 g, 28 mmol) and DMAP (43 mg, 0.35 mmol) were added and the was stirred for a further 12 hrs at rt. After evaporation the residue was suspended in Hexane: Et<sub>2</sub>O (50:50), sonicated and filtered. The mother liqueur was then evaporated and purification by silica gel column chromatography (0 - 10% EtOAc:DCM) gave a white solid (7.65 g, 81%).

$R_f$  (10% EtOAc:DCM) = 0.71.

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>31</sub>H<sub>52</sub>N<sub>5</sub>O<sub>6</sub>Si<sub>2</sub>S [M+H]<sup>+</sup>: 678.3177 Observed: 678.3254.

**3',5'-Bis-O-tert-butyldimethylsilyl-O<sup>6</sup>-(methoxycarbonyl-ethyl)-2'-deoxyguanosine (3)**



Compound **2** (7.87 g, 11.6 mmol) was dissolved in anhydrous 1,2-dimethoxyethane (44 mL) under a nitrogen atmosphere. To this was added DABCO (2.60 g, 23.2 mmol), 3 Å molecular sieves and methyl (L)-lactate (5.54 mL, 58.0 mmol) and the solution left to stir for 1 h. DBU (2.60 mL, 17.4 mmol) was then added and the reaction left to stir for 24 h. After evaporation the residue was suspended in hexane:Et<sub>2</sub>O (50:50), sonicated and filtered. The mother liquor was then evaporated and purification by silica gel column chromatography (0 – 30 % EtOAc:DCM) gave compound **2** as a white solid (5.40 g, 80 %).

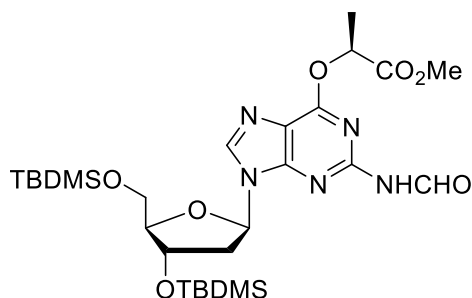
$R_f$  (5 % EtOAc:DCM) = 0.43.

<sup>1</sup>H NMR (250 MHz, CDCl<sub>3</sub>):  $\delta_H$  7.90 (1H, s, H8), 6.28 (1H, t, H1'  $J$  = 6.4 Hz), 5.42 (1H, q, O-CH(CH<sub>3</sub>)COOMe,  $J$  = 7.0 Hz), 5.04 (2H, bs, NH<sub>2</sub>), 4.53 (1H, m, H3'), 3.91 (1H, dt, H4',  $J$  = 3.1, 6.0 Hz), 3.80- 3.70 (2H, m, H5', H5''), 3.66 (3H, s, O-CH(CH<sub>3</sub>)COOCH<sub>3</sub>), 2.51- 2.22 (2H, m, H2', H2''), 1.63 (3H, d, O-CH(CH<sub>3</sub>)COOMe,  $J$  = 7.0 Hz), 0.86 (18H, d, SiCH<sub>2</sub>CH<sub>3</sub>,  $J$  = 1.8 Hz), 0.04 (12H, d, SiCH<sub>2</sub>CH<sub>3</sub>,  $J$  = 5.8 Hz).

<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>):  $\delta_C$  171.9, 159.7, 159.0, 153.7, 137.7, 115.0, 87.5, 83.5, 71.7, 70.4 66.6, 62.7, 52.2, 52.0, 41.0, 25.8, 25.6, 20.2, 18.3, 17.7, 17.3, 14.0, -4.7, -4.8, -5.4, -5.5.

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>26</sub>H<sub>48</sub>N<sub>5</sub>O<sub>6</sub>Si<sub>2</sub> [M+H]<sup>+</sup>: 582.3138 Observed: 582.3144.

***N*<sup>2</sup>-[(Dimethylamino)methylidene]-3',5'-bis-*O*-*tert*-butyldimethylsilyl-*O*<sup>6</sup>-  
(methoxycarbonylethyl)-2'-deoxyguanosine (4)**



Compound **3** (5.01 g, 8.6 mmol) was dissolved in anhydrous DMF (37 mL) with 3 Å molecular sieves under a nitrogen atmosphere. After 1 h *N,N*-dimethylformamide dimethylacetal (8.00 mL, 56 mmol) was added and the solution left to stir at rt overnight. After evaporation the residue was loaded onto a silica column and left for 1.5 hours before starting the process of elution (0 – 10 % EtOAc:DCM) to give compound **4** as a white solid (4.92 g, 90 %).

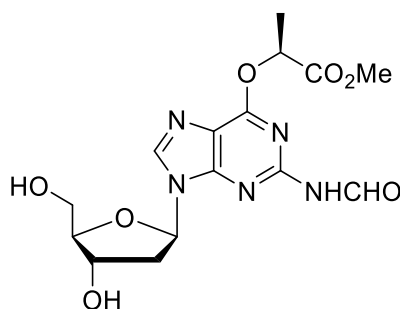
$R_f$  (5 % EtOAc:DCM) = 0.50.

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta_H$  9.37 (1H, d, NHCHO,  $J$  = 10.6 Hz), 8.20- 8.19 (2H, m, NHCHO, H8), 6.39 (1H, t, H1',  $J$  = 6.4 Hz), 5.43 (1H, q, O-CH(CH<sub>3</sub>)COOMe,  $J$  = 7.0 Hz) 4.58 (1H, dt, H3',  $J$  = 3.7, 5.4 Hz), 4.00 (1H, dd, H4',  $J$  = 3.1, 6.3 Hz), 3.88- 3.70 (2H, m, H5', H5''), 3.74 (3H, s, O-CH(CH<sub>3</sub>)COOCH<sub>3</sub>), 2.59- 2.36 (2H, m, H2', H2''), 1.73 (3H, d, O-CH(CH<sub>3</sub>)COOMe,  $J$  = 7.0 Hz), 0.91 (18H, d, SiCH<sub>2</sub>CH<sub>3</sub>,  $J$  = 2.8 Hz), 0.09 (12H, m, SiCH<sub>2</sub>CH<sub>3</sub>).

<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>)  $\delta_C$  171.5, 162.9, 159.5, 152.7, 151.2 140.1, 118.6, 87.9, 84.2, 71.7, 71.5, 62.7, 52.4, 41.6, 26.0, 25.7, 21.1, 18.4, 18.0, 17.4, 14.2, -4.7, -4., -5.4, -5.5.

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>27</sub>H<sub>48</sub>N<sub>5</sub>O<sub>7</sub>Si<sub>2</sub> [M+H]<sup>+</sup>: 610.3087 Observed: 610.3093.

***N*<sup>2</sup>-Formyl -*O*<sup>6</sup>-(methoxycarbonyl)ethyl)-2'-deoxyguanosine (5)**



Compound **4** (3.80 g, 6.0 mmol) was dissolved in anhydrous THF (35 mL) and to this tetra-n-butylammonium fluoride in THF (3.40 mL, 13.2 mmol) and the solution stirred at rt overnight. After evaporation the residue was purified by silica gel column chromatography (0 – 10 % MeOH:DCM) to give compound **5** as a white solid (1.71 g, 72 %).

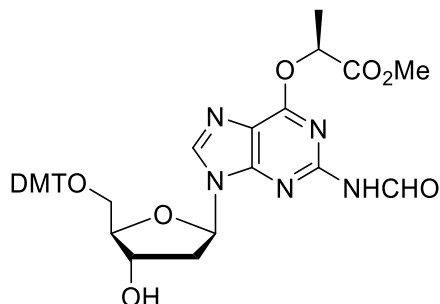
$R_f$  (10 % MeOH:DCM) = 0.26.

<sup>1</sup>H NMR (400 MHz; d<sup>4</sup>-MeOD)  $\delta_H$  9.36 (1H, s, NHCHO), 8.45 (1H, s, H8), 6.45 (1H, t, H1',  $J$  = 6.7 Hz), 5.53 (1H, q, O-CH(CH<sub>3</sub>)COOMe,  $J$  = 7.0 Hz), 4.62- 4.55 (1H, m, H3'), 4.02 (1H, dd, H4',  $J$  = 3.8, 7.6 Hz), 3.86- 3.72 (2H, m, H5', H5''), 3.76 (3H, s, O-CH(CH<sub>3</sub>)COOCH<sub>3</sub>), 2.87- 2.39 (2H, m, H2', H2''), 1.71 (3H, d, O-CH(CH<sub>3</sub>)COOMe,  $J$  = 7.0 Hz).

<sup>13</sup>C NMR (101 MHz, MeOD)  $\delta_C$  171.8, 164.1, 159.4, 152.7, 152.2, 141.3, 117.4, 87.9, 84.5, 71.4, 71.0, 61.6, 51.5, 39.9, 16.2.

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>15</sub>H<sub>20</sub>N<sub>5</sub>O<sub>7</sub> [M+H]<sup>+</sup>: 382,1357 Observed: 382.1367.

**5'-O-(4,4'-Dimethoxytriphenylmethyl)-N<sup>2</sup>-formyl-O<sup>6</sup>-(methoxycarbonyl-ethyl)-2'-deoxyguanosine (6)**



Compound **5** (1.50 g, 3.9 mmol) was dissolved in anhydrous pyridine (40 mL), under a nitrogen atmosphere with stirring. DMTCl (1.59 g, 4.7 mmol) was added. After 4 h the mixture was concentrated by evaporation and redissolved in ethyl acetate (30 mL) for extraction with sat. aq. sodium bicarbonate (30 mL), water (30 mL), brine (30 mL) followed by drying with magnesium sulfate. Evaporation and purification of the resulting residue by silica gel column chromatography (1 % Et<sub>3</sub>N/EtOAc) gave compound **6** as a white foam (2.08 g, 78 %).

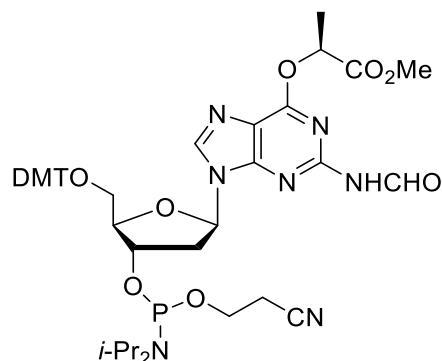
$R_f$  (1 % Et<sub>3</sub>N/EtOAc) = 0.31.

<sup>1</sup>H NMR (400 MHz; d<sup>4</sup>-MeOD)  $\delta_H$  9.26 (1H, s, NHCHO), 8.23 (1H, s, H8), 7.36- 7.04 (9H, m, DMT), 6.72 (4H, dd, DMT,  $J = 3.4, 8.9$  Hz), 6.39 (1H, t, H1',  $J = 6.4$  Hz), 5.48 (1H, q, O-CH(CH<sub>3</sub>)COOMe,  $J = 7.0$  Hz), 4.63 (1H, dd, H3',  $J = 5.1, 9.4$  Hz), 4.13 (1H, dd, H4',  $J = 4.2, 8.0$  Hz), 3.70- 3.65 (9H, m, O-CH(CH<sub>3</sub>)COOCH<sub>3</sub> & DMT-OMe), 3.41- 3.23 (2H, m, H5', H5''), 2.95- 2.43 (2H, m, H2', H2''), 1.66 (3H, d, O-CH(CH<sub>3</sub>)COOMe,  $J = 7.0$  Hz) .

<sup>13</sup>C NMR (63 MHz, d<sub>6</sub>-DMSO)  $\delta_C$  171.7, 163.9, 159.4, 158.6, 158.6, 152.6, 152.1, 148.7, 144.8, 141.3, 137.0, 135.9, 135.5, 129.9, 129.8, 127.9, 127.4, 126.5, 124.2, 117.8, 112.7, 86.7, 86.2, 84.7, 71.4, 71.2, 63.8, 60.2, 54.3, 51.6, 39.2, 19.6, 16.3, 13.2.

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>36</sub>H<sub>38</sub>N<sub>5</sub>O<sub>9</sub> [M+H]<sup>+</sup>: 685.2695 Observed: 685.2714.

**5'-O-(4,4'-Dimethoxytriphenylmethyl)-N<sup>2</sup>-formyl-O<sup>6</sup>-(methoxycarbonylethyl)-2'-deoxyguanosine-3'-O-(2-cyanoethyl-N,N-diisopropylamino)-phosphoramidite (7)**



Compound **6** (1.20 g, 1.7 mmol) was dissolved in dry DCM (12 mL) and stirred at rt under a nitrogen atmosphere. This was followed by the addition of anhydrous DIPEA (1.2 mL, 6.8 mmol) and 3 Å molecular sieves. After 30 min 2-cyanoethyl-*N*-diisopropylamine chlorophosphoramidite (460 µL, 2.0 mmol) was added dropwise over a 15 min period. After 3 h the mixture was concentrated by evaporation, dissolved in EtOAc (50 mL) and washed with sat. aq. sodium bicarbonate (50 mL), water (50 mL) and brine (50 mL). The organic layer was dried over magnesium sulfate and evaporated. Purification by silica gel column chromatography (30% DCM: Hexane containing 10% Et<sub>3</sub>N) gave compound **7** as a white foam (1.08 g, 72%).

$R_f$  (40% DCM: Hexane containing 10% Et<sub>3</sub>N) = 0.57, 0.47.

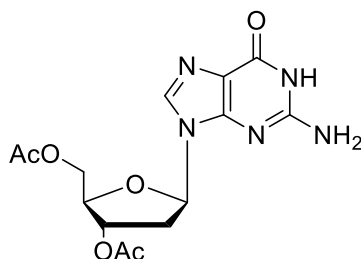
<sup>1</sup>H NMR (400 MHz; d<sup>4</sup>-MeOD) δ<sub>H</sub> 10.91 (1H, dd, NHCHO,  $J$  = 5.0, 9.6 Hz), 9.22 (1H, d, NHCHO,  $J$  = 9.5 Hz), 8.39 (1H, d, H8,  $J$  = 5.7 Hz), 7.24 (9H, m, DMT), 6.89- 6.71 (4H, m, DMT), 6.37 (1H, dd, H1',  $J$  = 6.4, 11.0 Hz), 5.55 (1H, q, O-CH(CH<sub>3</sub>)COOMe,  $J$  = 7.0 Hz), 4.78- 4.64 (1H, m, H3'), 4.15- 4.00 (1H, m, H4'), 3.71 (9H, m, O-CH(CH<sub>3</sub>)COOCH<sub>3</sub> & DMT-OMe), 3.56 (2H, m, H5', H5''), 3.23 (1H, dd, H2''  $J$  = 4.9, 9.3 Hz), 3.01 (1H, dt, H2'  $J$  = 6.0, 11.8 Hz), 2.77 (2H, t, OCH<sub>2</sub>CH<sub>2</sub>CN,  $J$  = 5.9), 2.66 (2H, t, OCH<sub>2</sub>CH<sub>2</sub>CN  $J$  = 6.0), 1.62 (3H, d, O-CH(CH<sub>3</sub>)COOMe,  $J$  = 7.0 Hz), 1.15 (14H, m, <sup>i</sup>Pr<sub>2</sub>N).

<sup>31</sup>P NMR (162 MHz, d<sub>6</sub>-DMSO) δ<sub>P</sub> + 148.1, +147.4.

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>46</sub>H<sub>56</sub>N<sub>7</sub>O<sub>10</sub>P [M+H]<sup>+</sup>: 884.3748 Observed: 884.3734.



### 3',5'-Bis-O-acetyl-2'-deoxyguanosine (9)



Prepared according to the literature procedure<sup>97</sup>.

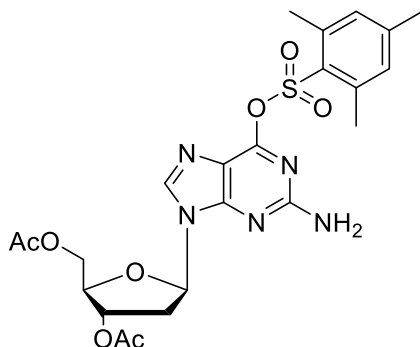
2'-Deoxyguanosine (100 mg, 0.37 mmol) was dissolved in anhydrous DMF (1 mL) to which was added Et<sub>3</sub>N (0.3 mL, 2.04 mmol), DMAP (16 mg, 0.13 mmol) and acetic anhydride (0.1 mL, 1.05 mmol) and the reaction stirred at rt for 2 h. The solvent was then removed under reduced pressure and the crude product dissolved in DCM. Sonication and vacuum filtration of the solution left a white precipitate (126 mg, 97 %).

$R_f$  (10 % MeOH/ EtOAc) = 0.40.

<sup>1</sup>H NMR (400 MHz, DMSO)  $\delta_H$  10.70 (1H, s, H1), 7.93 (1H, s, H8), 6.53 (2H, sbr, NH<sub>2</sub>), 6.14 (1H, m, H1'), 5.29 (1H, d, H3',  $J = 5.4$  Hz), 4.27 (2H, dd, H5', H5'',  $J = 7.0, 13.1$  Hz), 4.18 (1H, m, H4'), 2.93 (1H, dd, H2',  $J = 7.9, 15.5$  Hz), 2.46 (1H, dd, H2'',  $J = 5.9, 14.3$  Hz), 2.07 (3H, s, CH<sub>3</sub>), 2.04 (3H, s, CH<sub>3</sub>).

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>14</sub>H<sub>18</sub>N<sub>5</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 352.1632 Observed: 352.1473.

**6-Mesitylenesulfonyl-3', 5'-bis-O-acetyl-2'-deoxyguanosine (10)**

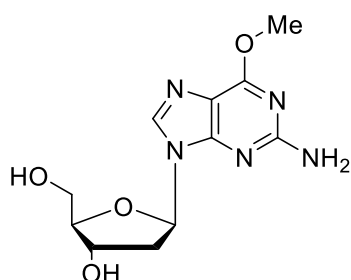


3',5'-Bis-O-acetyl-2'-deoxyguanosine (100 mg, 0.28 mmol) was dissolved in anhydrous chloroform (35 mL) and reacted with mesitylenesulfonyl chloride (196 mg, 0.90 mmol, 3.2 eq.), DMAP (16 mg, 0.013 mmol) and Et<sub>3</sub>N (0.6 mL, 4.08 mmol) and stirred at rt for 18 h. The solvent was evaporated and the crude product dissolved in EtOAc (20 mL) and then washed with acetic acid (pH 2.5, 20 mL) and then H<sub>2</sub>O (3 x 20 mL). Evaporation of the solvent left a yellow oil (132 mg, 88 %).

$R_f$  ( EtOAc) = 0.52.

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>23</sub>H<sub>28</sub>N<sub>5</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 534.1432 Observed: 534.1521.

### ***O*<sup>6</sup>-Methyl-2'-deoxyguanosine (13)**



6-Mesitylenesulfonyl-3', 5'-bis-O-acetyl-2'-deoxyguanosine (120 mg, 0.23 mmol) was dissolved in 1,2-dimethoxyethane (3 mL) and treated with DABCO (57 mg, 0.46 mmol, 2 eq) and methanol (45  $\mu$ L, 1.1 mmol, 5 eq). The reaction mixture was stirred at rt for 1 h after which DBU (56  $\mu$ L, 0.39 mmol, 1.5 eq) was added and the mixture stirred for a further 72 h. The solvent was evaporated and the crude product purified by preparative RP-HPLC (5-20 % acetonitrile: water) to afford a white solid (52 mg, 83%).

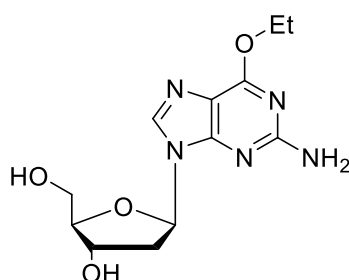
$R_f$  (10 % MeOH/ EtOAc) = 0.15.

<sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O)  $\delta_H$  7.93 (1H, s, H8), 6.21 (1H, dd, H1',  $J = 6.5, 7.4$  Hz), 4.51 (1H, dt, H3',  $J = 3.1, 6.1$  Hz), 4.02 (1H, m, H4'), 3.94 (3H, s, O-CH<sub>3</sub>), 3.67 (2H, qd, H5', H5'',  $J = 3.9, 12.6$  Hz), 2.69 (1H, ddd, H2'',  $J = 6.2, 7.6, 13.9$  Hz), 2.38 (1H, ddd, H2',  $J = 3.3, 6.3, 14.0$  Hz).

<sup>13</sup>C NMR (400 MHz, DMSO)  $\delta_C$  161.1, 160.2, 154.2, 138.2, 123.2, 88.0, 83.3, 71.2, 62.2, 53.7.

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>11</sub>H<sub>16</sub>N<sub>5</sub>O<sub>4</sub> [M+H]<sup>+</sup>: 282.1124 Observed: 282.1203.

### ***O*<sup>6</sup>-Ethyl-2'-deoxyguanosine (14)**



6-Mesitylenesulfonyl-3', 5'-bis-O-acetyl-2'-deoxyguanosine (120 mg, 0.23 mmol) was dissolved in 1,2-dimethoxyethane (3 mL) and treated with DABCO (57 mg, 0.46 mmol, 2 eq) and ethanol (64  $\mu$ L, 1.1 mmol, 5 eq). The reaction mixture was stirred at rt for 1 h after which DBU (56  $\mu$ L, 0.39 mmol, 1.5 eq) was added and the mixture stirred for a further 72 h. The solvent was evaporated and the crude product purified by preparative RP-HPLC (5-20 % acetonitrile: water) to afford a white solid (42 mg, 62%).

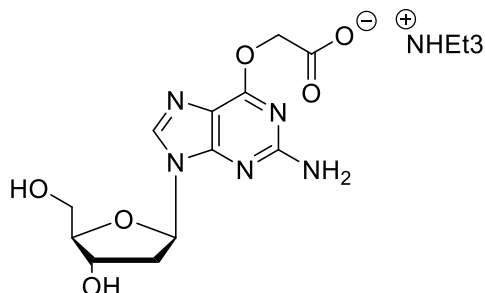
$R_f$  (10 % MeOH/ EtOAc) = 0.16.

<sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O)  $\delta_H$  7.95 (1H, s, H8), 6.21 (1H, t, H1',  $J$  = 6.8 Hz), 4.54 (1H, dt, H3',  $J$  = 3.2, 6.2 Hz), 4.42 (2H, q, O-CH<sub>2</sub>CH<sub>3</sub>,  $J$  = 7.1 Hz), 4.06 (1H, dd, H4',  $J$  = 3.4, 7.3 Hz), 3.71 (2H, qd, H5', H5'',  $J$  = 3.9, 12.6 Hz), 2.72 (1H, ddd, H2'',  $J$  = 6.0, 9.8, 13.9 Hz), 2.42 (1H, ddd, H2',  $J$  = 3.3, 6.3, 14.0 Hz), 1.34 (3H, t, O-CH<sub>2</sub>CH<sub>3</sub>,  $J$  = 7.1 Hz).

<sup>13</sup>C NMR (400 MHz, D<sub>2</sub>O)  $\delta_C$  161.2, 159.8, 152.6, 138.9 114.6, 87.2, 84.3, 71.2, 63.8, 61.7, 18.7, 13.7.

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>12</sub>H<sub>18</sub>N<sub>5</sub>O<sub>4</sub> [M+H]<sup>+</sup>: 296.1281 Observed: 296.1308.

### ***O*<sup>6</sup>-Carboxymethyl-2'-deoxyguanosine (15)**



6-Mesitylenesulfonyl-3', 5'-bis-O-acetyl-2'-deoxyguanosine (120 mg, 0.23 mmol) was dissolved in 1,2-dimethoxyethane (3 mL) and treated with DABCO (57 mg, 0.46 mmol, 2 eq) and methyl glycolate (84  $\mu$ L, 1.1 mmol, 5 eq). The reaction mixture was stirred at rt for 1 h after which DBU (56  $\mu$ L, 0.39 mmol, 1.5 eq) was added and the mixture stirred for a further 24 h. aq 0.5 M NaOH (3 mL) was then added and the reaction mixture was stirred for a further 72 h. The solvent was evaporated and the crude product purified by preparative RP-HPLC (5-20 % acetonitrile: 0.1 M TEAB) to afford a white solid (50 mg, 67%).

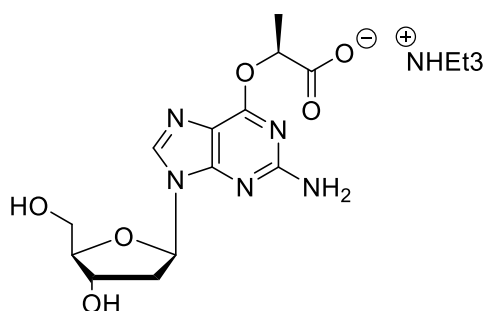
$R_f$  (10 % MeOH/ EtOAc) = 0.13.

<sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O)  $\delta_H$  7.97 (1H, s, H8), 6.24 (1H, t, H1',  $J$  = 7.2 Hz), 4.71 (2H, s, O-CH<sub>2</sub>COOH), 4.53 (1H, dt, H3',  $J$  = 3.1, 6.1 Hz), 4.05 (1H, dd, H4',  $J$  = 3.4, 7.2 Hz), 3.70 (2H, qd, H5', H5'',  $J$  = 3.9, 12.6 Hz), 2.72 (1H, ddd, H2'',  $J$  = 6.3, 7.5, 13.9 Hz), 2.41 (1H, ddd, H2',  $J$  = 3.3, 6.3, 14.0 Hz).

<sup>13</sup>C NMR (400 MHz, D<sub>2</sub>O)  $\delta_C$  176.0, 162.3, 160.5, 152.8, 139.0, 114.5, 87.2, 84.4, 71.4, 65.0, 61.7, 48.8, 38.7.

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>12</sub>H<sub>15</sub>N<sub>5</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 326.1022 Observed: 326.1132.

### ***O*<sup>6</sup>-Carboxyethyl-2'-deoxyguanosine (16)**



6-Mesitylenesulfonyl-3', 5'-*bis*-O-acetyl-2'-deoxyguanosine (120 mg, 0.23 mmol) was dissolved in 1,2-dimethoxyethane (3 mL) and treated with DABCO (57 mg, 0.46 mmol, 2 eq) and (S)-methyl lactate (104  $\mu$ L, 1.1 mmol, 5 eq). The reaction mixture was stirred at rt for 1 h after which DBU (56  $\mu$ L, 0.39 mmol, 1.5 eq) was added and the mixture stirred for a further 24 h. aq 0.5 M NaOH (3 mL) was then added and the reaction mixture was stirred for a further 72 h. The solvent was evaporated and the crude product purified by preparative RP-HPLC (5-20 % acetonitrile: 0.1 M TEAB) to afford a white solid (45.2 mg, 58%).

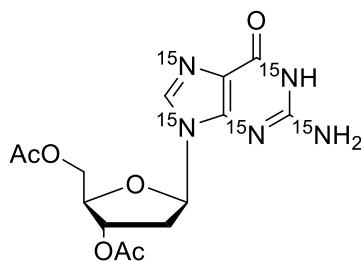
$R_f$  (10 % MeOH/ EtOAc) = 0.14.

<sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O)  $\delta_H$  7.99 (1H, s, H8), 6.27 (1H, t, H1',  $J = 7.6$  Hz), 5.06 (2H, q, O-CH(CH<sub>3</sub>)COOH,  $J = 7.0$  Hz), 4.54 (1H, dt, H3',  $J = 3.2, 6.2$  Hz), 4.06 (1H, dd, H4',  $J = 3.4, 7.4$  Hz), 3.71 (2H, qd, H5', H5'',  $J = 3.9, 12.6$  Hz), 2.73 (1H, ddd, H2'',  $J = 6.2, 7.5, 13.9$  Hz), 2.42 (1H, ddd, H2',  $J = 3.3, 6.3, 14.0$  Hz), 1.51 (3H, d, O-CH(CH<sub>3</sub>)COOH,  $J = 7.0$  Hz).

<sup>13</sup>C NMR (400 MHz, D<sub>2</sub>O)  $\delta_C$  179.7, 160.4, 159.8, 152.8, 138.9, 118.1, 87.2, 84.3, 73.9, 71.3, 61.7, 38.7, 17.1.

Acc. Mass [ESI+]: Calculated for C<sub>13</sub>H<sub>17</sub>N<sub>5</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 340.1179 Observed: 340.1201.

**<sup>15</sup>N<sub>5</sub>-Labelled-3',5'-bis-O-acetyl-2'-deoxyguanosine (17)**

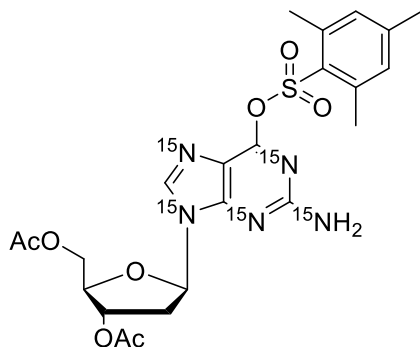


<sup>15</sup>N<sub>5</sub>-Labelled-2'-deoxyguanosine (100 mg, 0.37 mmol) was dissolved in anhydrous DMF (1 mL) to which was added Et<sub>3</sub>N (0.3 mL, 2.04 mmol), DMAP (16 mg, 0.13 mmol) and acetic anhydride (0.1 mL, 1.05 mmol) and the reaction stirred at rt for 2 h. The solvent was then removed under reduced pressure and the crude product dissolved in DCM. Sonication and vacuum filtration of the solution left a white precipitate (126 mg, 97 %).

$R_f$  (10 % MeOH/ EtOAc) = 0.40.

Acc. Mass [ESI+]: Calculated for C<sub>14</sub>H<sub>18</sub><sup>15</sup>N<sub>5</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 357.1031 Obtained: 357.1102.

**<sup>15</sup>N<sub>5</sub>-Labelled-6-mesitylenesulfonyl-3', 5'-bis-O-acetyl-2'-deoxyguanosine (18)**



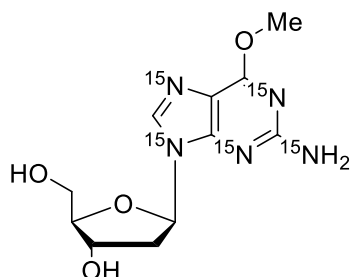
<sup>15</sup>N<sub>5</sub>-Labelled-3',5'-bis-O-acetyl-2'-deoxyguanosine (100 mg, 0.28 mmol) was dissolved in anhydrous chloroform (35 mL) and reacted with mesitylenesulfonyl chloride (196 mg, 0.90 mmol, 3.2 eq.), DMAP (16 mg, 0.013 mmol) and Et<sub>3</sub>N (0.6 mL, 4.08 mmol) and stirred at rt for 18 h. The solvent was evaporated and the crude product dissolved in EtOAc (20 mL) and then washed with acetic acid (pH 2.5, 20 mL) and then H<sub>2</sub>O (3 x 20 mL). Evaporation of the solvent left a yellow oil (132 mg, 88 %).

$R_f$  ( EtOAc) = 0.52.

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>23</sub>H<sub>28</sub><sup>15</sup>N<sub>5</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 539.1432 Observed: 539.1521.



**<sup>15</sup>N<sub>5</sub>-Labelled-*O*<sup>6</sup>-methyl-2'-deoxyguanosine (19)**



<sup>15</sup>N<sub>5</sub>-Labelled-6-mesitylenesulfonyl-3', 5'-*bis*-*O*-acetyl-2'-deoxyguanosine (12.0 mg, 0.023 mmol) was dissolved in 1,2-dimethoxyethane (3 mL) and treated with DABCO (5.7 mg, 0.046 mmol, 2 eq) and methanol (4.5 μL, 0.110 mmol, 5 eq). The reaction mixture was stirred at rt for 1 h after which DBU (5.6 μL, 0.039 mmol, 1.5 eq) was added and the mixture stirred for a further 72 h. The solvent was evaporated and the crude product purified by preparative RP-HPLC (5-20 % acetonitrile: water) to afford a white solid (5.8 mg, 88%).

$R_f$  (10 % MeOH/ EtOAc) = 0.15.

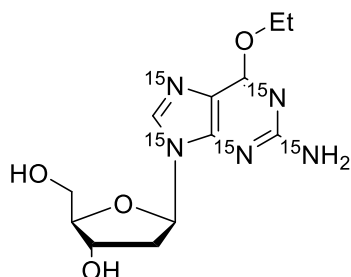
<sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O) δ<sub>H</sub> 7.93 (1H, dd, H8,  $J$  = 8.1, 11.0 Hz), 6.26 (1H, t, H1',  $J$  = 6.5 Hz), 4.55 (1H, dt, H3',  $J$  = 3.1, 6.2 Hz), 4.06 (1H, m, H4'), 3.99 (3H, s, O-CH<sub>3</sub>), 3.71 (2H, qd, H5', H5'',  $J$  = 3.9, 12.6 Hz), 2.73 (1H, m, H2''), 2.42 (1H, ddd, H2',  $J$  = 3.3, 4.9, 14.0 Hz).

<sup>13</sup>C NMR (400 MHz, D<sub>2</sub>O) δ<sub>C</sub> 87.3, 84.4, 84.3, 71.3, 61.7, 54.3, 54.3, 38.6.

<sup>15</sup>N NMR (400 MHz, D<sub>2</sub>O) δ<sub>N</sub> 227.9, 204.8, 182.5, 169.5, 73.4.

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>11</sub>H<sub>16</sub><sup>15</sup>N<sub>5</sub>O<sub>4</sub> [M+H]<sup>+</sup>: 287.0976 Observed: 287.1012.

**<sup>15</sup>N<sub>5</sub>-Labelled-*O*<sup>6</sup>-ethyl-2'-deoxyguanosine (20)**



<sup>15</sup>N<sub>5</sub>-Labelled-6-mesitylenesulfonyl-3', 5'-bis-O-acetyl-2'-deoxyguanosine (12.0 mg, 0.23 mmol) was dissolved in 1,2-dimethoxyethane (3 mL) and treated with DABCO (5.7 mg, 0.046 mmol, 2 eq) and ethanol (6.4 μL, 0.110 mmol, 5 eq). The reaction mixture was stirred at rt for 1 h after which DBU (5.6 μL, 0.039 mmol, 1.5 eq) was added and the mixture stirred for a further 72 h. The solvent was evaporated and the crude product purified by preparative RP-HPLC (5-20 % acetonitrile: water) to afford a white solid (5.2 mg, 75%).

$R_f$  (10 % MeOH/ EtOAc) = 0.17.

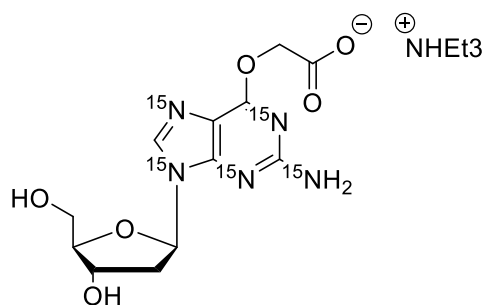
<sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O)  $\delta_H$  7.92 (1H, dd, H8,  $J$  = 8.0, 11.0 Hz), 6.21 (1H, t, H1',  $J$  = 6.4 Hz), 4.49 (1H, dt, H3',  $J$  = 3.1, 6.3 Hz), 4.40 (2H, q, O-CH<sub>2</sub>CH<sub>3</sub>,  $J$  = 7.1 Hz), 4.00 (1H, dd, H4',  $J$  = 3.3, 7.1 Hz), 3.65 (2H, qd, H5', H5'',  $J$  = 3.9, 12.6 Hz), 2.73-2.63 (1H, m, H2''), 2.42 (1H, d, H2',  $J$  = 9.2 Hz), 1.29 (3H, t, O-CH<sub>2</sub>CH<sub>3</sub>,  $J$  = 7.1 Hz).

<sup>13</sup>C NMR (400 MHz, D<sub>2</sub>O)  $\delta_C$  86.4, 83.5, 83.4, 70.5, 63.0, 62.9, 60.9, 37.8, 12.8.

<sup>15</sup>N NMR (400 MHz, D<sub>2</sub>O)  $\delta_N$  227.2, 201.1, 182.2, 169.0, 73.4.

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>12</sub>H<sub>18</sub><sup>15</sup>N<sub>5</sub>O<sub>4</sub> [M+H]<sup>+</sup>: 301.1132 Observed: 301.1202.

**<sup>15</sup>N<sub>5</sub>-Labelled-*O*<sup>6</sup>-carboxymethyl-2'-deoxyguanosine (21)**



<sup>15</sup>N<sub>5</sub>-Labelled-6-mesitylenesulfonyl-3', 5'-bis-*O*-acetyl-2'-deoxyguanosine (12.0 mg, 0.023 mmol) was dissolved in 1,2-dimethoxyethane (3 mL) and treated with DABCO (5.7 mg, 0.046 mmol, 2 eq) and methyl glycolate (8.4 μL, 0.110 mmol, 5 eq). The reaction mixture was stirred at rt for 1 h after which DBU (5.6 μL, 0.039 mmol, 1.5 eq) was added and the mixture stirred for a further 24 h. aq 0.5 M NaOH (3 mL) was then added and the reaction mixture was stirred for a further 72 h. The solvent was evaporated and the crude product purified by preparative RP-HPLC (5-20 % acetonitrile: 0.1 M TEAB) to afford a white solid (5.0 mg, 67%).

$R_f$  (10 % MeOH/ EtOAc) = 0.13.

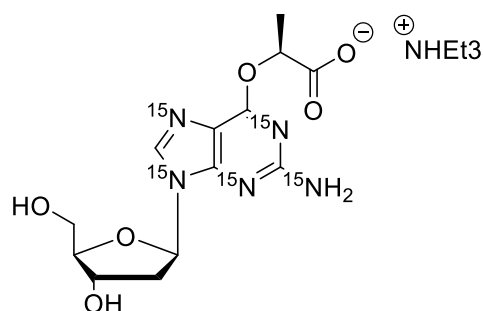
<sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O) δ<sub>H</sub> 8.00 (1H, dd, H8,  $J$  = 8.1, 10.7 Hz), 6.28 (1H, t, H1',  $J$  = 6.9 Hz), 4.73 (2H, s, O-CH<sub>2</sub>COOH), 4.55 (1H, dt, H3',  $J$  = 3.2, 6.2 Hz), 4.06 (1H, dd, H4',  $J$  = 3.5, 7.2 Hz), 3.71 (2H, qd, H5', H5'',  $J$  = 3.9, 12.6 Hz), 2.74 (1H, m, H2'), 2.50- 2.37 (1H, m, H2').

<sup>13</sup>C NMR (400 MHz, D<sub>2</sub>O) δ<sub>C</sub> 176.0, 87.3, 84.5, 84.3, 71.4, 65.1, 65.1, 61.8, 38.7, 30.2.

<sup>15</sup>N NMR (400 MHz, D<sub>2</sub>O) δ<sub>N</sub> 229.3, 217.5, 200.0, 182.4, 147.4

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>12</sub>H<sub>15</sub><sup>15</sup>N<sub>5</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 331.0874 Observed: 331.0934.

**<sup>15</sup>N<sub>5</sub>-Labelled-*O*<sup>6</sup>-carboxyethyl-2'-deoxyguanosine (22)**



<sup>15</sup>N<sub>5</sub>-Labelled-6-mesitylenesulfonyl-3', 5'-bis-O-acetyl-2'-deoxyguanosine (12.0 mg, 0.023 mmol) was dissolved in 1,2-dimethoxyethane (3 mL) and treated with DABCO (5.7 mg, 0.046 mmol, 2 eq) and (S)-methyl lactate (10.4 μL, 0.110 mmol, 5 eq). The reaction mixture was stirred at rt for 1 h after which DBU (5.6 μL, 0.039 mmol, 1.5 eq) was added and the mixture was stirred for a further 24 h. aq 0.5 M NaOH (3 mL) was then added and the reaction mixture was stirred for a further 72 h. The solvent was evaporated and the crude product purified by preparative RP-HPLC (5-20 % acetonitrile: 0.1 M TEAB) to afford a white solid (5.1 mg, 67%).

$R_f$  (10 % MeOH/ EtOAc) = 0.14.

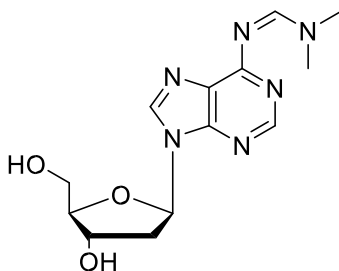
<sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O)  $\delta_H$  7.99 (1H, dd, H8,  $J = 8.0, 10.8$  Hz), 6.28 (1H, t, H1',  $J = 6.9$  Hz), 5.07 (2H, q, O-CH(CH<sub>3</sub>)COOH,  $J = 7.0$  Hz), 4.56- 4.52 (1H, m, H3'), 4.06 (1H, dd, H4',  $J = 3.5, 7.4$  Hz), 3.70 (2H, qd, H5', H5'',  $J = 3.7, 12.5$  Hz), 2.80- 2.68 (1H, m, H2'), 2.42 (1H, m, H2'), 1.51 (3H, d, O-CH(CH<sub>3</sub>)COOH,  $J = 7.0$  Hz).

<sup>13</sup>C NMR (400 MHz, D<sub>2</sub>O)  $\delta_C$  179.7, 87.2, 84.3, 84.2, 74.0, 73.9, 71.4, 61.7, 38.7, 17.1.

<sup>15</sup>N NMR (400 MHz, D<sub>2</sub>O)  $\delta_N$  223.5, 218.1, 200.1, 183.0, 59.5.

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>13</sub>H<sub>17</sub><sup>15</sup>N<sub>5</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 345.1031 Observed: 345.1127.

***N*<sup>6</sup>-[(Dimethylamino)methylene]-2'-deoxyadenosine (23)**



Prepared according to the literature procedure<sup>108</sup>.

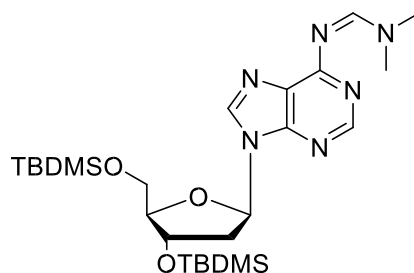
2'-Deoxyadenosine monohydrate (2 g, 7.42 mmol) was dissolved in methanol (86 ml) and *N,N*-dimethylformamide dimethylacetal (5 ml, 37.2 mmol) was added. The mixture was heated to 50 °C and left to stir for 24 hours. After the reaction was complete, the methanol was removed under reduced pressure to leave a white solid. This was washed with a mixture of DCM: hexane 1:1 (50 ml) which was removed by vacuum filtration to leave a white solid (2.2 g, 97%).

$R_f$  ( DCM/ 5 % MeOH) = 0.25.

<sup>1</sup>H NMR  $\delta$  (400 MHz, CD<sub>3</sub>OD)  $\delta_H$  8.92 (1H, s, N=CHN(Me)<sub>2</sub>), 8.44 (1H, s, H8), 8.43 (1H, s, H2), 6.48 (1H, dd, H1'  $J$  = 6.2, 7.8 Hz), 4.59 (1H, m, H3'), 4.07 (1H, m, H4'), 3.86 (1H, dd, H5',  $J$  = 3.1, 12.2 Hz), 3.76 (1 H, dd, H5'',  $J$  = 3.5, 12.2 Hz), 3.25 (3H, s, N-CH<sub>3</sub>), 3.24 (3H, s, NCH<sub>3</sub>), 2.80-2.88 (1H, ddd, H2', H2'',  $J$  = 5.9, 7.8, 13.5 Hz);

Mass Spectrometry [ESI<sup>+</sup>]: [M+H]<sup>+</sup>: 308.1

***N*<sup>6</sup>-[(Dimethylamino)methylene]-3',5'-bis-*O*-(*t*-butyldimethylsilyl)-2'-deoxyadenosine  
(25)**



Prepared according to the literature procedure<sup>108</sup>.

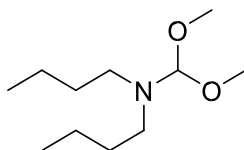
3',5'-Bis-*O*-(*t*-butyldimethylsilyl)-2'-deoxyadenosine monohydrate (2 g, 7.42 mmol) was dissolved in methanol (86 ml) and *N,N*-dimethylformamide dimethylacetal (5 ml, 37.2 mmol) was added. The mixture was heated to 50 °C and left to stir for 24 hrs. After the reaction was complete, the methanol was removed under reduced pressure to leave a white solid. This was washed with DCM: hexane 1:1 (50 ml) before vacuum filtration left a white solid (3.7 g, 80%).

$R_f$  (DCM) = 0.52.

<sup>1</sup>H NMR  $\delta$  (400 MHz, CD<sub>3</sub>OD):  $\delta_H$  8.61 (1H, s, N=CHN(Me)<sub>2</sub>), 8.47 (1H, s, H8), 8.41 (1H, s, H2), 6.48 (1H, t, H1'), 4.75 (1H, m, H3'), 4.0 (1H, m, H4'), 3.92 (1H, dd, H5'), 3.81 (1H, dd, H5''), 3.33 (3H, s, NCH<sub>3</sub>), 3.28 (3H, s, NCH<sub>3</sub>), 3.87 (1H, m, H2'), 2.48 (1H, m, H2''), 0.99 (9H, s, C(CH<sub>3</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>3</sub>), 0.91 (9H, s, C(CH<sub>3</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>3</sub>), 0.2 (6H, s, C(CH<sub>3</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>3</sub>), 0.1 (6H, d, C(CH<sub>3</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>3</sub>).

Mass Spectrometry [ESI<sup>+</sup>]: [M+H]<sup>+</sup>: 536.1

***N, N*-Di-*n*-butylformamide dimethyl acetal (26)**



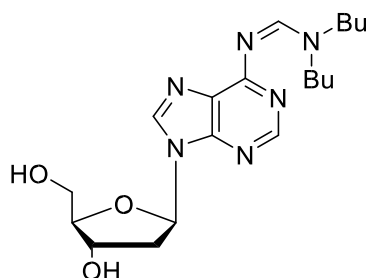
Prepared according to the literature procedure<sup>18</sup>.

Di-*n*-butylamine (16.8 ml, 100 mmol) was refluxed at 100 °C with *N, N*-dimethylformamide dimethylacetal (14.6 ml, 100 mmol) for 3 days under nitrogen. The reaction mixture was then purified by fractional distillation (90 °C, 10 mmHg).

<sup>1</sup>H NMR  $\delta$  (400 MHz, CDCl<sub>3</sub>):  $\delta_{\text{H}}$  4.35 (1H, s, NCH(OCH<sub>3</sub>)<sub>2</sub>), 3.13 (6H, s, NCH(OCH<sub>3</sub>)<sub>2</sub>), 2.44 (4H, m, N(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>2</sub>), 1.09-1.31 (4H, m, N(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>2</sub>), 0.72-0.78 (6H, m, N(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>2</sub>).

Mass Spectrometry: ESI<sup>+</sup> m/z 205.1 [M+H]<sup>+</sup>

***N*<sup>6</sup>-[(Dibutylamino)methylene]-2'-deoxyadenosine (27)**



Prepared according to the literature procedure<sup>18</sup>.

*N,N*-Di-*n*-butylformamide dimethylacetal (4 g, 20 mmol) was reacted with 2'-deoxyadenosine (3.35 g, 13 mmol) in dry DMF (50 ml) for 24 hrs. After 24 hrs the DMF was removed under reduced pressure and the mixture dissolved in ethyl acetate and subsequently washed with sodium bicarbonate (3 x 20 ml) and water (30 ml) before being dried with magnesium sulfate. The crude product was purified by silica column chromatography (DCM: MeOH, 95:5) to leave a yellow oil (5 g, 67%).

$R_f$  ( EtOAc) = 0.31.

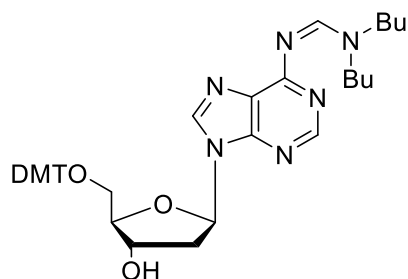
<sup>1</sup>H NMR (400 MHz, DMSO)  $\delta_H$  8.94 (1H, s, N=CHN(Bu)<sub>2</sub>), 8.51 (1H, s, H8), 8.43 (1H, s, H2), 6.46 (1H, t, H1',  $J$  = 6.8 Hz), 5.38 (2H, d, -OH,  $J$  = 48.0 Hz), 4.49 (1H, dt, H3',  $J$  = 4.6, 9.4 Hz), 3.96 (1H, dd, H4',  $J$  = 4.3, 8.7 Hz), 3.59 (2H, m, N-(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>2</sub>), 3.45 (2H, t, N-(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>2</sub>,  $J$  = 7.1 Hz), 3.37 (2H, d, H5', H5'',  $J$  = 4.9 Hz), 2.76 (1H, dt, H2'',  $J$  = 6.3, 12.9 Hz), 2.34 (1H, ddd, H2',  $J$  = 4.8, 6.6, 13.2 Hz), 1.53 (4H, tt, N-(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>2</sub>,  $J$  = 7.5, 14.6 Hz), 1.25 (4H, qt, N-(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>2</sub>,  $J$  = 8.4, 16.6 Hz), 0.85 (6H, td, N-(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>2</sub>,  $J$  = 5.4, 7.3 Hz).

<sup>13</sup>C NMR (400 MHz, DMSO)  $\delta_C$  159.9, 158.4, 152.2, 151.4, 141.7, 126.3, 88.5, 84.5, 71.5, 62.3, 51.5, 44.9, 40.1, 30.9, 29.1, 20.0, 19.6, 14.1, 13.9.

Mass Spectrometry [ESI<sup>+</sup>]: [M+H]<sup>+</sup>: 392.1



**9-[2'-Deoxy-5'-dimethoxytritryl-β-D-erythro-pentofuranoyl]-6-N',N'-n-dibutylformamidinepurine (28)**



9-[2'-Deoxy-β-D-erythro-pentofuranoyl]-6-N',N'-n-dibutylformamidinepurine (1 g, 2.55 mmol) was reacted with DMTCI (1 g, 3 mmol) and DMAP (15 mg, 0.12 mmol) in pyridine (25 mL) for 18 h at rt. The solvent was evaporated and the crude mixture redissolved in EtOAc (30 mL) for extraction with sat. aq. sodium bicarbonate (30 mL), water (30 mL), brine (30 mL) followed by drying with magnesium sulfate. Evaporation and purification of the resulting residue by silica gel column chromatography (1 % Et<sub>3</sub>N/EtOAc) gave a white foam (1.65 g, 93 %).

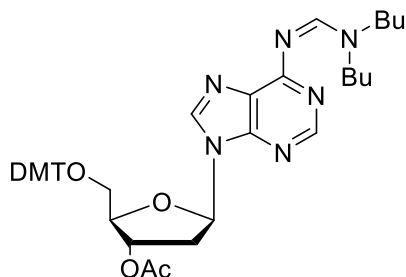
$R_f$  ( EtOAc) = 0.45.

<sup>1</sup>H NMR (400 MHz, DMSO)  $\delta_H$  8.92 (1H, s, N=CHN(Bu)<sub>2</sub>), 8.36 (1H, s, H8), 8.35 (1H, s, H2), 7.36- 7.15 (9H, m, DMT), 6.84- 6.73 (4H, m, DMT), 6.42 (1H, t, H1',  $J = 6.4$  Hz), 5.38 (1H, d, -OH,  $J = 4.6$  Hz), 4.50 (1H, dt, H3',  $J = 4.6, 9.4$  Hz), 3.99 (1H, dd, H4',  $J = 4.3, 8.7$  Hz), 3.71 (6H, d, DMT,  $J = 5.1$  Hz), 3.59 (2H, m, N-(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>2</sub>), 3.45 (2H, t, N-(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>2</sub>,  $J = 7.1$  Hz), 3.17 (2H, d, H5', H5'',  $J = 4.9$  Hz), 2.91 (1H, dt, H2'',  $J = 6.3, 12.9$  Hz), 2.35 (1H, ddd, H2',  $J = 4.8, 6.6, 13.2$  Hz), 1.61 (4H, tt, N-(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>2</sub>,  $J = 7.5, 14.6$  Hz), 1.32 (4H, qt, N-(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>2</sub>,  $J = 8.4, 16.6$  Hz), 0.93 (6H, td, N-(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>2</sub>,  $J = 5.4, 7.3$  Hz).

<sup>13</sup>C NMR (400 MHz, DMSO)  $\delta_C$  159.8, 158.5, 158.4, 158.3, 152.3, 151.6, 145.4, 141.8, 136.0, 130.2, 130.0, 128.2, 128.1, 127.0, 126.3, 113.5, 86.3, 85.8, 83.9, 71.1, 64.5, 60.2, 55.4, 51.4, 44.9, 39.0, 31.0, 29.2, 20.1, 19.6, 14.2, 14.1.

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>40</sub>H<sub>49</sub>N<sub>6</sub>O<sub>5</sub> [M+H]<sup>+</sup>: 693.3686 Observed: 693.3756.

**9-[2'-Deoxy-5'dimethoxytrityl-β-D-erythro-pentofuranoyl]-6-N',N'-n-dibutylformamidinepurine (29)**



9-[2'-Deoxy-5'dimethoxytrityl-β-D-erythro-pentofuranoyl]-6-N',N'-n-dibutylformamidinepurine (0.8 g, 1.12 mmol) was reacted with DMAP (6 mg, 0.06 mmol), acetic anhydride (0.2 mL, 2.10 mmol) and Et<sub>3</sub>N (1.0 mL, 6.86 mmol) in anhydrous DMF at rt for 2.5 h. The solvent was evaporated and the crude product purified by silica gel chromatography (1 % Et<sub>3</sub>N/EtOAc) to give a white foam (0.8 g, 94 %).

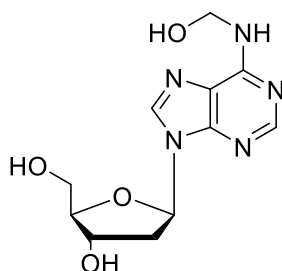
$R_f$  ( EtOAc) = 0.60.

<sup>1</sup>H NMR (400 MHz, DMSO) δ<sub>H</sub> 8.91 (1H, s, N=CHN(Bu)<sub>2</sub>), 8.37 (1H, s, H8), 8.31 (1H, s, H2), 7.15- 7.37 (9H, m, DMT), 6.81 (4H, dt, DMT,  $J$  = 6.0, 14.9 Hz), 6.43 (1H, t, H1',  $J$  = 7.0 Hz), 5.38 (1H, d, -OH,  $J$  = 4.6 Hz), 4.40- 4.45 (1H, m, H3'), 4.23- 4.16 (1H, m, H4'), 3.72 (6H, d, DMT,  $J$  = 4.0 Hz), 3.64- 3.57 (2H, m, N-(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>2</sub>), 3.45 (2H, t, N-(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>2</sub>,  $J$  = 7.1 Hz), 3.31- 3.17 (2H, m, H5', H5''), 2.08 (3H, s, C(O)CH<sub>3</sub>), 1.61 (4H, tt, N-(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>2</sub>,  $J$  = 7.5, 14.7 Hz), 1.32 (4H, qt, N-(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>2</sub>,  $J$  = 8.8, 17.5 Hz), 0.93 (6H, td, N-(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>2</sub>)<sub>2</sub>(CH<sub>3</sub>)<sub>2</sub>,  $J$  = 4.8, 7.4 Hz).

<sup>13</sup>C NMR (400 MHz, DMSO) δ<sub>C</sub> 170.4, 159.9, 158.5, 158.4, 158.3, 152.3, 151.6, 145.3, 142.0, 135.9, 130.2, 130.0, 128.2, 128.1, 127.1, 126.4, 113.6, 86.0, 84.3, 83.9, 75.1, 64.3, 60.2, 55.4, 51.4, 44.9, 35.8, 30.9, 29.2, 21.3, 20.1, 19.6, 14.2, 14.1.

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>42</sub>H<sub>51</sub>N<sub>6</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 735.3792 Observed: 735.3846.

**9-[2'-Deoxy-β-D-erythro-pentofuranosyl]-6-(hydroxymethyl)purine (31)**



2'-Deoxyadenosine (5.0 g, 0.02 mmol) was dissolved in water (50 mL) to which 3.7 % formaldehyde (46 μL, 1.20 mmol) was added. The reaction mixture was reacted at rt for 24 h. The solvent was evaporated and the crude mixture was purified by preparative RP-HPLC (5-20 % acetonitrile: water) to give a white solid (5.3 g, 52 %).

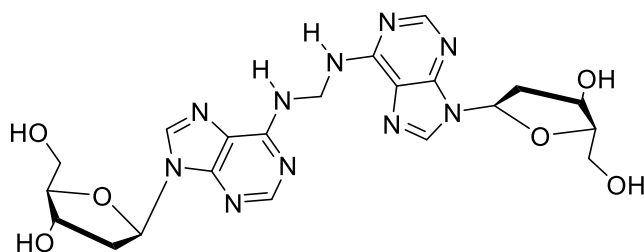
$^1\text{H}$  (600 MHz,  $\text{D}_2\text{O}$ )  $\delta_{\text{H}}$  8.35 (1H, s, H8), 8.33 (1H, s, H2), 5.17 (2H, s,  $\text{HNCH}_2\text{OH}$ ), 4.69- 4.62 (1H, m, H3'), 4.20 (1H, d, H4',  $J = 6.4$  Hz), 3.90- 3.75 (2H, m, H5', H5''), 2.92- 2.54 (2H, m, H2', H2'').

Acc. Mass [ESI+]: Calculated for  $\text{C}_{11}\text{H}_{16}\text{N}_5\text{O}_4$   $[\text{M}+\text{H}]^+$ : 282.1197 Observed: 282.1202.

Stability analysis of 9-[2'-deoxy-β-D-erythro-pentofuranosyl]-6-(hydroxymethyl)purine

9-[2'-deoxy-β-D-erythro-pentofuranosyl]-6-(hydroxymethyl)purine (0.5 mg, 97.2 μmol) was dissolved in phosphate buffer (pH 7.1, 1 mL). The solvent was removed under reduced pressure and the sample dissolved in  $\text{D}_2\text{O}$ .  $^1\text{H}$  NMR spectra of the reaction sample was run on a Bruker AV600 spectrometer at 0, 1, 7 and 8 days.

***Bis-(N<sup>6</sup>-2'-deoxyadenosyl) methane (32)***



2'-Deoxyadenosine (5.0 g, 0.02 mmol) was dissolved in water (50 mL) to which 3.7 % formaldehyde (46  $\mu$ L, 1.20 mmol) was added. The reaction mixture was stirred at rt for 1 day then heated to 50 °C. After a further 9 days the solvent was evaporated and the crude mixture was purified by preparative RP-HPLC (5-20 % acetonitrile: water) to give a white solid (1.2 g, 12 %).

<sup>1</sup>H (600 MHz, D<sub>2</sub>O)  $\delta$ <sub>H</sub> 8.35 (2H, s, H8), 8.31 (2H, s, H2), 6.48 (2H, t, H1', *J* = 6.9 Hz), 5.46 (2H, s, HNCH<sub>2</sub>NH), 4.64 (2H, dd, H3', *J* = 3.0, 6.0 Hz), 4.18 (2H, dd, H4', *J* = 3.4, 7.0 Hz), 3.81 (4H, qd, H5', H5'', *J* = 3.7, 12.6 Hz), 2.87-2.53 (4H, m, H2', H2'').

Acc. Mass [ESI<sup>+</sup>]: Calculated for C<sub>21</sub>H<sub>27</sub>N<sub>10</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 515.2110 Observed: 515.2124.

**Stability analysis of *bis-(N<sup>6</sup>-2'-deoxyadenosyl) methane***

*Bis-(N<sup>6</sup>-deoxyadenosyl) methane* (0.5 mg, 97.2  $\mu$ mol) was dissolved in phosphate buffer (pH 7.1, 1 mL). The solvent was removed under reduced pressure and the sample dissolved in D<sub>2</sub>O. <sup>1</sup>H NMR spectra of the reaction sample was run on a

Bruker AV600 spectrometer at 0, 1, 4, 15, 22 and 43 days.

## ODN Synthesis Chemistry

Table 16. ODN codes and corresponding sequences.

<i>ODN Code</i>	<i>ODN Sequence</i>
<i>AJ01 ClPur</i>	5' ATG CCT GCA ATT ACX TAT GTC GTA ATC ATG GT 3'
<i>AJ02 ClPur</i>	5' ACC ATG ATT ACG ACA TXT GGA ATC CTG ACG AAC 3'
<i>AJ03 CMG</i>	5' GAA CTX CAG CTC CGT GCT GGC CC 3'
<i>AJ04 CEG</i>	5' GAA CTX CAG CTC CGT GCT GGC CC 3'
<i>AJ05 CMG amide</i>	5' GAA CTX CAG CTC CGT GCT GGC CC 3'
<i>AJ06 CEG amide</i>	5' GAA CTX CAG CTC CGT GCT GGC CC 3'
<i>AJ07 MeG</i>	5' GAA CTX CAG CTC CGT GCT GGC CC 3'
<i>AJ08 JACS_FA1</i>	5' GCACCAAACAATTGG 3' 3' GTTTGTTAACCTCCC 5'
<i>AJ09 JACS_FA2</i>	5' GCACCAAACAATTGTTG 3' 3' GTTGTTAACAACCTCCC 5'
<i>AJ10 dA<sup>dbf</sup></i>	5' CCC TCC AXT TGT TTG 3'

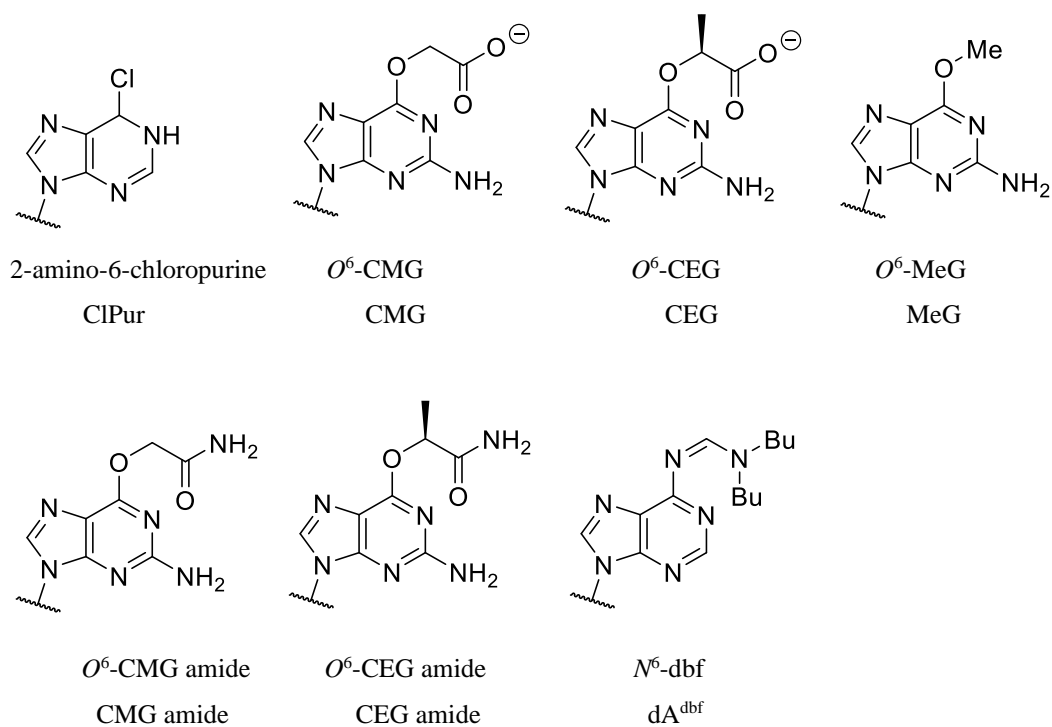


Figure 117. Structures of modifications contained in ODNs.

## RP-HPLC methods for purification of ODNs

Table 17. RP-HPLC methods and corresponding buffers and gradient.

<i>Method</i>	<i>Buffer A</i>	<i>Buffer B</i>	<i>Gradient (B %)</i>
<i>A</i>	Aq 0.1 M TEAB	Aq 0.1 M TEAB/ 50 % MeCN	5- 20 % over 20 mins
<i>B</i>	Aq 0.1 M TEAB	Aq 0.1 M TEAB/ 50 % MeCN	5- 40 % over 20 mins
<i>C</i>	Aq 0.1 M TEAB	Aq 0.1 M TEAB/ 50 % MeCN	5- 20 % over 30 mins
<i>D</i>	Aq 0.1 M TEAA	Aq 0.1 M TEAA/ 65 % MeCN	5- 18 % over 30 mins

### ODNs containing 6-chloropurine

CPG of ODNs AJ01 and AJ02 (1  $\mu$ mol) syntheses were treated with aq 0.5 M NaOH (1 mL) for 72 hrs at rt. The solutions were then desalted by gel filtration using a NAP<sup>TM</sup>-10 column (eluted with H<sub>2</sub>O, 1 mL).

The ODNs of AJ01 and AJ02 were purified by reverse-phase HPLC using method A and a Phenomenex Gemini C18 column (5  $\mu$ m 4.6 x 250 mm) at a flow rate of 1 mL/min. The UV detector was set to 260nm for analytical HPLC.

### ODN containing O<sup>6</sup>-(carboxyalkyl)guanine

CPG of ODNs AJ03 and AJ04 (1  $\mu$ mol) syntheses were treated with aq 0.5 M NaOH (1 mL) for 24 hrs at rt. The reaction mixtures were then treated with conc. aq ammonia for a further 72 hrs at rt. The solutions were then desalted by gel filtration using a NAP<sup>TM</sup>-10 column (eluted with H<sub>2</sub>O, 1 mL).

The ODNs were purified by reverse-phase HPLC using method B and a Phenomenex Gemini C18 column (5  $\mu$ m 4.6 x 250 mm) at a flow rate of 1 mL/min. The UV detector was set to 260nm for analytical HPLC.

ODNs containing  $O^6$ -(carboxyalkyl)guanine carboxamide

CPG of ODNs AJ05 and AJ06 (1  $\mu$ mol) syntheses were treated with conc. aq ammonia for 72 hrs at rt. The ammonia was then removed under reduced pressure before purification.

The ODNs of AJ05 and AJ06 were purified by reverse-phase HPLC using method B and a Phenomenex Gemini C18 column (5  $\mu$ m 4.6 x 250 mm) at a flow rate of 1 mL/min. The UV detector was set to 260nm for analytical HPLC.

ODN containing 6-*N*-((di-*n*-butylamino)-methylene)-2'-deoxyadenosine

CPG of ODN AJ10 (2  $\mu$ mol) was treated with 0.05 M  $K_2CO_3$ / MeOH (1 mL) for 4 hrs at rt. Once the solvent was removed under reduced pressure and the ODN redissolved in  $H_2O$ , the solution was desalted by gel filtration using a NAP<sup>TM</sup>-10 column (eluted with  $H_2O$ , 1 mL).

The ODN was purified by reverse-phase HPLC using method C and a Phenomenex Gemini C18 column (5  $\mu$ m 4.6 x 250 mm) at a flow rate of 1 mL/min. The UV detector was set to 260nm for analytical HPLC.

**General procedure for synthesis of ODNs containing formaldehyde-induced ICL**

ODNs AJ08 and AJ09 (150  $\mu$ L, 100 nmols) were added to a 1 mL solution of 50 mM phosphate buffer (pH 6.0), 25 mM NaCl, and were annealed at 95 °C for 5 min. 20  $\mu$ L of 25 mM aqueous formaldehyde was then added and solution incubated at 25 °C for 9 days. The solution was desalted by gel filtration using a NAP<sup>TM</sup>-10 column eluted with  $H_2O$ , 1 mL) and evaporated. For analysis, the ODN duplexes were dissolved in 1 mL of purified water and 5  $\mu$ L was taken and added with an equal amount of loading buffer (90% deionized formamide, 10 mM Tris (pH 7.3, 0.1 % xylene cyanol, and 0.1 mM EDTA). The resulting solution was heated at 90 °C for 2 min, placed in ice for 2

min, and analysed by denaturing PAGE. Denaturing PAGE was conducted on a 25 % gel (19: 1 acrylamide: bisacrylamide, 8 M urea, 0.35 mm thick, 16.5 x 13 cm, using a 14-toothed comb) at 10 W and at ca. 55 °C until xylene cyanol had run 7-9 cm.

The ICL-containing ODN duplex was purified by reverse-phase HPLC using method D and an AdvanceBio Oligonucleotides column (2.7 $\mu$ m 4.6 x 150 mm) flow rate 1 mL/min. The UV detector was set to 260 nm for analytical HPLC.

Table 18. ODN RP-HPLC conditions, retention times and MS [ESI-] observed masses.

<i>ODN Code</i>	<i>RP-HPLC Conditions</i>	<i>Retention Time</i>	<i>Calculated ESI-Mass (g mol<sup>-1</sup>)</i>	<i>Observed ESI-Mass (g mol<sup>-1</sup>)</i>
<i>AJ01</i>	Method A	16.3 mins	10764	10764
<i>AJ02</i>	Method A	16.7 mins	10767	10767
<i>AJ03</i>	Method B	14.1 mins	7060	7060
<i>AJ04</i>	Method B	13.3 mins	7073	7073
<i>AJ05</i>	Method B	14.5 mins	7059	7059
<i>AJ06</i>	Method B	13.5 mins	7072	7072
<i>AJ08</i>	Method D	24 mins	9853.4	9832.4
<i>AJ09</i>	Method D	14.7 mins	9235.9	9232.8
<i>AJ10</i>	Method C	10.9 mins	4632.5	4632.88



### **General procedure for enzymatic digestion of ODNs**

The ODN sample (2.5  $\mu$ L, 0.6 nmol) in water was incubated at 37 °C for 2 hr after adding DNA Degradase Plus™ (20  $\mu$ L, 100 units) in 10 X DNA Degradase Plus™ buffer (2.5  $\mu$ L) The DNA Degradase Plus™ was removed using a Nanosep 3K Omega spin column and the sample was analysed using a UHPLC Ultimate 3000 instrumentation with a Q Exactive HF Orbitrap mass spectrometer, flow rate 0.1 mL / min. UV detection was recorded at 260 nm.

Buffer A: 5 mM ammonium acetate (100% H<sub>2</sub>O)

Buffer B: 5 mM ammonium acetate/ 80 % MeCN. 20- 90 % B over 20 mins

Table 19. ODN nucleoside composition analyses data including RP-HPLC conditions, retention time of nucleoside peaks and the corresponding observed [ESI+] masses of the peak.

<i>ODN Code</i>	<i>RP-HPLC Conditions</i>	<i>Retention Time</i>		<i>Calculated ESI+ Mass</i> ( <i>g mol<sup>-1</sup></i> )		<i>Observed ESI+ Mass</i> ( <i>g mol<sup>-1</sup></i> )
AJ03	20-80% B, 20 mins	8.00,	9.58,	228.2,	242.2,	228.1, 242.2
		11.97,	13.63,	252.2,	268.2,	252.1, 268.1,
		16.29		326.1		326.1
AJ04	5-12 % B, 20 mins	6.98,	8.53,	228.2,	242.2,	228.1, 242.1,
		10.88,	12.76,	252.2,	268.2,	252.1, 268.1,
		13.29		340.1		340.1
AJ05	5-12 % B, 20 mins	8.03,	9.49,	228.2,	242.2,	228.1, 242.2
		11.95,	13.63,	252.2,	268.2,	252.1, 268.1,
		19.72		325.1		325.1
AJ06	5-12 % B, 20 mins	8.14,	9.55,	228.2,	242.2,	228.1, 242.1,
		11.97,	13.77,	252.2,	268.2,	252.2, 268.1,
		15.85		339.2		339.1
AJ07	5-12 % B, 20 mins	1.43,	5.52,	228.2,	252.2,	228.1, 252.1,
		6.52,	10.37,	268.2,	242.2,	268.1, 265.1,
		17.79		282.2		282.1
AJ09	5-12 % B, 20 mins	1.23,	7.09,	228.2,	252.2,	228.1, 265.1,
		8.95,	10.59,	268.2,	242.2,	252.1, 268.1,
		20.05		515.2		

### Stable-Isotope Dilution method validation

ODN AJ07 (14.7 nmol) was digested and desalted as stated above. The sample was divided into thirds and spiked with various amounts of <sup>15</sup>N<sub>5</sub>-labelled O<sup>6</sup>-MeG. The amounts were 34.4 pmol, 8.6 pmol and 0 pmol.

The samples were analysed using a UHPLC Ultimate 3000 instrumentation with a Q Exactive HF Orbitrap, flow rate 0.1 mL / min. UV detection was recorded at 260 nm.

Buffer A: 0.01 % Formic acid (100% H<sub>2</sub>O)

Buffer B: MeCN.

Table 20. LC-MS analysis of spiking ODN AJ07 with varying concentrations of <sup>15</sup>N<sub>5</sub>-labelled O<sup>6</sup>-MeG including observed masses of peak corresponding to O<sup>6</sup>-MeG and the ratio between the observed relative intensities of the observed masses.

<i>Amount of <sup>15</sup>N<sub>5</sub>-labelled O<sup>6</sup>-MeG (pmol)</i>	<i>Observed Masses at 18.08 mins (g mol<sup>-1</sup>)</i>	<i>Ratio of Observed Relative Intensity between <sup>14</sup>N: <sup>15</sup>N</i>
0	282.1,	100:0
8.6	282.1, 287.1	100:100
34.4	282.1, 287.1	25:100



## **Chapter 7- *References***

## 7. References

1. Watson, J. D. and Crick, F.H.C. (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.
2. Blackburn, G. M., Gait, M. J., Loakes, D., and Williams, D.M. (2006) Nucleic acids in chemistry and biology, 3<sup>rd</sup> Edn., *Royal Society of Chemistry*.
3. Voet, D. and Voet, J.G. (1995) Biochemistry 2<sup>nd</sup> Edn. *John Wiley & Sons Inc*.
4. Rhodes, D. and Lipps, H.J. (2015) Survey and summary G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.*, **43**, 8627–8637.
5. Moye, A.L., Porter, K.C., Cohen, S.B., Phan, T., Zyner, K.G., Sasaki, N., Lovrecz, G.O., Beck, J.L. and Bryan, T.M. (2015) Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.*, **6**, 1–12.
6. Hänsel-Hertsch, R., Di Antonio, M. and Balasubramanian, S. (2017) DNA G-quadruplexes in the human genome: Detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.*, **18**, 279–284.
7. Steitz, T.A. (1999) DNA polymerases: Structural diversity and common mechanisms. *J. Biol. Chem.*, **274**, 17395–17398.
8. Caruthers, M.H. (1985) Gene synthesis machines: DNA chemistry and its uses. *Science* **230**, 281–285.
9. Schulhof, J. C., Molko, D. and Teoule, R. (1987) The final deprotection step in oligonucleotide synthesis is reduced to a mild and rapid ammonia treatment by using labile base-protecting groups. *Nucleic Acids Res.*, **15**, 397–416.
10. Tanaka, T., Shimizu, M., Kochi, T. and Moriwaki, H. (2013) Chemical-induced Carcinogenesis. *J. Exp. Clin. Med.*, **5**, 203–209.
11. Drake, J.W., Charlesworth, B., Charlesworth, D. and Crow, J. F., (1998) Rates of Spontaneous Mutation John. *Genetics*, **148**, 1667–1686.

12. De Bont, R., van Larebeke, N. (2004) Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis*, **19**, 169–185.
13. Nakabeppu, Y. (2014) Cellular levels of 8-oxoguanine in either DNA or the nucleotide pool play pivotal roles in carcinogenesis and survival of cancer cells. *Int. J. Mol. Sci.*, **15**, 12543–12557.
14. Cavalieri, E., Saeed, M., Zahid, M., Cassada, D., Snow, D., Miljkovic, M. and Rogan, E. (2012) Mechanism of DNA depurination by carcinogens in relation to cancer initiation. *IUBMB Life*, **64**, 169–179.
15. Dahm, C.C., Keogh, R.H., Spencer, E.A., Greenwood, D.C., Key, T.J., Fentiman, I.S., Shipley, M.J., Brunner, E.J., Cade, J.E., Burley, V.J., (2010) Dietary fiber and colorectal cancer risk: A nested case-control study using food diaries. *J. Natl. Cancer Inst.*, **102**, 614–626.
16. Owens, C.W.I. and Padovan, W. (1976) Limitations of ultracentrifugation and in vivo dialysis as methods of stool analysis. *Gut*, **17**, 68–74.
17. Beranek, D.T. (1990) Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.*, **231**, 11–30.
18. Frohler, B. C. and Matteucci, M.D. (1983) Dialkylformamidines: depurination resistant N<sup>7</sup>-protecting group for deoxyadenosine. *Nucleic Acids Res.*, **11**, 8031–8036.
19. Zhao, C., Tyndyk, M., Eide, I. and Hemminki, K. (1999) Endogenous and background DNA adducts by methylating and 2-hydroxyethylating agents. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.*, **424**, 117–125.
20. Gates, K.S., Nooner, T. and Dutta, S. (2004) Biologically relevant chemical reactions of M-alkylguanine residues in DNA. *Chem. Res. Toxicol.*, **17**, 839–856.
21. Barnes, D.E. and Lindahl, T. (2004) Repair and Genetic Consequences of Endogenous

- DNA Base Damage in Mammalian Cells. *Annu. Rev. Genet.*, **38**, 445–476.
22. Huff,A.C. and Topal,M.D. (1987) DNA damage at thymine N-3 abolishes base-pairing capacity during DNA synthesis. *J. Biol. Chem.*, **262**, 12843–12850.
23. Fu,D., Calvo,J.A. and Samson,L.D. (2012) Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat. Rev. Cancer*, **12**, 104–120.
24. Monti,P., Broxson,C., Inga,A., Wang,R. wen, Menichini,P., Tornaletti,S., Gold,B. and Fronza,G. (2011) Effect of  $N^3$ -methyladenine and an isosteric stable analogue on DNA polymerization. *DNA Repair (Amst)*., **10**, 861–868.
25. Warren,J.J., Forsberg,L.J. and Beese,L.S. (2006) The structural basis for the mutagenicity of  $O^6$ -methyl-guanine lesions. *Proc. Natl. Acad. Sci.*, **103**, 19701–19706.
26. Margison,G.P., Santiba,M.F. and Povey,A.C. (2002) Mechanisms of carcinogenicity / chemotherapy by  $O^6$ -methylguanine ' n. **17**, 483–487.
27. Kuhnle,G.G.C. and Bingham,S.A. (2007) Dietary meat, endogenous nitrosation and colorectal cancer. *Biochem. Soc. Trans.*, **35**, 1355–1357.
28. Zhang,F., Tsunoda,M., Suzuki,K., Kikuchi,Y., Wilkinson,O., Millington,C.L., Margison,G.P., Williams,D.M., Czarina Morishita,E. and Takénaka,A. (2013) Structures of DNA duplexes containing  $O^6$ -carboxymethylguanine, a lesion associated with gastrointestinal cancer, reveal a mechanism for inducing pyrimidine transition mutations. *Nucleic Acids Res.*, **41**, 5524–5532.
29. J.L.,B. (1989) ras Oncogenes in human cancer: A review. *Cancer Res.*, **49**, 4682–4689.
30. Hall,C.N., Badawi,A.F., O'Connor,P.J. and Saffliill,R. (1991) The detection of alkylation damage in the DNA of human gastrointestinal tissues. *Br. J. Cancer*, **64**, 59–63.
31. Pollock,J., Bingham,S.A., Cross,A.J., Bailey,N., Bandaletova,T., Bowman,R., Lewin,M.H. and Shuker,D.E.G. (2006) Red meat enhances the colonic formation of the DNA adduct  $O^6$ -carboxymethyl guanine: implications for colorectal cancer risk. *Cancer*



- Res.*, **66**, 1859–1865.
32. García-Santos, M.D.P., Calle, E. and Casado, J. (2001) Amino acid nitrosation products as alkylating agents. *J. Am. Chem. Soc.*, **123**, 7506–7510.
  33. Shephard, S.E., Schlatter, C. and Lutz, W.K. (1987) Assessment of the risk of formation of carcinogenic N-nitroso compounds from dietary precursors in the stomach. *Food Chem. Toxicol.*, **25**, 91–108.
  34. Shibata, T., Glynn, N., McMurry, T.B.H., McElhinney, R.S., Margison, G.P. and Williams, D.M. (2006) Novel synthesis of *O*<sup>6</sup>-alkylguanine containing oligodeoxyribonucleotides as substrates for the human DNA repair protein, *O*<sup>6</sup>-methylguanine DNA methyltransferase (MGMT). *Nucleic Acids Res.*, **34**, 1884–1891.
  35. Lyer, R.R., Pluciennik, A., Burdett, V. and Modrich, P.L. (2006) DNA mismatch repair: Functions and mechanisms. *Chem. Rev.*, **106**, 302–323.
  36. Orren, D.K., Selby, C.P., Hearst, J.E. and Sancar, A. (1992) Post-incision steps of nucleotide excision repair in *Escherichia coli*. *J. Biol. Chem.*, **267**, 780–788.
  37. Margison, G.P., Povey, A.C., Kaina, B. and Santibáñez Koref, M.F. (2003) Variability and regulation of *O*<sup>6</sup>-alkylguanine-DNA alkyltransferase. *Carcinogenesis*, **24**, 625–635.
  38. Potter, P. M., Wilkinson, M. C., Fittont, J., Carrt, F.J., Brennand, J., Cooper, D. P., Margison, G.P., (1987) Characterisation and nucleotide sequence of *ogt*, the *O*<sup>6</sup>-alkylguanine-DNA-alkyltransferase gene of *E.coli*. *Nucleic Acids Res.*, **15**, 9177–9193.
  39. Rebeck, G.W., Smith, C.M., Goad, D.L. and Samson, L. (1989) Characterization of the major DNA repair methyltransferase activity in unadapted *Escherichia coli* and identification of a similar activity in *Salmonella typhimurium*. *J. Bacteriol.*, **171**, 4563–4568.
  40. Jeggo, P., Defais, M., Samson, L. and Schendel, P. (1977) An adaptive response of *E. coli* to low levels of alkylating agent: Comparison with previously characterised DNA repair

- pathways. *MGG Mol. Gen. Genet.*, **157**, 1–9.
41. Paalman,S.R., Sung,C. and Clarke,N.D. (1997) Specificity of DNA repair methyltransferases determined by competitive inactivation with oligonucleotide substrates: Evidence that *Escherichia coli* ada repairs  $O^6$ -methylguanine and  $O^4$ -methylthymine with similar efficiency. *Biochemistry*, **36**, 11118–11124.
  42. Lindahl, T., Demple, B. & Robins,P. (1982) *E. coli*  $O^6$ -methylguanine-DNA methyltransferase. **1**, 1359–1363.
  43. Senthong,P., Millington,C.L., Wilkinson,O.J., Marriott,A.S., Watson,A.J., Reamtong,O., Evers,C.E., Williams,D.M., Margison,G.P. and Povey,A.C. (2013) The nitrosated bile acid DNA lesion  $O^6$ -carboxymethylguanine is a substrate for the human DNA repair protein  $O^6$ -methylguanine-DNA methyltransferase. *Nucleic Acids Res.*, **41**, 3047–3055.
  44. Kaina,B., Christmann,M., Naumann,S. and Roos,W.P. (2007) MGMT: Key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. *DNA Repair (Amst)*., **6**, 1079–1099.
  45. Gerson,S.L. (2004) MGMT: Its role in cancer aetiology and cancer therapeutics. *Nat. Rev. Cancer*, **4**, 296–307.
  46. Pegg,A.E. and Byers,T.L. (1992) Repair of DNA containing  $O^6$ -alkylguanine. *FASEB J.*, **6**, 2302–2310.
  47. Middleton,M.R., Kelly,J., Thatcher,N., Donnelly,D.J., McElhinney,R.S., McMurry,T.B.H., McCormick,J.E. and Margison,G.P. (2000)  $O^6$ -(4-bromophenyl)guanine improves the therapeutic index of temozolomide against A375M melanoma xenografts. *Int. J. Cancer*, **85**, 248–252.
  48. Wilkinson,O.J., Latypov,V., Tubbs,J.L., Millington,C.L., Morita,R., Blackburn,H., Marriott,A., McGown,G., Thorncroft,M., Watson,A.J., (2012) Alkyltransferase-like protein (At1) distinguishes alkylated guanines for DNA repair using cation-

- interactions. *Proc. Natl. Acad. Sci.*, **109**, 18755–18760.
49. Pearson, S.J., Wharton, S., Watson, A.J., Begum, G., Butt, A., Glynn, N., Williams, D.M., Shibata, T., Santibáñez-Koref, M.F. and Margison, G.P. (2006) A novel DNA damage recognition protein in *Schizosaccharomyces pombe*. *Nucleic Acids Res.*, **34**, 2347–2354.
50. Latypov, V.F., Tubbs, J.L., Watson, A.J., Marriott, A.S., Thorncroft, M., Wilkinson, O.J., Senthong, P., Butt, A., Arvai, A.S., Millington, C.L., (2013). *Mol. Cell*, **47**, 50–60.
51. Klaene, J.J., Sharma, V.K., Glick, J. and Vouros, P. (2013) The analysis of DNA adducts: The transition from <sup>32</sup>P-postlabeling to mass spectrometry. *Cancer Lett.*, **334**, 10–19.
52. Tretyakova, N., Villata, P. W. and Kotapati, S. (2013) Mass Spectrometry of structurally modified DAN. *Chem. Rev.*, **113**, 2395–2436.
53. Kebarle, P. and Tang, L. (1993) From ions in solution to ions in the gas phase - the mechanism of electrospray mass spectrometry. *Anal. Chem.*, **65**, 972A-986A.
54. Buhrman, D.L., Price, P.I. and Rudewicz, P.J. (1996) Quantitation of SR 27417 in human plasma using electrospray liquid chromatography-tandem mass spectrometry: A study of ion suppression. *J. Am. Soc. Mass Spectrom.*, **7**, 1099–1105.
55. McLuckey, S.A. (1992) Principles of Collisional Activation in. *J. Soc. Mass Spectrom.*, **3**, 599–614.
56. Tretyakova, N., Goggin, M., Sangaraju, D. and Janis, G. (2012) Quantitation of DNA adducts by stable isotope dilution mass spectrometry. *Chem. Res. Toxicol.*, **25**, 2007–2035.
57. Wang, S., Cyronak, M. and Yang, E. (2007) Does a stable isotopically labeled internal standard always correct analyte response?. A matrix effect study on a LC/MS/MS method for the determination of carvedilol enantiomers in human plasma. *J. Pharm. Biomed. Anal.*, **43**, 701–707.
58. Mesaros, C., Lee, S. H., Blair, I.A. (2009) Targeted quantitative analysis of eicosanoid

- lipids in biological samples using liquid chromatography-tandem mass spectrometry. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, **26**, 2736–2745.
59. Blair, I.A. (2010) Analysis of endogenous glutathione-adducts and their metabolites. *Biomed. Chromatogr.*, **24**, 29–38.
60. Da Pieve, C., Sahgal, N., Moore, S. A. and Velasco-Garcia, M.N. (2013) Development of a liquid chromatography/tandem mass spectrometry method to investigate the presence of biomarkers of DNA damage in urine related to red meat consumption and risk of colorectal cancer. *Rapid Commun. Mass Spectrom.*, **27**, 2493–2503.
61. Bussche, J. Vanden, Moore, S.A., Pasmans, F., Kuhnle, G.G.C. and Vanhaecke, L. (2012) An approach based on ultra-high pressure liquid chromatography – tandem mass spectrometry to quantify *O*<sup>6</sup>-methyl and *O*<sup>6</sup>-carboxymethylguanine DNA adducts in intestinal cell lines. *J. Chromatogr. A*, **1257**, 25–33.
62. Kotandeniya, D., Murphy, D., Yan, S., Park, S., Seneviratne, U., Koopmeiners, J., S., Pegg, A., Kanugula, S., Kassiey, F. and N.T. (2013) Kinetics of *O*<sup>6</sup>-Pyridyloxobutyl-2'-deoxyguanosine repair by human *O*<sup>6</sup>-alkylguanine DNA alkyltransferase. *Biochemistry*, **52**, 4075–4088.
63. Tretyakova, N.Y., Wishnok, J.S. and Tannenbaum, S.R. (2000) Peroxynitrite-induced secondary oxidative lesions at guanine nucleobases: Chemical stability and recognition by the Fpg DNA repair enzyme. *Chem. Res. Toxicol.*, **13**, 658–664.
64. Malayappan, B., Johnson, L., Nie, B., Panchal, D., Matter, B., Jacobson, P. and Tretyakova, N. (2010) Quantitative high-performance liquid chromatography-electrospray ionization tandem mass spectrometry analysis of bis-N7-guanine DNA-DNA cross-links in white blood cells of cancer patients receiving cyclophosphamide therapy. *Anal. Chem.*, **82**, 3650–3658.
65. Lawley, P.D. and Phillips, D.H. (1996) DNA adducts from chemotherapeutic agents.

- Mutat. Res. - Fundam. Mol. Mech. Mutagen.*, **355**, 13–40.
66. Noll, D. M., Mason, T. M., Miller, P.S. (2006) Formation and repair of interstrand cross-links in DNA. *Chem. Rev.*, **106**, 277–310.
67. Panasci, L., Paiement, J.-P., Christodoulopoulos, G., Belenkov, A., Malapetsa, A. and Aloyz, R. (2001) Chlorambucil drug resistance in chronic lymphocytic leukemia. *Clin. Cancer Res.*, **7**, 454–461.
68. Cho, Y., Rizzo, C.J., Huang, H., Minko, I.G., Lloyd, R.S., Wang, H., Kozekova, A., Harris, T.M., Kozekov, I.D., Kim, H., (2008) Interstrand DNA cross-links induced by  $\alpha,\beta$ -unsaturated aldehydes derived from lipid peroxidation and environmental sources. *Acc. Chem. Res.*, **41**, 793–804.
69. McGhee, J.D. and von Hippel, P.H. (1975) Formaldehyde as a probe of DNA structure. 2. Reaction with endocyclic imino groups of DNA bases. *Biochemistry*, **14**, 1297–1303.
70. McGhee, J.D. and von Hippel, P.H. (1977) Formaldehyde as a probe of DNA structure. 3. Equilibrium denaturation of DNA and synthetic polynucleotides. *Biochemistry*, **16**, 3267–3276.
71. McGhee, J.D. and von Hippel, P.H. (1977) Formaldehyde as a probe of DNA structure. 4. Mechanism of the initial reaction of formaldehyde with DNA. *Biochemistry*, **16**, 3276–3293.
72. Chaw, Y.F.M., Eric Crane, L., Lange, P. and Shapiro, R. (1980) Isolation and identification of cross-links from formaldehyde-treated nucleic acids. *Biochemistry*, **19**, 5525–5531.
73. Huang, H. and Hopkins, P.B. (1993) DNA interstrand cross-linking by formaldehyde: nucleotide sequence preference and covalent structure of the predominant cross-link formed in synthetic oligonucleotides. *J. Am. Chem. Soc.*, **115**, 9402–9408.
74. Hodson, M.R.G., Silhan, J., Crossan, G.P., Garaycochea, J.I., Mukherjee, S., Johnson, C.M., Schärer, O.D. and Patel, K.J. (2014) Mouse SLX4 is a tumor suppressor

- that stimulates the activity of the nuclease XPF-ERCC1 in DNA crosslink repair. *Mol. Cell*, **54**, 472–484.
75. Pontel,L.B., Rosado,I. V, Burgos-barragan,G., Swenberg,J.A., Crossan,G.P., Pontel,L.B., Rosado,I. V, Burgos-barragan,G., Garaycoechea,J.I., Yu,R., (2015) Endogenous formaldehyde is a hematopoietic stem cell genotoxin and metabolic carcinogen article endogenous formaldehyde is a hematopoietic stem cell genotoxin and metabolic carcinogen. *Mol. Cell*, **60**, 1–12.
76. Knipscheer,P., Räschle,M., Smogorzewska,A., Enoiu,M., Ho,T.V., Schärer,O.D., Elledge,S.J. and Walter,J.C. (2009) The Fanconi Anemia pathway promotes replication-dependent DNA interstrand cross-link repair puck. *Science.*, **326**, 1698–1701.
77. Angelov,T., Guainazzi,A. and Schärer,O.D. (2009) Generation of DNA interstrand cross-links by post-synthetic reductive amination. *Org. Lett.*, **11**, 661–664.
78. Shibata,T., Dohno,C. and Nakatani,K. (2009) DNA cross-link generated by a novel modified DNA containing a formyl group. *Nucleic Acids Symp. Ser.*, **53**, 171–172.
79. Noll, David,M. (2004) Preparation of interstrand cross-linked DNA oligonucleotide duplexes. *Front. Biosci.*, **9**, 421.
80. Millard,J.T., Raucher,S. and Hopkins,P.B. (1990) Mechlorethamine cross-links 2'-deoxyguanosine residues at 5'-GNC sequences in duplex DNA fragments. *J. Am. Chem. Soc.*, **112**, 2459–2460.
81. Rink,S.M., Solomon,M.S., Taylor,M.J., Hopkins,P.B., Rajur,S.B. and McLaughlin,L.W. (1993) Covalent structure of a nitrogen mustard-induced DNA interstrand cross-link: an N7-to-N7 Linkage of 2'-deoxyguanosine residues at the duplex sequence 5'-d(GNC). *J. Am. Chem. Soc.*, **115**, 2551–2557.
82. Fischhaber,P.L., Gall,A.S., Duncan,J.A. and Hopkins,P.B. (1999) Direct demonstration in synthetic oligonucleotides that N , N ' deoxyguanosine to N3 of deoxycytidine on

- opposite strands of duplex DNA. *Cancer Res.*, **59**, 4363–8.
83. Raschle, M., Knipscheer, P., Enoiu, M., Angelov, T., Sun, J., Griffith, J. D., Eleenberger, T. E., Scharer, O. D. and Walter, J. C. (2008) Mechanism of replication coupled DNA interstrand crosslink repair. *Cell*, **134**, 969–980.
84. Ráz, M.H., Dexter, H.R., Millington, C.L., Van Loon, B., Williams, D.M. and Sturla, S.J. (2016) Bypass of mutagenic  $O^6$ -carboxymethylguanine DNA adducts by human Y- and B-family polymerases. *Chem. Res. Toxicol.*, **29**, 1493–1503.
85. Lees, N.P., Harrison, K.L., Hall, C.N., Margison, G.P. and Povey, A.C. (2004) Reduced MGMT activity in human colorectal adenomas is associated with K-ras GC→AT transition mutations in a population exposed to methylating agents. *Carcinogenesis*, **25**, 1243–1247.
86. Shuker, D.E.G. and Margison, G.P. (1997) Nitrosated glycine derivatives as a potential source of  $O^6$ -methylguanine in DNA. *Cancer Res.*, **57**, 366–369.
87. Pletsas, D., Wheelhouse, R.T., Pletsa, V., Nicolaou, A., Jenkins, T.C., Bibby, M.C. and Kyrtopoulos, S.A. (2006) Polar, functionalized guanine- $O^6$  derivatives resistant to repair by  $O^6$ -alkylguanine-DNA alkyltransferase: implications for the design of DNA-modifying drugs. *Eur. J. Med. Chem.*, **41**, 330–339.
88. Xu, Y.Z. (2000) Synthesis and characterization of DNA containing  $O^6$ -carboxymethylguanine. *Tetrahedron*, **56**, 6075–6081.
89. McBride, L.J., Kierzek, R., Beaucage, S.L. and Caruthers, M.H. (1986) Amidine protecting groups for oligonucleotide synthesis. *J. Am. Chem. Soc.*, **108**, 2040–2048.
90. Lakshman, M.K., Ngassa, F.N., Keeler, J.C., Dinh, Y.Q. V., Hilmer, J.H. and Russon, L.M. (2002) Facile synthesis of  $O^6$ -alkyl-,  $O^6$ -aryl-, and diaminopurine nucleosides from 2'-deoxyguanosine. *Org. Lett.*, **2**, 927–930.
91. Nos, C. and Units, E. DNA Degradase Plus <sup>TM</sup>. **2020**, 2020–2021.

92. Cavaluzzi, M.J. and Borer, P.N., (2004) Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res.*, **32**, 13e – 13.
93. Daniels, D.S. (2000) Active and alkylated human AGT structures: a novel zinc site, inhibitor and extrahelical base binding. *EMBO J.*, **19**, 1719–1730.
94. Daniels, D.S., Woo, T.T., Luu, K.X., Noll, D.M., Clarke, N.D., Pegg, A.E. and Tainer, J.A. (2004) DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nat. Struct. Mol. Biol.*, **11**, 714–720.
95. Churchwell, M.I., Beland, F.A. and Doerge, D.R. (2006) Quantification of *O*<sup>6</sup>-methyl and *O*<sup>6</sup>-ethyl-2'-deoxyguanosine adducts in C57BL/6N/Tk+/- mice using LC/MS/MS. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, **844**, 60–66.
96. Thomale, J., Hochleitner, K. and Rajewsky, M.F. (1994) Differential formation and repair of the mutagenic DNA alkylation product *O*<sup>6</sup>-ethylguanine in transcribed and nontranscribed genes of the rat. *J. Biol. Chem.*, **269**, 1681–1686.
97. Tintoré, M., Grijalvo, S., Eritja, R. and Fàbrega, C. (2015) Synthesis of oligonucleotides carrying fluorescently labelled *O*<sup>6</sup>-alkylguanine for measuring hAGT activity. *Bioorganic Med. Chem. Lett.*, **25**, 5208–5211.
98. Janeba, Z., Francom, P. and Robins, M.J. (2003) Efficient Syntheses of 2-Chloro-2' - deoxyadenosine. *J. Org. Chem.*, **68**, 989-992.
99. Guo, J., Villalta, P.W. and Turesky, R.J. (2017) Data-independent mass spectrometry approach for screening and identification of DNA adducts. *Anal. Chem.*, **89**, 11728–11736.
100. Shrivastava, A. and Gupta, V. (2011) Methods for the determination of limit of detection and limit of quantitation of the analytical methods. *Chronicles Young Sci.*, **2**, 21.
101. Reddy, N.R. (2018) Stable labelled isotopes as internal standards: a critical review. *Mod. Appl. Pharm. Pharmacol.*, **1**, 1–4.



102. Stokvis,E., Rosing,H. and Beijnen,J.H. (2005) Stable isotopically labeled internal standards in quantitative bioanalysis using liquid chromatography/mass spectrometry: Necessity or not? *Rapid Commun. Mass Spectrom.*, **19**, 401–407.
103. Varga,E., Glauner,T., Köppen,R., Mayer,K., Sulyok,M., Schuhmacher,R., Krska,R. and Berthiller,F. (2012) Stable isotope dilution assay for the accurate determination of mycotoxins in maize by UHPLC-MS/MS. *Anal. Bioanal. Chem.*, **402**, 2675–2686.
104. Yang,Y., Nikolic,D., Swanson,S.M. and Van Breemen,R.B. (2002) Quantitative determination of  $N^7$ -methyldeoxyguanosine and  $O^6$ -methyldeoxyguanosine in DNA by LC-UV-MS-MS. *Anal. Chem.*, **74**, 5376–5382.
105. Moore,S.A. and Shuker,D.E.G. (2011) Synthesis of deuterium and C-13-labelled ethyl glycolate and their subsequent use in the synthesis of labelled analogues of the DNA adduct  $O^6$ -carboxymethyl-2'-deoxyguanosine. *J. Label. Compd. Radiopharm.*, **54**, 855–858.
106. Brink,A., Lutz,U., Völkel,W. and Lutz,W.K. (2006) Simultaneous determination of  $O^6$ -methyl-2'-deoxyguanosine, 8-oxo-7,8-dihydro-2'-deoxyguanosine, and 1, $N^6$ -etheno-2'-deoxyadenosine in DNA using on-line sample preparation by HPLC column switching coupled to ESI-MS/MS. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, **830**, 255–261.
107. Kim,H.Y., Nechev,L., Zhou,L., Tamura,P., Harris,C.M. and Harris,T.M. (1998) Synthesis and adduction of fully deprotected oligodeoxynucleotides containing 6-chloropurine. *Tetrahedron Lett.*, **39**, 6803–6806.
108. Fu,Y., Jia,G., Pang,X., Wang,R.N., Wang,X., Li,C.J., Dai,Q., Bailey,K.A., Nobrega,M.A., Han,K., (2013) FTO - Mediated formation of  $N^6$  - hydroxymethyladenosine and  $N^6$ -formyladenosine in mammalian RNA. *Nat. Commun.*, **4**, 1–8.

109. Ohkubo,A., Sakamoto,K., Miyata,K.I., Taguchi,H., Seio,K. and Sekine,M. (2005) Convenient synthesis of N-unprotected deoxynucleoside 3'- phosphoramidite building blocks by selective deacylation of N-acylated species and their facile conversion to other N-functionalized derivatives. *Org. Lett.*, **7**, 5389–5392.
110. Caruthers,M.H. (2009) Chemical synthesis of DNA. *J. Chem. Educ.*, **66**, 577.
111. Drewt,H.R., Wingtt,R.M., Takanot,T., Brokat,C., Tanakat,S., Itakurii,K. and Dickersont,R.E. (1981) Biochemistry Structure of a B-DNA dodecamer: Conformation and dynamics\* (DNA structure/sugar puckering in DNA/principle of anticorrelation/torsion angles in DNA/thermal vibration in DNA). *Proc. Natl. Acad. Sci. USA*, **78**, 2179–2183.
112. Angelica,M.D. and Fong,Y. (2008) 50 years of DNA ‘Breathing’: Reflections on Old and New Approaches. *October*, **141**, 520–529.
113. Hoffman,E.A., Frey,B.L., Smith,L.M. and Auble,D.T. (2015) Formaldehyde crosslinking: A tool for the study of chromatin complexes. *J. Biol. Chem.*, **290**, 26404–26411.
114. Cheng,G., Wang,M., Upadhyaya,P., Villalta,P.W. and Hecht,S.S. (2008) Formation of formaldehyde adducts in the reactions of DNA and deoxyribonucleosides with  $\alpha$ -acetates of 4-(methylnitrosamino)-1-(3- pyridyl)-1-butanone (NNK), 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL), and N-nitrosodimethylamine (NDMA). *Chem. Res. Toxicol.*, **21**, 746–751.
115. Wang,M., Cheng,G., Balbo,S., Carmella,S.G., Villalta,P.W. and Hecht,S.S. (2009) Clear differences in levels of a formaldehyde-DNA adduct in leukocytes of smokers and nonsmokers. *Cancer Res.*, **69**, 7170–7174.

## ***Chapter 8- Appendix***